S & M 3117

Recognition Method for Earthquake-induced Building Damage from Unmanned-aerial-vehicle-based Images Using Bag of Words and Histogram Intersection Kernel Support Vector Machine

Ying Zhang,[1] Hong-Mei Guo,[1*] Wen-Gang Yin,[2]
Zhen Zhao,[1] Chang-Jiang Lu,[1] and Yang-Yang Yu[1]

[1]The Seismological Bureau of Sichuan Province,
No. 29, Section 3, Renmin South Road, Chengdu, Sichuan 610041, China
[2]Officers College of People's Armed Police,
No. 489, Section 1, Police School Road, Chengdu, Sichuan 610200, China

(Received July 29, 2022; accepted November 21, 2022; online published December 12, 2022)

Keywords: scale-invariant feature transform, bag of words model, eigenvector tag library, histogram intersection kernel support vector machine, unmanned-aerial-vehicle-based image, earthquake-induced building damage recognition

The commonly used artificial visual interpretation and existing object-oriented computer automatic recognition methods have some disadvantages, such as low efficiency and insufficient accuracy in recognizing earthquake-induced building damage in unmanned aerial vehicle (UAV)-based images. In this paper, we report the latest progress in research on machine learning algorithms in artificial intelligence, then propose a new method of recognizing earthquake-induced building damage. Using the bag of words (BoW) model, scale-invariant feature transformation (SIFT) characteristics were clustered to build an eigenvector tag library with K clustering centers as visual words. After images were expressed by visual words as eigenvectors with unified dimensions, a histogram intersection kernel (HIK) was then employed to construct the histogram intersection kernel support vector machine (HIK-SVM) to classify images and recognize earthquake-induced building damage. Building damage due to the magnitude 6.0 earthquake that occurred in Luxian, Sichuan Province on September 16, 2021 was analyzed as an example. When the proposed method was applied to recognize earthquake damage using UAV-based images, the average recognition accuracy reached 91.7%. The experimental results verified the feasibility and validity of the proposed method.

1. Introduction

China has frequent earthquakes and serious earthquake disasters. The risk of earthquake disasters has been further aggravated with the rapid economic and social development and a high concentration of population and wealth coupled with the accelerated development of energy resources and the continuous promotion of new urbanization. The results of investigations of the

[*]Corresponding author: e-mail: guohongmei@scdzj.gov.cn
https://doi.org/10.18494/SAM4060

ISSN 0914-4935 © MYU K.K.
https://myukk.org/

earthquake damage of previous destructive earthquakes at home and abroad show that casualties and economic losses are mostly caused by the destruction or collapse of various buildings and structures. It can be seen that information acquisition regarding building damage after an earthquake and swift disaster distribution assessment are critical for emergency rescue deployment.(1)

In a traditional method, on-site investigation requires a great amount of coordination of human and material resources, and the period of obtaining information is long and the efficiency is low. The investigation is also difficult to carry out because of natural environmental factors such as terrain or disaster factors such as traffic and communication interruptions. Consequently, information acquisition by remote sensing has attracted significant research interest and related attention. Among the various means of remote sensing, the unmanned aerial vehicle (UAV) remote sensing system has the advantage of a flexible, fast, and efficient operation. It also has a strong ability to recognize and distinguish damaged buildings using high-resolution images, which can reflect obviously different characteristics of the buildings that have different damage conditions. Therefore, the UAV has become an important means of conveniently obtaining disaster information such as the level of earthquake damage of buildings. To recognize the earthquake damage of buildings in remote sensing images, the commonly used manual visual interpretation method can accurately extract the earthquake damage information, but it is very time-consuming. The existing object-oriented and other automatic recognition methods can extract earthquake damage information quickly, but the accuracy is low. With these two methods, it is difficult to simultaneously accommodate the requirements of information timeliness and accuracy for handling earthquake disasters.

Recent machine learning developments and optimization allow new image classification techniques, such as random forest and the support vector machine (SVM), to be applied to information recognition. Remote-sensing-based seismic damage recognition for buildings is an image classification using damage levels based on various characteristics. UAVs are generally used in the most severely damaged areas to carry out earthquake damage surveys. However, the number of training samples on site is usually limited, and hyperspectral remote sensing classification has stringent requirements on the number of training samples.(2) Previous practical applications have shown that SVM classification can not only solve problems due to small samples sets, but also effectively solve problems of linear inseparability due to the large number of characteristic dimensions in hyperspectral remote sensing image classification.(3) Tuia *et al.* used a hierarchical tree to encode the output spatial structure and then added these relationships to the kernel function used to construct the SVM to design a structured output SVM that can be used in remote sensing image classification.(4) Patra and Bruzzone proposed an iterative active learning technology based on a self-organizing mapping neural network and an SVM classifier for remote sensing image classification.(5) Alimjan *et al.* proposed a new remote sensing image classification technology based on a combination of the SVM and K-nearest-neighbors algorithm, the separability of SVM classification, and the spatial and spectral characteristics of remote sensing images.(6) To improve the classification accuracy of remote sensing images using the SVM, Yu and Dong analyzed the impact of SVM parameters on the classification performance. They proposed an SVM parameter optimization method based on a dynamic coevolution algorithm with the characteristics of optimized particle swarm optimization and the genetic algorithm.(7) Alafandy *et al.* used a trained classical deep convolutional neural network to extract the characteristics of remote sensing images and employed them as the input of an SVM classifier to classify remote sensing images.(8) It can be seen that the SVM has been widely used in remote sensing image classification.

Therefore, we selected an SVM based on statistical learning theory to classify images and realize building damage recognition. To further improve the recognition speed and accuracy, building distribution data and UAV-based images were superimposed during preprocessing, and buildings were extracted from the images. Then, scale-invariant feature transformation (SIFT) was performed, and through the scale-space extremal detection and positioning of key points and the assignment of their directions, the image's SIFT descriptors were generated and SIFT characteristics extracted. However, different images have different numbers of SIFT characteristic points. Thus, the bag of words (BoW) model was used to cluster the SIFT characteristics of the images. Then, the eigenvector tag library of UAV-based images of earthquake-induced building damage was constructed using K-clustering centers as visual words. After the sample images were expressed as eigenvectors having unified dimensions of characteristics by visual words, a frequency histogram was used to calculate the frequency of each visual word appearing in the eigenvector vector. Subsequently, the histogram intersection kernel SVM (HIK-SVM) was constructed for earthquake-induced building damage recognition using UAV remote sensing. The damaged due to the Luxian earthquake (2021, magnitude 6.0) was used as an example case study. The results confirmed that the proposed method can rapidly and accurately recognize earthquake-induced building damage from UAV-based images.

2. Materials and Methods

2.1 Extracting SIFT characteristics from UAV images

2.1.1 Image preprocessing

UAV-based images acquired after an earthquake include various ground objects, some of which have similar shapes and textures to buildings, causing confusion in image feature extraction and building damage recognition. The first national comprehensive risk survey for natural disasters has recently been completed in China, and national key projects were implemented to reinforce buildings and facilities in earthquake prone areas to help prevent and control natural disasters. Therefore, we accessed the building distribution data and superimposed it over corresponding UAV-based images during preprocessing to accurately extract buildings from the images. Figure 1 shows the flowchart for the main preprocessing.

By this preprocessing, interference from similar ground objects is filtered and the information processing workload is reduced, effectively improving the building damage recognition efficiency and accuracy.

