

SCENE CLASSIFICATION BASED ON THE INTRINSIC MEAN OF LIE GROUP

Chengjun Xu¹, Guobin Zhu^{2,*}, Kejian Yang³

¹ School of Remote Sensing and Information Engineering, Wuhan University- (jjtu_001@163.com)

² School of Remote Sensing and Information Engineering, Wuhan University - (gbzhu@whu.edu.cn)

³ School of Geodesy and Geomatics, Wuhan University - (kjiang@sgg.whu.edu.cn)

KEY WORDS: Scene classification, Remote sensing, Intrinsic mean, Lie Group

ABSTRACT:

Remote Sensing scene classification aims to identify semantic objects with similar characteristics from high resolution images. Even though existing methods have achieved satisfactory performance, the features used for classification modeling are still limited to some kinds of vector representation within a Euclidean space. As a result, their models are not robust to reflect the essential scene characteristics, hardly to promote classification accuracy higher. In this study, we propose a novel scene classification method based on the intrinsic mean on a Lie Group manifold. By introducing Lie Group machine learning into scene classification, the new method uses the geodesic distance on the Lie Group manifold, instead of Euclidean distance, solving the problem that non-euclidean space samples could not be calculated by Euclidean distance directly. The experiments show that our method produces satisfactory performance on two public and challenging remote sensing scene datasets, UC Merced and SIRI-WHU, respectively.

1. INTRODUCTION

Remote sensing scene classification refers to distinguishing semantic objects with similar scene characteristics from multiple image categories and classifying them into scene types accordingly. In other words, different remote sensing images in the database are classified according to certain dominating features, which make the extraction of image features a key to scene classification. In the past, many classical methods for extracting image features have been developed, which are mainly divided into three categories: (1) extracting feature descriptors directly from images, by using methods such as the Scale-invariant feature transform(SIFT)(Dellinger et al., 2014, Sedaghat, Ebadi, 2015), Histogram of Oriented Gradients(HOG)(Qi et al., 2015, Kaâniche, Bremond, 2012), and Local Binary Pattern(LBP)(Ren et al., 2016, Xiao et al., 2018); (2) extracting continuous features on the basis of image blocks, by using the Bag-of-visual-words model (BoVW)(Zhao et al., 2014, Zhu et al., 2016) and sparse matrix(Ye et al., 2014, Zhang et al., 2018); and (3) extracting features by training the deep learning models(Yang, Newsam, 2010, Cheriyyadath, 2013, Yang et al., 2015, Huang, Yan, 2015, Zhang et al., 2015).

Each of the three methods mentioned above for extraction has its advantages and disadvantages. The first method is simple and easy to implement, but it contains very little information about the semantic characteristics of the scene. Although the classification accuracy of the second method is higher than that of the first method, the processing process is more complicated. The third method is developed in recent years, which does not need to extract feature descriptors manually, and the classification effect on scenes is very good if a learning network is well trained. Nevertheless, a deep network model needs a large amount of data for training, which usually takes a long time for computing and relies on highly configured device supports.

In addition to the above feature extraction methods, the construction of scene classifiers is also very important. For example, clustering is a powerful tool for data classification. It is

a kind of method for clustering a data set into groups with the most similarity in the same cluster and the most dissimilarity between different clusters(Aggarwal, 2014, Kaufman, Rousseeuw, 2009). Clustering methods mainly include a probability model-based approach and a non-parametric approach. Among non-parametric methods, partitional methods are most commonly used. The k-means algorithm, as an example, is the earliest and most famous method and has been proved efficient in remote sensing image classification fields(Pelleg et al., 2000, MacQueen et al., 1967, Jain, 2010, Kanungo et al., 2002).

Although the k-means model has been widely used in remote sensing classification, it still has shortcomings. First, the k-means method requires pre-setting the size of K value according to previous experience, and the size of K value largely determines the quality of subsequent classification results. Second and more importantly, this model relies on the hypothesis that all training samples satisfy the distribution on classic Euclidean space. However, when the sample is in the manifold space, the Euclidean spatial distance cannot accurately reflect the real distance of the sample, as shown in Figure 1.

In order to solve the above problems, a novel remote sensing scene image classification method, which is based on the intrinsic mean of Lie Groups, is proposed in this study. In a Lie Group, general means will be further treated as two types: intrinsic and extrinsic ones. Since the intrinsic mean is more able to reflect the commonness between one category and another, then an unknown sample will be closer to the intrinsic mean of a category than other categories. Thus it is considered that the unknown sample is most likely to belong to that category. To this end, the sample-set will be mapped to the Lie Group manifold space according to the category, and the intrinsic mean of each category in the Lie Group manifold space is calculated. For the unknown sample, only the geodesic distance of the intrinsic mean from the sample to each category is calculated, and the category to which the intrinsic mean with the shortest distance belongs is thus determined.

