Detection of Asphalt Pavement Potholes and Cracks Based on the Unmanned Aerial Vehicle Multispectral Imagery

Yifan Pan, Xianfeng Zhang, Guido Cervone, and Liping Yang

***Abstract*—Asphalt roads are the basic component of a land transportation system, and the quality of asphalt roads will decrease during the use stage because of the aging and deterioration of the road surface. In the end, some road pavement distresses may appear on the road surface, such as the most common potholes and cracks. In order to improve the efficiency of pavement inspection, currently some new forms of remote sensing data without destructive effect on the pavement are widely used to detect the pavement distresses, such as digital images, light detection and ranging, and radar. Multispectral imagery presenting spatial and spectral features of objects has been widely used in remote sensing application. In our study, the multispectral pavement images acquired by unmanned aerial vehicle (UAV) were used to distinguish between the normal pavement and pavement damages (e.g., cracks and potholes) using machine learning algorithms, such as support vector machine, artificial neural network, and random forest. Comparison of the performance between different data types and models was conducted and is discussed in this study, and indicates that a UAV remote sensing system offers a new tool for monitoring asphalt road pavement condition, which can be used as decision support for road maintenance practice.**

***Index Terms*—Artificial neural network (ANN), asphalt roads, multispectral imagery, pavement distress, random forest (RF), support vector machine (SVM), unmanned aerial vehicle (UAV).**

I. Introduction

ROADS, as one of the most important infrastructures, play a crucial role in supporting the development of an economy and human society. Therefore, it is imperative to maintain roads in a good condition. The quality of road pavement directly determines the lifetime of one road segment [1]. Asphalt road pavement is the most common road surface type in the road system. However, due to the combined effect of aging and deterioration of the road surface, some types of distresses would always appear on the pavement during the use of the road [2]. Potholes and cracks are the two most common categories of road surface damages and have significant influences on the running quality of vehicles [3]. Road condition inspection is the first work the road department should do before road maintenance and rehabilitation. However, nowadays, it is really a challenging work for the road management department to quickly obtain the large-scale technical condition of asphalt road pavement because of the rapid growth of the mileage of the road networks, especially the highway. According to the 2017 Statistical Communique of China Transport and Logistic Development [4], the total length of roads was about 4.77 million kilometers, and the highway reached 0.13 million kilometers. Moreover, the length of the roads that had been maintained and rehabilitated was up to 4.67 million kilometers, which accounted for 97.9% of the total road length. Previously, field investigations and manual measurements were the conventional methods to detect and assess the pavement distresses. For example, the falling weight deflectometer is used to check the road structure strength [5]. The 3-m ruler is used to measure the road roughness [6]. Pavement friction measurements can be tested by lateral force pendulum instrument [7]. Obviously, this kind of methods is time consuming and labor intensive, most of which are even destructive to the road surface meanwhile [8], [9].

Currently, with the support of computer and remote sensing technologies, many forms of remote sensing data and advanced pattern recognition algorithms have been introduced into the automated detection of pavement damages without doing harm to the road pavement [10]–[15]. For instance, light detection and ranging (LiDAR) technology can directly acquire the elevation information of the deteriorated pavement to measure the geometric dimensions of the road pavement damages [16], [17]. Ground penetrating radar utilizes radar pulses to image the subsurface profile to detect subsurface objects, changes in material properties, voids, and cracks [18]. Especially, in the previous studies, the digital images or videos with only single or RGB channels captured, based on the mobile vehicle, are the most frequently used data type to automatically detect the pavement damages [2], [19]. As for digital pavement images, image

Manuscript received February 9, 2018; revised May 2, 2018 and July 3, 2018; accepted August 4, 2018. Date of publication September 2, 2018; date of current version October 15, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant 41571331 and in part by the Xinjiang Corps Geospatial Information Technology Innovation under Grant 2016AB001. *(Corresponding author: Xianfeng Zhang.)*

Y. Pan is with the Institute of Remote Sensing and Geographic Information System, Peking University, Beijing 100871, China, and also with the Chinese Academy of Electronics and Information Technology, Beijing 100041, China (e-mail: yfpan@pku.edu.cn).

X. Zhang is with the Institute of Remote Sensing and Geographic Information System, Peking University, Beijing 100871, China (e-mail: xfzhang@pku.edu.cn).

G. Cervone and L. Yang are with the GeoInformatics and Earth Observation Laboratory, Pennsylvania State University, University Park, PA 16802 USA (e-mail: cervone@psu.edu; liping.yang@psu.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTARS.2018.2865528

processing and machine learning algorithms are the two primary methods to extract the surface defects [19]. Image processing can be used to isolate the defects from the background and create a binary image, such as histogram thresholding, morphological algorithms (opening, closing, dilation, and erosion), wavelet, and Fourier transformation [11], [20]–[23]. The identification of pavement distresses is then interpreted from the binary image [24]. For example, Georgopoulos detected the flexible pavement cracks using the digital image processing tools [21]. Kim utilized one simple unmanned aerial vehicle (UAV) system to capture the pavement images and identified cracks based on the image binarization method [20]. However, the automatic detection of pavement distresses based on an image processing method would become more complex and challenging for images with high variations of lighting and road surface texture [25], [26]. With the development of artificial intelligence in recent years, some machine learning algorithms were introduced into the automatic pavement defects detection field, such as a support vector machine (SVM) [27], [28], an artificial neural network (ANN) [29], [30], and a random forest (RF) [31]. For example, Xu segmented the pavement images into a number of square tiles and then extracted four customized features of tiles to train the BP neural networks to identify the crack tiles [30]. Mokhtaria first extracted the crack image objects using the morphological image processing technique and then fed six features including area, length, texture, intensity, location, and orientation to four classifiers (ANN, decision tree, *K*-nearest neighbors, and adaptive neurofuzzy inference system) to identify the crack damages [32]. An accurate and effective description of pavement damages features is the key factor to ensure the accuracy and reliability of machine learning algorithms to detect road surface defects. The common digital images only have a single or three RGB channels that can only provide limited spectral features of pavement damages [24], [33], [34]. In the remote sensing field, multispectral images (MSI) with wide wavelength coverage have been widely used to characterize the detailed spectral features of the objects, which would further improve the classification accuracy [35]–[37]. Enlightened by this characteristic of MSI, the multispectral pavement images were captured and first attempted to detect the road surface damages in our study.

The vehicle integrated with some types of sophisticated and heavy remote sensing devices is the commonly used platform to collect the remote sensing data for pavement monitoring by the majority of road departments, such as the pavement management system [11], [19]. However, the mobile vehicle has some potential risks to the normal traffic order and pedestrian safety. In addition, the camera mounted on the vehicle can only capture one small portion of the road surface at a time. Therefore, it is unable to obtain the full pavement of different lanes simultaneously, and consequently most of the studies just only focused on one kind of distresses (e.g., potholes or cracks) in their research because of the small coverage of the road surface, whereas one more type of damage could exist on the pavement at the same time. In last ten years, UAV taking advantage of carrying light-weight remote sensors and large spatial coverage has been applied to some fields, for instance, topographic mapping, natural disaster surveying, precision agriculture, and traffic monitoring [38]–[43]. To solve the above-mentioned problems of current vehicle platform, the flexible UAV platform configured with one multispectral imaging system was selected to collect the road pavement image data in our experiment.

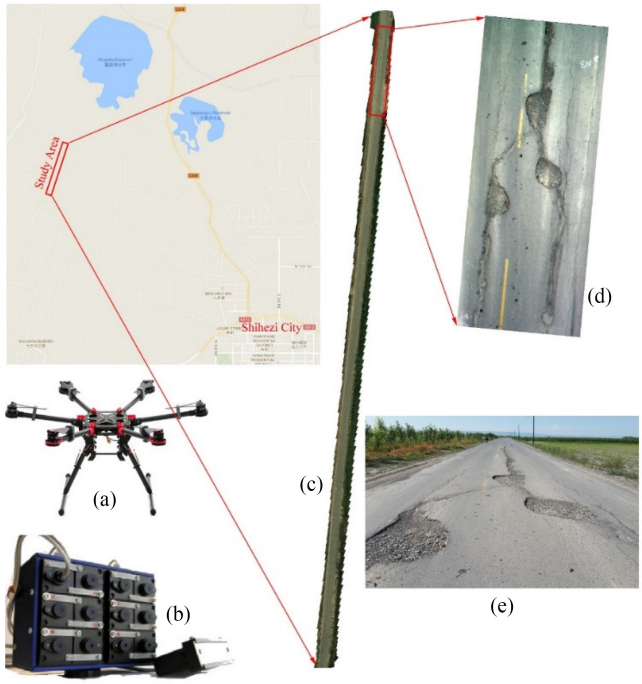


Fig. 1. (a) Six-spreading-wings UAV. (b) MCA snap12 camera. (c) Full MCA road image. (d) Sample MCA road image. (e) Spot photo of the study road.

