Intelligent identification of rock samples

Abstract

In oil and gas exploration, rock sample identification is a basic and important link. In order to accurately extract color, texture, granularity and other features in rock to identify lithology, the image is enhanced by histogram equalization transformation, and then multi-level Gaussian blur is adopted,SIFT algorithm is used to realize feature matching and extract feature points There is diversity in rock samples, which is affected by occlusion, illumination and visual angle,In this paper, Ada-Boost algorithm and prm algorithm combined with PCA are used to reduce the dimension of training set, test set and pictures to be tested,The algorithm can finally upgrade a weak classifier to a strong classifier. In order to calculate the percentage content of oil-bearing area of rock, homomorphic filtering is used to increase contrast and standardize brightness, and histogram equalization is used to enhance image contrast, make the illumination of details more obvious, and achieve the purpose of enhancing detail features,Moreover, the effect of histogram equalization is stronger than that of high-pass filter. Then, with the help of pixel calculation, the whole area of the picture is calculated, and then assisted by image segmentation algorithm, and then the area with oil content is calculated. In this paper, the recognition system makes full use of image recognition technology and deep learning algorithm, and introduces AdaBoost algorithm into the modeling of rock sample model, thus successfully realizing intelligent recognition of rock samples.

Key words: **intelligent recognition, histogram equalization, AdaBoost algorithm, PCA, SIFT algorithm, homomorphic filtering**

Catalogue

I Mining target _____ *1*

II Analysis method and process _____ *1*

2.1 Analytical methods and processes of questions: _____ **1**

2.1.1 Histogram equalization _____ 1

2.1.2 SIFT algorithm _____ 2

2.1.3 Summary _____ 3

2.2 Analysis and Process of Method 2 of Question 1 _____ **3**

2.2.1 Flow chart _____ 3

2.2.2 Adaboost algorithm and pca dimension reduction _____ 3

2.2.3 Prm random sampling _____ 5

2.2.4 Summary _____ 5

III Analysis method and process of Question 2 _____ *6*

3.1 Flow chart _____ **6**

3.1.1 Homomorphic filtering/histogram homogenization _____ 6

3.1.2 Calculation of pixel points _____ 8

3.2 Question 2, Method 2, Analysis and Process. _____ **8**

3.2.1 Image segmentation and Area calculation _____ 8

IV Conclusion _____ *8*

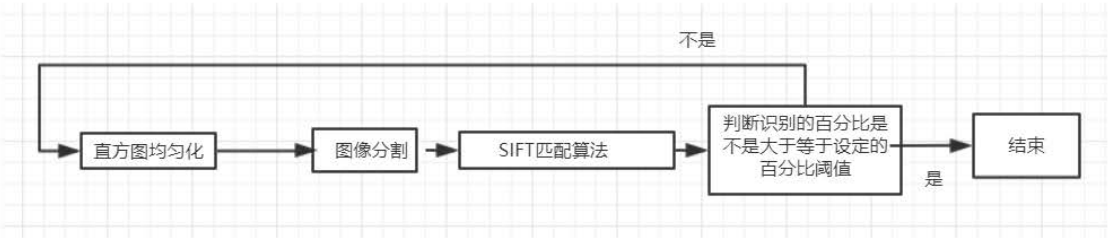
References _____ *9*

I Mining target

In the data set rock, the data with "white light/fluorescence" label "1" is the rock sample image data shot under the same white light environment, and the data with "white light/fluorescence" label "2" is the rock sample image data shot under the same fluorescence environment. This modeling mainly uses PCA-based Ada-Boost algorithm to achieve the following two goals:

1. Build an intelligent identification model for lithology of rock samples. Accord to that rock data set, the lithology intelligent identification and classification of rock samples are realize.
2. Oil has luminous characteristics under ultraviolet irradiation, and the green or yellow part of photos taken under fluorescent lamps contains oil,then calculate the percentage of oil-bearing area of rock.

II Analysis method and process



2.1 Analytical methods and processes of questions:

2.1.1 Histogram equalization

Affected by light, angle, visual angle, etc., the pictures taken by remote sensing technology will have the problems that the feature points of the pictures are not obvious enough and the features are not prominent,Facing this problem, this paper adopts histogram equalization method to enhance the pictures to achieve the purpose of enhancing the detail features. Histogram equalization algorithm is a common and classical effective image enhancement algorithm, which redefines the distribution of gray value by function, and then changes the contrast to achieve the purpose of image enhancement.

