

# AUTOMATIC BENGGANG RECOGNITION BASED ON LATENT SEMANTIC FUSION OF UHR DOM AND DSM FEATURES

Shengyu Shen <sup>1,2\*</sup>, Tong Zhang <sup>3</sup>, Yuanling Zhao <sup>1,2</sup>, Zhigang Wang <sup>1,2</sup>, Feng Qian <sup>1,2</sup>

<sup>1</sup> Department of Soil and Water Conservation, Changjiang River Scientific Research Institute (CRSRI), Wuhan, China - shenshengyu@mail.crsri.cn, yuanlingzhao@vip.qq.com, 371381624@qq.com, 402483315@qq.com

<sup>2</sup> Research Center on Mountain Torrent & Geologic Disaster Prevention of the Ministry of Water Resources, Wuhan, China

<sup>3</sup> State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, Wuhan, China - zhangt@whu.edu.cn

## TCIII, WG 7

**KEY WORDS:** Benggang, DOM, DSM, BoV-TW, LDA, Automatic Recognition

### ABSTRACT:

Benggang are characterized by deep-cut slopes with various shapes and depressions on the vast weathered crust slopes in southern China. The gully heads have been continuously collapsed and eroded to form a chair-like erosion landforms. It develops rapidly, and leads to large amounts of erosion, with the hazards of damaging land resources, destroying basic farmland, and deteriorating ecological environment. To study and manage Benggang, the primary task is to discover it. Traditional methods based on local in-situ investigations, which are not only labour-consuming but also inefficient. These methods are difficult to meet the needs of large-scale investigations of Benggang. This paper proposes a method for automatic Benggang recognition based on Ultra-High Resolution (UHR) DOM (Digital Orthophoto Map) and DSM (Digital Surface Model) obtained from UAV (Unmanned Aerial Vehicle) survey. This method adopts a Bag of Visual-Topographical Words (BoV-TW) model. The local features extracted from DOM and DSM are represented based on BoV-TW, and fused by Latent Dirichlet Allocation (LDA). Finally Support Vector Machine (SVM) is adopted as a supervised classifier to achieve high-precision automatic Benggang recognition. Experimental results prove that the total accuracy of our method can be maintained at about 95%, with recall and precision above 80% (the highest are 97.22% and 94.44%, respectively), which are significantly higher than the methods of using only DOM local features and using only BoV-TW.

## 1. INTRODUCTION

Benggang are characterized by deep-cut slopes with various shapes and depressions on the vast weathered crust slopes in southern China. The gully heads have been continuously collapsed and eroded to form a chair-like erosion landforms (Zeng, 1960), as shown in Figure 1. It develops rapidly, and leads to large amount of erosion (Liang, 2009), with the hazards of damaging land resources, destroying farmland, and deteriorating ecological environment, which directly threatens the safety of land, food, and ecology.



Figure 1. Typical Benggang in southern China

The discovery of Benggang is the primary task for large-scale investigation, control and analysis of erosion mechanism of Benggang. Traditional methods based on local in-situ investigations and manual interpretation of high-resolution satellite remote sensing images have low levels of automation. These methods are not only labour-consuming but also inefficient and difficult to meet the needs of a large-scale study. For example, in December 2005, the Ministry of Water Resources organized the Yangtze River Water Conservancy Commission's Soil and Water Conservation Bureau, the Zhuhai Committee, and the Water Conservation Division of the Taihu Basin Management Bureau to accomplish the investigation on status of Benggang in southern China after 16 months (Feng, 2009).

With the rapid development of remote sensing and photogrammetry technology, especially the automation of UAV survey, Ultra-High Resolution (centimetre or decimetre) DOM and DSM are increasingly easy to obtain.

Efficient and automatic Benggang recognition using these data will greatly save cost, and it is very important for large-scale Benggang investigation and research on the mechanism of Benggang erosion.

According to remote sensing image classification and recognition methods, this paper proposes a novel BoV-TW model combining UHR DOM and DSM local features to represent Benggang areas. Then LDA is leveraged to implement latent semantic analysis to construct low-dimensional high-level semantic representations. Finally, SVM is used as a supervised classifier to achieve high-precision and fast automatic Benggang recognition.

## 2. RELATED WORKS

It has been 60 years since Zeng proposed the concept of Benggang in his academic monograph in 1960 (Zeng, 1960). Chinese scholars have studied the mechanism of occurrence and development of Benggang from different perspectives such as geomorphology, geology, ecology, and soil science. Extensive research has been carried out for the prevention and control of Benggang (Qiu, 2017). To study or manage the Benggang, the primary task is to discover it. However, traditional methods of discovering Benggang are based on local in-situ investigation. Currently, manual interpretation of high-resolution satellite remote sensing images based on expert experience is the most used method in practice. But in the plane remote sensing image, the interior appearance of Benggang is similar to flat bare soil. Due to vegetation restoration, the Benggang boundaries are not obvious. All these factors make the detection of Benggang areas a challenging task.

\* Corresponding author

With the rapid development of information technology in the 21st century, 3S technology, especially UAV survey, has been increasingly used as a high-precision spatial information collection method for quantitative research on Benggang investigation and monitoring (Shen, 2018; Jiang, 2019; Zhou, 2019). But such applications are based on the premise that the location of the Benggang is known. The discovery procedure is still fulfilled manually in the early stage. Meanwhile, the potential value of the UHR DOM and DSM has not been fully utilized. Considering the characteristics of Benggang, how to recognize Benggang automatically based on UHR DOM and DSM with image analysis and mining techniques will be an interesting research direction.