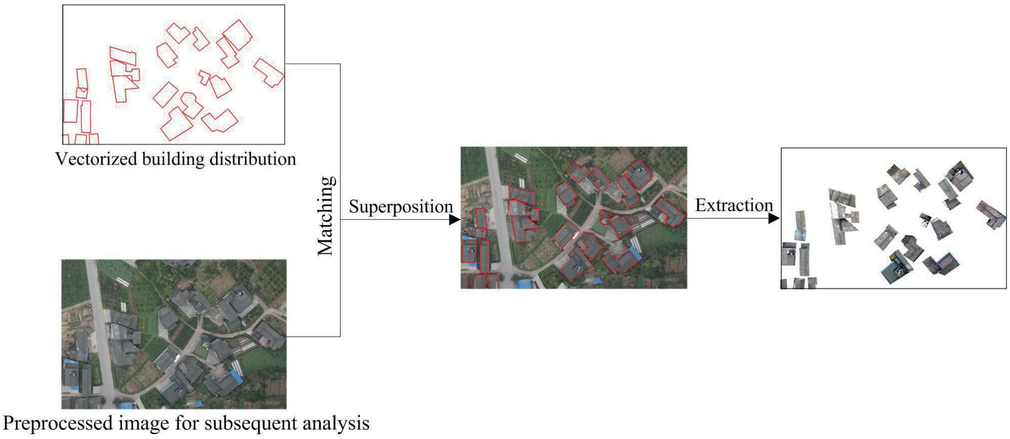

Vectorized building distribution
Matching
Superposition
Extraction
Preprocessed image for subsequent analysis

Fig. 1. (Color) Preprocessing of UAV-based images.

2.1.2 Extraction of SIFT characteristics

The recognition of earthquake-induced building damage from UAV-based images is mainly based on texture, edge, gray scale, and other basic characteristics to accurately describe an image. The image is classified after extraction to recognize earthquake damage, but it is often difficult for basic characteristics to accurately describe the image owing to differences in, for example, lighting conditions and the shooting method used, resulting in poor classification adaptability and low recognition accuracy.(9) The key to improving recognition accuracy lies in the selection of image characteristics. As much as possible, the selected characteristics should not be affected by lighting, shooting angle, or scale transformation and should have reasonable robustness to noise.

SIFT extraction can handle local characteristics well and hence satisfy the above requirements. Thus, we adopted SIFT to extract the image characteristics. Many UAV-based images were collected from buildings with recognized damage incurred during past earthquakes. Each image was preprocessed to extract any buildings, and the following four steps were applied to accurately describe image characteristics for buildings of different damage levels.

(1) Detecting extreme values in scale space

In the computer vision analysis of unknown scenes, the computer cannot predict the object's scale in the image. Only by considering the image multiscale description can the optimal scale of the target object be known. The scale space of the image is the image description at all scales.(10) Scale-space extremal detection is the process of detecting the candidate SIFT key points of the image. These key points are the basis of image description and they are the same at different scales. The Gaussian convolution kernel is the only linear kernel that can generate the scale space and simulate multiscale characteristics of image data.(11) Therefore, the SIFT method defines the scale space $L(x, y, \sigma)$ of the UAV-based image as the convolution of the original image $I(x, y)$ and the two-dimensional Gaussian filter function $G(x, y, \sigma)$,(12)

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y), \tag{1}$$

and

$$G(x,y,\sigma)=\frac{2}{2\pi\sigma^2}e^{\frac{-(x^2+y^2)}{2\sigma^2}}, \tag{2}$$

where (x, y) are spatial coordinates for pixels within the image and σ is the scale-space factor. A large scale corresponds to the general appearance of the image, whereas a small scale corresponds to image details, which determine image smoothness. To better detect extreme points in the scale space, we also define the Gaussian difference operator, DOG, as the difference between two adjacent scale spaces,(13) i.e.,

$$D(x,y,\sigma)=L(x,y,k\sigma)-L(x,y,\sigma), \tag{3}$$

where k is a constant multiple for two adjacent scale spaces. Figure 2 shows the Gaussian difference pyramid obtained by subtracting the images from every second adjacent layer in each pyramid group [Eq. (3)]. Each detection point is sequentially compared with adjacent points in the scale space to obtain extreme points for $D(x, y, \sigma)$.

The detection point is a possible SIFT key point. It is an extreme among its surrounding 26 adjacent points. The local extreme points form a set of candidate SIFT key points.

(2) Key point positioning

All candidate SIFT key points can be obtained from the UAV images through scale-space extreme value detection, which requires a two-step detection process. First, the key points must differ significantly from surrounding pixels. Therefore, key points with low contrast can be eliminated. Second, DOG has a strong edge response, and hence, unstable edge response points should be discarded. Thus, various extreme points can be eliminated by subpixel interpolation and edge response elimination depending on the key point location and scale to realize final optimal key point locations.(14) In this way, a limited number of characteristic points are obtained.

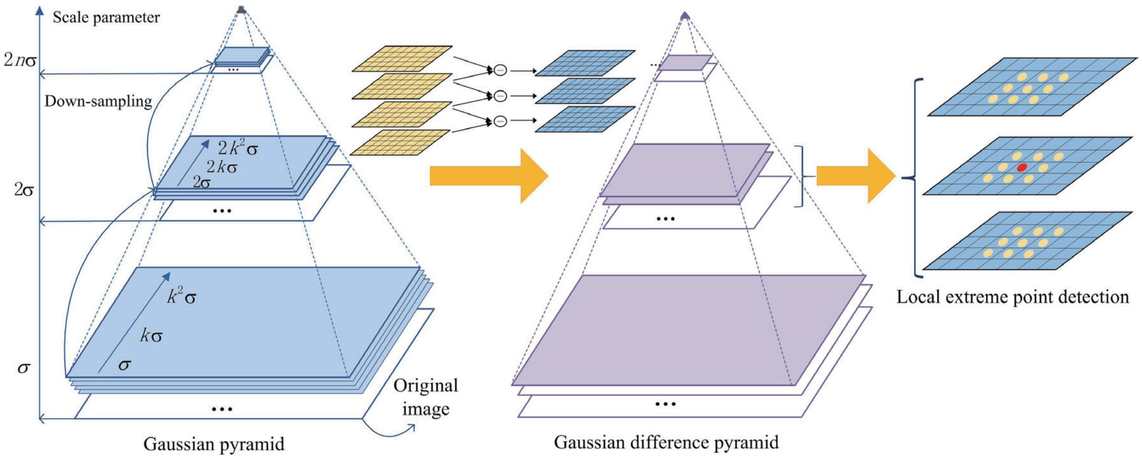


Fig. 2. (Color) Gaussian difference pyramid construction and local extreme point detection.

DOG from Eq. (3) can be expanded in a Taylor series as

$$D(\hat{x}) = D + \frac{1}{2}\frac{\partial D^T}{\partial x}\hat{x}\,. \tag{4}$$

Then, function values are calculated for extreme points, and subpixel interpolation is performed. We experimentally verified that removing extreme points with $D\left(\hat{x}\right) < 0.03$ can effectively eliminate unstable extreme points from low-contrast areas.

According to Harris' corner point theory,(15) pixel values in the local window change significantly when the key point at the edge moves in any direction. Such a key point has a large principal curvature, whereas stable key points do not. Therefore, edge responses can be eliminated on the basis of the magnitude of the principal curvature,

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix}, \frac{Tr(H)^2}{Det(H)} < \frac{(\gamma+1)^2}{\gamma}. \tag{5}$$

Thus, principal curvature γ can be calculated from the Hessian matrix H at key point positions, where D_{xx} represents the double derivative in the x direction for a key point scale. Experimental results confirmed that extreme point stability can be effectively enhanced to eliminate edge responses when $\gamma = 10$ and only the extreme points satisfying Eq. (5) are retained. Extreme points retained after the detection steps are identified as stable key points in the building target area.

There are three common types of building structures. The brick-and-concrete structure has vertical-load-bearing walls made of bricks or block masonry. The roof and the floor slab can be one of two types, namely, prefabricated boards or cast in place. There are usually three to six floors in the brick-and-concrete structure. The brick-and-wood structure also has load-bearing walls made of bricks or block masonry, while the floor slab and the roof frame are both constructed from wood. The roof cover is made of tile materials. As a flat structure, there are usually one to three floors in the brick-and-wood structure. The frame structure is built with reinforced concrete, the beams and columns are connected with steel bars, and the roof is cast in place. There are no more than 10 floors. Among these three types of structures, the plane and facade of the frame structure are the most regular.

