Research paper

A trace map comparison algorithm for the discrete fracture network models of rock masses

Shuai Han [a], Gang Wang [b], Mingchao Li [a,*]

[a] *State Key Laboratory of Hydraulic Engineering Simulation and Safety, Tianjin University, Tianjin 300072, China*
[b] *Chengdu Engineering Corporation Limited, PowerChina, Chengdu 410014, China*

ARTICLE INFO

ABSTRACT

Discrete fracture networks (DFN) are widely used to build refined geological models. However, validating whether a refined model can match to reality is a crucial problem, concerning whether the model can be used for analysis. The current validation methods include numerical validation and graphical validation. However, the graphical validation, aiming at estimating the similarity between a simulated trace map and the real trace map by visual observation, is subjective. In this paper, an algorithm for the graphical validation of DFN is set up. Four main indicators, including total gray, gray grade curve, characteristic direction and gray density distribution curve, are presented to assess the similarity between two trace maps. A modified Radon transform and loop cosine similarity are presented based on Radon transform and cosine similarity respectively. Besides, how to use Bézier curve to reduce the edge effect is described. Finally, a case study shows that the new algorithm can effectively distinguish which simulated trace map is more similar to the real trace map.

1. Introduction

Rock mass consists of fractures and blocks cut by the fractures. Fractures have a significant impact on the mechanical and hydraulic characteristics of a rock, usually make a rock mass instable (Park, 2013; Rossmanith, 2014). In different geologic environments, they vary greatly in many aspects, such as size, orientation and spatial distribution. In geology, they can reflect the discontinuity, heterogeneity and anisotropy of rocks and have a significant influence on their mechanical characteristics and hydraulic characteristics. Thus, characterization of rock fractures in terms of the orientation, spacing, size, aperture, etc., is a basic step in rock engineering design (Han et al., 2016).

With the development of mathematical geology and computer technology, researchers tend to incorporate fractures with terrain models and stratum models to make a more elaborate model, called discrete fracture network (DFN) model (Herbert, 1996). For example, Cédric Lambert et al. (2012) simulated a DFN model based on the Baecher disk model to analyze rockfall hazard. Li et al. (2017a) proposed an enhanced polyhedral simulation to identify complex blocks and block-groups. Li et al. (2017b) set up an algorithm for calculating the connectivity parameters of fractures in rock masses based on three-dimensional fracture networks. Ni et al. (2017) also used DFNs to create a more elaborate model, which estimates the representative elementary volume of a rock mass.

Currently, fracture definitions and DFN modelling are recognized as an effective approach to study rock mass structures (Xu and Pruess, 2001; Li et al., 2017a), and the later are used more widely. In this research, we set our sight on the validation of DFN modelling. Although DFN modelling has been put forward for a long time, it is always based on stochastic simulation. Thus, the DFN model will not always match reality. For this reason, validation methods should be employed to ensure the validity of the models. Generally, there are two kinds of methods to validate the models: numerical validation and graphical validation. Most numerical validations are based on statistics. In numerical validations, data from the stochastic models are collected after statistical analysis and are compared with the measured data; it includes the orientations, density, and trace lengths. During the analysis, stochastic models should have the same distribution characteristics with measured data (Merrien-Soukatchoff et al., 2012; Guo et al., 2015a, 2015b; Mendoza-Torres et al., 2017; Wang et al., 2004). As for the graphical validation, the analysis procedure lacks theoretical support, in most cases. Graphical validation is simply based on visual observations: by directly comparing the real trace map and the corresponding simulated trace map to obtain a qualitative judgment (Guo et al., 2015a, 2015b; Li et al., 2017a). Since there are no standards to determine the similarity between a real trace map and a simulated trace