In recent years, in the field of remote sensing image classification and recognition, in order to improve the accuracy and solve the problem "semantic gap" (Bratasanu, 2011), that is, the relationship between the low-level visual features and high-level semantic information, researchers have adopted BoW (Bag of Words) and PTM (Probabilistic Topic Model) (Blei, 2012) which come from NLP (Natural Language Processing). And they

achieved good performance (Bratasanu, 2011; Lienou, 2010; Tang, 2015; Zhao, 2018; Zhu, 2018). In the applications of soil and water conservation field, an automatic landslide detection method based on Bag of Visual Words (BoVW) and Probabilistic Latent Semantic Analysis (PLSA) was proposed (Cheng, 2013). It is a scene classification method, which monitors landslides automatically from remote sensing images, and can identify landslides. Generally speaking, the above methods are based on ortho or plane remote sensing images, which are more suitable for artificial features (such as roads, houses, etc.) with obvious corners, edges, and texture features. But Benggang has obvious topographical characteristics, with indistinct boundaries and texture features. The accuracy of Benggang recognition is poor. DSM, which contains information from the elevation dimension, can describe the characteristics of the complex terrain and the sharp changes in the edges of Benggang. It can effectively distinguish Benggang from flat terrain. The fusion data of DSM and DOM, just like 3D model, will greatly improve the accuracy of Benggang recognition

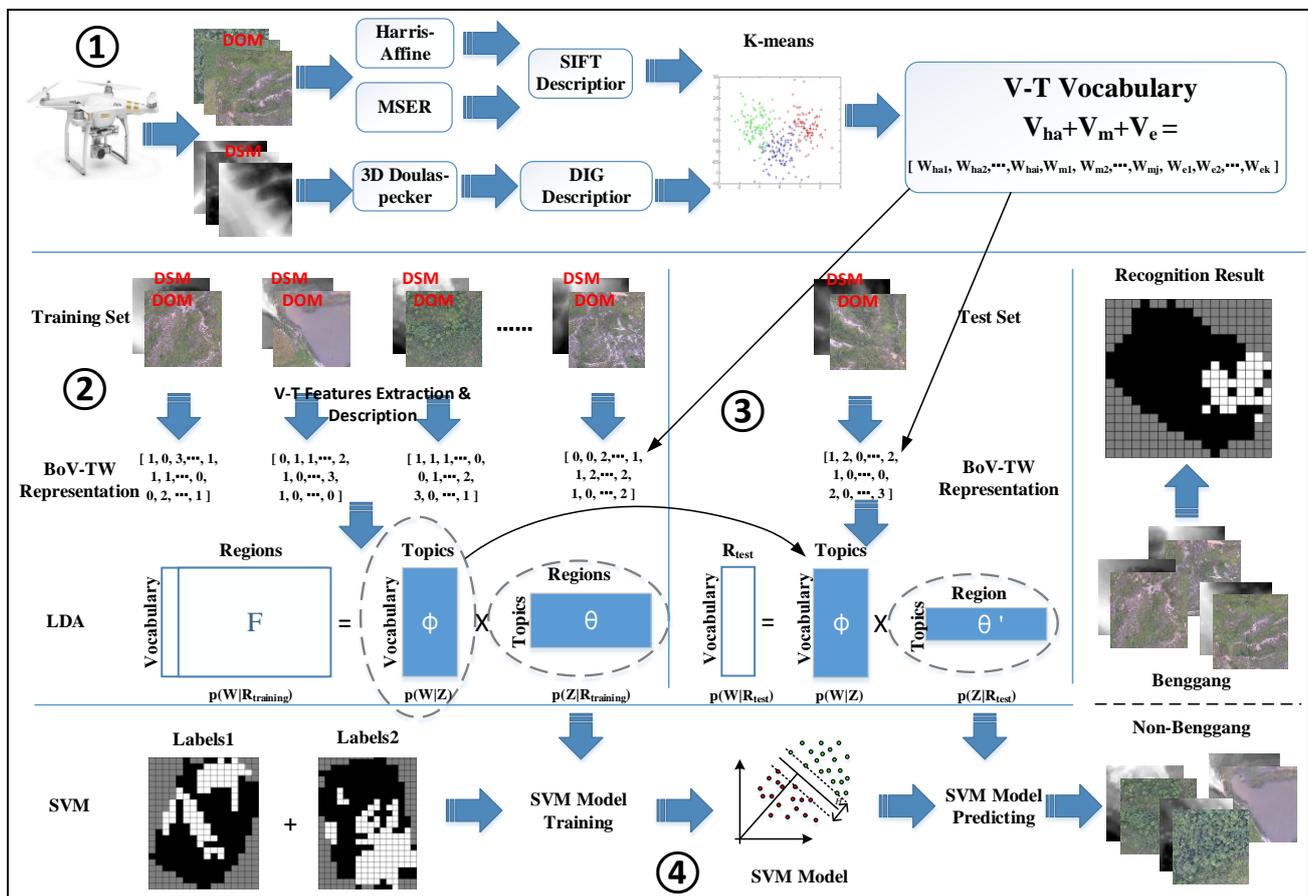


Figure 2. Overall workflow of automatic Benggang recognition

### 3. THEORETICAL APPROACH

The method workflow is divided into four steps: 1) Data preparation. This step includes DOM and DSM data acquisition and pre-processing, local feature extraction and description, and Visual-Topographical Vocabulary generation; 2) Training data modelling. This step consists of feature extraction, word frequency statistics of the Benggang training dataset, building BoV-TW representation, and building an LDA model through latent semantic analysis; 3) Test data modelling. Comprising feature extraction, word frequency statistics of the Benggang test

dataset, building BoV-TW representation, and generating a probabilistic topic distribution representation based on the existing LDA model; 4) Benggang recognition. This final step needs to perform Benggang sample labelling, SVM classifier training, prediction of the Benggang test dataset, and automatic recognition.