In China's national standard "Classification of earthquake damage to buildings and special structures" (GB/T 24335-2009), the building damage in the macro-ground investigation is classified into five levels that range from light to heavy, with level 1 being "no damage" and level 5 being "collapsed". However, in remote sensing images, the building damage is generally divided into three levels, i.e., "not collapsed", "partially collapsed", and "collapsed". The "not collapsed" level corresponds to the "no damage" level of the ground survey, meaning that the outline of the building in the image is clearly visible and neatly arranged with uniform gray levels and complete imaging and geometric forms. Fine cracks can only be observed after several times of magnification. The "partially collapsed" level corresponds to the "damaged" level of the ground survey, meaning that the building's image contour can be recognized, but its corners fall off in the form of bright fragments. Additionally, the building's geometric form is no longer complete. The "collapsed" level corresponds to the "collapsed" level of the ground survey, meaning that the building's image contour has basically disappeared leaving a messy image texture and bright piled-up debris covering the entire building. It also does not have any geometric features.

The buildings in different damage conditions are classified in accordance with the above features and the associated damage levels using the UAV-based images. Figure 3 shows some typical images of brick-and-wood houses common to rural areas with marked key points extracted from the target area. The red circles represent the characteristics of roof corners. The yellow circles represent the characteristics of the roof surface. The green circles represent the characteristics of the roof and wall edges. The blue circles represent the characteristics of the wall.

These key points describe building condition differences well for different damage levels, and provide stable "vocabulary" information to help construct the eigenvector tag library for UAV-based images.

(3) Assignment of key point direction

A key point direction is assigned by considering extreme point detection, key point position, and scaling, as well as gradient direction distribution characteristics for key points on neighboring pixels. Key points are invariant to rotation, and their gradient magnitude $m(x, y)$ and direction $\theta(x, y)$ can be expressed as[16]

$$m(x,y)=\sqrt{(L(x+1,y)-L(x-1,y))^2+(L(x,y+1)-L(x,y-1))^2}, \tag{6}$$

and

$$\theta(x,y)=\arctan\frac{L(x,y+1)-L(x,y-1)}{L(x+1,y)-L(x-1,y)}, \tag{7}$$

respectively. The neighborhood and gradient histogram for key points can be obtained using the gradient amplitude and angle, as shown in Fig. 4.

Figure 4(a) shows typical key point neighborhood ranges, and Fig. 4(b) shows the gradient histogram with the horizontal axis representing amplitude angle and the vertical axis representing accumulated amplitude. The histogram is divided into eight directions at 45°

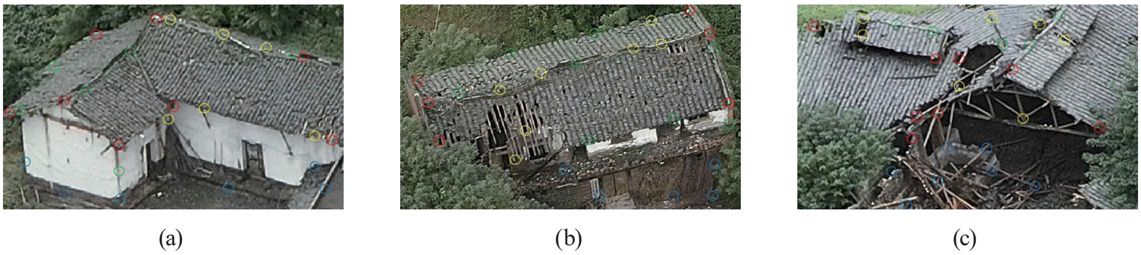
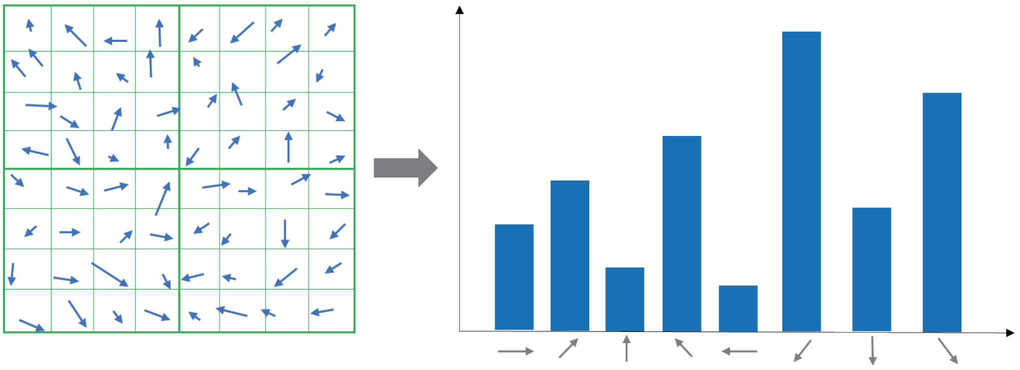

(a) (b) (c)

Fig. 3. (Color) Extracted key points for typical buildings in the target area. (a) Not collapsed. (b) Partially collapsed. (c) Collapsed.



(a) (b)

Fig. 4. (Color online) Typical neighborhood and gradient histogram for key points.

intervals, and the direction with the maximum accumulated amplitude is taken to be the principal key point direction.

(4) SIFT descriptor generation

Key points located by scale-space extreme point detection are actual pixel points that only contain the image spatial position information. Once each key point direction has been assigned, the point is further converted into an eigenvector to generate a SIFT descriptor, i.e., the mathematical expression for the key point. Thus, the pixel gradient and direction at the key point can be fully reflected in the UAV image, and an accurate description of the image can be realized. Figure 5 shows the process of generating the SIFT descriptor for each key point.

We select an 8 × 8 pixel block adjacent to the key point, divide the pixel block into 16 2 × 2 sub-blocks, and derive gradients for the four pixel points in each 2 × 2 sub-block using Gaussian weighting projected to eight directions. Each 2 × 2 sub-block is described as a 1 × 8 eigenvector; hence, if there are 16 sub-blocks in total, each key point is described by a 1 × 128 eigenvector.

2.2 Eigenvector tag library construction

Each UAV-based image extracted by the SIFT algorithm is a collection of many SIFT descriptors of the key points. Each key point is a multidimensional eigenvector describing partial building information, which is inconvenient for automatic recognition and judgment, and various key points are highly similar.

As a feature description method that is closer to the semantic expression of information, the BoW model was initially applied to natural language processing and information retrieval. It uses the frequency of keywords in the document to express the document's content. Csurka *et al.* introduced the BoW model into the study of image classification.(17) Their idea was to count the distribution information of different local features of each image block by extracting the different local features of the image to correlate the image block to the words in the text, eventually obtaining a bag of visual words as the image model. The BoW can be regarded as the aggregation and integration of the low-level features of the image. As new and more stable extraction algorithms have been proposed to obtain such low-level features, such as the SIFT, the

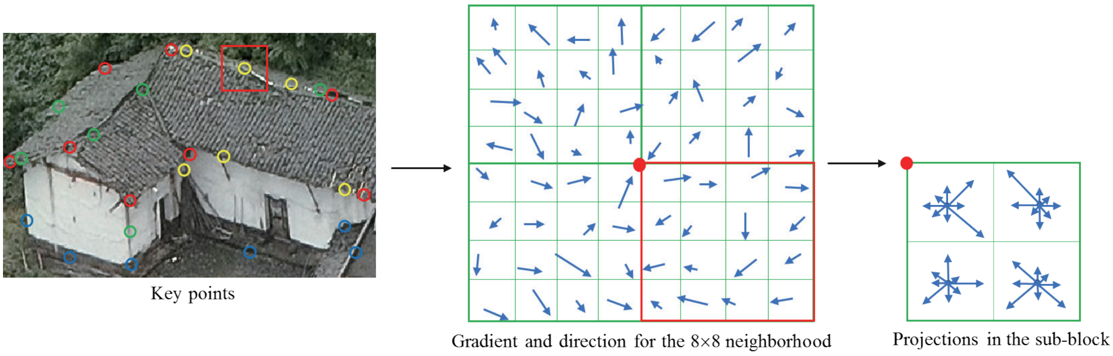

Key points
Gradient and direction for the 8×8 neighborhood
Projections in the sub-block

Fig. 5. (Color) Generating the SIFT descriptor for the key point.