Consider a discrete grayscale image {x}

n_i Let И show the number of occurrences of gray I, and the probability of occurrence

of pixels with gray I in images is P_x（i）=P（x=i）=$\frac{n_i}{n}$ 0≤i<L

L is the number of all gray levels in the image

The cumulative integral function corresponding to P_x is defined as:

c df_x(i)=$\sum_{j=0}^{i} P_x$ (j) x is the cumulative normalized histogram of images

We create a transformation in the form of y=T(x), which produces a y for each value of the original image, so that the cumulative probability function of y can be linearized in all values,The transformation formula is defined as:

c df_y(i)=ik

For the constant k, the property of K,CDF allows us to make such a transformation, which is defined as: c df_x($y^{'}$)= c df_y(T(k))= c df_x(k) k ∈ [0, L) note that T maps different gray levels to (0,1),In order to map these values to the initial domain, the following simple transformation needs to be applied to the results:
y '= y*(max{x}−min{x})+min{x}

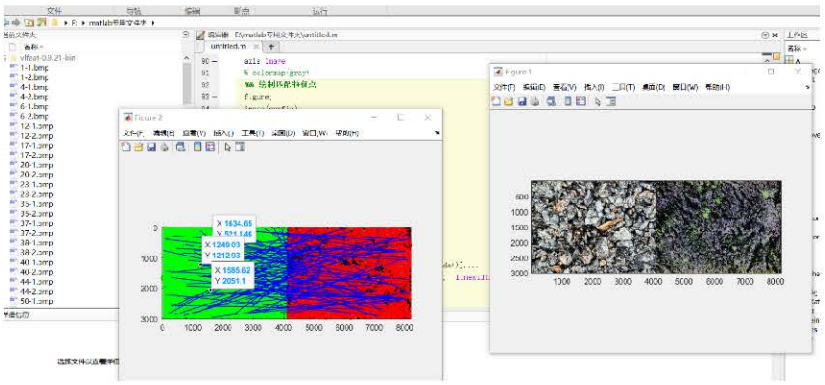
In order to extract feature points of pictures, we first use SIFT algorithm to realize feature point matching. First, we locate the key points and determine the feature direction, and then find several pairs of matching feature points by comparing the feature vectors of each key point in pairs.

2.1.2 SIFT algorithm

Overview of SIFT algorithm:
SIFT (Scale Invariant Feature Transform) is the full name of Scale Invariant Feature Transform,SIFT operator describes the feature points detected in an image with a 128-dimensional feature vector, so an image is represented as a 128-dimensional feature vector set after SIFT algorithm,This feature vector set has the characteristics of invariant image scaling, translation and rotation, and is also invariant to illumination, affine and projection transformation, which is a very excellent local feature.
There are the following steps

1. Scale space pole detection
2. Precise positioning of key points
3. Determine the direction of key points
4. Generating feature vectors

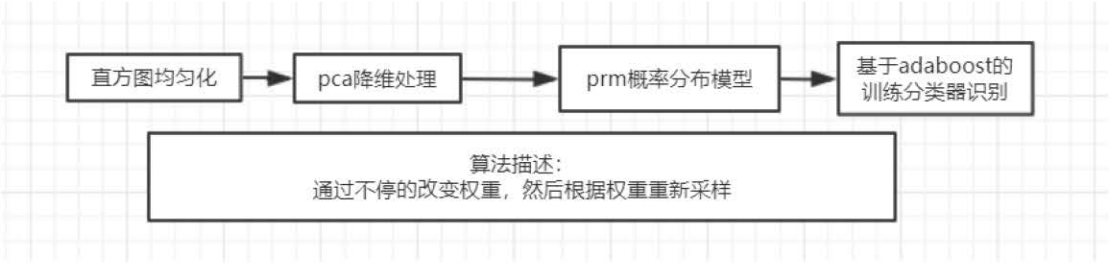
2.1.3 Summary



It can be preliminarily identified by calculating the matching points per unit rock area, but the calculation unit is a bit large and the time is slow, which is not suitable for large-scale identification.