The four critical techniques and models, including local features of DOM and DSM, BoV-TW representation, LDA model, SVM model will be introduced as follows.

### 3.1 Local Features

#### 3.1.1 Local Features of DOM

According to relevant research in the field of computer vision, Affine Covariant Regions (Sivic, 2005), such as Harris-Affine (Mikolajczyk, 2002), MSER (Matas, 2004), etc., have the following three advantages: 1) Multi-scale viewpoint description is invariant; 2) Good description stability for the specified area; 3) Good tolerance to radiation changes. They are suitable choices for describing high-resolution remote sensing images. Therefore, this study uses Harris-Affine and MSER visual features, and then uses SIFT feature descriptors (Lowe, 2004) to describe the visual features.

#### 3.1.2 Local Feature of DSM

Following the principles of geology, in order to preserve the main geomorphological morphology and to synthesize local broken micro-geomorphic features, we use the 3D Douglas-Peucker algorithm (Fei, 2006) to extract the feature points of the terrain. In recursion way, the algorithm calculates the distance to the base plane determined by corner points of each points, and judges whether to retain the maximum value point as the feature point according to a threshold value. If the maximum value point is retained, the algorithm continues division and calculation.

A simple but direction-invariant DSM feature descriptor was designed. It takes a circular window of the neighbourhood of the key points, calculates the gradient size and gradient direction of each pixel. The gradient histogram is acquired after being weighted by Inverse Distance Weighted (IDW). DSM description is generated by normalizing and rotating histograms to the main direction, as shown in Figure 3.

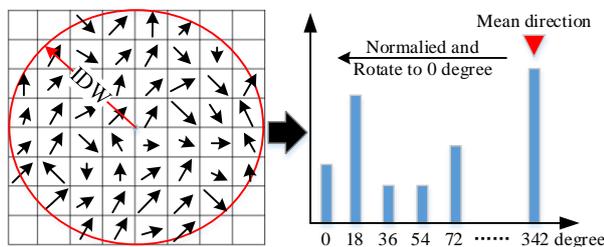


Figure 3. DSM feature descriptor

### 3.2 BoV-TW Representation

The Bag of Words (BoW) model (Sivic, 2003) ignores the word order, grammar, syntax, and paragraph links in the document. It is assumed that the appearance of each word is an independent event, and the document is regarded as a collection of words, only counting the number of occurrences of terms in the document. Similar to BoW, Bag of Features (Keypoints) and Bag of Visual Words have been proposed by researchers in the field of image and computer vision (Sivic, 2005; Lazebnik, 2006). They treat images as documents, and cluster the centers of the lower visual features of the image as a vocabulary to describe the image. Based on this method, the DSM local features are further integrated, and the centers of the clustered local feature of the terrain are used as a vocabulary describing the terrain. The local DOM feature is combined to form a BoV-TW to realize the BoV-TW representations of Benggang.

### 3.3 Latent Dirichlet Allocation

LDA is a probabilistic generation model for document sets. It organizes three levels of words, topics, and documents through a finite mixture based on probability. Each document can be

represented as a finite probability mixture of multiple topics, and each topic corresponds to a polynomial distribution on the vocabulary, and the topics are shared by all documents in the corpus.

The generation process is described as a graph model as shown in Figure 4 below. Among them, hollow circles represent latent variables, while solid circles represent observed variables; two rectangular boxes represent repeated processes. The inner rectangle represents the process of generating  $N$  words in a document from a polynomial distribution with parameter  $\beta$ , and the outer rectangle represents the process of generating topics from  $M$  documents in a document set from a Dirichlet distribution with parameter  $\alpha$ .

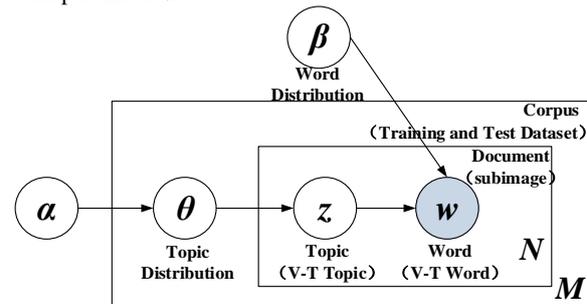


Figure 4. LDA graphic Model (Blei, 2003)

In this method, we considered Visual-Topographical words as words in LDA, Benggang subimages as Documents in LDA, training dataset and test dataset as Corpus in LDA. Then LDA was leveraged to implement latent semantic analysis to construct low-dimensional high-level semantic Benggang representations.

### 3.4 Support Vector Machine

The support-vector network is a learning machine for two-group classification problems (Cortes, 1995). It finds the best separating hyperplane in the very high-dimension feature space to maximize the interval between positive and negative samples in the training dataset. After introducing kernel functions, SVM can also be used to solve nonlinear problems.

In this study, we labelled the training dataset and test dataset, with Benggang and non-Benggang labels. Then an SVM classifier was trained to classify test dataset obtained for automatic Benggang recognition. The results were compared with the labels in the test dataset.

## 4. DATA AND METHOD

### 4.1 Raw Data Acquisition

The experimental data were acquired in Tongcheng County, Hubei Province, China in September 2016. The county has a subtropical monsoon climate, with an average annual rainfall of 1450 to 1600 mm. The soil in the area is mostly red soil developed from granite parent materials. The soil structure is loose, the soil and water loss is serious, and the erosion of Benggang is widespread.

The three Benggang areas are near Pingshan Village, Daping Town, Tongcheng County, approximately  $113^{\circ} 48' E$  and  $29^{\circ} 18' N$ . The experimental data was acquired using DJI Phantom 3P. The flights are designed to be 500m in length, 200m in width, and 150m in height. The round-trip overlap rate is 70% and the lateral overlap rate is 60%. DOM and DSM are shown in Figure 5 below.