The major contributions of this paper are as follows:

* Corresponding author (G. Zhu)

1. We propose a strategy to classify the samples of Lie Group in the space of Lie Group manifold, and implement the intrinsic mean classification algorithm within the Lie Group according to this strategy.
2. We demonstrate that the intrinsic mean within a Lie Group is more likely to reflect the commonality between one category and another. When the sample is projected onto the manifold space of the Lie Group, it has obvious advantages such as few parameters and strong understanding.
3. Due to Lie Group samples do not belong to vector space, Euclidean distance cannot be used as a classification metrics. Therefore, we propose a novel method of calculating the distance from geodesics based on the manifold space, which can better reflect the spatial distance between samples and has a better classification effect.

The rest of this paper is organized as follows. Section 2 presents the proposed scene classification method based on the intrinsic mean of Lie Group. In Section 3, the experimental results are provided. Finally, the conclusion is given in Section 4.

2. METHODOLOGY

2.1 Definition of Matrix Lie Group

Let $M_{mn}(K)$ be a matrix of $m \times n$, in which each element belongs to K , and K represents an (exchange) field. In most cases, K is defined as \mathbf{R} (real number) or $K = \mathbf{C}$ (complex number). In this study, K is restricted as \mathbf{R} (real number). A_{ij} or a_{ij} is used to represent the i^{th} row and j^{th} column of $m \times n$ matrix A , so,

$$A = (a_{ij}) = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \quad (1)$$

In matrix theory, the determinant of a matrix is a mapping $\det : M_{mn}(K) \rightarrow K$, which has the following properties:

1. for any $A, B \in M_n(K)$, there is $\det(AB) = \det A \det B$.
2. $\det(I_n) = 1$.
3. If and only if $A \in M_n(K)$, $\det A \neq 0$, A is reversible.

There exists two very important matrix groups, namely general linear group $GL_n(K) = \{A \in M_n(K) | \det A \neq 0\}$ and special linear group $SL_n(K) = \{A \in M_n(K) | \det A = 1\} \subseteq GL_n(K)$. The set $GL_n(K)$ and $SL_n(K)$ constitute the group under matrix multiplication. Further, there is $SL_n(K) \leq GL_n(K)$, that is, $SL_n(K)$ is a subgroup of $GL_n(K)$.

In this study, the matrix groups applied later are all subgroups of $GL_n(K)$, and are all subgroups of $GL_n(\mathbf{R})$.

2.2 Intrinsic mean of Lie Group for remote sensing image scene classification

The flowchart of the intrinsic mean of Lie Group for remote sensing image scene classification is shown in Figure 2, where this study divides the scene classification process into three portions, i.e. 1)mapping the sample data set to a Lie Group manifold space, 2)calculating the intrinsic mean of each category in the Lie Group manifold space, 3)calculate the geodesic distance between the unknown sample and the intrinsic mean of each category, and finally, the category of the unknown sample is given.

2.2.1 Mapping operations The training sample set is mapped to the Lie Group manifold space according to different categories, and the Lie Group sample data set on the Lie Group manifold space is obtained after the mapping. Accordingly, we get $\{M_{ij} | i = 1, 2, \dots, c; j = 1, 2, \dots, n_i\}$, where M_{ij} represents the j^{th} sample of the i^{th} category in the sample training set, and n_i represents the number of training samples in the i^{th} category, with a total of c categories.

$$x_{ij} = \exp(M_{ij}) \quad (2)$$

where x_{ij} represents a sample mapped to a Lie Group manifold space.

2.2.2 Calculating intrinsic mean in the sample of Lie Group

As we can see, the intrinsic mean μ on manifold S^1 is shown in Figure 1-(a). The distance from each data point to μ is the geodesic distance (curve length) on S^1 , and the intrinsic mean obtained is also on S^1 . Figure 1-(b) is the average value obtained by direct calculation of Euclidean distance, where μ obtained by such calculation is called the extrinsic mean, and obviously, the extrinsic mean is not in the manifold space, i.e. μ shown in the Figure 1-(b) is in \mathbf{R}^2 instead of S^1 . In \mathbf{R}^d , the average value of $\{x_1, x_2, \dots, x_n\} \subseteq \mathbf{R}^d$ set of n data points can be calculated by the following equation,

$$\mu = \frac{1}{n_c} \sum_{i=1}^{n_c} x_i \quad (3)$$

where the sample x_i can be a vector or matrix, and the obtained μ can maintain the minimum sum of the Euclidean distance squared of each point in $\{x_1, x_2, \dots, x_n\}$, that is, μ can be expressed as:

$$\mu = \arg \min_{x \in \mathbf{R}^d} \sum_{i=1}^n \|x - x_i\|^2 \quad (4)$$

In general, an d -dimensional manifold M^d is not a vector space, so the Euclidean distance cannot be used to represent the distance between two points on the manifold. Therefore the above equation (4) cannot be directly applied to manifolds. But M^d is locally homeomorphic in \mathbf{R}^d , so in the local domain of M^n , by looking for a homeomorphic mapping, $\Phi : M^d \rightarrow \mathbf{R}^d$, the local part of M^d is embedded in \mathbf{R}^d , and then the calculation can be made in \mathbf{R}^d by $\mu_\Phi = \arg \min_{x \in M^d} \sum_{i=1}^n \|\Phi(x) - \Phi(x_i)\|^2$.