In our study, we proposed one new application procedure that applied SVM, ANN, and RF classifiers to detect the asphalt road surface cracks and potholes from the UAV multispectral pavement images. The central objective and contribution of this study is to investigate the feasibility of the proposed model, and offer a new tool for monitoring asphalt road pavement condition to improve the efficiency of road maintenance practice. This paper is organized as follows. After this introduction, Section II will introduce the data and methods that include image acquisition and processing, image segmentation, sample data preparation, feature selection, and three classifiers introduction. In Section III, practical experiment and results will be presented to prove the feasibility of the proposed procedure and find out the best model on detection of asphalt pavement damages. In Section IV, the selectable spatial resolution of UAV pavement images for detecting pavement damages and the importance of every feature will be discussed. The conclusion is delivered in the end.

II. Data and Methods

A. Image Acquisition and Segmentation

The asphalt road segment is located in the rural farmland area of Shihezi City, Xinjiang, China (see Fig. 1). This road was built in 2006, and the east half-side of the road was repaved in 2010. The total length of the road, we captured, was about 500 m. This road was mainly used to connect villages and for

TABLE I
SPECIFICATIONS OF THE MINI-MCA12 CAMERA

Band Number	Wavelength(nm)	Band Number	Wavelength(nm)
1	490	7	700
2	520	8	720
3	550	9	800
4	570	10	840
5	671	11	900
6	680	12	950

(a) (b)

Fig. 2. (a) GCPs landmark. (b) Huace LT600T Client.

crops transportation by tractors or heavy trucks. Based on the field investigation, the pavement was in a worse condition with a variety of severe potholes and cracks [see Fig. 1(e)]. The field investigation shows that the mean width of cracks is about 2.8 cm, and the diameter of potholes is about 97 cm. A multispectral camera microminiature multiple camera array system (MCA) designed by Tetracam Inc., Chatsworth, GA, USA, was mounted on a six-wing UAV [see Fig. 1(a) and (b)] to capture the pavement images on July 29, 2017. MCA configures 12 bands spanning from blue to near-infrared wavelength (450–1000 nm) (see Table I). The flight task was conducted at noon without clouds to avoid the shadow influence, and the wind speed was less than 0.2 m/s, all of which guaranteed the safety and success of the flight experiment. The UAV flied along with the road at the height of 25 m above the ground level, and as a result one image pixel corresponded to about 13.54*13.54 mm^2 area of the road pavement. Two hundred one pavement image tiles were acquired with 70% of overlapping between two sequential images. The DN images were calibrated using a 100% reflectance white board to produce the reflectance images. The Pix4D software was used to mosaic all the image tiles into one full image [see Fig. 1(c)]. Fourteen ground control points (GCPs) were set at both sides of the road to rectify the geometric distortion of the mosaic image [see Fig. 2(a)]. The Huace LT600T client integrated with continuously operating reference stations was used to measure the GPS location of GCPs within the error 0.05 m [see Fig. 2(b)]. The road area was extracted manually by one shape file using ArcGIS tools.

Given the high resolution of the UAV pavement images, multiresolution segmentation (MS) algorithm integrated in eCognition Developer software was adopted to conduct the segmentation of the pavement images in a multiresolution manner. MS identifies individual image objects of one pixel in size and merges them with their neighbors based on relative homogeneity criteria. This homogeneity criterion is a combination of spectral and shape criteria that are calculated through a comprehensive scale parameter. Higher values for the scale parameter result in larger image objects and smaller values in smaller ones [44].

Consider that potholes and cracks are in different scales and the pixel values of the pavement images in distress (e.g., potholes and cracks) areas present bigger variation than those in the nondistressed pavement. Therefore, one fixed scale parameter cannot segment the potholes, cracks, and normal pavement correctly at the same time. In order to obtain intact and practical image objects, the segmentation will follow the following procedure. First, one small-scale parameter is used to extract the small and linear crack objects. Second, the contrast value of every image object, calculated based on the gray-level cooccurrence matrix (GLCM) [45], is selected as the threshold to determine which small objects will be merged into one full object. GLCM will also be used to extract other textural features in feature selection (see Table II).

B. Sample Dataset Preparation and Feature Selection

Sufficient sample data are essentially necessary for training and validating machine learning algorithms [32]. Three classes were defined in our study, i.e., pothole, crack, and nondistressed pavement. The nondistressed pavement includes a normal pavement and yellow traffic lines. It can be observed from the comparison of two sequential images that the pixel value in the same location has bias because of the illumination difference caused by the different solar incident angle (see Fig. 3). Consequently, this will lead to some degree of difference between the segmentation results of the same target in the sequential images. In this study, the same target in two of sequential images is seen as two different objects, and they will be extracted based on the two images, respectively.

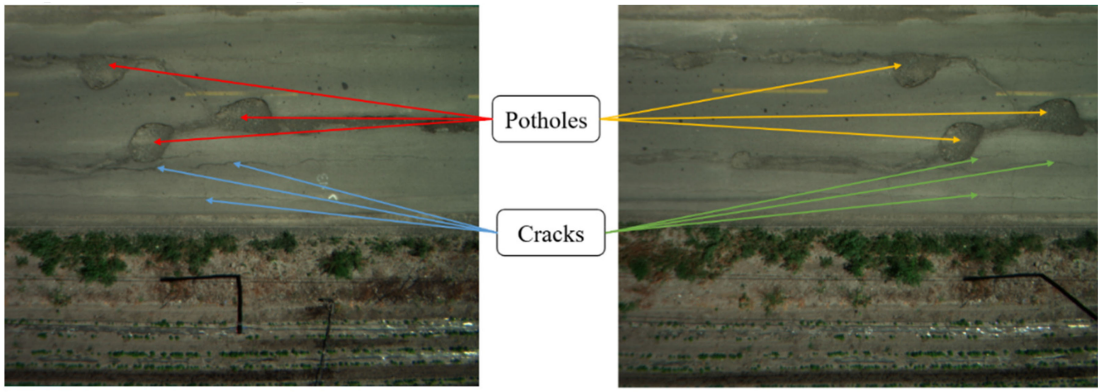
Due to the fact that feature selection has a great influence on the performance of learning algorithms, reasonable numbers and types of features are able to increase the accuracy of algorithm while decreasing the computation time [46]. Generally speaking, three types of image features can be extracted from digital imagery, i.e., spectral, geometric, and textural features. In our study, based on the prior knowledge of feature value distribution of every category of image objects, ten types of features were extracted to train and validate the learning algorithms (see Table III). Descriptions of every feature are listed as follows.

1) *Mean:* The mean of pixel values in every band for every image object. Generally, the mean of distressed area is lower than nondistressed pavement area.

TABLE II
FORMULA OF GLCM FEATURES

Type	Formula	Type	Formula
Contrast	$\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} P_{i,j}(i-j)^2$	Dissimilarity	$\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} P_{i,j}\lvert i-j\rvert$
Homogeneity	$\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} \frac{P_{i,j}}{1+(i-j)^2}$	Correlation	$\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} \frac{(i-\mu_I)(j-\mu_J)}{\sigma_I \sigma_J}$

Note: *i*, *j* are the row and column number of GLCM, respectively. $P_{i,j}$ is the value in the cell *i*, *j*. *N* is the number of rows or columns. $\mu_J = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} jP_{i,j}, \sigma_J = \sqrt{\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} P_{i,j}(j-\mu_j)^2}$.



(a) (b)

Fig. 3. Same targets in two sequential images. (a) Left. (b) Right.