BoW was introduced into the automatic recognition of computer information in the field of machine vision for the classification of remote sensing satellite images.(18) It has achieved good application results. The bag of visual words model uses K-means clustering to gather all SIFT characteristic points into K clusters. It then classifies key point eigenvectors such that internal elements in the cluster are highly similar while ensuring low similarity among different clusters. Therefore, K cluster centers are treated as visual words, i.e., eigenvector tags, and the eigenvector tag library can be constructed using K visual words. After the sample images were expressed as eigenvectors having unified dimensions of characteristics by visual words, a frequency histogram was used to calculate the frequency of each visual word appearing in the eigenvector. By using these classified visual words as tags to describe the UAV-based images of the earthquake-induced building damage, the differences in key points between buildings of different damage levels can be highlighted. Additionally, the image eigenvector's dimension can be reduced to effectively improve the operation efficiency of automatic computer recognition. Figure 6 shows the construction process of the eigenvector tag library.

First, K points were selected as the initial clustering centers because the self-organizing incremental learning neural network (SOINN) can automatically discover the number of appropriate categories in the clustering application.(19) We utilized this feature to conduct incremental learning on the low-level features of the image extracted by SIFT. In this way, the problem of repeatedly adjusting the initial clustering center K, as in the experiment where the traditional bag of visual words method directly uses K-means for clustering, is avoided.

SOINN is a competitive neural network with a two-layered structure that excludes the input layer. It uses a self-organizing method to conduct clustering and topological representation for the input data.(20) The first layer is used to accept the input of the original SIFT characteristic points and adaptively generate prototype neurons to represent the input data. The distribution of the SIFT characteristic points is reflected by the nodes and their connections. The second layer estimates the distances between and within the original SIFT characteristic point classes in accordance with the results of the network's first layer. Then, the distances are taken as parameters that are to be combined with the neurons generated by the first layer as input data for relearning to output stable results and serve as the K-means initial clustering center.

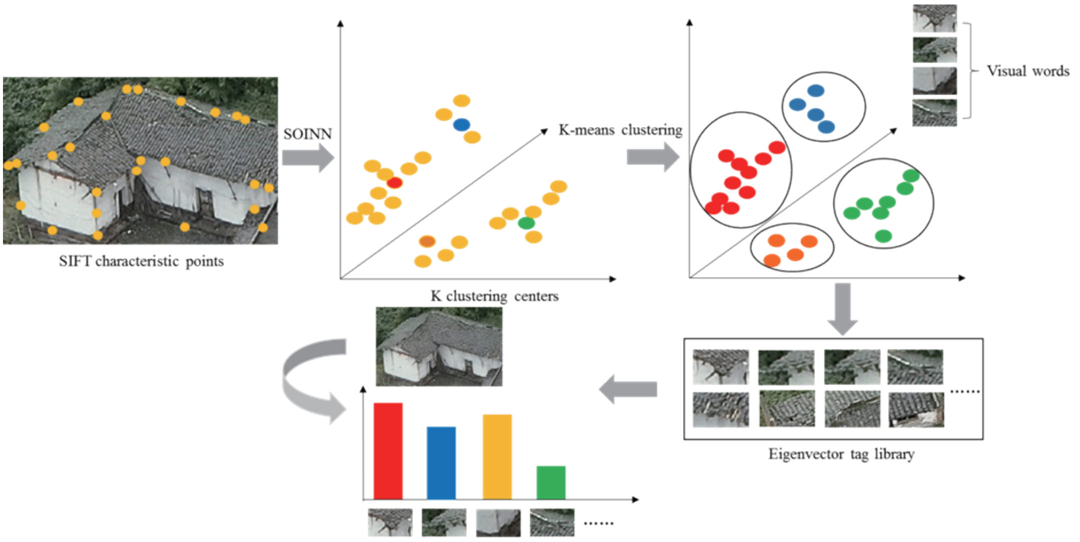


Fig. 6. (Color) Construction process of the eigenvector tag library.

We then calculated the Euclidean distance between eigenvector *X* for each characteristic point and the *i*th clustering center using Eq. (8). The cluster center closest to the characteristic point is then identified and put into the corresponding cluster as

$$D_i = \sqrt{\sum_{j=1}^{128}\left(x_j - k_{ij}\right)^2}, \tag{8}$$

where x_j is the *j*th dimension for vector *X* and k_{ij} is the *j*th dimension for the *i*th cluster center.

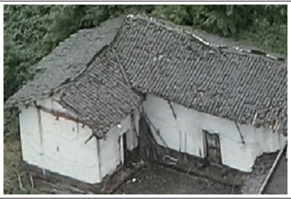
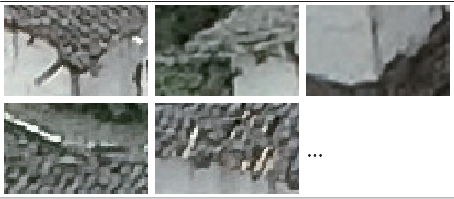
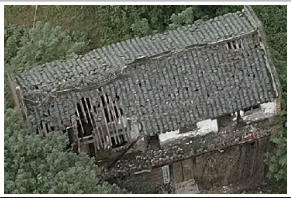
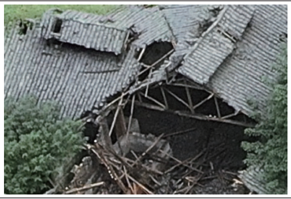
The centroid for each cluster is recalculated after all characteristic points are classified into *K* clusters and taken to be the new cluster center. If the distance between the new and original cluster centers is less than a preset threshold, clustering has achieved the expected effect and calculation is terminated; otherwise, steps (2) and (3) are iterated if the distance exceeds the preset threshold.

The eigenvector tag library is finally obtained after K-means clustering. The library comprises *K* visual words expressed as $L = (l_1, l_2, \ldots, l_k)$, where the visual word l_i is the eigenvector tag.

Key points with similar descriptions are combined into a class that shares similar characteristics through clustering. Eigenvectors describing the key points are determined from pixel points in the neighborhood of the key points. Visual words formed by the clustering key point eigenvectors can be regarded as morphological damaged building features at different damage levels. Table 1 shows several examples.

Thus, each UAV-based image *P* can be represented by a set of visual words $(f_1, f_2, \ldots, f_k)$ formed by clustering and is a numerical vector of 1 × *K* dimensions in the eigenvector tag library, where f_i is the frequency of visual word l_i when describing a specific image.

Table 1
(Color online) Visual word examples.

Damage level		Visual words
Not collapsed		 ...
Partially collapsed		...
Collapsed		...

2.3 SVM classifier design

After the eigenvectors are obtained by constructing the eigenvector tag library, a classifier is required to classify the image. SVM is a supervised machine learning model that has been shown to be very effective for practical applications of solving linear inseparability problems caused by large characteristic dimensions for images with small training sample sets and hyperspectral remote sensing features. SVM has been widely used in image classification and recognition.