At the same time, we need to define another projection mapping $\pi : \mathbf{R}^d \rightarrow M^d$, $\mu = \pi(\mu_\Phi)$, for mapping μ_Φ back to M^d and eliminating μ_Φ . The calculation equation can be obtained:

$$\mu = \pi \left(\arg \min_{x \in M^d} \sum_{i=1}^n \|\Phi(x) - \Phi(x_i)\|^2 \right) \quad (5)$$

In practical application, it is difficult to find the right two maps at the same time, and M^d doesn't have to be one Φ to be all embedded in \mathbf{R}^d . A more reasonable solution is to use Riemann distance on M^d instead of ϕ function in equation (5) to calculate the distance between any $x_i, x_j \in M^d$ directly on M^d , rather than in Euclidean space.

Therefore, similar to the case in the Euclidean space R^d , the average value of a set point on M^d can be achieved by calculating the point \bar{x} , which must meet two conditions: (1) $\bar{x} \in M^d$,

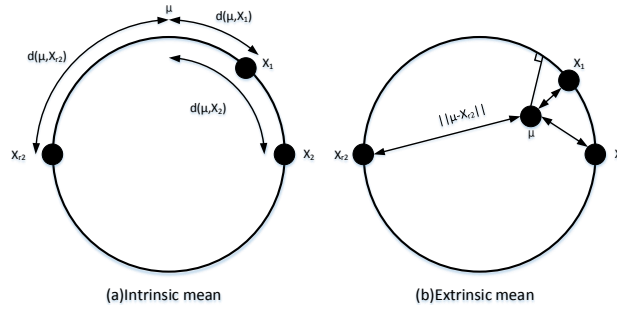


Figure 1. Difference of intrinsic mean and extrinsic mean.

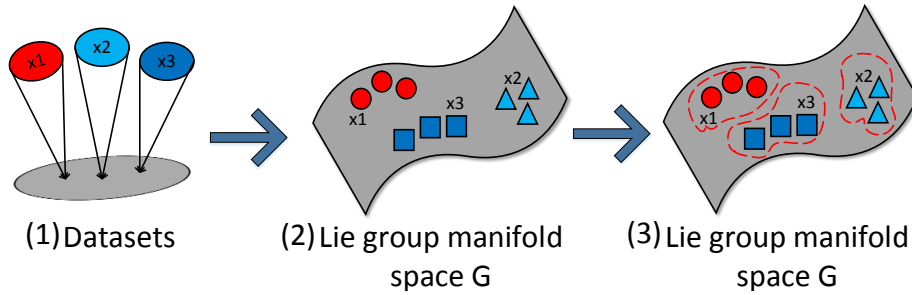


Figure 2. Proposed.

(2) the geodesic line of each point in \bar{x} and $\{x_i\}_{i=1}^n \in M^d$ is the minimum sum of squares. In a tight flow, an intrinsic mean consisting of a collection of $\{x_i\}_{i=1}^n$ data points n can be defined as:

$$\mu = \bar{x} = \arg \min_{x \in M^d} \sum_{i=1}^n d(x, x_i)^2 \quad (6)$$

where $d(\cdot, \cdot)$ represents the geodesic distance between two points in parentheses.

From the above contents, it can be obtained that the distance between any two points in the Lie Group is given by equation,

$$d(x_1, x_2) = \|\log_2(x_1^{-1}x_2)\| \quad (7)$$

substitute it into equation(6), and obtain:

$$\mu = \arg \min_{x \in G} \sum_{i=1}^n \|\log_2(x^{-1}x_i)\|^2 \quad (8)$$

where G represents the manifold of Lie Groups.