---

* Corresponding author.
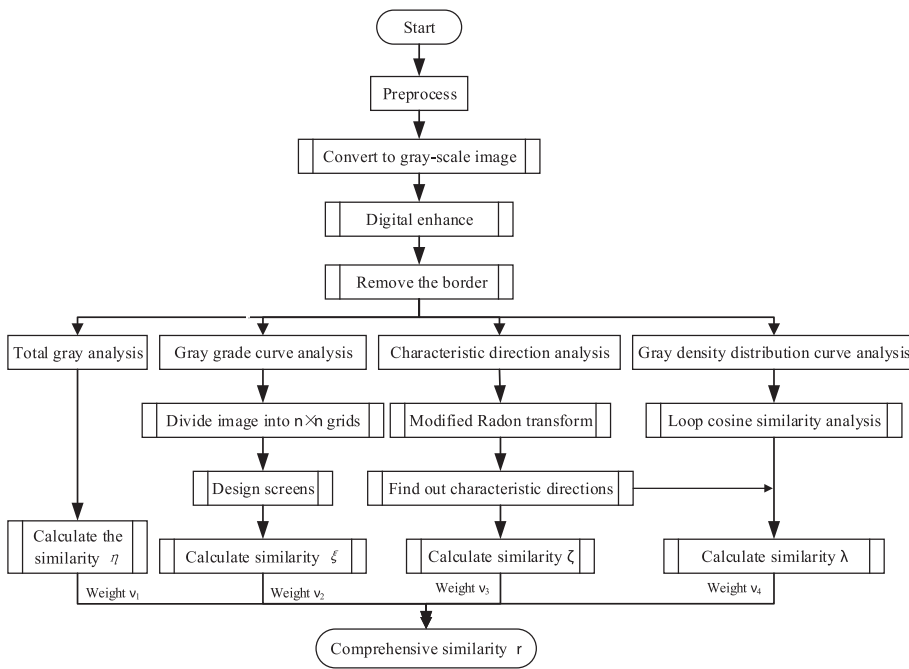 *E-mail address:* lmc@tju.edu.cn (M. Li).
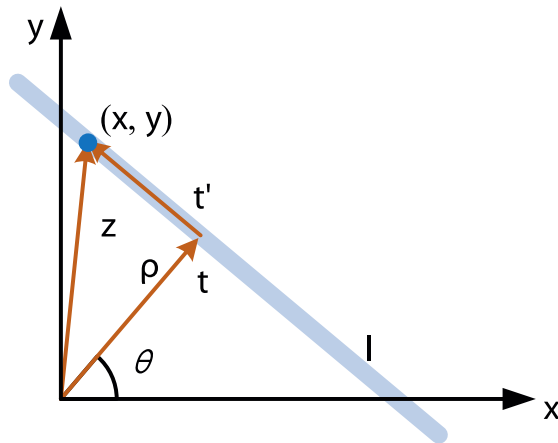
Fig. 1. Flow chart of the algorithm.

Fig. 2. Coordinate system for defining the Radon transform.
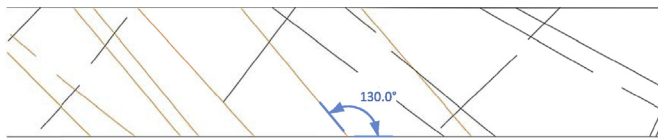

Fig. 3. A real trace map example.

map, different observers may give different validation results. However, few researchers have set their sights on this topic and been working on improving the graphical validation method (Cacas et al., 1990). With the development of digital image processing technology and artificial intelligence, we put our focus on this method.

Many researchers have studied trace maps with the help of digital image processing technology. For example, Fitton and Cox (1998) developed a general and robust procedure to extract linear features of all scales from geoscientific datasets. Kemeny and Post (2003) described a computer approach for estimating three-dimensional fracture orientations from two-dimensional fracture trace information gathered from digital images of exposed rock faces. Lemy and Hadjigeorgiou (2003) presented a digital face mapping methodology that can generate trace maps according to photographs of surfaces of rock mass. Ferrero and Umili. (2011) proposed a method to estimate characteristics of rock mass from orthophotographs. Gigli and Casagli (2011) presented a Matlab tool that can calculate the parameters and features of rock mass just according to Light Detection and Ranging (LiDAR) point clouds. Riquelme et al. (2014) also proposed a method for the semiautomatic calculation of the orientations and position of rock mass discontinuities from LIDAR data. Zeeb et al. (2013) presented the software FraNEP (Fracture Network Evaluation Program), which automatically analyses the statistical properties of 2D fracture networks based on trace maps. Cao et al. (2017) set up a method to identify structural fractures directly from three-dimensional point clouds and evaluate the fracture distribution. Clearly, most of the methods focused on the detection of fractures from digital images, rather than for geological models. Aydin and Caers (2013) set up a method based on image transform to analysis geological models. However, its purpose is to determine the complexity of a model that used in flow analysis, different from graphic validation. Umili et al. (2013) provided an automatic method for discontinuity trace mapping and sampling with a rock mass digital model to automatically identify discontinuity traces from digital surface models. Li et al. (2016) also set up a similar method. However, these can't be used for validating the model, either.

