Application of Improved K-means Algorithm in Microvadose Image Segmentation

Dahui Qin
School of Civil Engineering and Architecture, Southwest Petroleum University, China
e-mail: qindahui@qq.com

Haijun Liu
School of Civil Engineering and Architecture, Southwest Petroleum University, China
e-mail: swpulhj@163.com

Jianjun Liu
State Key Laboratory of Oil and Gas Reservoir Geology and Exploitation (Southwest Petroleum University), Chengdu, China;
e-mail: 315245690@qq.com

ABSTRACT

Water flooding microscopic seepage experiment is an effective method to study the microvadose mechanisms as well as the distributions of remaining oil. Throughout the experiment the quantitative description of the porous medium fluid flow parameters such as porosity, oil saturation and so forth is crucial. A quantitative analysis method is proposed via image processing, including image preprocessing, image feature extraction, improvement of K-means algorithm, segmentation of flooding water, binding water, remaining oil and pores along with the calculation of seepage parameter, etc. An example is provided to demonstrate the validity of this method. Furthermore, microscopic seepage images can be divided automatically, accurately and efficiently.

KEYWORDS: Color Segmentation; K-means; Lab color space; porosity; microscopic seepage

INTRODUCTION

Hydrocarbon reservoir porosity is the main research object in petroleum geology, oil and gas field exploration or development. With the development of the petroleum industry, Hydrocarbon reservoir porosity play a more and more important role, because it not only determines the amount of oil and gas reserves, but also directly affect the capacity and ultimate recovery factor of the oil and gas wells [1]. Master the reservoir porosity is the basis to understand reservoir storage, confirm the boundaries of main layer, effective thickness and interlayer physical, estimate reserves, analyze oilfield production situation. Therefore, porosity is the fundamental constants which must be
grasped for the oilfield exploration and development. Porosity not only is critical for the calculation of oil saturation, but also has very vital significance for reserve estimates of oil, natural gas and other resources, especially future development plans of the country.\textsuperscript{[2]}

Whereas the porosity is so important, many domestic and foreign scholars have studied on porosity, and also have put forward many ways to calculate porosity. There are many traditional methods such as Gas Bohr's law method, weighing method and theoretical calculations. The first one is mainly via detecting the change in pressure or volume of gas that pass through the rock to determine rock porosity. But it has some disadvantages: measuring instruments requires high sealing, the apparatus is expensive and bulky, measured slowly, high power consumption. The weighing method for measuring is manual, using electronic or mechanical balance, by manually recording and calculation to obtain the porosity. Generally, this method has complicated procedures, large errors and low efficiency \textsuperscript{[3]}. The last method is combined logging data and empirical formulas to calculate porosity, but it also has large error \textsuperscript{[4]}.

In oil exploitation, microscopic flow simulation technology plays a very important role because it needs to research the microscopic structure of the rock in order to study the seepage condition of subsurface fluid and make clear the distribution of remaining oil. The traditional means adopts the microscopic observation, using experience and the theory of mineral petrology for qualitative analysis. Nevertheless, this approach is still insufficient, namely, unable to accurately identify the pores as well as detailed describe or characterize the porous structure \textsuperscript{[5]}. Using image processing technology is able to accurately identify the pores which cannot recognize by naked eye, extract the effective pore feature. Moreover, it can accurately obtain porosity, oil saturation and other important parameters.

The image processing technology has been applied to micro seepage image, which started in the 1980s. For micro seepage image, the segmentation of remaining oil and pores, both at home and abroad, are generally done by setting threshold manually. However, the research to achieve automatic segmentation through some involved algorithms is rare \textsuperscript{[6]}. Jiucheng Zhu et al. have introduced the image processing technology into multiphase flow experiments, and studied on seepage image processing system \textsuperscript{[7]}. Zimin Wang and Shouyu Xu et al. have studied the applications of image processing in the rock pore structure analysis. Nevertheless, the gray-level threshold segmentation algorithm used applies only to gray images and the results are not good \textsuperscript{[5]}. Shunkang Zhang and Jian Hou et al. have achieved automatic segmentation of micro seepage images by image processing, but the segmentation algorithm also applies to gray images and the segmentation accuracy is not high \textsuperscript{[6]}. Jian Hou and Zhenquan Li have done research on real-time acquisition and quantitative description of flowing-foam texture images in porous media, which used indicator kriging method for image segmentation \textsuperscript{[8]}. Ying Zhang and Puchun Chen have applied K-means algorithm to separate the rock casting image \textsuperscript{[9]}. Although the algorithm is suitable for color image, the segmentation accuracy is not high, even the algorithm is lack of stability. Those image segmentation algorithms most researchers used are aimed at gray image segmentation, so it is difficult to divide color flow images effectively and accurately.