According to the first order BCH equation approximate equation (8) on the right side of the $\|\cdot\|$ content, can obtained:

$$\log_2(x^{-1}x_i) = -\log_2(x) + \log_2(x_i) \quad (9)$$

where x and x_i represent Lie algebras of points in Lie Groups, respectively. Relatively accurate results can be obtained by approximating the first n -th derivative, which can be obtained through experimental calculation. The latter approximate calculation is very large, but the contribution to the results is very small and can be ignored. Considering the substitution of first-order approximation, the intrinsic mean equation can be ob-

tained:

$$\begin{aligned} \hat{\mu} &= \arg \min_{x \in G} \sum_{i=1}^n \|\log_2(x) + \log_2(x_i)\|^2 \\ &= \arg \min_{m \in G} \sum_{i=1}^n \|\log_2(x) + m_i\|^2 \end{aligned} \quad (10)$$

2.2.3 Algorithm Find the $x \in G$ that meets the criteria according to equation (9) and (10). In (Fletcher et al., 2003), the process of solving $\hat{\mu}$ by gradient descent method is given, and the optimization function is $f(x) = \frac{1}{2n} \sum_{i=1}^n d(x, x_i)^2$. The gradient of function f is $\nabla f(x) = -\frac{1}{n} \sum_{i=1}^n \log_2(x^{-1}x_i)$. When estimate of the intrinsic mean for a given k iteration is μ_k , the $k+1$ iteration equation is $\mu_{k+1} = \exp(\frac{\tau}{n} \sum_{i=1}^n \log_2(\mu_k^{-1}x_i))$, where τ is the step length. The specific solution algorithm of the intrinsic mean in Lie Group is given by Algorithm 1.

Algorithm 1 The intrinsic mean algorithm of n category

Input: $\{x_{ij}\}_{i=1,2,\dots,c}^{j=1,2,\dots,n_i} \in G$, x_{ij} represents the i^{th} sample in the j^{th} class distributed on the Lie Group G , n_i represents the number of training samples in the i^{th} category, with a total of c classes.
Output: $\mu_i, i = 1, 2, \dots, c$, the intrinsic mean for each category.

```

1 Do  $i = 1$ 
2    $k = 0$ 
3    $\mu = x_{i1}$ 
4   Do
5      $\nabla \mu = \frac{\tau}{n_i} \sum_{j=1}^{n_i} \log_2(\mu^{-1}x_{ij})$ 
6      $\mu = \exp(\nabla \mu)$ 
7      $k = k + 1$ 
8   While  $\|\nabla \mu\| > \xi$  and  $k < Max\_Iters$ 
9      $\mu_i = \mu$ 
10     $i = i + 1$ 
11 While  $i \leq c$ 

```

The gradient descent method is only local convergence, the finding is not necessarily the globally optimal $\hat{\mu}$. In practice, we can get a better effect by modifying the initial estimate μ_0

and step τ . The choice of τ is related to the manifold structure of Lie Group G . In (Buss, Fillmore, 2001) demonstrates that $\tau = 1$ is appropriate for spherical manifolds. When the manifold structure of G is a vector space, the gradient descent of $\tau = 1$ is equivalent to the linear average, and if it is R^+ , R^d , $SO(3)$, $\tau = 1$ is equivalent to the geometric average, and the algorithm 1 can converge in a one-step iteration. For Lie Groups under general circumstances, if the algorithm 1 cannot converge when $\tau = 1$, a smaller positive number can be set for τ appropriately. According to the characteristics of gradient descent method, when the value of τ is too large, it may cross the extreme point; while if the value is too small, the convergence speed will be too slow.

2.2.4 Classification Owing the intrinsic mean is more able to reflect the commonness between one category and another, if an unknown sample is closer to the intrinsic mean of one of the categories than other categories, it can be determined that the unknown sample is most likely to belong to the closest category. Therefore, a Lie-mean algorithm is designed based on the intrinsic mean, by comparing the space manifold distance between the unknown sample and the trained sample Lie Group intrinsic mean. The unknown sample is determined to the category with the shortest distance. The specific discriminant equation is:

$$i^* = \arg \min_{i=1,2,\dots,c} \sum_{j=1}^n \|\log_2(\mu_i^{-1} x)\| \quad (11)$$

3. EXPERIMENTAL RESULTS AND ANALYSIS

3.1 Experiment setup

To evaluate the performance of the proposed method, we selected two datasets, i.e. UC Merced dataset and the SIRI-WHU dataset. In this study, each of the datasets was randomly split into 75% for training and 25% for testing. And the experiment on each data set was repeated five times, the average classification accuracy was recorded finally.

3.2 Experiment on UC Merced dataset

The 21-class UCM land-use dataset (Yang, Newsam, 2012) was designed from large optical images by the U.S. Geological Survey. The UCM dataset covers all typical regions of the United States and includes 21 scene categories, each with 100 scene images. Each scene image consists of 256×256 pixels with a spatial resolution of 1 foot per pixel, as shown in Figure 3.

