Soil Erosion Image Segmentation Based on Improved K-means clustering method

Xuanzhang Song[a],Liujianqiang and QiongyanLi[b]*

School of Engineering,Beijing Forestry University, Beijing, 100083, China

Email: [a]381365211@qq.com,[b]liqiongyan@bjfu.edu.cn

* Corresponding authors

Keywords:*soil erosion; images segmentation; improved K-means clustering method*

Abstract:A method was developed for soil erosion image segmentation based on improved K-means clustering in order to solve some problem with the traditional K-means clustering method. First, the peaks was detected in the smoothed histogram of a gray-scale image, and sort peaks in descending order; then the number of clusters K is determined according to the number of main peaks in the smoothed histogram, while the grey value of a peak is selected as a cluster center; finally, the weighted Euclidean distance was applied to measure the similarity instead of simple Euclidean distance. Experiment results show that the improved K-means clustering method can not only shorten the process of convergence to the object but also get a more reasonable clustering result. It is effective and suitable for soil erosion images segmentation.

Introduction

Soil erosion is a major environmental and agricultural problem worldwide. The establishment of a scientific and reasonable monitoring mechanism for the effective measurement of soil erosion, is the key to prevent and control the soil erosion. But the traditional measurement methods such as tape, erosion needle, wind erosion circle, topography needle plate, are time-consuming, labor-intensive, low efficiency, and the measurement accuracy is low subject to the complexity of erosion. In recent years, researchers have built a variety of remote sensing platform for aerial photography level monitoring.

Ries and other researchers in 2003 used the hot air balloon for remote sensing for the Spanish Eblo Basin, at different hot air balloon flight height (10 ~ 300 m) and multi lens focal length (50,28 mm) [1].The method can be used to monitor soil erosion in a range of 1: 100 to 1: 10000 scale. Marzolff[2] and other researchers used flying a kite at a speed of 40m as a remote sensing platform. A semi-arid areas of soil erosion ditch in Spain were monitored by using photography equipment,and high-resolution DEM information was obtained, and the monitoring error is within 0.5 grid cells.

The method of soil erosion measurement based on remote sensing technology has the advantages of high speed and high efficiency of data acquisition, which is convenient for long-term monitoring. In addition, the measurement accuracy of remote sensing technology can be quantified, but the technology of indoor work is heavy and technically demanding. However, high-resolution remote sensing technology is greatly affected by climate and environmental factors, and the quality of the image information obtained by remote sensing cannot meet the demand. And the aviation remote sensing technology is highly specialized, involving hardware, software facilities, technical content is very high, so the measurement cost will increase[3-5].

Image-based measurement has been used in several area, Image-based soil erosion monitoring can improves the monitoring efficiency and reduces the intensity of the work[6]. Soil erosion image segmentation is to extract the target features from images. There are many types of soil erosion, and each erosion type exists under various terrain condition, covered by vegetation and grass, so that it is difficult to be identified. This poses a challenge to the image segmentation.

Clustering techniques are widely used in image segmentation [7-10]. Image segmentation based on clustering algorithm divides the spatial pixels into spatial data points, then divides the dataset according to the distance metric function. Finally, the set of points is mapped back to the image space to get the segmented images.

In the clustering algorithm, K-means mean method, fuzzy C-means clustering (FCM) algorithm are commonly used. In this paper, an improved K-means clustering algorithm is applied to the soil erosion image segmentation in order to get a good segmentation result.

Method

K-means Clustering.K-means clustering[7] (MacQueen, 1967) is a method commonly used to automatically group a data set into k groups by selecting k initial cluster centers and then iteratively running them as follows:

First, each instance di is assigned to its closest cluster center;

Second, each cluster center Cj is updated to be the mean of its constituent instances;

Finally, the algorithm converges when there is no further change in assignment of instances to clusters.

K-means algorithm is to find k centers of a dataset. The minimized objective functioncan be described as (1):

$$J(X,C)=\sum_{i=1}^{k}\sum_{x_j\in S_i} d(x_j,c_i)$$

$$d(x_j,c_i)=\left\|x_j-c_i\right\|_2=\left(\sum_{l=1}^{d}\left|x_{jl}-c_{il}\right|^2\right)^{\frac{1}{2}} \qquad c_i=\frac{1}{n_i}\sum_{x_j\in c_i} x_j$$

$$X=\{x_1,x_2,\mathbf{L}\,x_n\}\in R^d \qquad x_j=\{x_{j1},x_{j2},\mathbf{L}\,x_{jd}\}^T \qquad C=\{c_1,c_2,\mathbf{L}\,c_k\}^T \tag{1}$$

Where Xis a dataset with n samples Xj , which are vectorsin d dimension, Ciare the centers of X.is the Euclidean distance between the sample Xj and the center of cluster Ci.