In this paper, we set our sight on the graphic validation, aiming at automatically distinguishing which simulated trace map is more similar to the real trace map. In this research, a set of image preprocessing methods are firstly described. Then, four main indicators, including total gray, gray grade curve, characteristic direction and gray density distribution curve, are presented to assess the similarity between two trace maps. For analyzing the characteristic direction, a modified Radon transform equation is presented based on Radon transform, and Bézier curve is adopted to reduce the edge effect. For analyzing the gray density distribution curve, the loop cosine similarity is put forward based on cosine similarity. Finally, a case study is carried out. The result shows
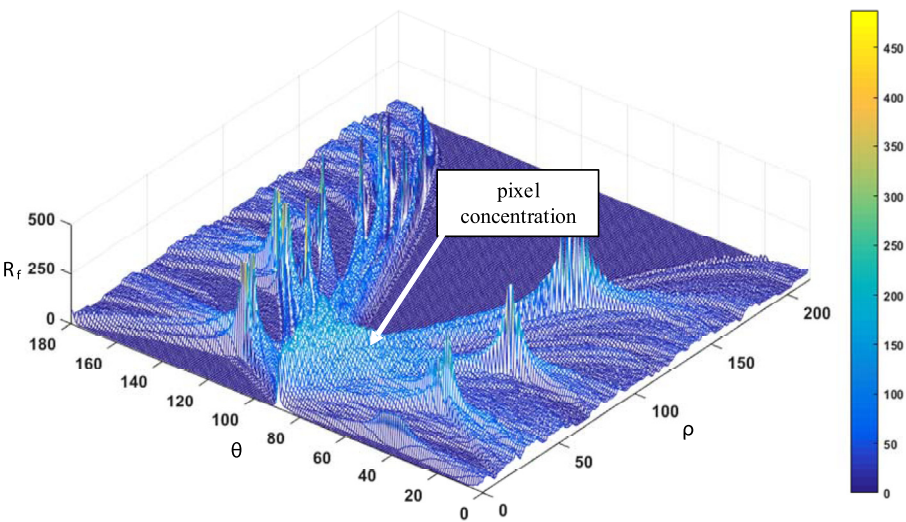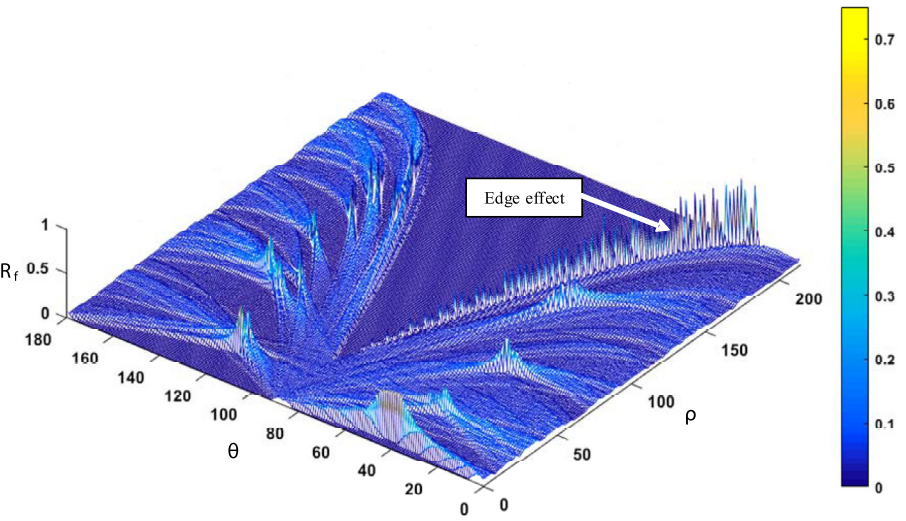
Fig. 4. Radon transform.
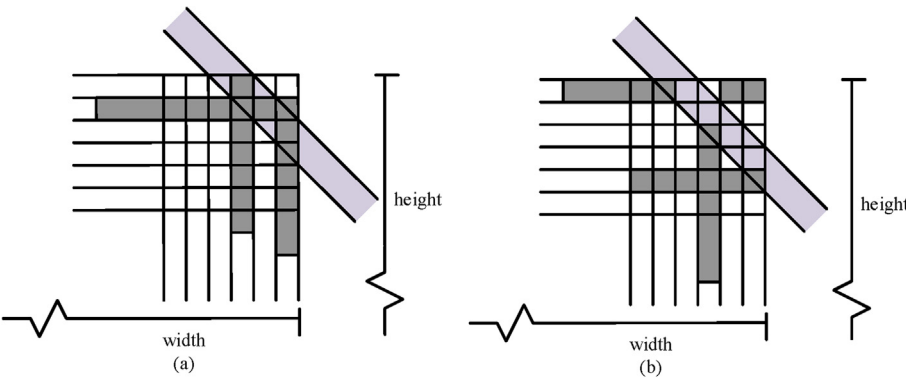

Fig. 5. Density Radon transform.