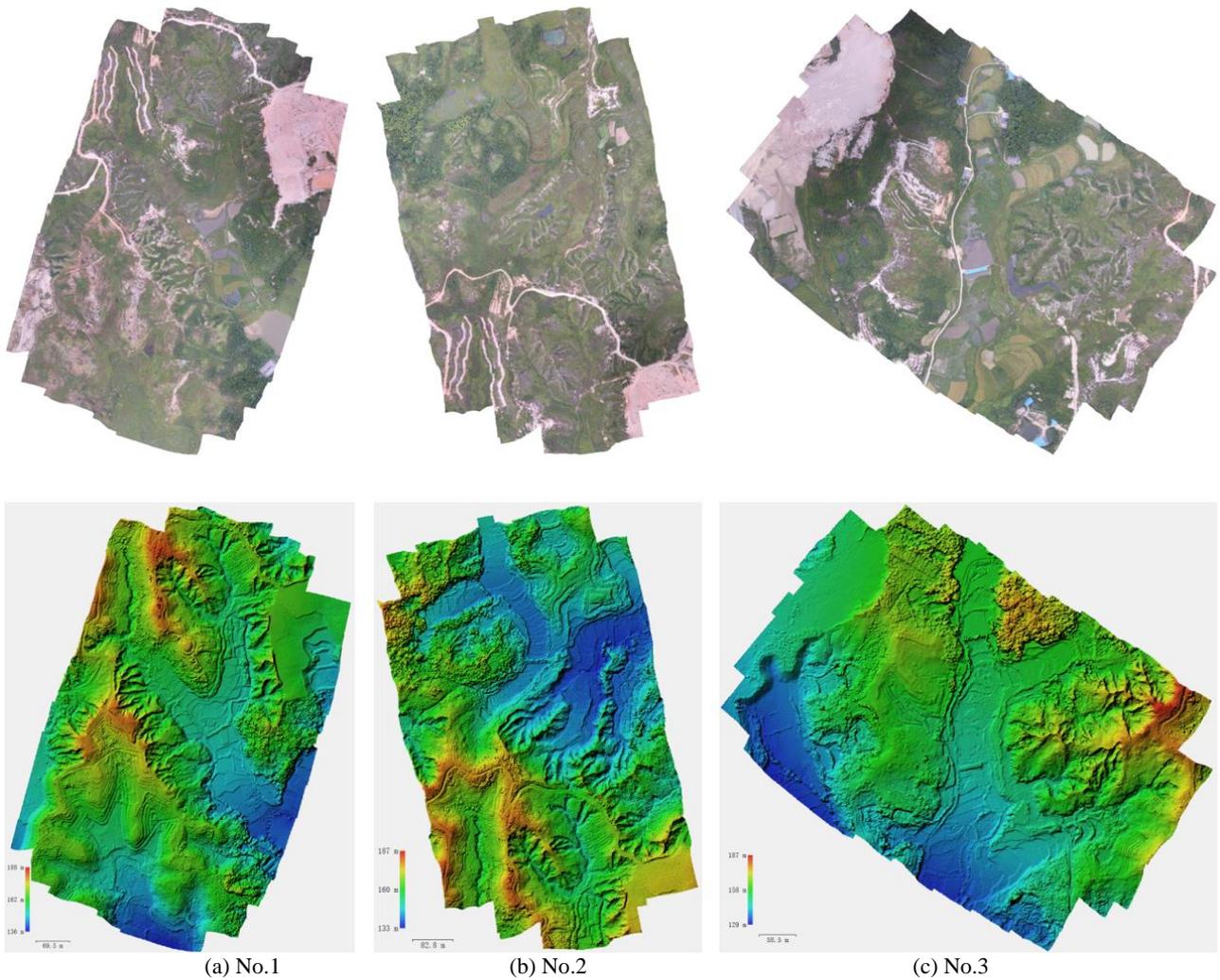
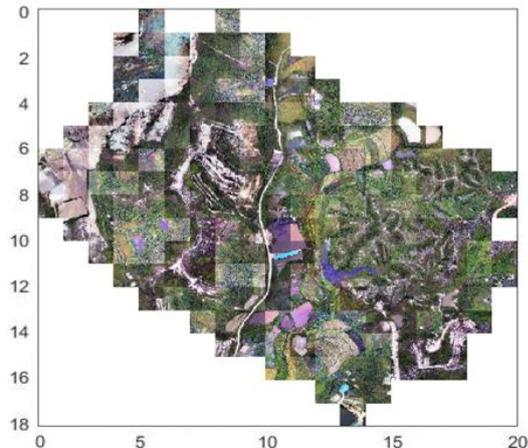


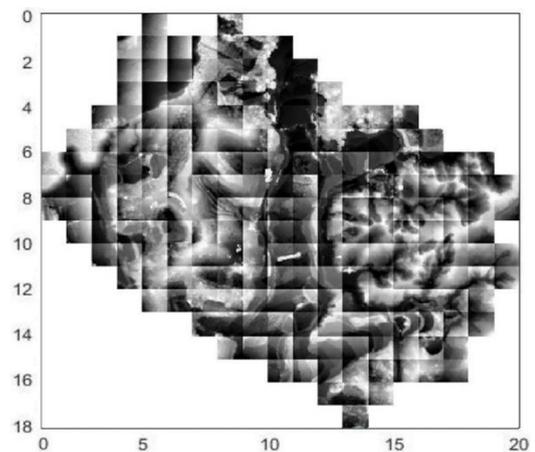
Figure 5. DOM and DSM of three Benggang areas