2.2 Analysis and Process of Method 2 of Question 1

2.2.1 Flow chart



2.2.2 Adaboost algorithm and pca dimension reduction

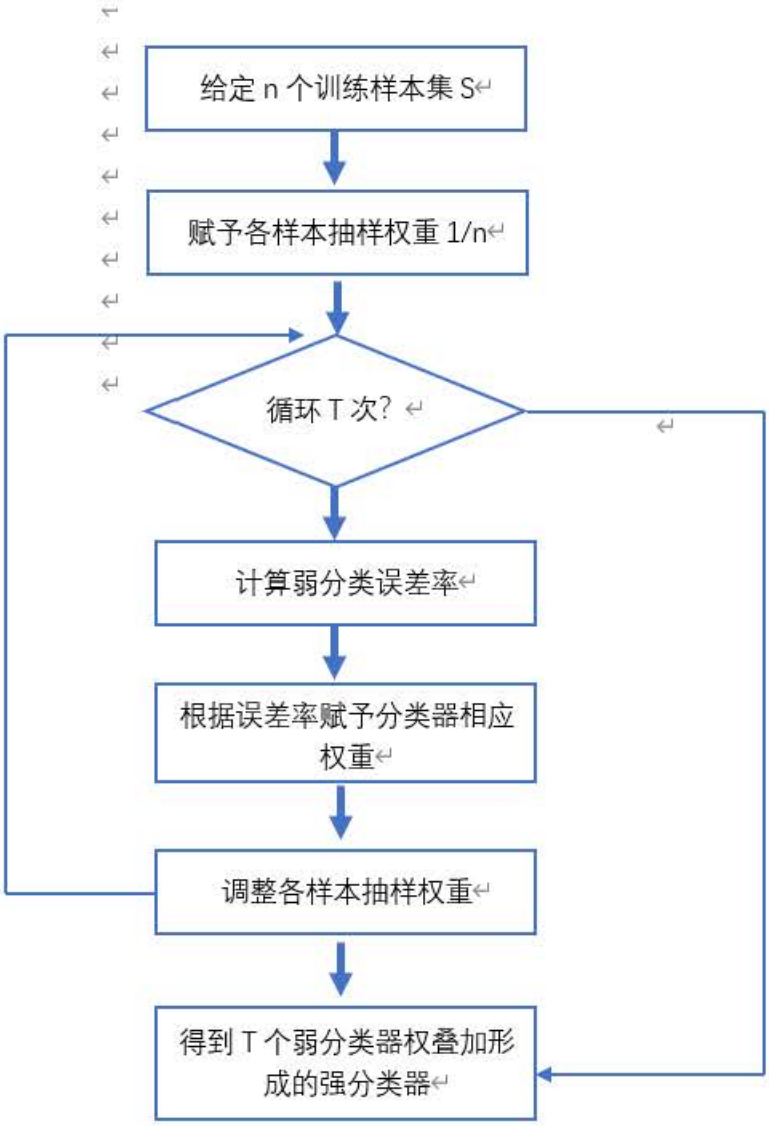
We use PCA algorithm and Adaboost algorithm to train samples.
We now transform the picture into the corresponding data matrix, and then into the corresponding vector representation,This vector has high dimension and great complexity. In order to reduce the dimension without losing most information of the picture, we use PCA algorithm to reduce the dimension of the picture. We selected M-span pictures and recorded them as $X_1X_2\ldots\ldots X_m$.Using K-L transformation to get a low-dimensional subspace representing the original picture,

$$\overline{X} = \frac{1}{M}\sum_{i=1}^{m} X_i$$

A new vector set can be obtained from the sample, A=($A_1A_2\cdots\cdots A_m$)The mean value of set A is zero, and its covariance matrix is:

$$C=\cos(A)=\frac{1}{M}AA^T$$

The matrix in the formula retains some orthogonal bases according to the eigenvalues, and finally it is restored to an image by computer and displayed. We use prm to build probability distribution model. Firstly, read the image, set up the starting point, target point and random number k, display the starting point and ending point in the form of random number, divide the image into m areas, and scatter the same number of random sections in each area. End when the summary points reach K .. Read the line segments between all nodes. According to the A^* algorithm, the cost of initial track inspection is defined, and the travel time and the shortest track length of the last reading sequence are determined. Then, Ada-Boost algorithm is used to recognize the image. Ada-Boost algorithm adjusts the sampling weight of each sample according to a certain strategy, and accumulates these weak classifiers according to the weight, so as to obtain a strong classifier with each weak classifier accumulated. Ada-Boost algorithm specific steps:



Finally, the strong classifier is used to detect the projected test picture and output the judgment result.

2.2.3 Prm random sampling

In the above figure, prm random sampling is used to find the best weight for identification.