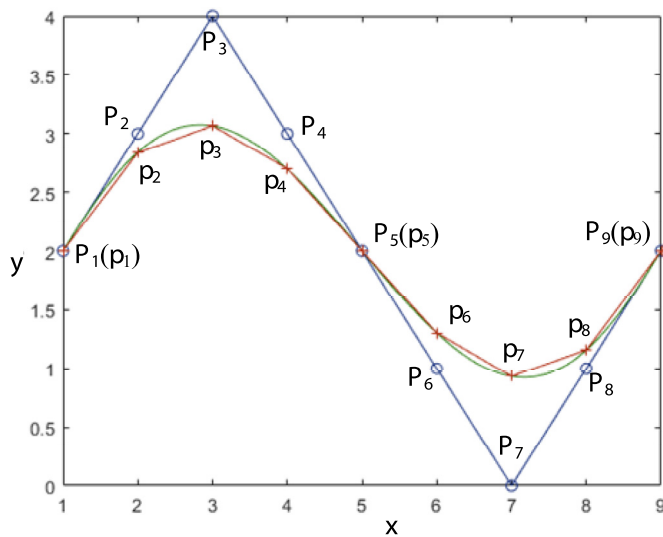**Fig. 6.** Schematic diagram of the edge effect.

Fig. 7. Method for smoothing the fluctuations based on the Bézier curve.
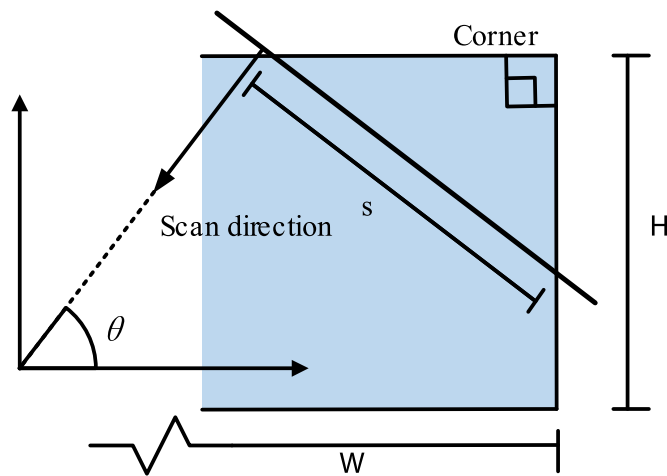

Fig. 8. The scope of influence of the edge effect.

that the algorithm is indeed able to accurately assess the similarity between two trace maps. In summary, this method provides a new idea for graphic validation, and will be helpful to select the optimal 3D model from repeated simulating.

2. Methodology

In this algorithm, the inputs are two trace maps needed to be compared. The simulated trace map is obtained from a 3D model, the real trace map is from the walls of adits or surface of slopes, and should be drawn in CAD format or other format that can be imported to the 3D modelling software. In this way, the images can be ensured in the same size and resolution. It begins with preprocessing to match the specifications of the two trace maps, which includes converting the two images to gray-scale, enhancing the traces, removing the borders and changing the sizes to be equal, as illustrated in Fig. 1. The preprocessing step is necessary and is the basis of the next steps. The next steps include four parts:

(1) Total gray analysis. The amount of black pixels represents the total quantity of joints. This step is for analyzing the similarity of the global fracture intensity of the two images.

(2) Gray grade curve analysis. This idea is from the particle size distribution used in soil mechanics. First, divide the two pre-processed trace maps into $n \times n$ grids. Each of the cells can be regarded as a soil particle. Meanwhile, design a set of screens. Then, compute the similarity of the gray grade curve analysis of the two images, see section 3.3.

(3) Characteristic direction analysis. Because the directions of the traces follow certain laws, a modified Radon transform method is proposed based on Radon transform to detect the characteristic trace directions. In addition, a Bézier curve is used to reduce the edge effect. After that, the characteristic directions will be calculated, and the characteristic directions will be used in step (4), see section 3.4.

(4) Gray density distribution curve analysis. After step (3), the characteristic directions of the two images have been determined. Their gray density distribution curves along the same characteristic directions will be similar. By the loop cosine similarity method, the similarity of the gray density distribution curves of the two images can be calculated, see section 3.5.

Based on the four similarities from steps (1)~(4), a comprehensive similarity, which can be calculated from a weighted average, is presented. When simulating a rock mass, the simulation process may be repeated many times. The comprehensive similarity can be used to measure which simulation is the closest to the reality, see section 3.6.