Problems with K-means Clustering.K-means clustering algorithm normallyneeds userto specify the number of clusters before clustering.However, in most cases, the actual number of clusters is unknown, and it is not clear to the user how many clusters should be generated due to lack of experience or other reasons. If the given number of clusters is too large, it is easy to make the clustering result too complex and difficult for further processing. If it is too small, it will also causeclusteringeffect is too simple to interpret. Therefore, a rational choice the number of clusters is critical.

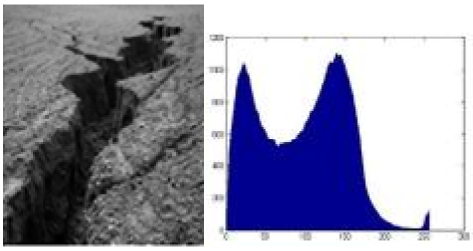
K-means clustering algorithm randomly chooses its initial clustering center, which leads to the possibility of different clustering results in unstable clustering results, and may not get the desired clustering; and it may also result in convergence of clustering only to local optimal solutions. Therefore, an improved algorithm is proposed to solve this problem.

The similarity measurement directly determines the quality of clustering. It is key to choose the appropriate metric function. The degree of similarity is usually described by the physical distance between objects, which is spatial, temporal, and density. If we only cluster the data according to the spatial distance, it is difficult to get a reasonable clustering result. So the distance should be a polysemy distance rather than a single. Ordinary K-means clustering algorithm uses only Euclidean distance to evaluate the similarity without considering the density attribute.

In summary, traditional K-means clustering algorithm is easy to fall into the local optimum, but not the global optimal solution because of the problems mentioned above. In order to solve the problems, the improved K-means clustering algorithm are proposed. The major contributions of the current work are twofold. First, we have developed a method to determine the number K and centers of cluster based on the histogram of an image. Second, the similarity measure method is improved by using the weighted Euclidean distance.

Improved K-means Clustering method. The basic idea of the improved method is to detect the peaks in the histogram of the gray-scale image, and sort peaks in descending order. The number of clusters K is determined according to the number of main peaks in the histogram; while the grey value of a peak is selected as a cluster center.

As shown in Fig. 1, Fig. 1(a) is the original image, which is first grayed and then its histogram can be obtained, Fig. 1(b) is its smoothed histogram. The cluster number is determined as 2 according to its histogram, the grey value of the peaks are selected as the initial cluster centers.



(a) (b)

Figure 1. Histogram ofan erosion image

For the improved similarity measure method The improved K-means clustering algorithm uses the weighted Euclidean distance to calculate the similarity.The weighted Euclidean distance is given as (2):

$$d(x,y)=\sqrt{w_1(x_1-y_1)^2+w_2(x_2-y_2)^2+\mathbf{L}+w_n(x_n-y_n)^2} \quad (2)$$

Where $x_1, x_2, \mathbf{L}, x_n$ and $y_1, y_2, \mathbf{L}, y_n$ are two n dimensional vectors. The weight w_i is calculated as (3):

$$w_i = v_i \Big/ \sum_{i=1}^{j} v_i \qquad v_x = S_x / |\overline{x}| \qquad S_x = \left(\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{x})^2 \right)^{\frac{1}{2}} \qquad \overline{x} = \frac{1}{n} \sum_{i=1}^{n} X_i \quad (3)$$

Where v_i is thecoefficient of variation. $\overline{X}$, $\boldsymbol{S_i}$are the mean and variance respectively.

As mentioned above, the improved K-means clustering method selects the cluster number and centers based on the histogram of image at the beginning, and the weighted Euclidean distance is applied to measure the similarity instead of simple Euclidean distance, which can not only shorten the process of convergence to the object but also get a more reasonable clustering result.