Figure 7 shows the basic SVM principle of finding a classification hyperplane that maximizes the minimum classification interval such that the points in the training sample can be correctly divided into two categories.(21)

The sole classification hyperplane that can maximize the minimum classification interval can be expressed as

$$\omega^T x + b = 0, \tag{9}$$

where ω is the normal vector for the hyperplane and b is the intercept. The point closest to the hyperplane is the support vector. For a training dataset $T = \{(x_1, y_1), \ldots, (x_n, y_n)\}$, $y_i \in \{-1, +1\}$, $i = 1, 2, \ldots, n$, x_i is an n-dimensional eigenvector represented by a visual word in the tag library and y_i is the eigenvector classification tag. The geometric distance on the hyperplane is $1/\|\omega\|$. The search for the hyperplane can be converted to the optimization of convex quadratic programming with inequality constraints:

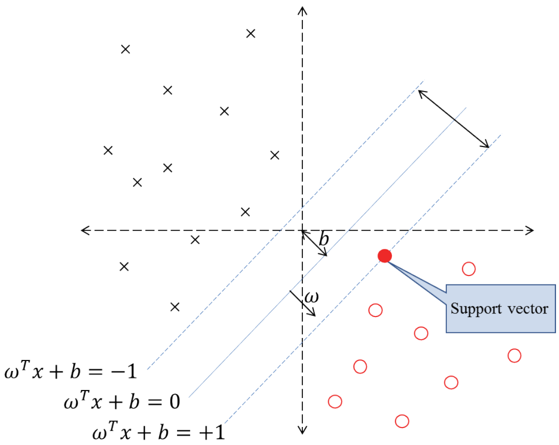


Fig. 7. (Color online) SVM classification principle.

$$\begin{cases} \min_{\omega,b} \dfrac{1}{2}\|\omega\|^2 \\ \text{s. t. } y_i(\omega^T x_i + b) \geq 1, i = 1,2,\ldots,n. \end{cases} \tag{10}$$

Optimal solutions ω* and b* can be obtained by dual optimization by the Lagrange multiplier method. For linear inseparability in a practical application problem, SVM maps a low-dimensional sample space to a high-dimensional characteristic space through a kernel function to obtain the hyperplane in the high-dimensional characteristic space.(22) Then, a discrimination function $f(x)$ is constructed, where attributes for each sample are determined using the discrimination function,

$$f(x) = \text{sgn}\left(\sum_{i=1}^{l} a_i^* y_i K(x, x_i) + b^*\right), \tag{11}$$

where a_i^* is the Lagrangian optimal solution and $K(x, x_i)$ is the kernel function. Different kernel functions correspond to different discrimination functions, forming different SVM classifiers.

The histogram intersection kernel (HIK) uses statistical histograms of the image characteristics to determine whether two images belong in the same category. Moreover, the HIK has high robustness and low computational complexity with respect to target background interference, viewing angle changes, blockage, and image resolution changes.(23) It is also efficient when converting low-dimensional-space linearly inseparable problems of the SVM to high-dimensional-space linearly separable problems and has a better recognition performance than linear and radial basis function (RBF) kernels in image classification, especially for images expressed by histogram characteristics.(24) Therefore, we adopted the HIK to construct the classification decision-making function. The HIK is defined as

$$K_{HI}(x, y) = \sum_{j=1}^{d} \min(x_j, y_j), \tag{12}$$

where x and y are a pair of histograms. Each histogram contains d stripes. In the eigenvector tag library, x and y represent the frequency histograms of the image, and each histogram is composed of d visual words. Parameters x_j and y_j are the values in each category of frequency histograms x and y. By substituting Eq. (12) into Eq. (11), the following classification decision-making function is obtained:

$$f(x) = \operatorname{sgn}\left(\sum_{i=1}^{l} a_i^* y_i \left[\sum_{j=1}^{d} m(x_j, x_{ij}) \right] + b^* \right). \tag{13}$$

In this way, a HIK-SVM classifier, which can be used for the UAV remote sensing recognition of earthquake-induced building damage, is formed.

3. Experiment and Results

3.1 Experimental data

After the magnitude 6.0 earthquake occurred in Luxian (29.20°N,105.34°E), a Dajiang M300 RTK UAV equipped with a PSDK 102S five-lens tilting camera was launched to take aerial photographs of the severely damaged Datian Community of Fuji Town (29.21°N,105.37°E) and Tuanshanbao Village of Jiaming Town (29.23°N,105.32°E) near the epicenter, as shown in Fig. 8.

The main specifications of the UAV and the flight parameters, which were set in accordance with the requirements of Luxian's terrain and image resolution, are presented in Table 2. The UAV images and digital surface model (DSM) data collected by aerial photography are displayed in Fig. 9.

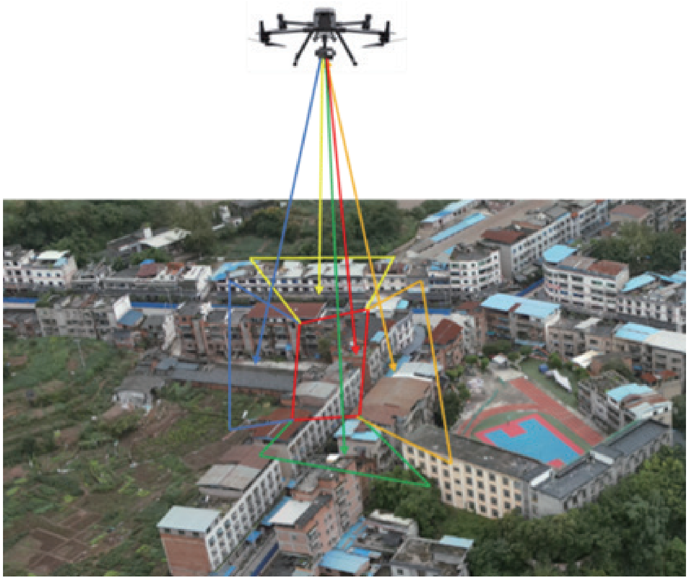
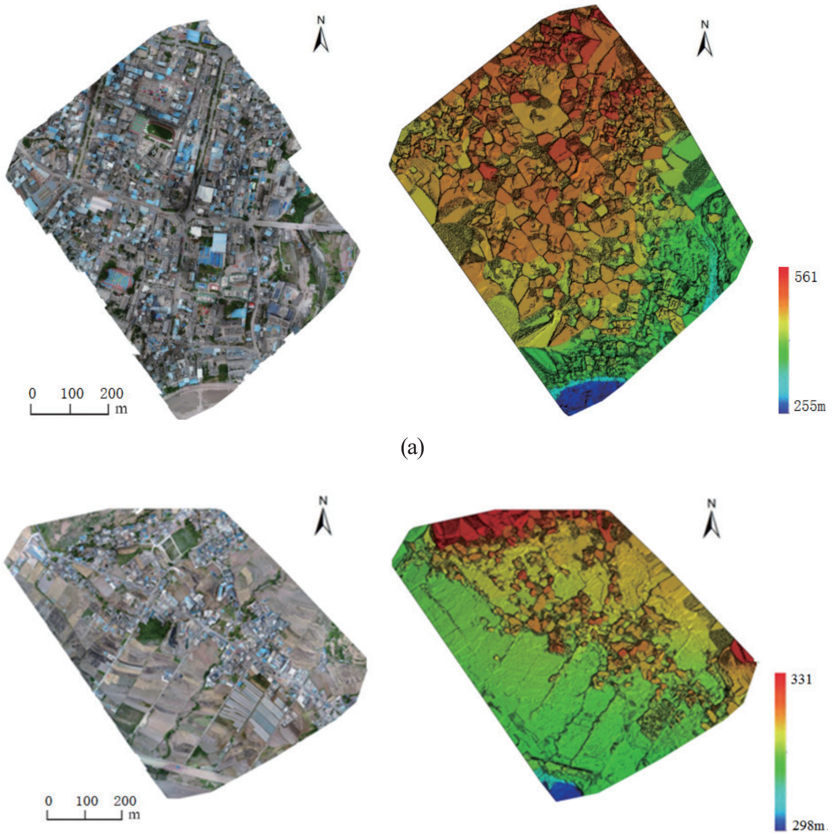


Fig. 8. (Color) Schematic diagram of UAV aerial photography at the earthquake site.