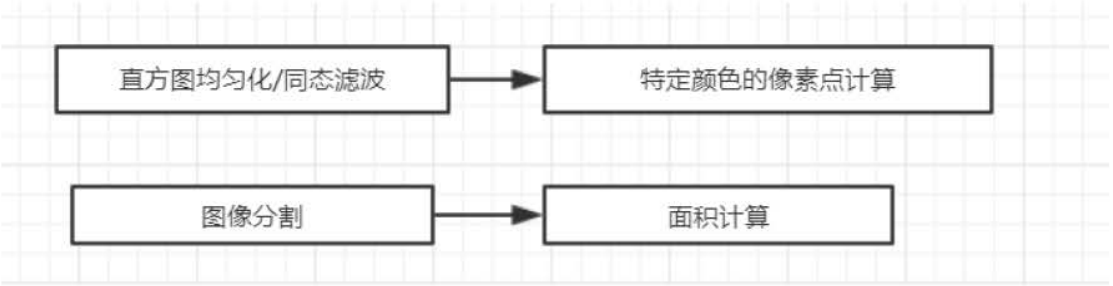
2.2.4 Summary

```
命令行窗口
Warning: numvecs is 50; only 3 exist.
Warning: numvecs is 50; only 3 exist.
CLASSNUM = 2, t=1, nRightCount=7, pseudo_loss=1.250000e-01, bt = 1.428571e-01
CLASSNUM = 2, t=2, nRightCount=5, pseudo_loss=2.142857e-01, bt = 2.727273e-01
t=1, DIM= 50, RightCount=7, RightRatio=8.750000e-01
t=2, DIM= 50, RightCount=7, RightRatio=8.750000e-01
t=2, DIM= 50, RightCount=7, RightRatio=8.750000e-01
```

First, 30 groups are set up for recognition, each group has eight pictures,By setting weights through continuous sampling, the accuracy rate is still very high, and the error rate is 10%,If more clear processing is added, it is estimated that it can be 6-7% higher.

III Analysis method and process of Question 2

3.1 Flow chart



3.1.1 Homomorphic filtering/histogram homogenization

When we process the image, there will be a situation that the dynamic range is very large, but a part of the gray scale of the image is very dark and the range is very small,In order to compress the brightness range of the image and enhance the contrast of the image in the frequency domain, we introduce homomorphic filtering to eliminate the influence of uneven illumination without losing the image details.

Firstly, the image is expressed as the product of the illumination component and the reflection component

$m(x,y)=i(x,y)\cdot r(x,y)$ m is the image, i is the illumination component and r is the reflection component

The logarithm $\ln(m(x,y))=\ln(i(x,y))+\ln(r(x,y))$

Then Fourier transform is performed to linearly separate the illumination component and the reflection component.

$F\{\ln(m(x,y))\}=F\{\ln(i(x,y))\}+F\{\ln(r(x,y))\}$

In this paper, F{ln(m(x,y))} is defined as M(u,v)

High-pass filtering is carried out on the image, so that the illumination of the image is uniform, the high-frequency component increases, and the low-frequency component decreases

$N(u,v)=H(u,v)\cdot M(u,v)$ H is a high-pass filter

In the next step, inverse Fourier transform is performed on N, and the image is transferred from frequency domain to time domain

$n(x,y)=F-1\{N(x,y)\}$

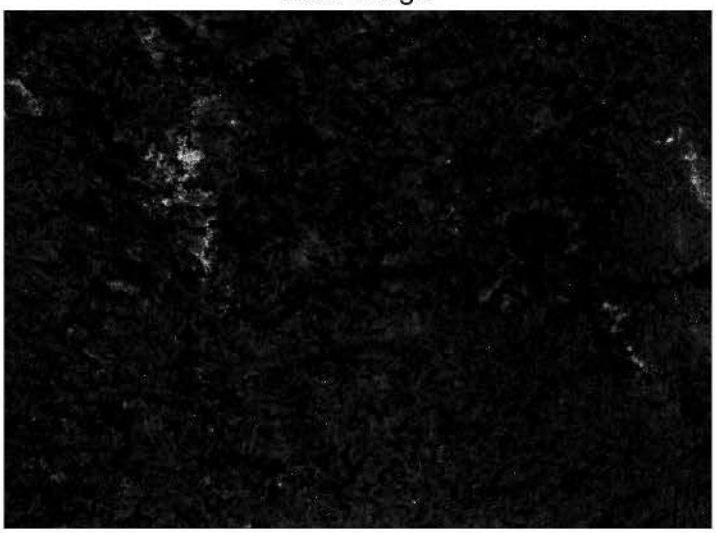
Finally, the logarithm taken at the beginning is restored by exponential function

$m'(x,y)=\exp\{n(x,y)\}$

In the end,we get the result.