3. Algorithm implementation

3.1. Image preprocessing

There are four main steps to preprocess the images:

(1) Convert to a gray-scale image. The lines in some trace maps may be colored to represent different groups of dominant orientations (Li et al., 2017a, 2017b). However, the grouping is contrived, and the joints on an adit's wall have no color. So, the colors are redundant to comparing, and sometimes the colored joints are not clear enough to be recognized. Therefore, they should be converted to gray-scale images and enhanced in step (2).

(2) Digitally enhance. Sometimes, the joints on the trace maps are not clear enough to be detected. Then the algorithm will enhance them: gray values that less than a threshold are directly set to 0; gray values that greater than a threshold are directly set to 255 (where 0 represents black and 255 represents white). In this research, we propose to set the threshold to 230 ($0.9 \times 255 \approx 230$). The well-known thresholding method, Otsu method (Otsu, 1979), is not suitable in this study, because the colors of the lines are varying, and the lines with light colors may be detected as the background automatically.

(3) Remove the border. Generally, a trace map is rectangular, and there is a rectangular frame surrounding the traces. It is necessary to remove the rectangle to ensure that there is nothing else except for the traces in the image. Otherwise, the borders will be regarded as a part of the joints by the algorithm. The process is as follows: 1) progressively scan the gray matrix from the top edge to bottom until a row is found with few black pixels (relative to the previous row); 2) take the number of rows that have been scanned as the thickness of the top border and remove them; 3) remove the bottom, left and right border in the same way that removing the top border.

3.2. Total gray calculation

After the preprocessing, the two trace maps are converted into binary images. These lines represent the joints. Therefore, the number of black pixels can represent the total quantity of joints. If the two trace maps are
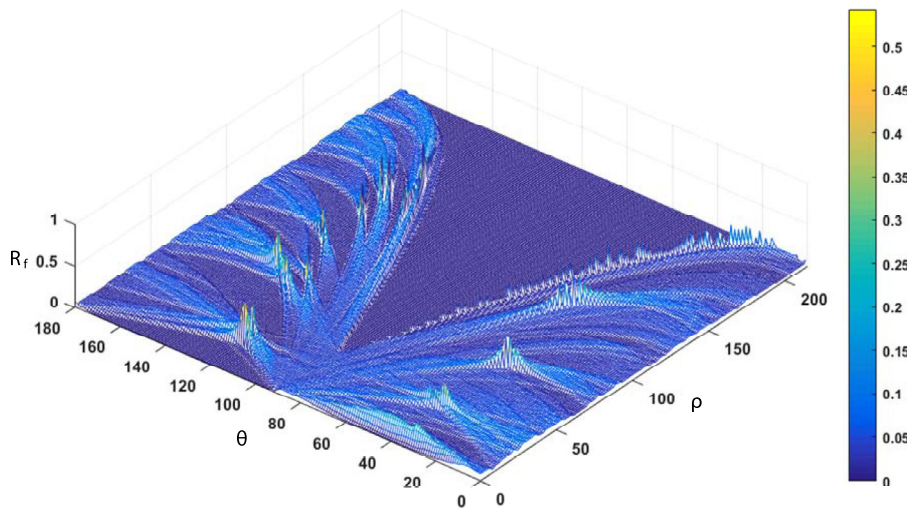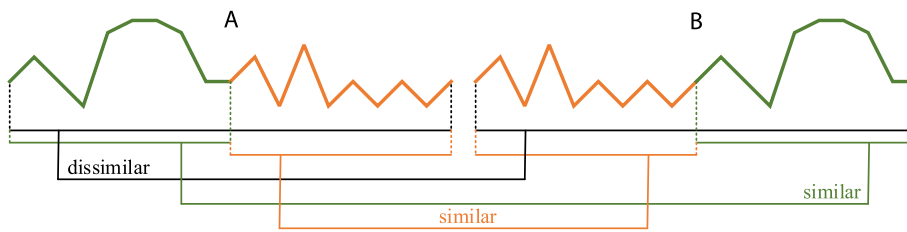
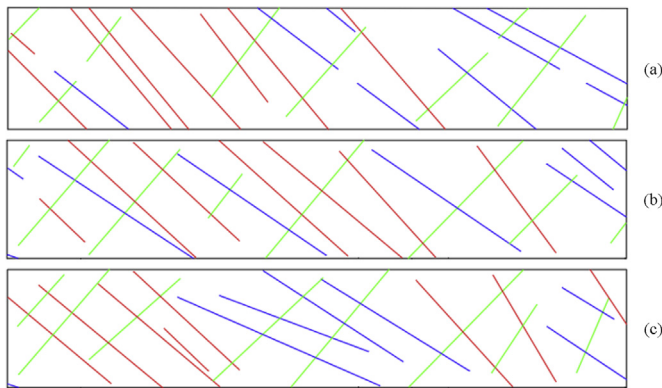**Fig. 10.** Schematic diagram of the relative similarity.