The scene classification results are reported in Table 1. The results of previous scene classification methods, such as BoVW, pLSA, LDA, SPM+SIFT, SIFT+SC (Cheriyadat, 2013), SPCK++ (Yang, Newsam, 2011), S-UTF (Zhang et al., 2014), GBRCN (Zhang et al., 2015), SAL-PTM (Zhong et al., 2015), CCNN, SRSCNN (Liu et al., 2016), and SRSCNN were compared. To keep things simple, CCNN denotes CNN without random-scale stretching and voting, and SRSCNN denotes CNN with random-scale stretching but not voting. Table 1 shows that our proposed method performs best among all compared methods, with an overall classification accuracy of 96.71%, on average 24.66% and 1.61% higher than the BoVW and SRSCNN, respectively. Figure 4 shows the confusion matrix established on the UC Merced data set, indicating that the scene mis-classification between categories has relative small percentage.

From Figure 4, we can see that our proposed method identifies most of the scene categories, except dense residential, freeway and medium residential. From this, we find that there is a major confusion between the two categories of medium residential and dense residential because the intrinsic means of the Lie Groups of medium residential and dense residential are very close.

Methods	OA(%)
BoVW	72.05
pLSA	80.71
SPCK ++	76.05
LDA	81.92
SPM+SIFT	82.30
SIFT+SC	81.67
S-UTF	82.72
SAL-PTM	88.33
GBRCN	94.53
CCNN	90.07
SRSCNN	92.70
SRSCNN	95.10
PROPOSED	96.71

Table 1. Comparison of the previous reported accuracies with the UC Merced Dataset.

3.3 Experiment on SIRI-WHU dataset

The SIRI-WHU dataset (Zhao et al., 2015) consists of 2,400 scene sample images mainly collected from China (see Figure 5). The dataset includes 12-classes, and each scene class contains 200 images with a spatial resolution of 2 meters and a size of 200×200 pixels. Twelve categories cover agriculture, commerce, ports, idleland, industry, grassland, overpasses, parks, ponds, residential buildings, rivers, and water.

Methods	OA(%)
SPM-Sift	80.26
BOVW	73.93
LLC	70.89
LDA	66.85
SRSCNN	90.07
CCNN	89.60
SRSCNN	93.40
VGG-VD16	96.88
VGG-M(IFK)	95.10
Caffe features	93.42
PROPOSED	97.16

Table 2. Comparison with the previous reported accuracies with the Google Dataset of SIRI-WHU

Table 2 compares the scene classification results by some common methods on SIRI-WHU datasets. Compared with the state-of-art methods, the proposed method has the highest accuracy with the overall classification accuracy as 97.16%, on average 0.28%, 2.06%, 3.74%, and 3.76% higher than the VGG-VD16 (Hu et al., 2015), VGG-M(IFK) (Hu et al., 2015), Caffe features (Penatti et al., 2015), and SRSCNN, respectively. The proposed method is still competitive in terms of low requirement for computational cost. Figure 6 shows the confusion matrix established on the SIRI-WHU data set. Likewise, the results based on the SIRI-WHU dataset can precisely recognize the correct scenes with very low mis-classification percentage, compared with the confusion matrix derived from the UC Merced dataset classification results.

From Figure 6, the classification accuracy of our proposed method reaches 97%. In the category of misclassified scenes, such as,



Figure 3. Some category examples from the UC Merced dataset: (a)agriculture, (b)airplane, (c)baseballdiamond, (d)beach, (e)buildings, (f)chaparral, (g)denseresidential, (h)forest,(i)freeway, (j)golfcourse, (k)harbor, (l)intersection, (m)mediumresidential, (n)mobilehomepark, (o)overpass, (p)parkinglot, (q)river, (r)runway, (s)sparseresidential, (t)storagetanks, (u)tenniscourt.