TABLE III
FEATURES SELECTED TO RUN RANDOM FOREST

Category	Name	Category	Name	Category	Name
Spectral	Mean of each band	Geometric	Area	Textural	Contrast
			Length/Width		Homogeneity
			Elliptic Fit		Dissimilarity
	STD of each band		Roundness		Correlation

2) *Standard deviation:* (STD) The STD of pixel values for every image object in every band. Because the internal material of distressed area is exposed to the pavement, the STD would be higher than that of nondistressed pavement objects.
3) *Area:* The area of image object is calculated based on the spatial resolution and pixel numbers. Generally, the size of four classes follows this sequence: Normal pavement >Pothole> Traffic line>Crack.
4) *Length/width:* The ratio of length to width of one image object. Obviously, the length/width of a crack is higher than other classes.
5) *Elliptic fit:* The feature is used to evaluate the similarity between an image object and ellipse. Potholes are more similar to ellipse; as a result, the value will be higher than other classes.
6) *Roundness:* The feature is used to evaluate the similarity between an image object and a circle. Some potholes are also similar to the circle.

Contrast, homogeneity, dissimilarity, and correlation features are calculated based on the GLCM to evaluate the value variation inside the image object. Nondistressed pavement objects have lower variation than damage areas (see Table II).

Furthermore, considering the different value distribution of every feature, feature normalization was implemented based on

$$X_{\mathrm{Norm}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{1}$$

where X_{Norm} is the normalized feature vector. $X_{\max}$ and $X_{\min}$ are the maximum and minimum values of the feature X, respectively. All features are in the same range from 0 to 1, which could speed up the convergence efficiency of learning algorithms.

C. Support Vector Machine

An SVM is a classification system derived from statistical learning theory. It separates the classes with a decision surface that maximizes the margin between the classes [27]. The surface is often called the optimal hyperplane, and the data points closest to the hyperplane are called support vectors. The support vectors are the critical elements of the training set. An SVM is one of the nonprobabilistic binary classifiers to assign new examples to one category or the other [28]. It means that one

SVM can only solve the two-class problems. An SVM can also perform the multiclass problems by combining several binary SVM classifiers together, based on the logic classification procedure of one-versus-one or one-versus-all. One special feature of an SVM is the kernel function, which is introduced to deal with nonlinear classification problems. The kernel function can map the original examples into a high-dimensional feature space, in which the nonlinear classification problem will become the linear case. There are several types of kernel model with different performance for different applications, such as linear kernel, polynomial kernel, Gaussian kernel, etc. In our study, the performance of four types of kernel models on the detection of potholes and cracks were evaluated, i.e., linear, quadratic, cubic, and Gaussian.

D. Artificial Neural Network

An ANN mimics the way human brain solves problems with a large number of neurons [29], [30]. An ANN is composed, typically, of three kinds of layers, i.e., the input layer, the hidden layer, and the output layer. Every layer comprises a certain number of nodes, similar to the neurons in the brain. The number of nodes in the input layer is determined by the number of features in the example data, whereas the number of output classes decides the number of nodes in the output layer. The number of hidden layers and associated nodes could vary for different applications. Moreover, every node corresponds to a kind of activation function that defines the output of that node given a set of inputs. Sigmoid, Softmax, and rectified linear unit are commonly used in an ANN. Which of them should be used depends on the objective of the application. Back propagation is a widely used training procedure for an ANN to adjust the weights and bias between the nodes. In our study, a three-layer feed-forward network with one input layer, one Sigmoid hidden layer, and one Softmax output layer was constructed to classify the potholes and cracks. The network will be trained with the conjugate gradient method to minimize the difference between the output node activation and the output. In order to find out the appropriate number of nodes in the hidden layer for pavement distress detection, a series of numbers from 1 to 15 was evaluated one by one.

E. Random Forest

RF is a member of ensemble learning algorithms, which combines a certain number of decision tree classifiers together as a forest to predict the class of new examples [47]. Every tree in the forest is trained with a subset training sample set, which is resampled from the original training dataset. The resampling is implemented with replacement and follows the bootstrap sampling procedure, i.e., the number of subset examples is same as the original examples. In addition to the resampling of training samples for every tree, the features used to find the best split at each node of tree are resampled from the original feature set as well. The class of new sample is predicted by every tree in forest, and is assigned based on a majority vote of them. The bootstrap and feature resampling tactics strengthen the robust and generalization of RF compared with other machine learning algorithms, ANN, KNN, etc. Moreover, RF can also generate the importance of every feature based on the resampling strategy [48]. Based on the prior knowledge of previous work [31], an RF classifier was selected to detect the potholes and cracks from the UAV MSI in this study and to estimate the contribution of every feature on this application. The number of trees has a significant effect on the accuracy and computation time of RF. In this study, a series of tree number will be set to evaluate what size of forest will perform best on the pavement distress detection. The RF classification accuracy will be validated using the out-of-bag (OOB) error [47].

The above-mentioned three machine learning algorithms would run on a PC configured with Core i7-6700HQ CPU@ 2.6 GHz, Nvidia Quadro M1000M GPU, and 16 GB RAM. The running time was also recorded to be as one of the important indicators of the algorithm performance.

III. Experiments and Results

To evaluate the advantage of UAV MSI in the detection of road pavement damages, all the images acquired by the UAV remote sensing system in the suburb of Shihezi City were divided into two subdatasets: One only includes the RGB channels (Band5, Band3, and Band1) of every image as the common digital image (RGB images); and the other one includes all the 12 image bands (Band1–Band12) to represent the MSI Two subdatasets will be used to extract the pavement damages, respectively. The flowchart of our experiment is illustrated in Fig. 4.

A. Two-Step Image Segmentation

As mentioned in Section II-A, it is difficult to choose one appropriate scale parameter to extract intact potholes and cracks simultaneously. For instance, the pavement image in Fig. 5(a) contains one crack and one pothole. Fig. 5(b) and (c) shows the segmentation results by the scale parameter in eCognition application 50 and 150, respectively. It can be seen that smaller values for scale parameter will partition one whole pothole into several small image objects because of the heterogeneous material inside the pothole, whereas larger scale parameters will result in missing of some small cracks. Following the procedure in Section II-A, the GLCM contrast value was chosen to measure the variations within the distress area and the nondistressed area in this study. Fig. 5(d) shows the contrast values of every image object. Obviously, the contrast values of the distress area (e.g., pothole and cracks) are significantly higher than the nondistressed pavement. Hence, in order to obtain an intact pothole object, a merge operation was conducted based on the contrast values of objects over the initial segmentation result from the smaller scale parameter. All image objects, the contrast value of which exceed the given threshold of 0.58, will be merged into one image object. As Fig. 5(f) shows, the pothole became one intact object after a merge operation while the crack object remains unchanged. The RGB and MSI would all be segmented following this two-step segmentation procedure.

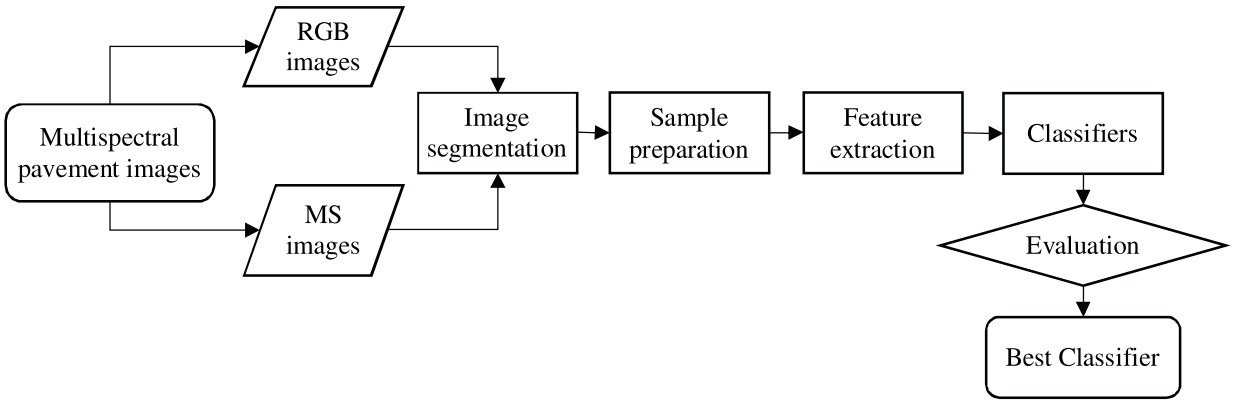


Fig. 4. Flowchart of the experiment of pavement damages detection.