Fig. 11. Original trace maps. (a) The original real trace map. (b) The first original simulation trace map. (c) The second original simulation trace map.

similar, the number of black pixels of the two images will be similar. Scan both from the left to the right and calculate the distribution of black pixels. Then, calculate the difference value according to the equation:

$$\eta = 1 - |S_A - S_B|/S_A \tag{1}$$

where η is the similarity value, S_A is the area of the gray histogram of image A, S_B is the area of gray histogram of image B. S_A and S_B also represent the number of black pixels of image A and B, respectively.

3.3. Gray grade curve analysis

Total gray is used to examine the global features of the two trace maps, and to examine their local features, the gray grade curve is presented. The concept of the gray grade curve is similar to the particle size distribution in soil mechanics (Tegen and Lacis, 1996). At first, the images should be evenly divided into $n \times n$ grids. In each grid, the numbers of black pixels are different. The cells that contain fewer black pixels correspond to the soil particles with smaller diameters, and the cells that contain more black pixels correspond to the soil particles with larger diameters. In some other algorithms, images may also be divided into grids, such as the perceptual hash algorithm and gray level co-occurrence matrix (Tahir et al., 2003; Monga and Evans, 2006). Generally, these algorithms divide an image into $2^m \times 2^m$ grids, usually 8×8 or 16×16. In this algorithm, 64 grids are too few to analyze the gray grade. Therefore, we suggest setting $m = 4$ and $n = 16$.

Next is designing the screens to sieve the grids. For each grid, calculate the proportion of black pixels and use the percentages, d_{ij}, as the diameters of the gray particles, i is the row number and j is the column
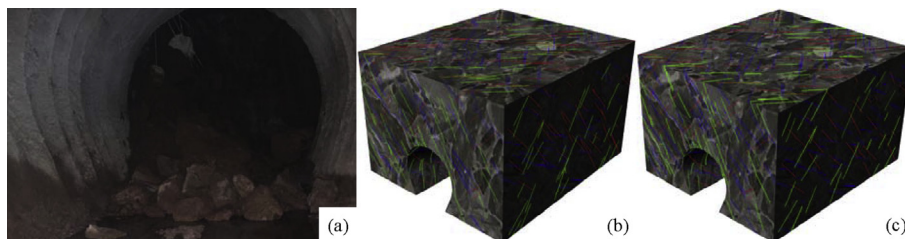

Fig. 12. Adits. (a) Real adit. (b) The adit of the first 3D model. (c) The adit of the second model.
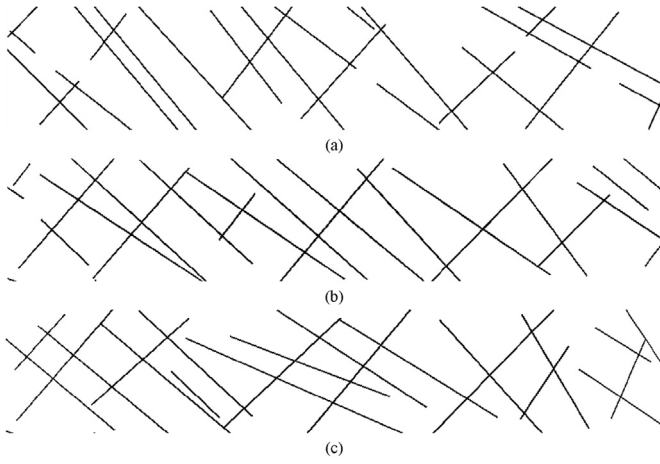
Fig. 13. Preprocessing results. (a) Preprocessing result of the real trace map. (b) Preprocessing result of the first simulation trace map. (c) Preprocessing result of the second simulation trace map.
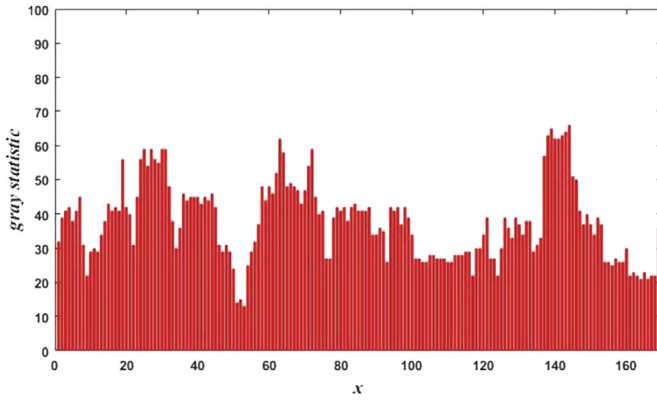