Table 2
Main specifications of the UAV.

Technical parameter	Specification	Flight parameter	Specification
Unfolded size	810 × 670 × 430 mm^3	Number of flights	4
Empty vehicle mass (including two batteries)	6.3 kg	Flight altitude	280 m
Battery capacity	4920 m·Ah	Heading overlap ratio	70%
Camera	24.3 million pixels	Side overlap ratio	70%
Maximum horizontal flight speed	23 m/s	Ground average resolution	>3 cm
Maximum rising and falling speeds	6 m/s, 5 m/s	Imaging area	1.5 km^2
Maximum flight time	55 min	Total number of images	587 sheets
Maximum flight altitude	5000–7000 m	Image format	TIFF or JPEG



(a)

(b)

Fig. 9. (Color) UAV images and DSM diagrams. (a) Datian Community of Fuji Town. (b) Tuanshanbao Village of Jiaming Town.

3.2 Experimental process

To verify the effectiveness of the method proposed in this paper, the images obtained from areas affected by the Luxian M6.0 earthquake were further processed and analyzed. The main process is shown in Fig. 10.

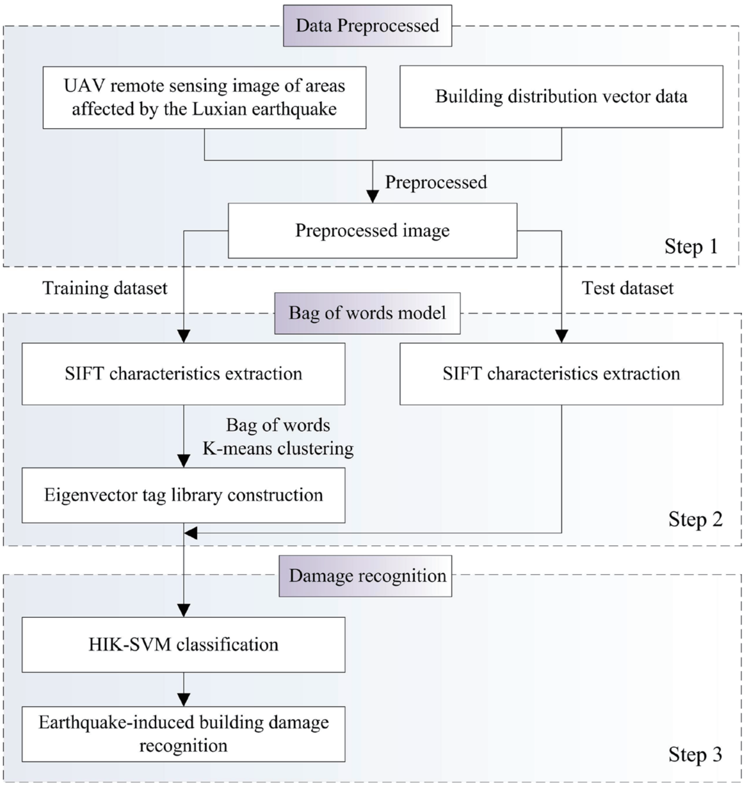

Data Preprocessed
UAV remote sensing image of areas affected by the Luxian earthquake
Building distribution vector data
Preprocessed
Preprocessed image
Step 1
Training dataset
Test dataset
Bag of words model
SIFT characteristics extraction
SIFT characteristics extraction
Bag of words
K-means clustering
Eigenvector tag library construction
Step 2
Damage recognition
HIK-SVM classification
Earthquake-induced building damage recognition
Step 3

Fig. 10. (Color online) Flowchart of experiment.

First, the UAV-based image and building distribution vector data were superimposed, and building images were extracted and preprocessed, producing 2014 building images with three damage levels. We randomly selected 1400 images of the three damage levels to build the training set to construct the eigenvector tag library and train the HIK-SVM. The remaining 614 images were used as test samples to validate the accuracy of the algorithm for classification recognition.

Then, a Python algorithm program was developed in the Spyder environment of Anaconda3. The SIFT function was called through the SIFT characteristic extraction interface in the Python 3.7 OpenCV extension module to detect the characteristic points of 2014 images. The positions of the characteristic points were marked and the SIFT descriptors were generated. After the SIFT characteristics were extracted from the images, SOINN was first created through the third-party library MiniSom for the incremental learning of the low-level features extracted from the SIFT image. Then, the K-means function was called to conduct a clustering analysis on the SIFT characteristic points of the 1400 images in the training set through the K-means interface in the OpenCV extension module. The K clustering centers were taken as visual words to obtain the eigenvector tag library composed of K visual words. We calculated the distance between each characteristic point and visual word in the eigenvector tag library for the 2014 images. The characteristic point was replaced by the nearest visual word and the frequency histogram was used to count the number of visual words in the feature vector; then, the image was expressed as a $1 \times K$-dimensional numerical vector.

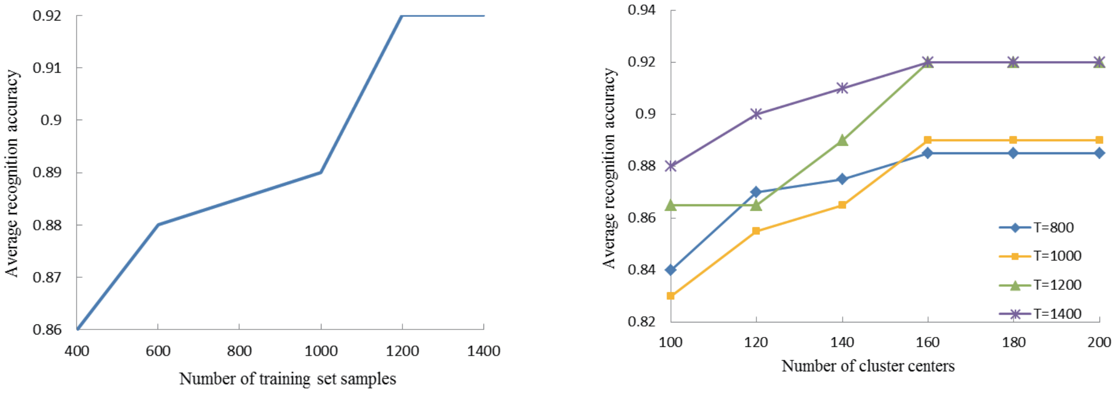
Finally, the Python 3.7 data analysis tool sklearn was employed to construct a HIK-SVM, and the images expressed by the multidimensional eigenvectors in the training set were used as samples to train it. Images in the test set were input to the well-trained HIK-SVM to recognize building damage. The recognition accuracy of the classification algorithm was then tested.

On the basis of practical experience, when the bag of visual words was used to build the eigenvector tag library in this process, the number of K-means clustering centers (K) was determined from the set threshold and the number of training samples (T) used to train the HIK-SVM affected the recognition results. Therefore, experiments were conducted using different parameter settings, and then, the optimal average recognition accuracy of the buildings at three damage levels was adopted as the criterion for selecting the parameter settings.

3.3 Experimental results

For the training set, the values of T were 400, 600, 800, 1000, 1200, and 1400 images, and K was determined to be 160 from the set threshold. The variation curve of the average recognition accuracy of building damage obtained by HIK-SVM training for different numbers of samples is presented in Fig. 11(a). To further test the effects of the numbers of samples and cluster centers on the algorithm, different thresholds were set to obtain the number of cluster centers with different numbers of samples for K values of 100, 120, 140, 160, 180, and 200. The variation curve of the average recognition accuracy of building damage obtained by HIK-SVM training is presented in Fig. 11(b).