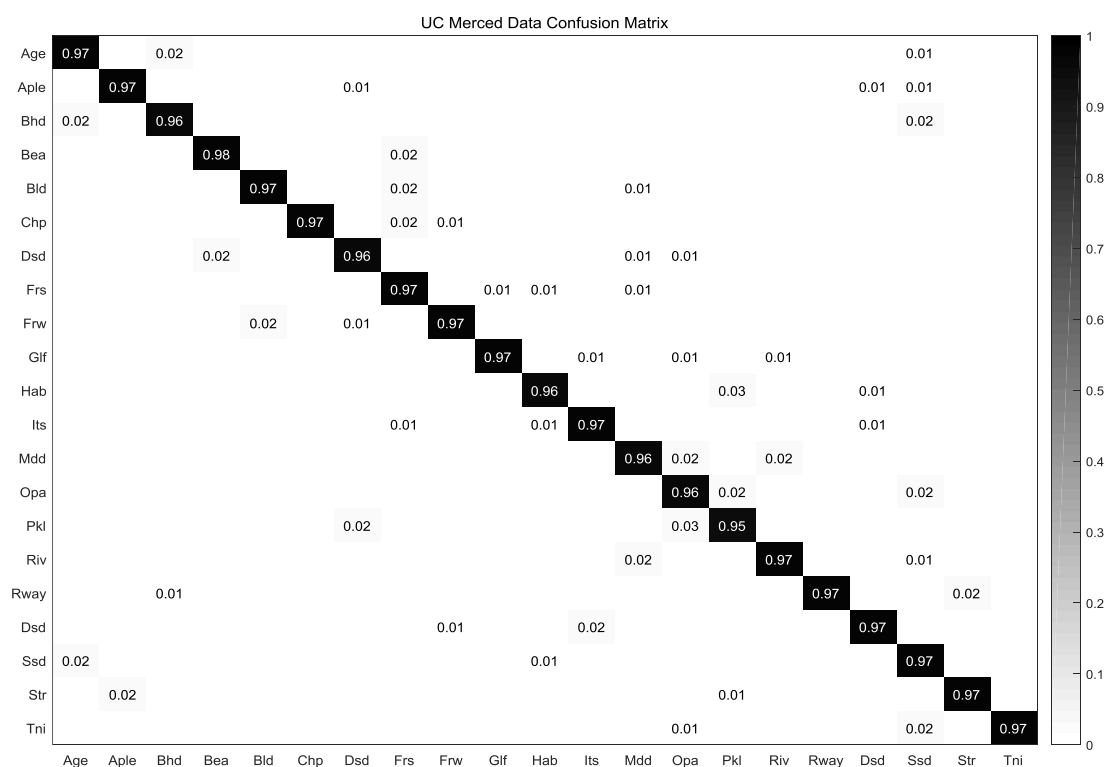


Figure 4. Confusion matrix of classification based on UC Merced dataset.

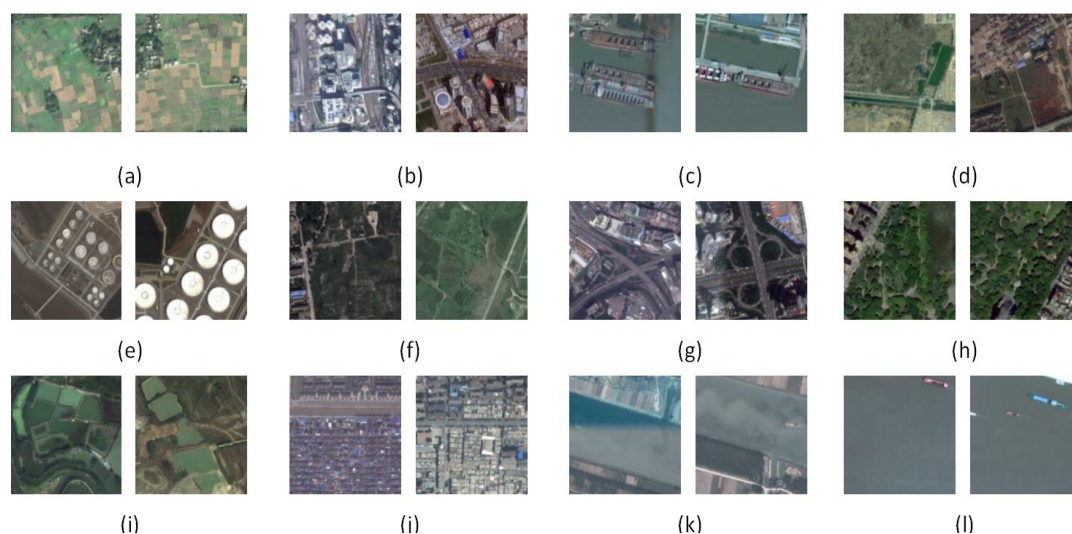


Figure 5. Some example from the SIRI-WHU dataset: (a)agriculture, (b)commercial, (c)harbor, (d)idle_land, (e)industrial, (f)meadow, (g)overpass, (h)park, (i)pond, (j)residential, (k)river, (l)water.