#### 4.2 Partition and Labelling

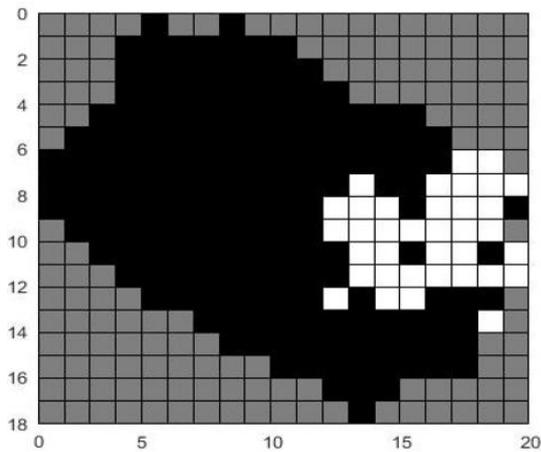
Taking the DOM and DSM of the No. 3 Benggang area in Figure 5 as test dataset as an example, the partition and labelling of the data are introduced. This method uses the uniform grid partition method, which is divided DOM and DSM into corresponding  $256 \times 256$  pixels subimages. The horizontal resolution of each subimage is 0.15 meters, that is, each subimage is  $38m \times 38m$ . Figure 6 shows the partition mosaic of the test dataset, and the labels of the Benggang and non-Benggang areas (white is Benggang areas, black is non-Benggang areas, and grey is empty).



(a) Partition mosaic of DOM of test dataset



(b) Partition mosaic of DSM of test dataset



(c) Labels of the Benggang and non-Benggang areas  
Figure 6. Partition and Labels of test dataset

After removing subimages with excessive edge deformation and abnormal elevation changes in each Benggang area, the training and test data and labels are shown in Table 1 below.

Type	Data	Benggang	Non-Benggang	Total
Training dataset	No.1	69	171	240
	No.2	109	182	291
	Subtotal	178	353	531
Test dataset	No.3	36	172	208

Table 1. Training dataset, Testing set and Labels

### 4.3 Vocabulary Generation

In order to represent each Benggang subimage into a vector of finite equal length and perform potential semantic analysis at a later stage, a vocabulary of a BoV-TW model must be established. This study uses Harris-Affine feature, MSER feature extraction algorithm and 72-dimensional SIFT descriptor developed by Oxford University Visual Geometry Group (VGG) for DOM. When extracting local features of DSM, an elevation threshold that is 4 times the horizontal resolution is used for 3D Douglas-Peucker terrain feature points detection. Describing features, the neighbourhood range radius is set to 12 pixels, and the number of gradient statistical direction is set to 24. There are 531 blocks of the training dataset DOM and DSM. 250,000 Harris-Affine feature descriptions, 340,000 MSER feature descriptions, and 100,000 3D Douglas-Peucker feature descriptions are acquired. After making K-means clustering with 500 clusters each, a 1500-dimensional Visual-Topographical Vocabulary is generated by concatenating them together.

### 4.4 Benggang and Non-Benggang Scene Classification

The scene classification of this method is divided into three steps: LDA training, SVM training and testing. 1) LDA training. Through the Visual-Topographical Vocabulary, 531 training dataset regions can be represented as a co-occurrence matrix  $p(W|R_{training})$  of 513 columns by 1500 rows using the k-Nearest Neighbor (KNN) algorithm. Then, by setting the number of different topics, LDA training is performed to conduct potential semantic analysis and dimensionality reduction, from which  $p(W|Z)$  and  $p(Z|R_{training})$  are obtained. 2) SVM training. Using  $p(Z|R_{training})$  and the Benggang and non-Benggang labels of training dataset, the SVM classifier is trained with Radial Basis Function (RBF) as the kernel. 3) Testing. DOM and DSM feature extraction and representation  $p(W|R_{test})$  are performed on the test

dataset by kNN and Visual-Topographical Vocabulary. And then  $p(Z|R_{test})$  is calculated based on  $p(W|Z)$  of LDA training by Gibbs-resampling. Finally, the predictions of the test dataset are obtained by inputting  $p(Z|R_{test})$  into the SVM classifier trained in 2).

## 5. EXPERIMENTAL RESULTS AND DISCUSSION

### 5.1 Experimental Results

In order to prove the advantage of this method, we implemented two experiments in this paper: 1) Method of using only DOM local features (Harris-Affine feature and MSER feature with SIFT descriptor) vs. our method, and 2) Method of using only BoV-TW vs. our method.

The main performance metrics of this experiment include *Total Accuracy*, *Benggang Recall* and *Precision*.

$$Total Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

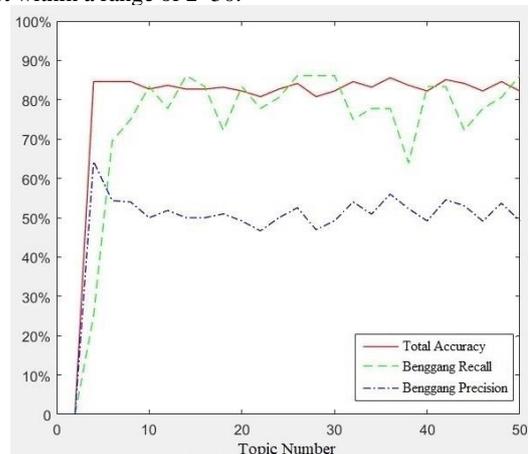
$$Benggang Recall = \frac{TP}{TP + FN} \quad (2)$$

$$Benggang Precision = \frac{TP}{TP + FP} \quad (3)$$

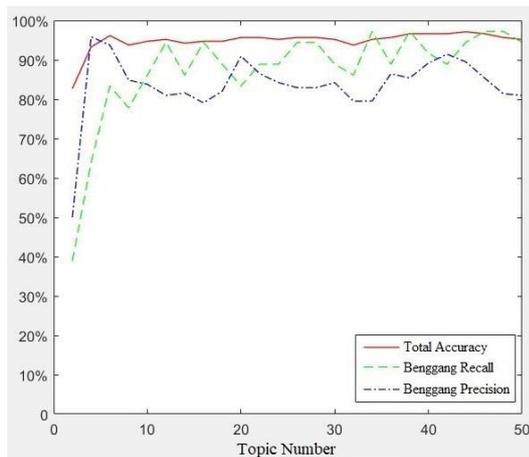
Where  $TP$  is the number of true positives,  $FN$  is the number of false negatives,  $FP$  is the number of false positives,  $TN$  is the number of true negatives.