Figure 11(a) indicates that the average recognition accuracy increased with T in the training set. At T = 1200, the average recognition accuracy became stable and no further changes occurred. Figure 11(b) shows that the average recognition accuracy also increased with the number of cluster centers. This is because a larger K value results in more visual words for describing the images in the eigenvector tag library. Although a more detailed description can lead to a higher recognition performance, the average recognition accuracy was stable when



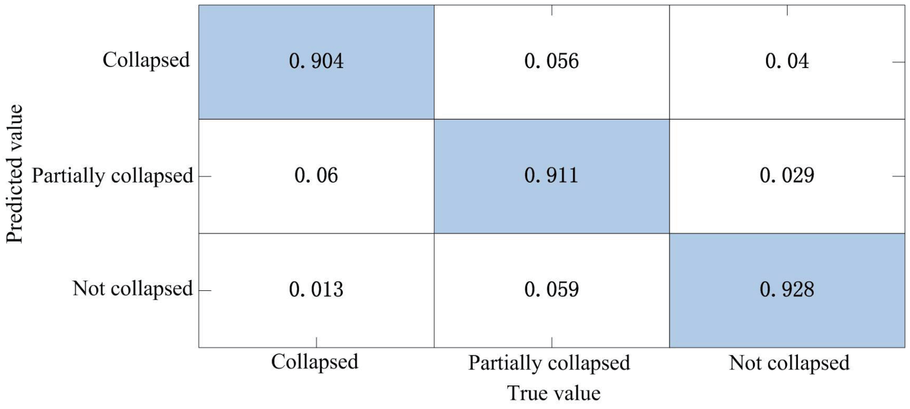
(a) (b)

Fig. 11. (Color) Average recognition accuracy. (a) Number of training set samples. (b) Number of cluster centers.

$K \geq 160$. For T values of 1200 and 1400, the curves overlapped and the average recognition accuracy reached a maximum. However, the larger the K value, the more complex the calculation. Therefore, $K = 160$ and $T = 1400$ were selected as the best parameter settings.

The images expressed by the multidimensional eigenvectors in the test set were input to the HIK-SVM, which was trained using the optimal parameter settings, to perform building damage recognition. The confusion matrix of the recognition results is displayed in Fig. 12. It indicates that errors mainly occurred when recognizing buildings that were either partially collapsed or collapsed.

The recognition accuracies of the HIK-SVM are presented in Table 3. These accuracies were compared with those of the commonly used pixel-based and object-oriented classification methods in automatic recognition methods, and the results are shown in Table 4. The pixel-based method cannot fully utilize the texture and structure information of remote sensing images, which leads to a low recognition accuracy. The object-oriented method can fully utilize the texture and structure information of the images; hence, it can better identify ground objects with



	Collapsed	Partially collapsed	Not collapsed
Collapsed	0. 904	0. 056	0. 04
Partially collapsed	0. 06	0. 911	0. 029
Not collapsed	0. 013	0. 059	0. 928

Fig. 12. (Color online) Confusion matrix of the recognition results.

Table 3
Recognition accuracies of three damage levels

Building damage level	Number of sample buildings	Number of correctly recognized buildings	Recognition accuracy (%)
Not collapsed	251	233	92.8
Partially collapsed	269	245	91.1
Collapsed	94	85	90.4

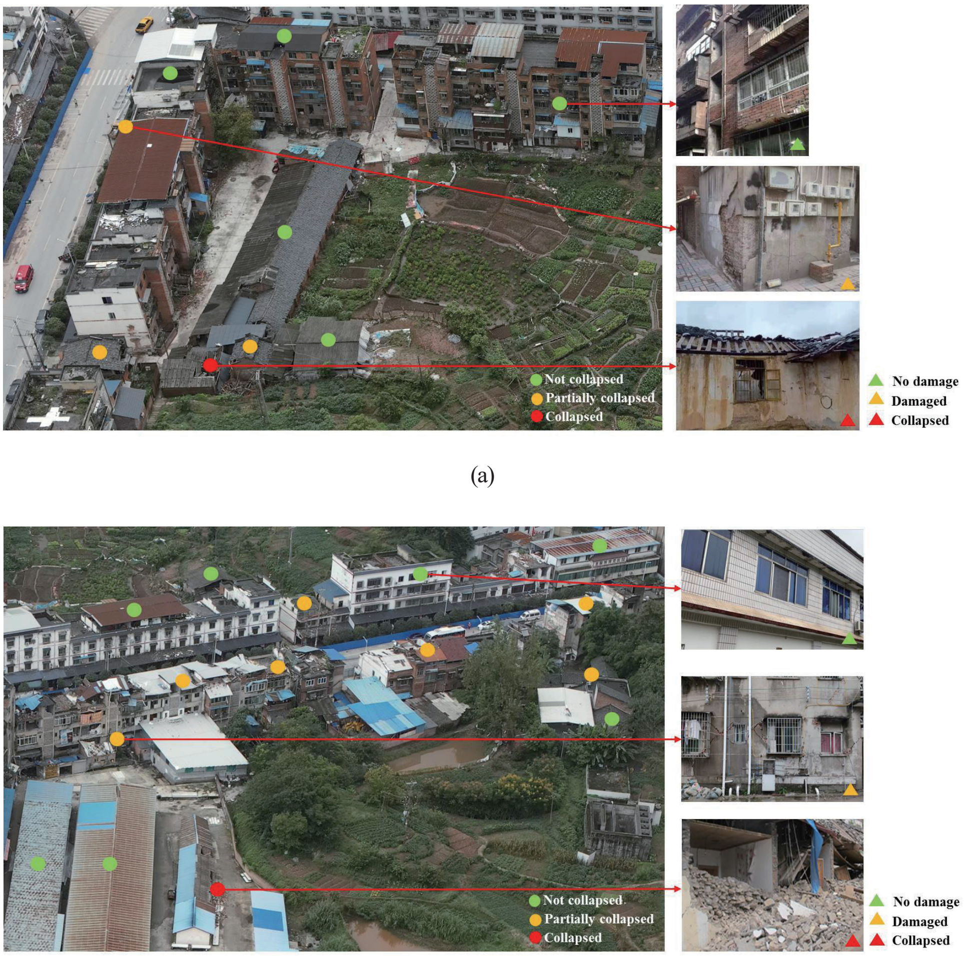
Table 4
Recognition accuracies of three classification methods.

Building damage level	Recognition accuracy		
	Pixel-based method (%)	Object-oriented method (%)	HIK-SVM (%)
Not collapsed	80.8	97.2	92.8
Partially collapsed	84.8	85.1	91.1
Collapsed	85.1	82.9	90.4
Total accuracy	83.2	89.7	91.7

regular shapes, and its recognition accuracy for not collapsed buildings is the highest. However, in complex scenes such as those with collapsed buildings, it is difficult to accurately segment the ground objects, which affects the classification accuracy. In this study, after the image was expressed as a feature vector of unified dimensions by using visual words, the HIK-SVM was used for image classification to ensure that all types of complex scenes in the image can be better processed, and its recognition accuracy was improved compared with those of the previous methods.

To further analyze the causes of recognition errors, some earthquake-induced building damage recognition results for Datian Community and Tuanshanbao Village in the sample were compared with the actual situation, as shown in Fig. 13.

The actual damage of the building can be clearly seen in the ground photographs of the buildings taken in the field investigation. Table 5 shows examples of recognition error.