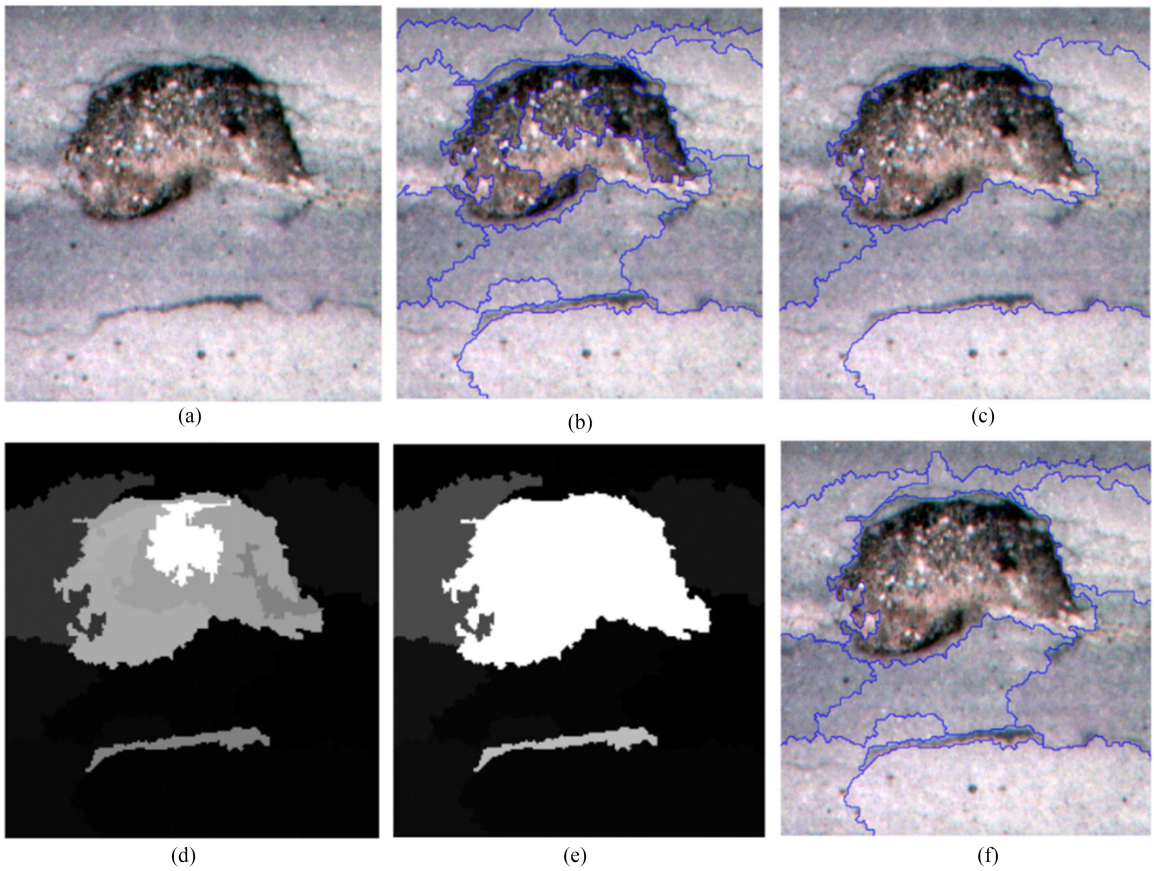


Fig. 5. (a) Original image. (b) Segmentation with a scale parameter 50. (c) Segmentation with a scale parameter 150. (d) Contrast values of image objects. (e) Merge of image object based on the contrast. (f) Improved segmentation.

B. Performance of SVM, ANN, and RF

Based on the rules in Section II-B, the sample data for training and validating were extracted, and total 1760 pieces of sample image objects containing 538 potholes, 753 cracks, and 469 nondistressed pavements (305 damage-free pavement and 164 yellow traffic lines) were collected from the UAV images finally (see Fig. 6). The total feature types of RGB and MSI are 14 and 32, respectively, because of the different number of bands (6 spectral features for RGB images and 24 spectral features for MSI). In order to verify the performance of each type of features toward the detection of potholes and cracks, seven combinations of three kinds of features were individually used to produce seven classification experiments, i.e., spectral features (C1), geometrical features (C2), textural features (C3), spectral features and geometrical features (C4), spectral features and textural features (C5), geometric features and textural features (C6), and spectral, geometrical features, and textural features (C7).

Fig. 7 indicates the performance of an SVM configured with different types of kernel functions and seven feature

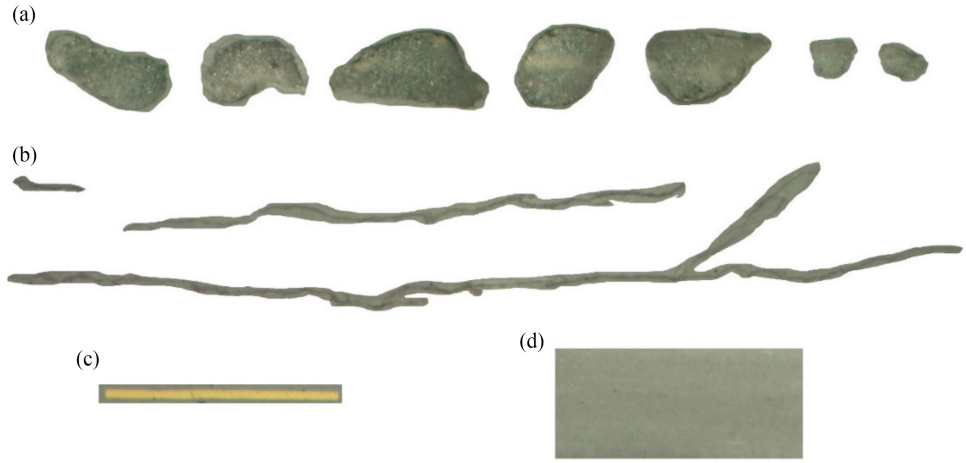


Fig. 6. Sample objects of (a) potholes, (b) cracks, (c) yellow traffic line, and (d) normal pavement.

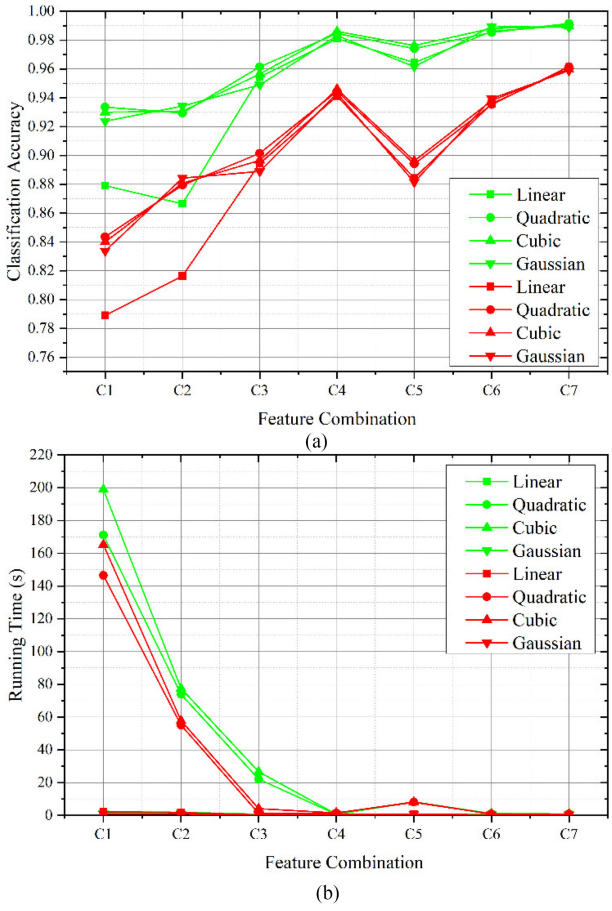


Fig. 7. (a) Classification accuracy and (b) running time of SVM algorithms over four types of kernel functions and seven feature combinations. The curves in red represent the result of the RGB images and the curves in green indicate the performance of the MSI.

combinations. Fig. 7(a) shows that the SVM with linear kernel presented lower classification accuracy when it is trained and validated only using either spectral features or geometry features individually. Along with introducing texture features or more types of features, the four kinds of SVM models (linear, quadratic, cubic, and Gaussian) almost performed similarly on feature combination C3, C4, C5, C6, and C7, and the highest accuracy was achieved by using three types of features together, i.e., C7. It can also be concluded that the accuracies derived from MSI are generally higher than from RGB images. Fig. 7(b) indicates the running time of different SVM models. It can be seen that the SVM with polynomial kernels (quadratic and cubic) cost most time on the feature sets of C1, C2, and C3. For C4, C5, C6, and C7, all types of SVM models performed similarly on the running time. It is interesting that the SVM models with linear and Gaussian kernel took almost the same time on each of six feature combinations. Considering the as higher accuracy as possible and relative less running time, Fig. 7 shows that the SVM model configured with linear kernel and the feature combination C7 could achieve the highest classification accuracy of 98.78% while cost the least running time of 0.63 s from the MSI.