From the above, in order to realize automatic segmentation of flooding water, binding water and remaining oil, this paper will introduce the improved K-means algorithm, and then calculate seepage parameters.
COLOR IMAGE SEGMENTATION

Image segmentation is the first step in image analysis, is also one of the oldest and the most difficult issues in image processing, and the results determine the final analysis quality \[10\]. The so-called image segmentation is to represent images as a collection of connected areas that is physically meaningful, namely, of which the gray or color is substantially same, or albedo, texture and other properties is similar \[11\].

In recent years, the commonly color image segmentation method used can be roughly divided into the following six kinds: (1) Based on the histogram threshold value, (2) Segmentation method based on region, (3) Model-based segmentation method, (4) Segmentation method based on edge detection, (5) Segmentation method based on clustering, (6) Based on a specific theory. With the continuous development of science and technology, the idea, research methods and techniques of different disciplines are mutual integrated. Consequently, technologies of other scientific areas can also be used in the field of color image segmentation. Resulting in more color segmentation method are proposed, including these ways that based on the theory of fuzzy set and logic, artificial neural network technology, genetic algorithm, physical technology as well as the theory of mathematical morphology, etc. \[12\]. Among these methods, clustering algorithm has been widely applied. Many scholars have used clustering algorithm to segment the image and achieved good results. K-means clustering algorithm is a relatively mature clustering analysis method, which has been successfully applied to lots of fields such as medical image processing, facial image segmentation, remote sensing image analysis and so on \[13\].

K-MEANS CLUSTERING ALGORITHM IMPROVED

In the clustering methods, K-means clustering algorithm is one of the most widely used clustering analysis method, besides, it has been widely used in image segmentation due to the characteristics of intuitive, fast and easy to implement \[14\].

To K-means algorithm, the conventional optimization algorithms are mainly focused on the selection of clustering number and class center, that is, through some tests clustering validity function to calculate optimal cluster number \( K \), and then to optimize the segmentation results. In recent years, some studies have shown that considering multiple image features is conducive to obtain better segmentation results. As presented in \[15\], taking into account the spatial characteristics of the pixel, the algorithm has better robustness.

Image Features Extraction

There are a lot of image features, but this paper mainly discusses the color and texture which most affect the segmentation results.

The Extraction of Color Features

Color features are the most widely used visual features in image segmentation, yet k-means clustering algorithm is feature space clustering. Therefore, choosing a different color space, there must be different segmentation results as adopt k-means algorithm for image segmentation \[16\].

The role of several color space in image segmentation are described and compared in \[9, 10, 17\], such as RGB, YUV, UIQ, CIELAB, HIS and CIELUV, each one has advantages and disadvantages. In this paper, we adopt the Lab color model. In this mode, the brightness and color information is stored separately, which means the brightness channel will remain unchanged when
adjust color channels. Thus L channel can be seen a gray-scale version of the image, which holds
details of the image, so using L channel can readily distinguish the particulars between light and
shade in natural images. What is more, Lab model has a wide color gamut, not only contains all
the color gamut of RGB, but also makes up for the uneven color distribution of the RGB color
model. Conversion from RGB to Lab can be expressed as [9, 13]:

\[
\begin{bmatrix}
X \\
Y \\
Z \\
\end{bmatrix} = 
\begin{bmatrix}
0.490 & 0.310 & 0.200 \\
0.177 & 0.812 & 0.011 \\
0 & 0.010 & 0.990 \\
\end{bmatrix}
\begin{bmatrix}
R \\
G \\
B \\
\end{bmatrix}
\] (1)

\[
L = \begin{cases} 
116(Y/Y_0)^{1/3} - 16, & Y/Y_0 > 0.008856 \\
903.3(Y/Y_0)^{1/3}, & Y/Y_0 \leq 0.008856 
\end{cases}
\]

\[
a = 500[f(X/X_0) - f(Y/Y_0)] \\
b = 500[f(Y/Y_0) - f(Z/Z_0)]
\] (2)

where \( f(x) = \begin{cases} 
7.787x + 0.138, & x \leq 0.008856 \\
x^{1/3}, & x > 0.008856 
\end{cases} \), and \( X_0, Y_0, Z_0 \) are the primary values of reference white.