Homomorphic filtering

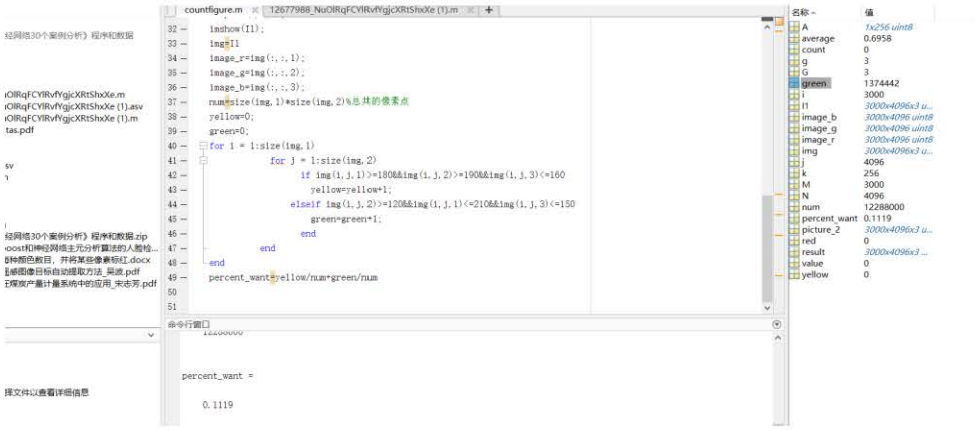
Raw Image



Histogram equalization

It can be seen from this that when looking at the distribution of rock oil area, histogram equalization can show the green part for calculation.

3.1.2 Calculation of pixel points



The picture shows that the oil area in this picture accounts for 11.9%

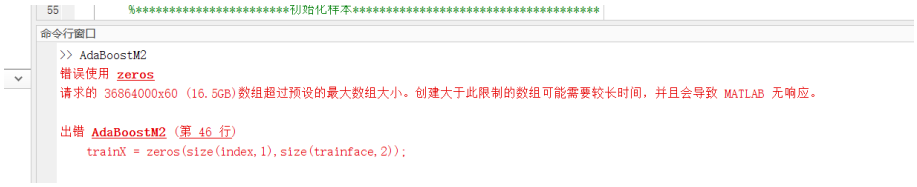
3.2 Question 2, Method 2, Analysis and Process.

3.2.1 Image segmentation and Area calculation

Image segmentation is one of the most critical steps in image processing,By searching every boundary in the image area and detecting according to the gradient change of the image, false edges are avoided,In order to solve the problem of uneven gray scale, we add fuzzy membership function into the offset field model, and integrate Canny operator and offset field into the double level set model, so as to realize the segmentation of images with complex boundaries and uneven gray scale, and extract useful information. Canny operator has high accuracy in processing image edge information, which is a standard multi-level detection algorithm, in which Gaussian filter is used to remove noise, so as to improve the segmentation accuracy, and then the gradient intensity and direction are calculated,Here, we use horizontal and vertical direction templates to ensure the detection accuracy,After template detection, the gradient edge of the image has multiple pixels wide,We use non-maximum suppression method to keep the maximum gradient, so as to get a clear gradient map.
If such treatment can be used, the calculation of oil area capacity will be more accurate.

IV Conclusion

Nowadays, compression and recognition algorithms are inextricably linked with the new generation of hardware,With the advent of cloud computing, faster and better algorithms will emerge constantly, and the situation that the array in the figure below is too large to handle will not appear.



References

[1] Chengjun Liu and Harry Wechsler, "Robust Coding Schemes for Indexing and Retrieval from Large Face Database", IEEE Trans. Image Processing, vol.9, 132-137, 2000

[2] Yoav Freund and Robert E.Schapire, "A Decision-Theoretic Generalization of On-line Learning and an Application to Boosting, Journal of computer and system sciences,55.119-139 (1997)

[3] Song Zhifang Shanxi Institute of Automation [A] Application of Raw Coal Identification Mode in Coal Output Measurement System

[4] Hu Song, Huang Zhiyuan, Deng Lei, Li Zhuanghao, Fan Jieying, Zeng Wei (Chengdu University of Technology, Chengdu, Sichuan 610000)

[5] Li Xiangjian，Zhu Jiaming，Xu Tingyi。Bilevel Set Medical Image Segmentation Based on Improved Canny Opera tor[J].Radio Communications Technology,2021,47(2):226-231