Fig. 8 shows the variation of classification accuracy and running time of an ANN with respect to different numbers of neurons in the hidden layer. Specifically, when the number of hidden neurons was set to one, it means that only one abstract feature in hidden layer was used to classify the objects, which was not sufficient to distinguish between the pavement and distresses (cracks and potholes). Moreover, it took the most time to train and validate an ANN in this case. With increasing the number of hidden neurons, the classification accuracy could benefit a lot from the more abstract features learned by the ANN, and the running time decreased generally [see Fig. 8(b)]. Fig. 8(a) shows that the ANN models with more than one type of features (C4, C5, C6, and C7) and two more hidden neurons could always result in a higher accuracy. It also can be observed that when the number of hidden neurons was set over two, the classification accuracy did not change so much. Moreover, it can also be concluded that the accuracies derived from MSI are generally higher than RGB images. Taking in account the running time and accuracy together, the ANN with four hidden neurons and feature combination C4 was the best model to classify the

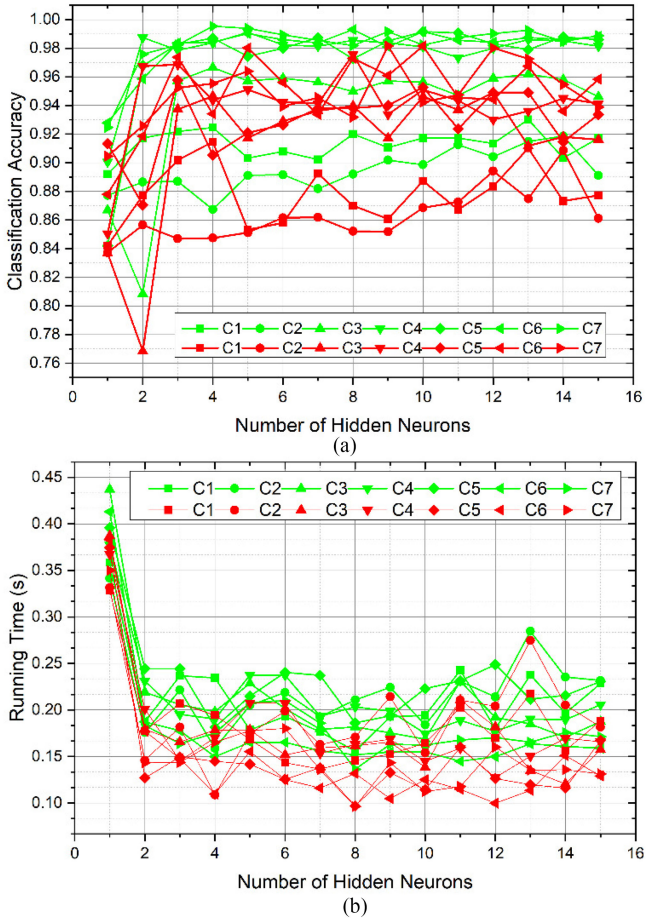


Fig. 8. (a) Classification accuracy and (b) running time of ANN algorithms over a series of numbers of hidden neurons and seven feature combinations. The curves in red represent the result of the RGB images and the curves in green indicate the performance of the MSI.

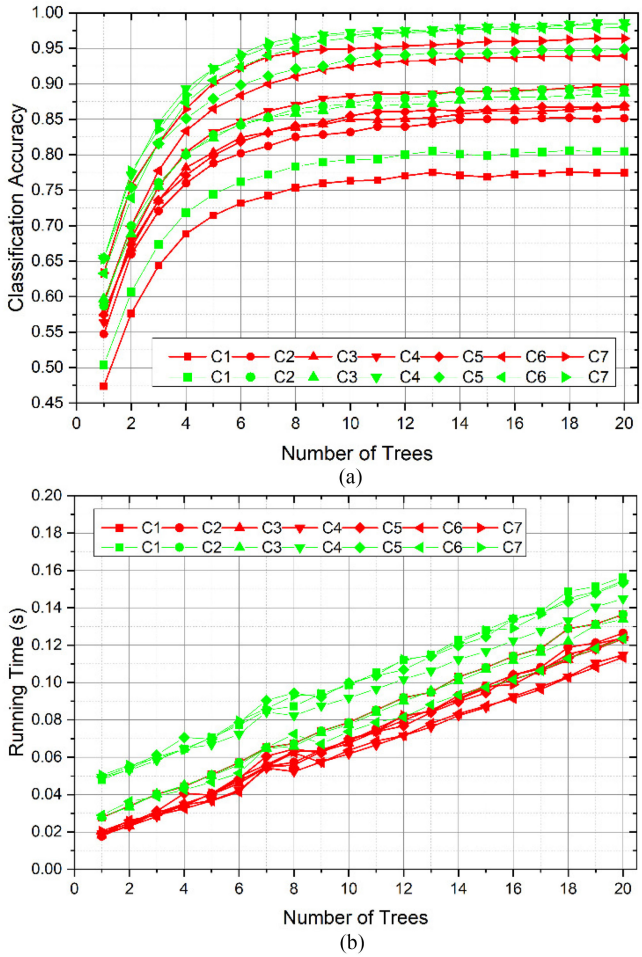


Fig. 9. (a) Classification accuracy and (b) running time of RF algorithms over a series of numbers of trees and seven feature combinations. The curves in red represent the result of the RGB images and the curves in green indicate the performance of the MSI.

pavement and distresses with the overall accuracy 98.46% and the corresponding running time was 0.21 s, which is also based on the MSI.

Appropriate number of decision trees in the forest greatly contributes to the accuracy and efficiency of RF classification. As a result, we selected a series of numbers of trees to optimize the best forest size for pavement damages detection. The performance of RF with different number of trees in the forest and two types of images is shown in Fig. 9(a). Obviously, the accuracy of RF classification maintained increases along with the growth of quantity of trees till a flat trend. The feature combinations with one more type of features (i.e., C4, C5, C6, and C7) always achieved better classification. Spatial features, namely geometry and texture, played a crucial role in improving the classification. Compared with the RGB images, the RF classification of the multispectral pavement images has higher accuracy when using the same feature combinations. The running time of RF algorithms for different number of trees and the two types of images is illustrated in Fig. 9(b), which has demonstrated that the more trees were used in the forest and the longer time would be spent in the experiment. Because MSI contain more spectral features, longer running time was also consumed. Nevertheless,

TABLE IV
CLASSIFICATION ACCURACY AND RUNNING TIME OF SVM, ANN, AND RF CLASSIFIERS

Models	SVM	ANN	RF
Accuracy	98.78%	98.46%	98.83%
Running Time(s)	0.63s	0.21s	0.09s

the difference of running time between the RGB and MSI is relatively small (<0.02 s). Considering the accuracy and running time together, the best model should get the highest classification accuracy and consume relatively less running time. As a result, the MSI are selected because of the better performance than RGB images (see Fig. 9). Moreover, the best size of forest is 18 trees with all three types of features (RF18-C7), which can achieve the highest accuracy (98.83%) with minimum running time (0.09 s) (see Table IV).

In a brief summary, among the three models, the RF18-C7 could get the best performance together with the highest classification accuracy and least running time for the detection of the pavement distresses (see Table IV). Table V presents the confusion matrix of the classification using the

TABLE V
CONFUSION MATRIX OF RF18-C7

Number of Observations		Predicted Class			Accuracy (%)
		Crack	Pothole	Non-distressed	
True Class	Crack	748	4	1	99.34
	Pothole	5	531	2	98.69
	Non-distressed	8	10	451	96.16
Reliability (%)		98.29	97.43	99.34	98.3 (OA)

Note: OA—Overall accuracy.