#### 5.1.1 Method of using only DOM Local Features vs. Our Method

In this experiment, we compared the performance of the proposed method and the method of using only DOM local features. The later uses Harris-Affine features, MSER features to build a 72-dimensional SIFT descriptor for representing DOM local features. BoW is adopted to integrate these features for Benggang representation. Then keeping feature extraction methods, the vocabulary size, and the SVM classifier parameters unchanged, we compared it with our method when the topic number of LDA is set within a range of 2~50.



(a) Recognition performance of method of using only DOM local features



(b) Recognition performance of our method

Figure 7. Method of using only DOM local features vs. our method

When using only DOM local features, the total accuracy maintains at about 85% after 4 or more topics, the Benggang recall remained between 70-85%, and the Benggang precision fluctuated around 50%.

The total accuracy of our method maintained at about 95% after 4 or more topics; the Benggang recall was above 80%, and the highest was 97.22% for 34, 46, and 48 topics; The Benggang precision was above 80%, and the highest was 94.44% at 12 topics. The optimal recognition effect appeared on 38 topics, the Benggang recall was 94.44%, and the Benggang precision was 85.00%, as shown in Figure 7. The Benggang recognition labels of our method is shown in Figure 8, where  $\circ$  means  $TP$ ,  $\times$  means  $FN$ .

The overall performance of our method is better than method of using only DOM local features.

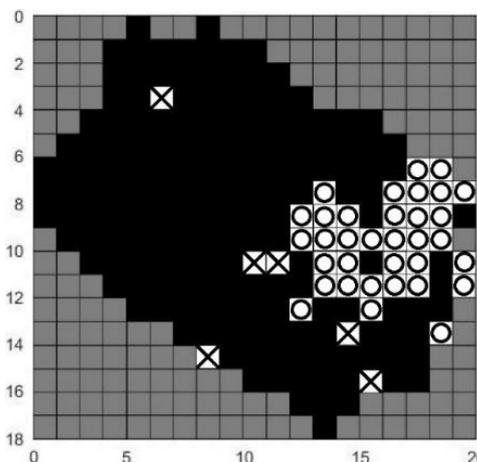


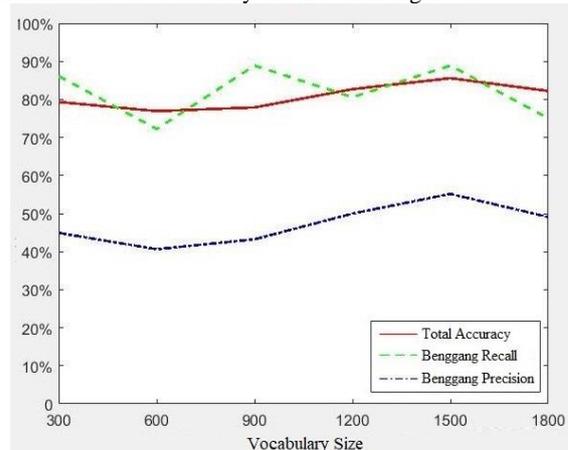
Figure 8. Benggang recognition labels of our method

### 5.1.2 Method of using only BoV-TW vs. Our Method

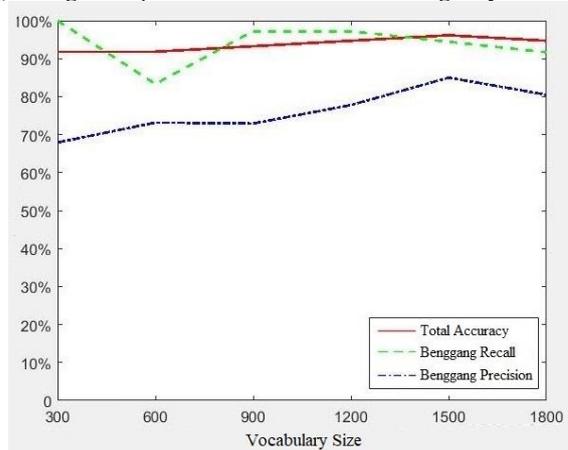
In the method of using only BoV-TW, the Visual-Topographical co-occurrence matrix  $p(W|R_{\text{training}})$  is normalized, the sum of each subimage is 1, which meets the requirement of probability distribution. Then SVM training and classification were carried out.

Keeping the feature extraction methods of , 40 topics of LDA, and the parameters of SVM classifier unchanged, the results of two methods were derived when the sizes of vocabulary are set as 300, 600, 900, 1200, 1500 and 1800(The number of each local feature are 100, 200, 300, 400, 500 and 600 respectively).

When using only BoV-TW, the total accuracy maintained at about 80%. Although the Benggang recall can be maintained at about 80%, the average Benggang precision is less than 50%. However, the total accuracy of our method is more than 90%, the highest is 96.1%. The Benggang recall is about 90%, the highest is 100%. The Benggang precision gradually increases with the increase of vocabulary size, reaching the highest 85% with 1500 words. The Benggang recognition labels of our method with six different sizes of vocabulary are shown in Figure 10.