The Extraction of Texture Features

Usually, texture refers to the local patterns recurring in the image and their arrangement rules,
with the characteristic that it is not dependent on color or illumination and can reflect the
homogeneity phenomenon of image. To extract image texture features, the method based on gray
co-occurrence matrix is taken in this paper. According to co-occurrence matrix, 16 kinds of
characteristic statistics used to extract texture information, such as entropy, contrast, energy,
correlation, variance and others, can be calculated. The contrast (CON), energy (ENG), correlation
(COR) and homogeneity (HOM) have been chosen to describe the texture features. Extracting
steps are as follows:

1. Change the color image to low-level gray image, and the gray level is 8.
2. Based on gray co-occurrence matrix, calculate the four texture namely contrast, energy,
correlation and homogeneity according to formulas (3) - (6), at last return the values to the
   corresponding pixel center.
3. Based on gray co-occurrence matrix, calculate the four texture namely contrast, energy,
correlation and homogeneity according to formulas (3) - (6), at last return the values to the
   corresponding pixel center.

\[
f = \text{CON} = \sum_{n=0}^{G-1} n^2 \left[ \sum_{i=1}^{G} \sum_{j=1}^{G} p(i,j) \right] \cdots |i - j| = n
\] (3)

\[
f = \text{ENG} = \sum_{i=1}^{G} \sum_{j=1}^{G} p^2(i,j)
\] (4)

\[
f = \text{COR} = \frac{\sum_{i=1}^{G} \sum_{j=1}^{G} (i \cdot j)p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}
\] (5)

\[
f = \text{HOM} = \sum_{i=1}^{G} \sum_{j=1}^{G} p(i,j) \frac{1}{1 + |i - j|}
\] (6)
CON could be considered as image sharpness; ENG is the measure of gray distribution uniformity, and the image with coarse textures corresponds to larger energy value; COR is the metric of gray linear relationship, reflecting the length some gray values extending along a direction; HOM is the measurement of local gray uniformity, if the local gray is uniform, homogeneous value is large. In addition, the paper take the regional average algorithm, because it is unable to get complete window information at image edge which will not gain texture eigenvalue, and considering the relationship between eigenvalue and the pixels involved in calculation. The radius d of matrix window is 3, for example, the values of the edge pixels can be approximately replaced by the average values of all texture characteristics associated. It can be expressed as:

\[
X = \text{mean}(\sum_{i=1-d}^{i+d} \sum_{j=j-d}^{j+d} p(i,j))
\]

where \((i, j)\) is the space coordinates.

Algorithm to Improve

The selection of the initial cluster centers is an important step in K-means clustering algorithm. Generally, these K samples are randomly selected from specimens to be clustered, and the clustering performance is relevant to the selection of the initial cluster centers. Incidentally, the results and the sample location have great relevance. Once these K samples selected irrational, it will increase the complexity of operation, mislead the clustering process and obtain unreasonable results \([14]\). How to reasonably choose cluster centers is the key to K-means clustering algorithm, as well as needs to be improved. In this paper, the desired mean and number of initial classes for K-means clustering is provided through rough set theory, which will enhance the clustering efficiency and classification accuracy.

Assuming that \(x_p\) is the gray value of pixel, \(Q_j^{(i)}\) is pixels set for class \(j\) after iterating \(i\) times, and \(\mu_j^{(i)}\) is the mean value of class \(j\).

Specific steps are as follows:

1. Take the L center points P provided by rough set theory as the initial class mean values, \(\mu_1^{(1)}, \mu_2^{(2)}, \mu_3^{(3)}, \ldots, \mu_l^{(l)}\).

2. In the \(i\) times iteration, calculate the spacing between every gray-scale mean value and each pixel, in the other words, the distance \(D\) each pixel to clustering centers. Put each pixel to the nearest category. As follows:

\[
D \left| x_p - \mu_i^{(i)} \right| = \min \left\{ D \left| x_p - \mu_i^{(i)} \right|, (j = 1, 2, \ldots, l) \right\}
\]

where \(x_p \in Q_j^{(i)}\).

3. Calculate the new cluster centers, and update the class mean: \(\mu_j^{(i+1)} = 1/N_j \cdot \Sigma_{x_p \in Q_j^{(i)}} x_p\), where \(N_j\) is the number of pixels in \(Q_j^{(i)}, j = 1, 2, \ldots, l\).
(4) Examining all the pixels one by one. If $\mu_i^{(i+1)} = \mu_i^{(i)}$ when $i = 1, 2, \ldots, K$, then the algorithm is convergent and will be terminated; otherwise, return to (2) and continue the next time iteration.

(5) After the clustering process above, in order to enhance the display effect, regard gray value of cluster center as the final gray of this class in the segmentation results.

---

**APPLICATION EXAMPLE**

To verify if the algorithm proposed can accurately and effectively segment seepage color image as well as the stability of the algorithm, a water-flooding microscopic seepage image with the size $1148 \times 708$ has been processed to divide the remaining oil and porosity, furthermore, to achieve the flow parameters.

**Image Preprocessing**

The outcome is not optimistic to directly separate images since the collected images will be affected by noise and other reasons. Consequently, it is necessary to conduct some related image enhancement and de-noise processing before image segmentation.