Experiment result

Pipeline of segmentation.The pipeline of segmentation is shown in Fig.2:

(1) Gray the input image to a grayscale image;

(2) Get the histogram of the grayscale image, and smooth the histogram;

(3) Detect the peaks of the histogram, and sort the peak set in descending order and the initial clustering centers and the cluster number are obtained.

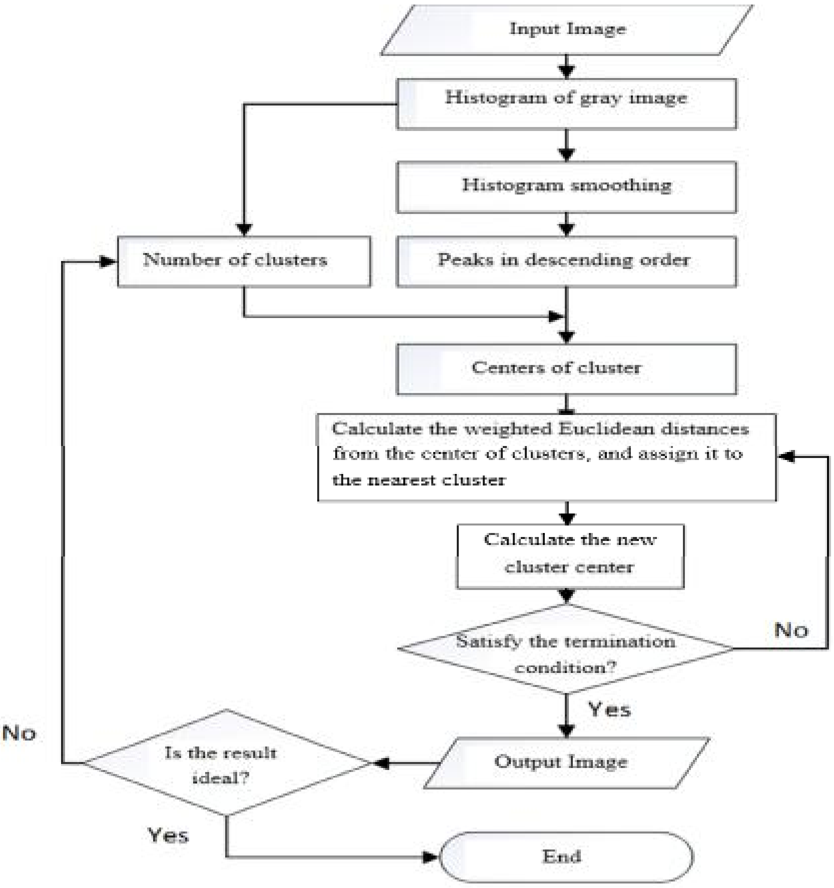

Input Image
Histogram of gray image
Histogram smoothing
Number of clusters
Peaks in descending order
Centers of cluster
Calculate the weighted Euclidean distances from the center of clusters, and assign it to the nearest cluster
Calculate the new cluster center
Satisfy the termination condition?
No
Yes
No
Is the result ideal?
Output Image
Yes
End

Figure 2. Pipeline of erosion image segmentation based improved K-means clustering method

(4) According to the initial clustering center, the K-means clustering algorithm with weighted Euclidean distance as the criterion is applied in clustering the image in order to separate the target from the background.

(5 After the iteration, the final segmentation result is obtained by clustering.