Fig. 14. Total gray of the real trace map.

number of the grid. Obviously, d_{ij} is in the range [0, 100%], or [0, 1]. Determine the maximum, d_{max}, from all the d_{ij} and use it as the maximum size of the screens. Then, design a set of screens with sizes in the range [0, d_{max}]. The difference between two adjacent screens is 0.001. For example, if d_{max} equals to 0.5, there will be 500 screens.

According to the grids and screens, create two gray grade curves: [a_1, a_2, ..., a_n] and [b_1, b_2, ..., b_n]. [a_1, a_2, ..., a_n] corresponds to the real trace map, and a_i is the number of grids screened out by the ith screen. [b_1, b_2, ..., b_n] corresponds to the simulated trace map, and b_i is the number of grids screened out by the ith screen. Then, calculate the similarity, ξ. The proposed equation is as follows:

$$\begin{cases} f(\xi) = \sum_{i=1}^{n} \sigma_i \Big/ n & \sigma_i = \begin{cases} 1, & 1 - |a_i - b_i| > \xi \\ 0, & 1 - |a_i - b_i| \le \xi \end{cases} \\ g(\xi) = \xi \end{cases} \quad (2)$$

Evidently $f(\xi)$ is a non-increasing function of ξ and $g(\xi)$ is an increasing function of ξ. Let

$$f(\xi) = g(\xi) \quad (3)$$

To solve Eq. (3), the ξ can be calculated. Clearly, if all the a_i is close to b_i, the ξ will close to 1, meaning that the grades of the two gray grade curves are similar.

3.4. Characteristic direction

Different from any other images, trace maps consist of many lines

with a certain regularity. Because of the dominant joint sets, the traces will be generally oriented in certain directions that are called the characteristic directions. If two trace maps are similar, they will have the same characteristic directions.

3.4.1. Radon transform

In this paper, the Radon transform is used to determine the characteristic directions.

J. Radon set up the Radon transform and this method has been widely used in the fields of pattern recognition and image processing (Toft and Sørensen, 1996; Schultz and Gu, 2013). The mathematical transformation is described as follows:

$$R_f(p, \theta) = \int_{-\infty}^{\infty} f(x, y) dl \quad (4)$$

where R_f represents Radon transform, ρ is the vertical distance from the coordinate system origin to the line to be detected, θ is the angle between the coordinate system axis x and ρ, $f(x, y)$ represents the gray values in the image at the pixel (x, y). l can be described as:

$$L(p, \theta) = \{(x, y) : x \cos \theta + y \sin \theta = \rho\} \quad (5)$$

With the delta function, Eq. (3) can also be described as:

$$R_f(\rho, \theta) = \iint_D f(x, y) \delta(\rho - x \cos \theta - y \sin \theta) dx dy \quad (6)$$

where, D is the area of the image, δ is a unit pulse function:

$$\delta = \begin{cases} 0, & \rho - x \cos \theta - y \sin \theta \ne 0 \\ 1, & \rho - x \cos \theta - y \sin \theta = 0 \end{cases} \quad (7)$$

In Eqs. (5)–(7), the function δ is obtained at any point (ρ, θ) and integrated over the straight line, l, as shown in Fig. 2.

3.4.2. Modified radon transform

As mentioned above, an example has been analyzed by the Radon transform and shown in Fig. 3 and Fig. 4.

Fig. 3 shows that a set of traces have been marked as orange and they are all tilted around $130°$ relative to the horizontal direction. Meanwhile, it can be found from Fig. 4 that when $\theta = 130°$, the surface shows its greatest fluctuation. It means that there is a clear link between the characteristic direction and the fluctuation of the surface: the characteristic direction equals the θ along the surface with the greatest fluctuation. In addition, the fluctuation can be represented by variance.