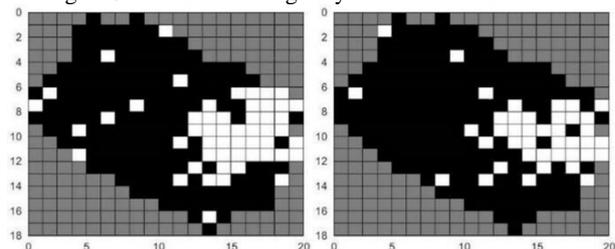


(a) Recognition performance of method of using only BoV-TW



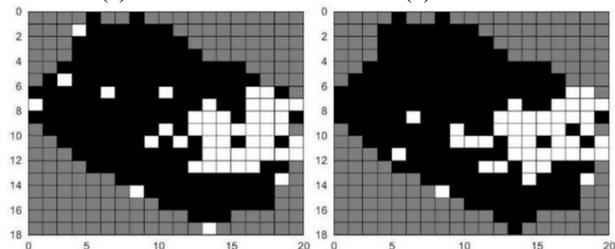
(b) Recognition performance of our method

Figure 9. Method of using only BoV-TW vs. our method



(a) 300 words

(b) 600 words



(c) 900 words

(d) 1200 words

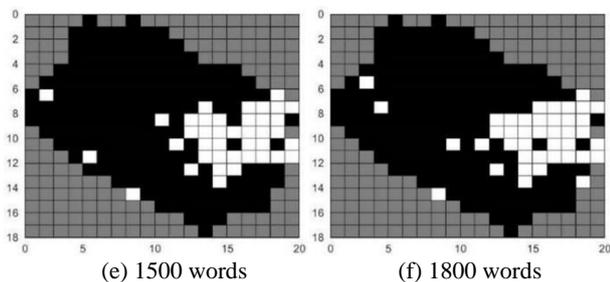
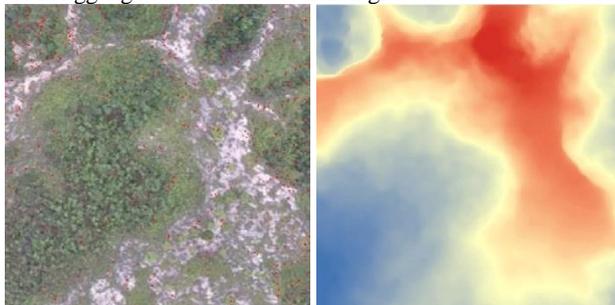


Figure 10. The Benggang recognition labels of our method with six different size of vocabulary

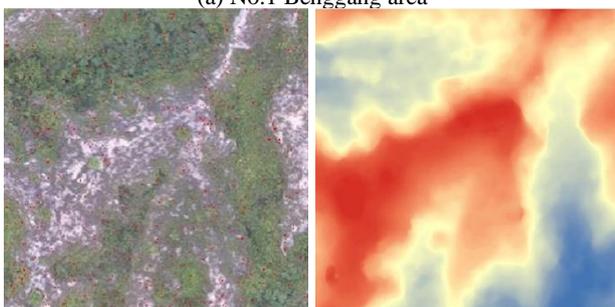
## 5.2 Discussion

### 5.2.1 Advantages

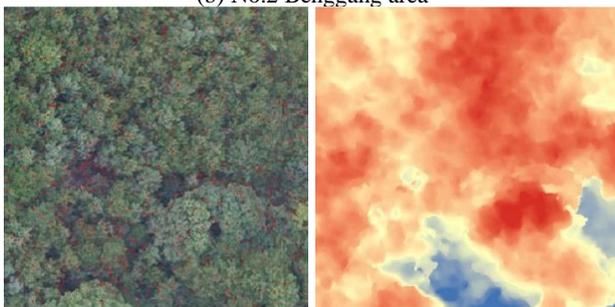
Benggang is difficult to recognize only by using DOM. Even manual interpretation by experts is difficult. Therefore, methods based on DOM local features has a low Benggang recall and precision. This paper proposed to combine DSM to recognize Benggang, which is equivalent to using DOM and DSM for constructing a three-dimensional model. By accounting for elevation information, our method facilitates recognize Benggang areas with obvious terrain features, contributing to significant improvement of recall and precision. Figure 11 shows the highest frequency samples of DOM and DSM of Benggang and non-Benggang when using our method with different number of LDA topics. Among them, local feature points of DSM, marked on DOM, are mainly distributed in areas with complex and abrupt elevation changes. The local feature points of DSM are widely distributed in Benggang areas. It can better represent the Benggang landform based on their gradient statistical features.



(a) No.1 Benggang area



(b) No.2 Benggang area



(c) No.1 non-Benggang area



(d) No.2 non-Benggang area

Figure 11. Samples of Benggang and non-Benggang

### 5.2.2 Limitation

Although the recall and precision of our method to recognize Benggang have been relatively high, there are still some misclassifications. Figure 12 shows the four regions that were misclassified as having a high frequency of Benggang. The misclassification rates were 92.53%, 80.05%, 62.69%, and 52.23%. The common reason is that there are bare soil and vegetation texture structures similar to the Benggang on the DOM, and there are obvious elevation changes caused by vegetation or houses on the DSM.



Figure 12. The top four misclassifications

### 5.2.3 Complexity

The process of our method is relatively complicated, and there are a number of parameters in each processing step, which all have effects on the results, such as the detection and description of visual and topographical features, the size of various feature vocabularies, the number of LDA topics, and the setting of SVM parameters. The experiment in this paper were implemented after limiting most parameters, and only by changing the number of LDA topics and the vocabulary size. It has achieved good recognition results. Based on this framework, by further optimizing the parameters of each step, better results can be obtained.