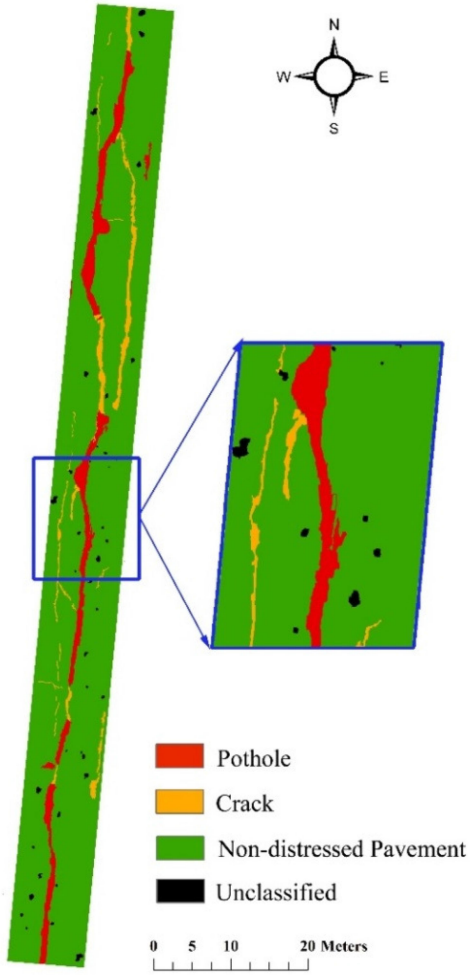


Fig. 10. Classification result of a sample road segment using the RF18-C7 model.

RF18-C7 model and multispectral pavement images, and the overall accuracy of 98.3% was achieved. The accuracy (also known as producer's accuracy) presents the accuracy of the classification, which is the fraction of correctly classified image objects with regard to all image objects of that ground truth class. The reliability (also known as user's accuracy) presents the reliability of classes in the classified image, which is the fraction of correctly classified image objects with regard to all image objects classified as this class in the classified image. The classification result of one sample road based on the RF18-C7 model is illustrated in Fig. 10. It can be seen that most of the potholes and cracks were classified correctly,

TABLE VI
STATISTICS OF THE NUMBER AND AREA OF POTHOLES AND CRACKS IN THE STUDY ROAD

Class	Potholes	Cracks
Number	17	34
Area(m^2)	46.86	57.19

but some locations of the pavement were polluted by oil spills or artificial traffic symbols so that they were unable to be classified using the RF model. Based on the classification result, the total number and area of cracks and potholes can be calculated (see Table VI).

IV. DISCUSSION

The incorporation of spectral, geometrical, and textural features extracted from the UAV MSI into the random forest classification can produce better delineation of the potholes and cracks from the regular pavement on the UAV images. However, the algorithm is still sensitive to the spatial resolution of the UAV images, thus we will discuss about the impact of UAV image spatial resolution and feature importance on the classification.

A. Selectable Resolution of Pavement Images

The setting of UAV flight altitude can change the spatial resolution of pavement images acquired during the fly. Higher spatial resolution can provide rich features of road pavement objects, especially the spatial and textural features. However, higher spatial resolution also mean that the large storage memory should be used to store the volume data, and lower flight altitude has a degree of potential risk to the public traffic and people. Considering above-mentioned problems, appropriate spatial resolution should be determined prior to the flight task. In this experiment, the multispectral pavement images were resampled to a series of images with different spatial pixel sizes. Following the methods mentioned in Section II, these images were segmented into image objects first, and then the features of every object were calculated. The optimized model RF18-C7 was applied to classify the potholes, cracks, and nondistressed pavements (see Fig. 11). The test samples were collected in advance to evaluate the classification accuracy.

The classification accuracies using a series of multispectral pavement images with different resampled pixel sizes are shown in Fig. 12. It can be seen that the classification accuracy generally decreased while reducing the spatial resolution or pixel size. Due to the fact, the mean width of cracks is only 2.8 cm in our experiment, coarser image pixel size would lead to that some

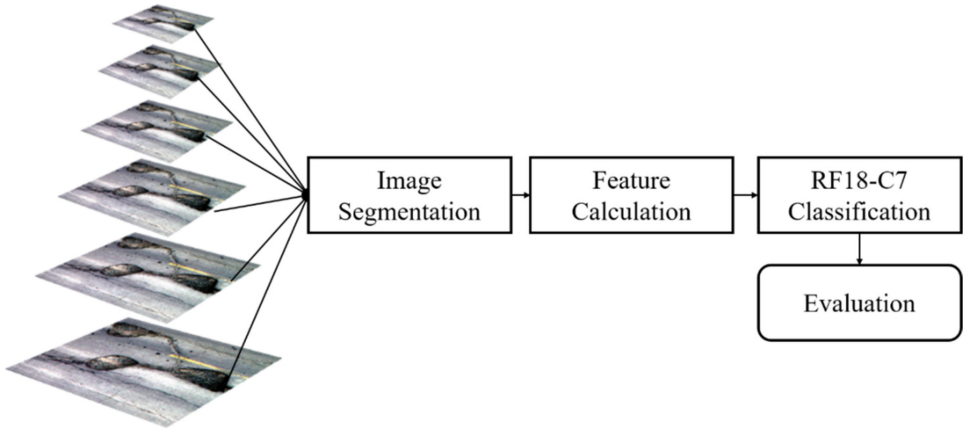


Fig. 11. Flowchart of evaluation of pavement image spatial resolutions.

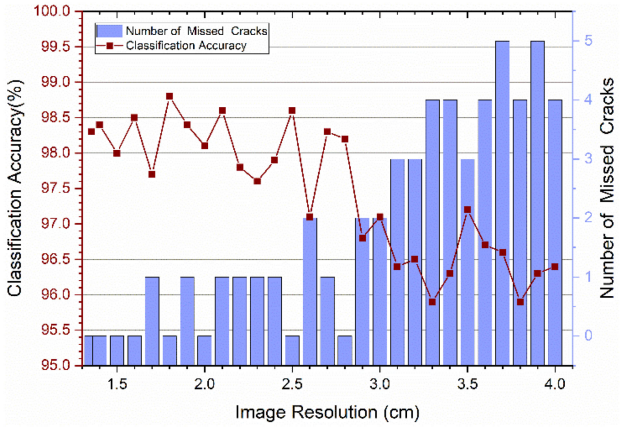


Fig. 12. Classification accuracy of MSI of pavement with different spatial resolutions and the number of missed cracks.

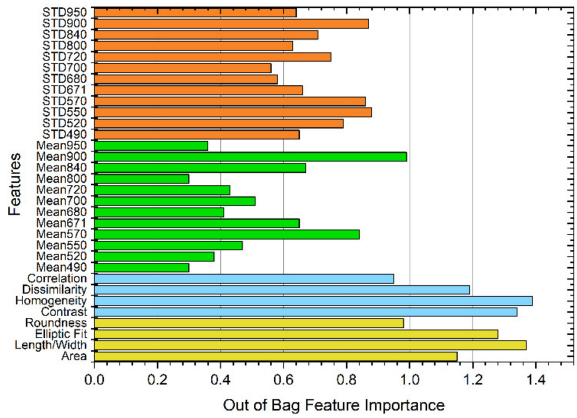


Fig. 13. Out-of-bag feature importance of three types of features (Red: Geometry, Green: Texture, Blue: Spectral-Mean, and Yellow: Spectral-STD).

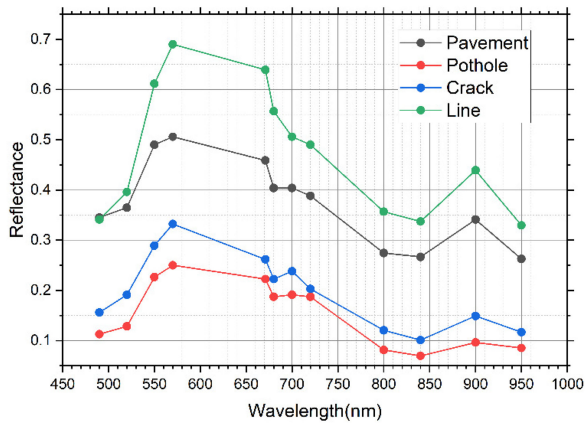


Fig. 14. Mean spectra of object of pavement, potholes, cracks, and traffic lines.

cracks could not be segmented successfully when the resolution exceeded the minimum scale of the cracks, and consequently they would be missed finally in the classification. It is also can be seen in Fig. 12 that the number of missed cracks would increase when the pixel size was over about 3 cm. Therefore, the appropriate spatial resolution should not exceed the minimum scale of pavement objects, which requires that sufficient field works should be conducted before the flight task.