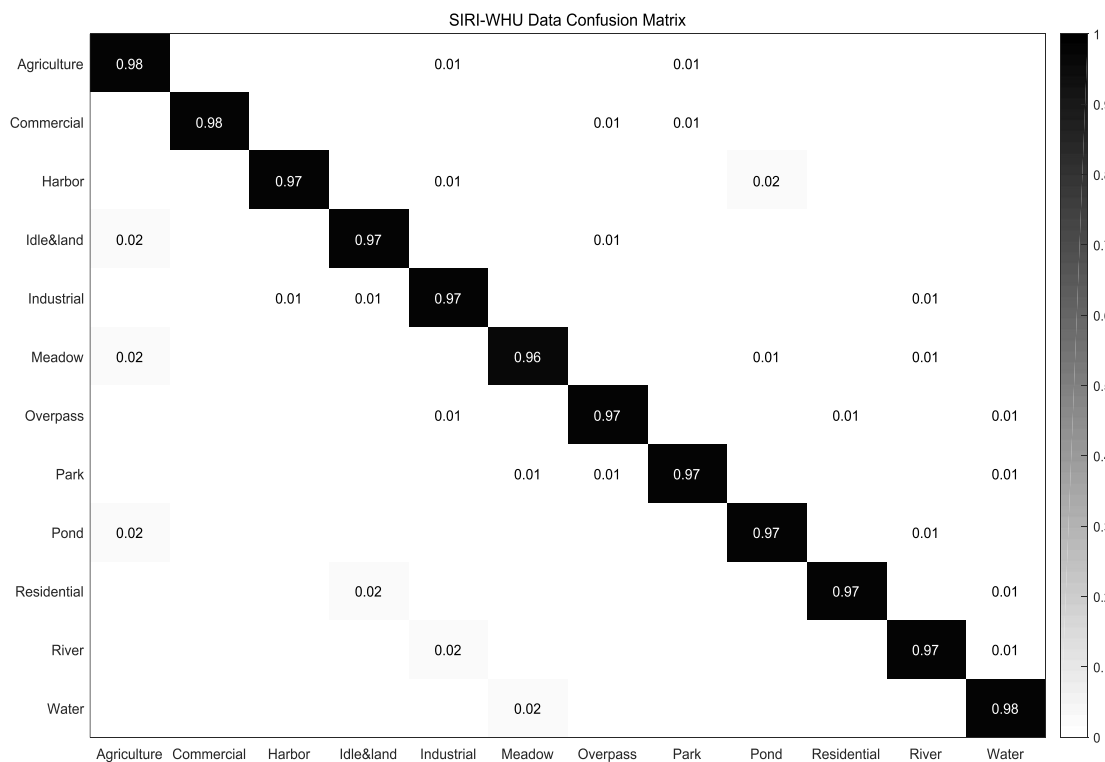


Figure 6. Confusion matrix of classification based on SIRI-WHU dataset.

industrial scenes are not well classified, and some river scene images are misclassified as water, which is likely because the river scene images contain more water. Therefore, these scenes may be misclassified as the water scene.

3.4 Comparison between two results

The UC Merced dataset contains 21 categories, and the SIRI-WHU dataset contains 12 categories. For the UC Merced dataset, the average accuracy of our method can reach more than 96%, while for the SIRI-WHU dataset, the average accuracy can reach more than 97%, and the classification accuracy of individual categories can reach 98%, which has obvious advantages. From the experiment, we find that the classification accuracy is higher when there is a large difference between categories in the data set, on the contrary, the classification accuracy is lower. For example, in the UC merged dataset, the classification accuracy of dense residential and medium residential is lower than other categories, and it is easy to misclassification. Further analysis, calculation, and analysis of the intrinsic mean difference within the two categories of Lie Groups, we found that the difference between the two categories is very small, but it is quite different from other categories, which is the root cause of their confusion. In addition, the method we proposed is based on the design and development of Lie Group manifold space. Therefore, only when the data samples meet the requirements of manifold space can the method have more advantages and achieve higher classification accuracy.

4. CONCLUSION

In this study, we proposed a high-resolution remote sensing image scene classification method based on the intrinsic mean of the Lie Group. Compared with the existing methods, which have a complex network structure, multiple parameters, and complex calculations, this method has the advantages of high accuracy, good computing performance, and low characteristic dimensions. In addition, the method solves the problem that non-euclidean space samples cannot be calculated by Euclidean distance. Finally the experiments on UCM and Google Dataset of SIRI-WHU demonstrate the effectiveness of the proposed method.

In the future, we will continue to study the scene classification method based on Lie Group machine learning to further improve the accuracy of the classification and maintain high computational performance.