## 6. CONCLUSION

In order to solve the problem of rapid discovery of Benggang, this paper proposed an automatic Benggang recognition method based on latent semantic fusion of local features extracted from UHV DOM and DSM. By combining BoV-TW representations of Visual-Topographical features, and then using LDA to mining high-level semantic information, automatic recognition performance of Benggang is improved with SVM classifier. Experimental results prove that the total accuracy of our method in this paper can be maintained at about 95%, the Benggang recall and precision are above 80%, and the highest are 97.22% and 94.44% respectively, which are significantly higher than the methods only by using local DOM features and only by using BoV-TW.

## ACKNOWLEDGEMENTS

This work was supported in part by the National Key Research and Development Program of China under Grant 2016YFC050230305 and 2016YFC0503705, in part by the National Natural Science Foundation of China under Grant 41601298, Grant 51909011, and in part by the Basic Scientific Research Operations Fund of Central Public Welfare Institutions under Grant CKSF2019196/TB.

## REFERENCES

- Blei, D.M., Ng, A.Y., Jordan, M.I., 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3: 993-1022.
- Blei, D.M., 2012. Probabilistic topic models. *Communications of the ACM*, 55(4): 77-84.
- Bratasanu, D., Nedelcu, I., Datcu, M., 2011. Bridging the Semantic Gap for Satellite Image Annotation and Automatic Mapping Applications. *IEEE Journal Selected Topics Applied Earth Observations and Remote Sensing*, 4(1):193-204.
- Cheng, G., Guo, L., Zhao, T., Han, J., Li, H., Fang, J., 2013. Automatic Landslide Detection from Remote-Sensing Imagery using a Scene Classification Method based on BoVW and pLSA. *International Journal of Remote Sensing*, 34(1): 45-59.
- Cortes, C., Vapnik, V., 1995. Support-Vector Networks[J]. *Machine Learning*, 20: 273-297.
- Fei, L., He, J., Ma, C., Yan, H., 2006. Three Dimensional Douglas-Peucker Algorithm and the Study of Its Application to Automated Generalization of DEM. *Acta Geodaetica et Cartographica Sinica*, 35(3): 278-284. (in Chinese)
- Feng, M., Liao, C., Li, S., Lu, S., 2009. Investigation on status of hill collapsing and soil erosion in southern China. *Yangtze River*, 4(8): 66-68, 75. (in Chinese)
- Jiang, X., Liang, Z., Zhou, H., Dai, Y., Chen, Y., 2019. Dynamic Change monitoring of Collapse Based on UAV Oblique Photograph Technique. *Tropical Geography*, 39(4): 597-603. (in Chinese)
- Lazebnik, S., Schmid, C., Ponce, J., 2006. Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR'06)*, New York, USA.
- Liang, Y., Ning, D., Pan, X., Li, D., Zhang, B., 2009. Characteristics and management of Benggang in the red soil region of South China. *Soil and Water Conservation in China*, 1: 31-34. (in Chinese)
- Lienou, M., Maitre, H., Datcu, M., 2010. Semantic Annotation of Satellite Images Using Latent Dirichlet Allocation. *IEEE Geoscience and Remote Sensing Letters*, 7(1): 28-32.
- Lowe, G.D., 2004. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2): 91-110.
- Matas, J., Chum, O., Urban, M., Pajdla, T., 2004. Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 22(10): 761-767.
- Mikolajczyk, K., Schmid, C., 2002. An affine invariant interest point detector. In *Proceedings of the 7th European Conference on Computer Vision*, Copenhagen, Denmark, vol. 1: 128-142.
- Qiu, J., Liu, X., 2017. Status and comprehensive analysis of benggang research in China based on knowledge maps. *Science of Soil and Water Conservation*, 15(3):139-148. (in Chinese)
- Shen, S., Zhao, Y., Cheng, D., Zhang, P., 2018. An Efficient Method of Investigating Slope Collapse by Unmanned Aerial Vehicle Remote Sensing. *Journal of Yangtze River Scientific Research Institute*, 35(4): 43-47, 53. (in Chinese)
- Sivic, J., Zisserman, A., 2003. Video Google: A Text Retrieval Approach to Object Matching in Videos. In *Proceedings of the International Conference on Computer Vision (ICCV'03)*, Nice, France.
- Sivic, J., Russell, B.C., Efros, A.A., Zisserman, A., Freeman W. T., 2005. Discovering objects and their location in images. In *Proceedings of the International Conference on Computer Vision (ICCV'05)*, Beijing, China.
- Tang, H., Shen, L., Qi, Y., Chen, Y., Shu, Y., Li, J., Clausi, D.A., 2015. A Multiscale Latent Dirichlet Allocation Model for Object-Oriented Clustering of VHR Panchromatic Satellite Images. *IEEE Transactions on Geoscience and Remote Sensing*, 51(3): 1680-1692.
- Zeng, Z., 1960. *Topographic Principles: Book I*. South China Normal University Press, Guangzhou. (in Chinese)
- Zhao, B., Zhong, Y., Xia, G., Zhang, L., 2015. Dirichlet-derived Multiple Topic Scene Classification Model for High Spatial Resolution Remote Sensing Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 54(4): 2108-2123.
- Zhou, X., Yu, H., Wei, Y., Hu, J., Cai, C., 2019. Method for monitoring change in Benggang erosion based on oblique aerial images of UAV. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 35(21):51-59. (in Chinese)
- Zhu, Q., Zhong, Y., Wu, S., Zhang, L., Li, D., 2018. Scene Classification Based on the Sparse Homogeneous-Heterogeneous Topic Feature Mode. *IEEE Transactions on Geoscience and Remote Sensing*, 56(5): 2689-2703.