B. Feature Importance Evaluation

One of the main aims in this study is to determine appropriate feature set for detecting asphalt potholes and cracks. Three types of features were introduced in the RF classification in this study, and generated the relative importance of every feature. In the RF framework, the most widely used score of importance of a given feature is the increase in mean of the error of a tree in the forest when the observed values of this variable are randomly permuted in the OOB samples. The larger this score is, the more important the feature is. In Fig. 13, the importance value of the three types of features is presented. It can be seen that the geometrical and textural features played an important role in the application, especially the homogeneity and length\width features. Obviously, some small pebbles and other material always filled potholes and cracks, which result in high variation inside the pothole and crack image objects. The importance of standard deviation is almost same because it is also a measurement of pixel value variation. For the spectral mean feature, the bands centered at the wavelength of 570 nm and 900 nm made

a great contribution to the detection of potholes and cracks. Combining with the detailed spectra of the pavement objects (see Fig. 14), it indicated that the reflectance of potholes and cracks was lower than nondistressed pavement and traffic line in the band wavelengths. The band reflectance centered at 570 nm and 900 nm has great difference between pothole, cracks, pavement, and traffic line, which is helpful to distinguish between them from each other.

V. Conclusion

This study presents an approach for the detection of asphalt road pavement distresses from UAV MSI using SVM, artificial neural networks, and RF learning algorithms. The case study in the suburb of Shihezi City indicated that spatial features (i.e., texture and geometry) contributed much more to the accuracy of the cracks and potholes detection than the spectral ones for both RGB and 12-band MSI. The three types of features extracted from the UAV MSI can achieve best classification and less running time if using an 18-tree RF classifier. The overall accuracy of the classification of cracks, potholes, and nondistressed pavements is 98.3% with the UAV MSI. In addition to the feature set and classifier, the spatial resolution of pavement imagery is also a decisive factor for the performance of the RF classifier. The comparison study of the simulated multiple resolution imagery showed that the spatial resolution should not exceed the minimum scale of pavement distress objects; otherwise some small damages (i.e., cracks) may be missed even in the segmentation procedure. In conclusion, the flexible UAV platform configured with multispectral remote sensors provide a valuable tool for the monitoring of asphalt pavement condition.

In future work, more UAV pavement images of different road areas and segments could be used to further evaluate the performance of these models and parameters on the detection of potholes and cracks. Because of the spatial resolution limitation, the UAV pavement images used in the paper still cannot capture the cracks that width are less than 13.54 mm. Therefore, higher resolution pavement images should be obtained to further increase the accuracy of pavement condition evaluation. In this paper, we just focus on the two common asphalt pavement damages, potholes and cracks. The next work will consider more types of road surfaces and pavement damages, such as cement road, gravel road, rutting, and road roughness. Other remote sensing data including LiDAR and radar by UAV also have a great potential on the pavement condition monitoring. For example, LiDAR can directly acquire the elevation information of the road surface. By means of it, the depth of potholes, cracks, and ruttings can be characterized directly by the LiDAR data. Additionally, other advanced learning algorithms could also be introduced into the pavement distresses detection, such as convolutional neural networks. Moreover, the integrated software and algorithms by individual coding could be used to speed up the detection of pavement distresses in the future.

Acknowledgment

The authors would like to thank Dr. Q. Z. Zhao for providing the Pix4D software to process the road images and analysis, and also Beijing Panorama Space Technology Co., Ltd., for providing the eCognition Developer Software to do the image analysis. The authors would like to extend their appreciation to Dr. Q. Du, Editor, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing and the anonymous reviewers for their meaningful comments and suggestions during the review process.

References

[1] Y. Pan, X. Zhang, J. Tian, X. Jin, L. Luo, and K. Yang, "Mapping asphalt pavement aging and condition using multiple endmember spectral mixture analysis in Beijing, China," *J. Appl. Remote Sens.*, vol. 11, no. 1, 2017, Art. no. 016003.

[2] C. Koch, K. Georgieva, V. Kasireddy, B. Akinci, and P. Fieguth, "A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure," *Adv. Eng. Inform.*, vol. 29, no. 2, pp. 196–210, 2015.

[3] A. Tedeschi and F. Benedetto, "A real-time automatic pavement crack and pothole recognition system for mobile Android-based devices," *Adv. Eng. Inform.*, vol. 32, pp. 11–25, 2017.

[4] "2017 Statistical Communique of China Transport and Logistic Development," China Transp. Newspaper, Ministry Transport PRC, Beijing, China, p. 002, 2018.

[5] A. Claessen, C. Valkering, and R. Ditmarsch, "Pavement evaluation with the falling weight deflectometer," in *Proc. Assoc. Asphalt Paving Technol.*, 1976, vol. 45, pp. 122–157.

[6] J. Wambold, L. Defrain, R. Hegmon, K. Macghee, J. Reichert, and E. Spangler, "State of the art of measurement and analysis of road roughness," *Transp. Res. Rec.*, vol. 836, pp. 21–29, 1981.

[7] G. G. Balmer, R. A. Zimmer, and R. D. Tonda, "Pavement friction measurements and vehicle control reparations for nontangent road sections," in *Pavement Maintenance and Rehabilitation*, Philadelphia, PA, USA: ASTM International, 1985.

[8] J. Eriksson, L. Girod, B. Hull, R. Newton, S. Madden, and H. Balakrishnan, "The pothole patrol: Using a mobile sensor network for road surface monitoring," in *Proc. 6th Annu. Int. Conf. Mobile Syst. Appl. Services*, 2008, pp. 29–39.

[9] M. Y. Shahin, *Pavement Management for Airports, Roads, and Parking Lots*. New York, NY, USA: Springer-Verlag, 2007.

[10] S. Zhang and S. M. Bogus, "Use of low-cost remote sensing for infrastructure management," in *Construction Research Congress 2014*, Atlanta, GE, USA: ASCE, 2014, pp. 1299–1308.

[11] E. Schnebele, B. Tanyu, G. Cervone, and N. Waters, "Review of remote sensing methodologies for pavement management and assessment," *Eur. Transp. Res. Rev.*, vol. 7, no. 2, pp. 1–19, 2015.

[12] C. Mettas, K. Themistocleous, K. Neocleous, A. Christofe, K. Pilakoutas, and D. Hadjimitsis, "Monitoring asphalt pavement damages using remote sensing techniques," *Proc. SPIE*, vol. 9535, 2015, Art. no. 95350S.

[13] W. Emery and M. C. Singh, "Large-area road-surface quality and land-cover classification using very-high spatial resolution aerial and satellite data," Tech. Rep., 2013.

[14] G. D. Cline, M. Y. Shahin, and J. A. Burkhalter, "Automated data collection for pavement condition index survey," in *Proc. Annu. Meeting Transp. Res. Board*, Washington, DC, USA, 2003, pp. 1–16.

[15] K. C. Wang, Z. Hou, Q. B. Watkins, and S. R. Kuchikulla, "Automated imaging technique for runway condition survey," in *Proc. FAA Worldwide Airport Technol. Transfer Conf. Expo.*, 2007.

[16] J. Choi, L. Zhu, and H. Kurosu, "Detection of cracks in paved road surface using laser scan image data," *Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. 41, pp. 559–562, 2016.

[17] W. Uddin and E. Al-Turk, "Airborne LiDAR digital terrain mapping for transportation infrastructure asset management," in *Proc. 5th Int. Conf. Manag. Pavements*, Seattle, WA, USA, 2001, pp. 11–14.

[18] H. M. Jol, *Ground Penetrating Radar Theory and Applications*. Edinburgh, U.K.: Elsevier, 2008.

[19] H. Zakeri, F. M. Nejad, and A. Fahimifar, "Image based techniques for crack detection, classification and quantification in asphalt pavement: A review," *Arch. Comput. Methods Eng.*, vol. 24, no. 4, pp. 935–977, 2017.

[20] H. Kim, S. Sim, and S. Cho, "Unmanned aerial vehicle (UAV)-powered concrete crack detection based on digital image processing," in *Proc. Int. Conf. Adv. Exp. Structural Eng.*, 2015.