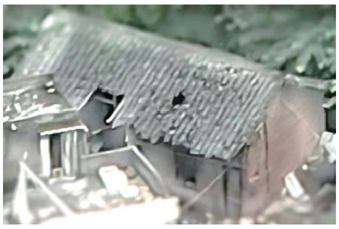
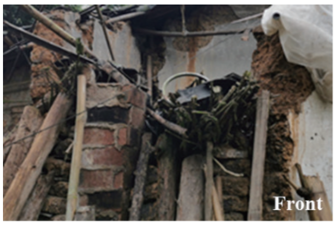
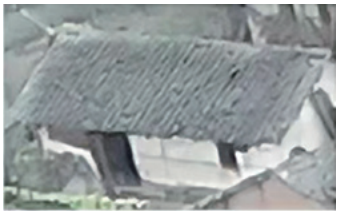
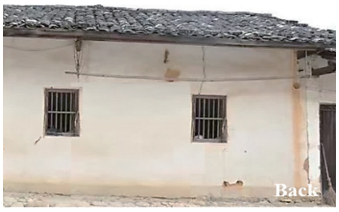
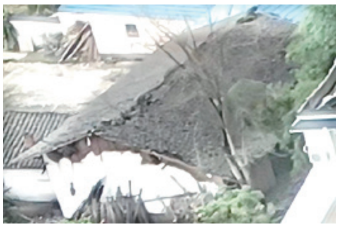
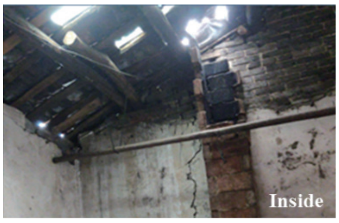


(a)

(b)

Fig. 13. (Color) Damage recognition results of some buildings shown with field investigation results. (a) Datian Community. (b) Tuanshanbao Village.

Table 5
(Color online) Examples of recognition error.

UAV-based image	Damage level in image recognition	Field investigation photo	Actual damage level
	Not collapsed		Collapsed
	Partially collapsed		No damage
	Collapsed		Damaged

Compared with the accurately identified images in Fig. 13, it can be observed in the incorrectly identified images of Table 5 that when there was interference by light illuminating the characteristic area of building damage classification or the external area, the effective SIFT key points extracted from the image were changed. This caused significant interference in recognizing building damage; this was particularly obvious when the damaged and destroyed buildings had high fragment brightness in the image.

4. Discussion

As seen in the experimental process, the proposed method only employs SIFT characteristics for classification, and hence, dimensionality remains large. This affects computing resource requirements for SVM image classification and recognition. Moreover, in the eigenvector tag library of UAV-based images constructed using the bag of visual words for earthquake-induced building damage, the scale of the visual word is relatively simple when the order and relationship between the words are not considered. This problem affects the expression of some image features, causing changes in the effective SIFT key points extracted from the images. Therefore, classification and recognition errors arise when external interference exists.

5. Conclusions

In this paper, we proposed a recognition method for earthquake-induced building damage from UAV-based images using the BoW model and the HIK-SVM. After the building image range was determined in advance using existing building distribution data, the selected SIFT characteristics exhibited a strong antinoise capability. The eigenvector tag library was created using the BoW model to express the SIFT characteristic point sets of the image with visual words as eigenvectors with unified dimensions. Hence, the HIK-SVM can be used to perform image classification and recognition. The building damage due to the Luxian earthquake (2021, magnitude 6.0) was used as a sample case study to validate the proposed method. The experimental results verified that the proposed method is feasible and accurate for UAV-based image recognition.

In future research, the SIFT features can be combined with the features that have less dimensions or the SIFT features can be processed to reduce their number of dimensions to further improve the speed of image recognition and better meet the demand for the real-time handling of earthquake disasters. Moreover, we are considering establishing visual words at different scales, forming more robust image feature expression schemes such as multiscale eigenvector tag libraries, reducing the effects of external factors, such as the imaging environment, on image recognition, and improving the adaptability of the method.

Acknowledgments

The authors are grateful for the financial support received from the National Key Research and Development Program of China (No. 2020YFA0710603), the National Natural Science Foundation of China (No. 42061073), and the Sichuan Key Research and Development Plan (No. 2020YFS0451). The authors are also grateful to the anonymous reviewers for their helpful comments.

References

1 W. H. Fan, W. K. Wang, C. K. Liang, M. L. Yang, W. L. Hsu, and Y. C. Shiau: Sens. Mater. **33** (2021) 1231. https://doi.org/10.18494/SAM.2021.3160

2 K. G. Nikolakopoulos and I. Koukouvelas: Proc. 2021 Society of Photo-Optical Instrumentation Engineers Conf. Series (SPIE, 2021) 14. https://ui.adsabs.harvard.edu/abs/2021SPIE11863E.04N/abstract

3 Y. Chen, X. Zhao, and Z. Lin: IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **7** (2014) 1295. https://doi.org/10.1109/JSTARS.2014.2307356

4 D. Tuia, J. Muñoz-Marí, M. Kanevski, and G. Camps-Valls: J. VLSI Sig. Process. Syst. **65** (2011) 301. https://doi.org/10.1007/s11265-010-0483-8

5 S. Patra and L. Bruzzone: IEEE Trans. Geosci. Remote Sens. **52** (2014) 6899. https://doi.org/10.1109/TGRS.2014.2305516

6 G. Alimjan, T. Sun, Y. Liang, H. Jumahun, and Y. Guan: Int. J. Pattern Recognit Artif Intell. **32** (2018) 1859012.1. https://www.worldscientific.com/doi/abs/10.1142/S0218001418590127

7 X. Yu and H. Dong: J. Intell. Fuzzy Syst. **35** (2018) 1. https://doi.org/10.3233/JIFS-169593

8 K. A. Alafandy, H. Hicham, M. Lazaar, and M. A. Achhab: Advan. Sci. Tec. Eng Syst. **5** (2020) 5. https://doi.org/10.25046/AJ050580

9 W. Zhai, C. Huang, and W. Pei: Remote Sens. **11** (2019) 8. https://doi.org/10.3390/rs11080897

10 R. Hess: Proc. 18th ACM Int. Conf. Multimedia (ACM, 2010) 1493. https://doi.org/10.1145/1873951.1874256

11 T. Lindeberg: J. Appl. Stat. **21** (1994) 225. https://doi.org/10.1080/757582976

12 Y. Bazi and F. Melgani: IEEE Trans. Geosci. Remote Sens. **48** (2010) 186. https://ieeexplore.ieee.org/document/5204216

13 S. Bukhari and S. Iqbal: Proc. 2014 12th Int. Conf. Frontiers of Information Technology (IEEE, 2014) 302. https://ieeexplore.ieee.org/document/7118424

14 Z. Hossein-Nejad, H. Agahi, and A. Mahmoodzadeh: Pat. Anal App. **24** (2021) 669. https://link.springer.com/article/10.1007/s10044-020-00938-w

15 J. B. Ryu, C. G. Lee, and H. H. Park: Electron. Lett. **47** (2011) 180. https://www.researchgate.net/publication/260615841_Formula_for_Harris_corner_detector

16 M. Lourenco: IEEE Trans. Rob. **28** (2012) 752. https://ieeexplore.ieee.org/document/6151178

17 G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray: Proc. Workshop on Statistical Learning in Computer Vision (ECCV 2004). https://www.researchgate.net/publication/228602850

18 A. Radoi and M. Datcu: Proc. 2018 IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS 2018) 111. https://ieeexplore.ieee.org/document/8519432

19 S. Kawai and S. Yamaguchi: IEEJ Trans. Electro Info. Syst. **136** (2016) 945. https://doi.org/10.1541/ieejeiss.136.945

20 F. Shen and O. Hasegawa: Neural Networks **21** (2008) 1537. https://doi.org/10.1016/j.neunet.2008.07.001

21 Q. Li and X. Wang: Proc. ACIS 17th Int. Conf. Computer and Information Science (ICIS 2018) 691. https://ieeexplore.ieee.org/document/8466432

22 L. Cheng and W. Bao: Telk. Ind. J. Elec. Eng. **12** (2013) 1037. https://doi.org/10.11591/telkomnika.v12i2.4325

23 P. Li, Y. Liu, G. Liu, M. Guo, and Z. Pan: Neurocomputing **184** (2016) 36. https://doi.org/10.1016/j.neucom.2015.07.136

24 J. Wu: IEEE Trans. Image Process. **21** (2012) 4442. https://doi.org/10.1109/TIP.2012.2207392