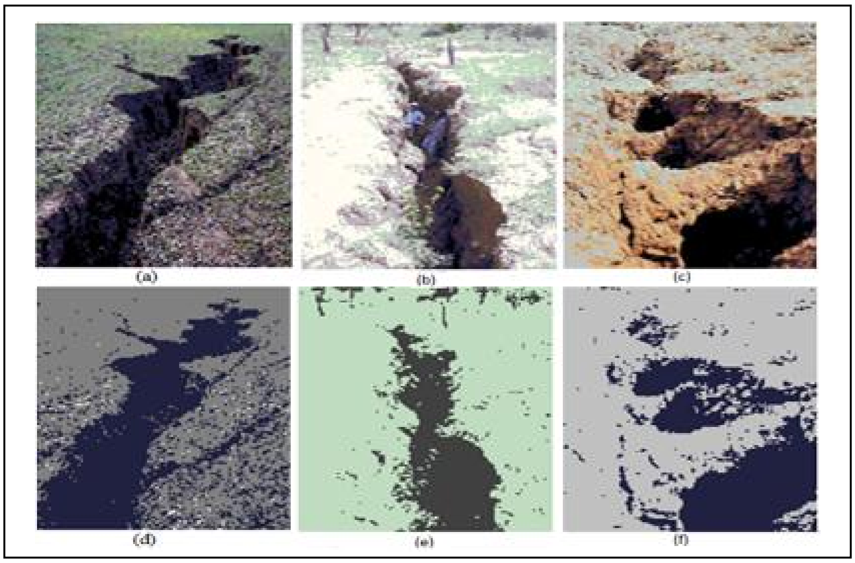


Figure.3samples of erosion image segmentation based on improved K-means clustering method

Experiment sample results. Fig. 3 shows few examples of erosion image segmentation based on our improved K-means clustering method. The original erosion images are shown in Fig. 3(a-c), which are from internet, and Fig. 3(d-f) are their segmentation results respectively based on improved K-means clustering method mentioned above. The cluster number is selected as 2 for the images based on their histograms. Experiment shows that the improved K-means clustering method for soil erosion image segmentation is reliable and effective, can catch target features well from the images.

Our experiments also approve that the improved K –means clustering method reduces the computing time, avoids the instability caused by random selection of cluster centers, and eliminates the possibility that the algorithm obtains the local optima rather than the global optimal solution. After a large number of image processing validation, the improved algorithm can reduce the number of iterations, is effective for soil erosion images segmentation.

Conclusion

An improved K-means clustering algorithm is proposed, which solves some problems encountering with the traditional K-means clustering method: the number of given clusters, the problem of randomly selecting the initial cluster centers and the simple similarity measure function. The main improvements of our method are the follows:

(1) The number of clusters is determined according to the number of peaks inthe gray smoothed histogram of an image.

(2) The gray levelsof the main peaksinthe smoothed histogram of image are selected as the initial clustering centers.

(3) The similarity measure function of traditional K-means clustering algorithm is improved, and the weight Euclidean distance is used for the similarity measure.

Experiments show that the improved K-means clustering method is effective and suitable for soil erosion images segmentation. In the future, we will use more intelligent method to complex soil erosion images.

Acknowledgements

This work is financial supported by the project: Innovative Training Program for College Students in Beijing.

References

[1] Ries J B，Marzolff I. Monitoring of gully erosion in the Central Ebro Basin by large-scale aerial photography taken from a remotely controlled blimp[J]. Catena，2003，50(2-4)：309-328.

[2] Marzolff I，Poesen J. The potential of 3D gully monitoring with GIS using high-resolution aerial photography and a digital photogrammetry system［J］. Geomorphology，2009，111(1-2)：48-60.

[3] Perroy R L，Bookhagen B，Asner G P，et al. Comparison of gully erosion estimates using airborne and ground-based Li-DAR on Santa Cruz Island，California[J]. Geomorphology，2010，118(3-4)：288-300.

[4] T.J.Toy, G.R.Foster，K.G.Renard. Soil Erosion: Processes, Prediction, Measurement and Control [M].American: John Wiley & Sons，2002:50-200.

[5] Lichun S. Processing of laser scanner data and extraction of structure lines using methods of the image processing[J]. Acta Geodaetica et Cartographica Sinica，2004,33(1) :63-70.

[6] Jianqiang Liu**, Qiongyan Li*,** Zhongdong Yin*, Min Chen and Dunmin Lu, A Primary Study on Gully Erosion Area Estimation Based on Images, International Journal of Earth Sciences and Engineering, v 9, n 2, p 574-588, 2016

[7] J. MacQueen. (1967) Some methods for classification and analysis of multivariate obser-vations. In Proc. 5th Berkeley Symp. Math. Stat. Prob. 3:281.

[8] Dong J, Qi M. K-means Optimization Algorithm for Solving Clustering Problem[C]. In: Proceedings of the 2nd International Workshop on Knowledge Discovery and Data Mining, Moscow, 2009: 52-55.

[9] Klein R W, Dubes R C. Experiments in Projection and Clustering by Simulated Annealing [J]. Pattern Recognition, 1989, 22(2): 213-220.

[10] Laszlo M，Mukherjee S. A Genetic Algorithm that Exchanges Neighboring Centers for K-means Clustering[J]. Pattern Recognition Letters, 2007, 28(16): 2359-2366.