[21] A. Georgopoulos, A. Loizos, and A. Flouda, "Digital image processing as a tool for pavement distress evaluation," *ISPRS J. Photogrammetry Remote Sens.*, vol. 50, no. 1, pp. 23–33, 1995.
[22] S. Chambon and J.-M. Moliard, "Automatic road pavement assessment with image processing: Review and comparison," *Int. J. Geophys.*, vol. 2011, pp. 1–20, 2011.
[23] P. P. Singh and R. Garg, "Automatic road extraction from high resolution satellite image using adaptive global thresholding and morphological operations," *J. Indian Soc. Remote Sens.*, vol. 41, no. 3, pp. 631–640, 2013.
[24] H. Oliveira, "Crack detection and characterization in flexible road pavements using digital image processing," Ph.D. dissertation, Universidade Técnica de Lisboa, Lisboa, Portugal, 2013.
[25] A. Zhang *et al.*, "Automated pixel-level pavement crack detection on 3D asphalt surfaces using a deep-learning network," *Comput.-Aided Civil Infrastructure Eng.*, vol. 32, no. 10, pp. 805–819, 2017.
[26] K. C. Wang, A. Zhang, J. Q. Li, Y. Fei, C. Chen, and B. Li, "Deep learning for asphalt pavement cracking recognition using convolutional neural network," in *Proc. Int. Conf. Airfield Highway Pavements*, 2017, pp. 166–177.
[27] M. R. Schlotjes, M. P. Burrow, H. T. Evdorides, and T. F. Henning, "Using support vector machines to predict the probability of pavement failure," *Proc. Inst. Civil Eng., Transport*, vol. 168, no. 3, pp. 212–222, 2015.
[28] S. Siddharth, P. Ramakrishnan, G. Krishnamurthy, and B. Santhi, "Pavement crack classifiers: A comparative study," *Res. J. Appl. Sci. Eng. Technol.*, vol. 4, no. 24, pp. 5434–5437, 2012.
[29] T. Saar and O. Talvik, "Automatic asphalt pavement crack detection and classification using neural networks," in *Proc. IEEE 12th Biennial Baltic Electron. Conf.*, 2010, pp. 345–348.
[30] G. Xu, J. Ma, F. Liu, and X. Niu, "Automatic recognition of pavement surface crack based on BP neural network," in *Proc. IEEE Int. Conf. Comput. Elect. Eng.*, 2008, pp. 19–22.
[31] Y. Pan, X. Zhang, M. Sun, and Q. Zhao, "Object-based and supervised detection of potholes and cracks from the pavement images acquired by UAV," *presented at the Tehran's Joint ISPRS Conf. GI Res., SMPR and EOEC 2017*, 2017.
[32] S. Mokhtaria, L. Wub, and H.-B. Yunc, "Comparison of supervised classification techniques for vision-based pavement crack detection," in *Proc. 95th Annu. Meeting, Transp. Res. Board*, Washington, DC, USA, 2016, pp. 1–17.
[33] P. P. Singh and R. Garg, "Study of spectral reflectance characteristics of asphalt road surface using geomatics techniques," in *Proc. IEEE Int. Conf. Adv. Comput. Commun. Inform.*, 2013, pp. 516–520.
[34] P. P. Singh and R. Garg, "Road detection from remote sensing images using impervious surface characteristics: Review and implication," *Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. 40, no. 8, pp. 955–959, 2014.
[35] M. R. Resende, L. L. B. Bernucci, and J. A. Quintanilha, "Monitoring the condition of roads pavement surfaces: Proposal of methodology using hyperspectral images," *J. Transp. Literature*, vol. 8, no. 2, pp. 201–220, 2014.
[36] A. Mei, R. Salvatori, N. Fiore, A. Allegrini, and A. D'Andrea, "Integration of field and laboratory spectral data with multi-resolution remote sensed imagery for asphalt surface differentiation," *Remote Sens.*, vol. 6, no. 4, pp. 2765–2781, 2014.
[37] M. Herold and D. Roberts, "Spectral characteristics of asphalt road aging and deterioration: Implications for remote-sensing applications," *Appl. Opt.*, vol. 44, no. 20, pp. 4327–4334, 2005.
[38] C. Zhang and A. Elaksher, "An unmanned aerial vehicle-based imaging system for 3D measurement of unpaved road surface distresses," *Comput.-Aided Civil Infrastructure Eng.*, vol. 27, no. 2, pp. 118–129, 2012.
[39] W. Feng, W. Yundong, and Z. Qiang, "UAV borne real-time road mapping system," in *Proc. IEEE Urban Remote Sens. Joint Event*, 2009, pp. 1–7.
[40] S. Hinz, *Airborne and Spaceborne Traffic Monitoring*. Amsterdam, The Netherlands: Elsevier, 2006.
[41] A. Puri, "A survey of unmanned aerial vehicles (UAV) for traffic surveillance," Dept. Comput. Sci. Eng., Univ. South Florida, Tampa, FL, USA, 2005.
[42] R. Workman, A. Otto, and A. Irving, "The use of appropriate high-tech solutions for road network and condition analysis, with a focus on satellite imagery," Tech. Rep., 2016.
[43] K. Themistocleous, K. Neocleous, K. Pilakoutas, and D. G. Hadjimitsis, "Damage assessment using advanced non-intrusive inspection methods: Integration of space, UAV, GPR, and field spectroscopy," in *Proc. 2nd Int. Conf. Remote Sens. Geoinf. Environ.*, 2014, vol. 9229, pp. 101–105.
[44] A. Darwish, K. Leukert, and W. Reinhardt, "Image segmentation for the purpose of object-based classification," in *Proc. Int. Geosci. Remote Sens. Symp.*, 2003, vol. 3, pp. 2039–2041.
[45] W. Su *et al.*, "Textural and local spatial statistics for the object-oriented classification of urban areas using high resolution imagery," *Int. J. Remote Sens.*, vol. 29, no. 11, pp. 3105–3117, Jun. 10, 2008.
[46] H. Oliveira and P. L. Correia, *Supervised Crack Detection and Classification in Images of Road Pavement Flexible Surfaces*. Rijeka, Croatia: InTech, 2009.
[47] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[48] R. Genuer, J.-M. Poggi, and C. Tuleau-Malot, "Variable selection using random forests," *Pattern Recognit. Lett.*, vol. 31, no. 14, pp. 2225–2236, 2010.

Yifan Pan received the B.E. degree from the Shandong University of Science and Technology, Qingdao, China, in 2013, and the Ph.D. degree from Peking University, Beijing, China, in 2018.

He is currently an Engineer with the China Academy of Electronics and Information Technology, Beijing, China. His research focuses on the application of hyperspectral remote sensing and machine learning, especially exploring the application of deep learning in remote sensing data mining.

Xianfeng Zhang received the B.S. degree from Beijing Normal University, Beijing, China, in 1990, the M.Sc. degree from Peking University, Beijing, China, in 1993, and the Ph.D. degree from the University of Western Ontario, London, ON, Canada, in 2005.

He is currently an Associate Professor with the Institute of Remote Sensing and GIS, Peking University. His research interests include methods for information extraction from multisensor imagery and its applications in urban, wetland, and agricultural cropland environments.

Guido Cervone received the B.S. degree from the Catholic University of America, Washington, DC, USA, in 1998, and the master's and Ph.D. degrees in computer science from the George Mason University, Fairfax, VA, USA, in 2000 and 2005, respectively.

He is currently an Associate Professor in geoInformatics with the Department of Geography and Institute for CyberScience, Pennsylvania State University, State College, PA, USA. He is also an Associate Director of the Institute for CyberScience. His research sits at the intersection of geospatial science, atmospheric science, and computer science. His research focuses on the development and application of computational algorithms for the analysis of spatiotemporal remote sensing, numerical modeling, and social media Big Data related to environmental hazards and renewable energy.

Liping Yang received the Ph.D. degree in spatial information science and engineering with the School of Computing and Information Science, University of MaineOrono, ME, USA, in 2015.

She is currently a Postdoctoral Researcher with the GeoVISTA Center, Department of Geography and Institute for CyberScience, Pennsylvania State University, State College, PA, USA. Her research interests include geographic information science (GIScience) and especially applying visual analytics and machine/deep learning techniques for tackling interesting GIScience problems.