(1) Image enhancement

The purpose of image enhancement is emphasizing overall or local characteristics of images. According to the fuzzy image, a variety of special technologies are used to highlight some information, weaken or eliminate irrelevant information, which will achieve this aim. There are still no unified theory methods for image enhancement, the techniques commonly used are histogram modification, image smoothing, image sharpening, etc. Image enhancement technology is mainly divided into frequency and spatial region enhancement. The first method is adopting a variety of frequency domain filter for image smoothing or sharpening, and then conducting the transformation or inverse conversion. Space field enhancement method is processing the image gray directly against the pixels.

At present, color image enhancement algorithms are mainly three kinds: based on RGB color space, based on conversion space and based on the theory of color constancy \[18\].

(2) Removing the noise

An important issue cannot be ignored that the noise will increase at the same time of considering the color image enhancement. The un-conspicuous noise would stand out after image enhancing which is bound to impact the image quality. So suppressing image noise should be considered while enhancing the image. Common methods of suppressing noise are neighborhood average and filtering, such as median filter, matched filtering, the minimum mean square error filter and low-pass filter. However, it would cause the loss of information as well as small targets when implement filtering before the enhancement of image, which will reduce the effect of enhancement. Moreover, the general filtering algorithm is unable to eliminate the noise, because it will become remarkable when doing filtering after image enhance. In recent years, partial differential equation \[19-21\] performs outstandingly in the field of image processing. It not only has rigorous mathematical foundation, but also has the very good development prospect if combined with traditional image processing technology. This means could achieve good smoothing result at the same time as maintain edge detail perfectly for image de-noising \[18\].
Image Segmentation

Figure 1 is the color flow images after image preprocessing. Using the algorithm presented in this paper to divide the image, the results are shown in Figure 2-4. Figure 2 displays the segmentation result of red yellow flooding water, Figure 3 indicates the result of blue bound water, and Figure 4 presents the result of black oil.

Figure 1: Water-flooding microscopic seepage image

Figure 2: The segmentation result of flooding water
Calculation of Seepage Parameters

Porosity and fluid saturation are the macroscopic representation of microscopic parameters in micro flow experiments. Assuming that the number of pixels in the rectangle is $N_t$, the corresponding number for pore is $N_p$, the oil is $N_o$ among $N_p$. Therefore, the porosity $K$ and oil saturation are expressed as:

$$
\begin{align*}
K &= N_p / N_t \\
W_o &= N_o / N_p
\end{align*}
$$

The porosity and the percentage of each component as shown in Table 1 can be obtained, based on the formula above and the amount of pixels shared by driving split water, bound water and oil. The results achieved by the manual inspection method are shown in Table 2.
Table 1: Results of each parameter obtained by experiment

<table>
<thead>
<tr>
<th>parameters</th>
<th>porosity percentage</th>
<th>Driving water saturation</th>
<th>Bound water saturation</th>
<th>Oil saturation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>30.2228</td>
<td>36.8094</td>
<td>11.2136</td>
<td>51.9770</td>
</tr>
</tbody>
</table>

Table 2: Results of each parameter obtained by manual inspection

<table>
<thead>
<tr>
<th>parameters</th>
<th>porosity percentage</th>
<th>Driving water saturation</th>
<th>Bound water saturation</th>
<th>Oil saturation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>30.2137</td>
<td>36.7603</td>
<td>11.2387</td>
<td>52.0010</td>
</tr>
</tbody>
</table>

From Tables 1 and 2, the values of these parameters obtained by the proposed algorithm are very consistent with the results of artificial inspection, which proves the algorithm is accurate and effective for microvadose image segmentation. The artificial detection method is time and effort consuming, however, the use of the proposed method can automatically, quickly and accurately get results without introducing random error.

CONCLUSION

Hydrocarbon reservoir porosity is one of the important parameters for oil-gas prediction and reserves calculation. So, to obtain the porosity accurately is the basis and key for stratigraphic interpretation and establishing the geological model. This paper presents a K-means clustering algorithm that could automatically calculate these flow parameters for instance porosity, etc., and gives a detailed exposition for the image feature extraction and improvement of algorithm. Segmentation and calculation results show that the improved K-means clustering segmentation algorithm can achieve automatic segmentation of flooding water, bound water, remaining oil and pore in water flooding microscopic seepage color image and realize quantitative calculation for water flooding microscopic seepage experiment results. The algorithm is on the basis of pixel gray value statistics in image. It has greater superiority than the manual segmentation because of these characteristics, implementation simple, fast convergence speed, high efficiency and good robustness.

ACKNOWLEDGEMENT

This paper is financially supported by Natural Science Foundation of China (Grant No.50874082, 51174170).

REFERENCES