REFERENCES

- Aggarwal, C. C., 2014. *Data classification: algorithms and applications*. CRC press.
- Buss, S. R., Fillmore, J. P., 2001. Spherical averages and applications to spherical splines and interpolation. *ACM Transactions on Graphics (TOG)*, 20(2), 95–126.
- Cheriyadat, A. M., 2013. Unsupervised feature learning for aerial scene classification. *IEEE Transactions on Geoscience and Remote Sensing*, 52(1), 439–451.
- Dellinger, F., Delon, J., Gousseau, Y., Michel, J., Tupin, F., 2014. SAR-SIFT: a SIFT-like algorithm for SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 53(1), 453–466.
- Fletcher, P. T., Lu, C., Joshi, S., 2003. Statistics of shape via principal geodesic analysis on lie groups. *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, 1, IEEE, I–I.
- Hu, F., Xia, G.-S., Hu, J., Zhang, L., 2015. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. *Remote Sensing*, 7(11), 14680–14707.
- Huang, F., Yan, L., 2015. Hull vector-based incremental learning of hyperspectral remote sensing images. *Journal of Applied Remote Sensing*, 9(1), 096022.
- Jain, A. K., 2010. Data clustering: 50 years beyond K-means. *Pattern recognition letters*, 31(8), 651–666.
- Kaâniche, M.-B., Bremond, F., 2012. Recognizing gestures by learning local motion signatures of HOG descriptors. *IEEE transactions on pattern analysis and machine intelligence*, 34(11), 2247–2258.
- Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R., Wu, A. Y., 2002. An efficient k-means clustering algorithm: Analysis and implementation. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 881–892.
- Kaufman, L., Rousseeuw, P. J., 2009. *Finding groups in data: an introduction to cluster analysis*. 344, John Wiley & Sons.
- Liu, Y., Zhong, Y., Fei, F., Zhang, L., 2016. Scene semantic classification based on random-scale stretched convolutional neural network for high-spatial resolution remote sensing imagery. *2016 IEEE international geoscience and remote sensing symposium (IGARSS)*, IEEE, 763–766.
- MacQueen, J. et al., 1967. Some methods for classification and analysis of multivariate observations. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, 1number 14, Oakland, CA, USA, 281–297.
- Pelleg, D., Moore, A. W. et al., 2000. X-means: Extending k-means with efficient estimation of the number of clusters. *Icml*, 1, 727–734.
- Penatti, O. A., Nogueira, K., Dos Santos, J. A., 2015. Do deep features generalize from everyday objects to remote sensing and aerial scenes domains? *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 44–51.
- Qi, S., Ma, J., Lin, J., Li, Y., Tian, J., 2015. Unsupervised ship detection based on saliency and S-HOG descriptor from optical satellite images. *IEEE Geoscience and Remote Sensing Letters*, 12(7), 1451–1455.
- Ren, J., Jiang, X., Yuan, J., 2016. Lbp-structure optimization with symmetry and uniformity regularizations for scene classification. *IEEE Signal Processing Letters*, 24(1), 37–41.
- Sedaghat, A., Ebadi, H., 2015. Remote sensing image matching based on adaptive binning SIFT descriptor. *IEEE transactions on geoscience and remote sensing*, 53(10), 5283–5293.
- Xiao, B., Wang, K., Bi, X., Li, W., Han, J., 2018. 2D-LBP: an enhanced local binary feature for texture image classification. *IEEE Transactions on Circuits and Systems for Video Technology*.

Yang, W., Yin, X., Xia, G.-S., 2015. Learning high-level features for satellite image classification with limited labeled samples. *IEEE Transactions on Geoscience and Remote Sensing*, 53(8), 4472–4482.

Yang, Y., Newsam, S., 2010. Bag-of-visual-words and spatial extensions for land-use classification. *Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems*, ACM, 270–279.

Yang, Y., Newsam, S., 2011. Spatial pyramid co-occurrence for image classification. *2011 International Conference on Computer Vision*, IEEE, 1465–1472.

Yang, Y., Newsam, S., 2012. Geographic image retrieval using local invariant features. *IEEE Transactions on Geoscience and Remote Sensing*, 51(2), 818–832.

Ye, M., Qian, Y., Zhou, J., 2014. Multitask sparse nonnegative matrix factorization for joint spectral–spatial hyperspectral imagery denoising. *IEEE Transactions on Geoscience and Remote Sensing*, 53(5), 2621–2639.

Zhang, F., Du, B., Zhang, L., 2014. Saliency-guided unsupervised feature learning for scene classification. *IEEE Transactions on Geoscience and Remote Sensing*, 53(4), 2175–2184.

Zhang, F., Du, B., Zhang, L., 2015. Scene classification via a gradient boosting random convolutional network framework. *IEEE Transactions on Geoscience and Remote Sensing*, 54(3), 1793–1802.

Zhang, J., Suo, Z., Li, Z., Zhang, Q., 2018. DEM Generation Using Circular SAR Data Based on Low-Rank and Sparse Matrix Decomposition. *IEEE Geoscience and Remote Sensing Letters*, 15(5), 724–728.

Zhao, B., Zhong, Y., Xia, G.-S., 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.

Zhao, L.-J., Tang, P., Huo, L.-Z., 2014. Land-use scene classification using a concentric circle-structured multiscale bag-of-visual-words model. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(12), 4620–4631.

Zhong, Y., Zhu, Q., Zhang, L., 2015. Scene classification based on the multifeature fusion probabilistic topic model for high spatial resolution remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 53(11), 6207–6222.

Zhu, Q., Zhong, Y., Zhao, B., Xia, G.-S., Zhang, L., 2016. Bag-of-visual-words scene classifier with local and global features for high spatial resolution remote sensing imagery. *IEEE Geoscience and Remote Sensing Letters*, 13(6), 747–751.

Revised January 2020