It is also notable that there is a pixel concentration at $\theta = 90°$. This is because l is perpendicular to the horizontal direction, in that case. When calculating the $R_f(\rho, \theta)$, the length of integrations will change with θ, unless the trace map is a circle. This leads to a problem: fluctuations cannot reflect characteristic directions exactly, for example, when the bound is longer, the pixels will distribute more widely with an invariant total quantity, leading to a smaller variance. It means that the fluctuations of the surface are not only affected by the characteristic directions, but also the shape of the trace map. To solve this problem, a modified Radon transform is set up as follows:

$$R_f'(\rho, \theta) = \frac{\iint_D f(x, y) \delta(\rho - x \cos \theta - y \sin \theta) dx dy}{\iint_D \delta(\rho - x \cos \theta - y \sin \theta) dx dy} \quad (8)$$

This transform can be named the density Radon transform. The numerator is the integral equation of Radon transform, and the denominator is the length of l. Even if the length of l will be changed, the density of pixels at any position of the trace map will not be affected. The result is shown in Fig. 5.

Additionally, because the l described in Eq. (5) will always pass through the area D, the denominator of Eq. (8) is nonzero.
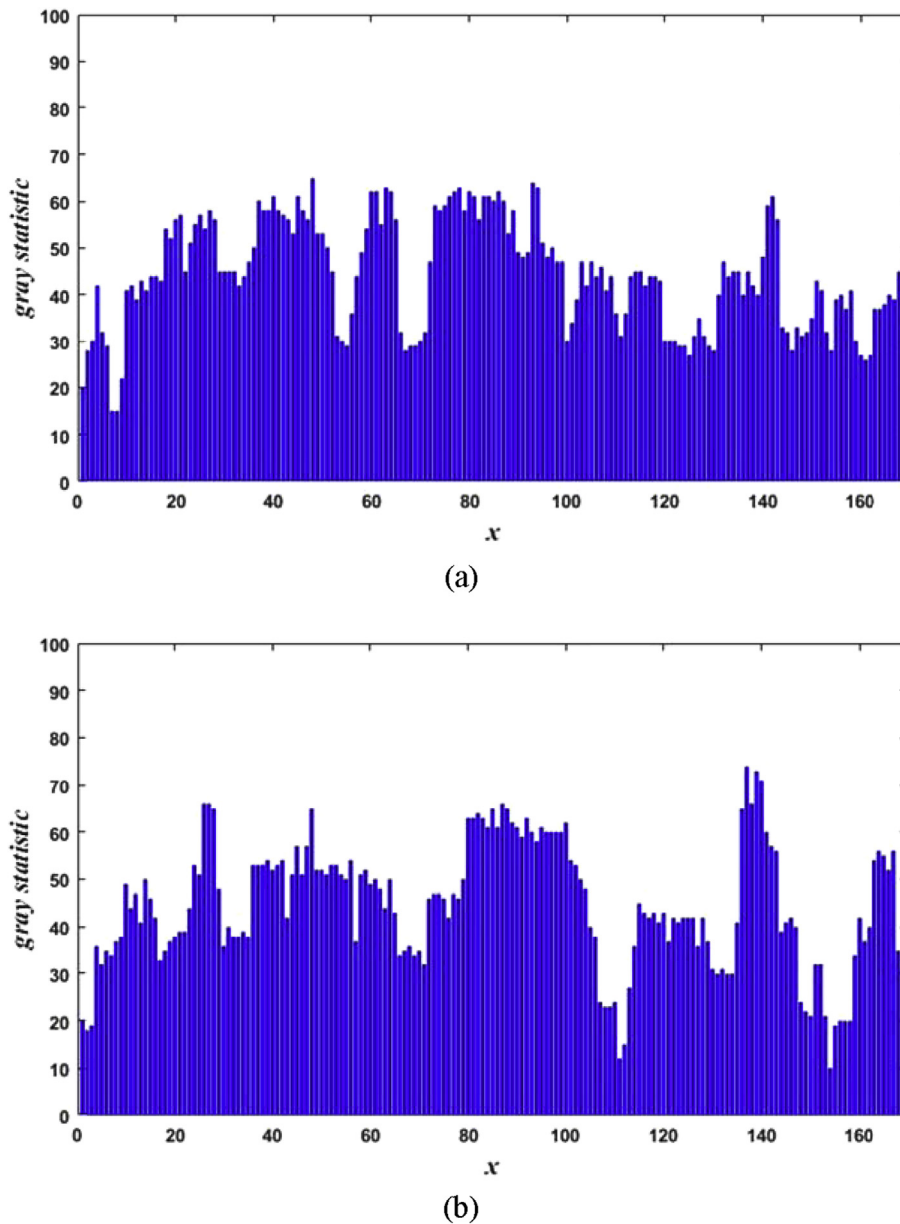
3.4.3. Symmetry of modified radon transform

Considering Eq. (9)

$$R_f'(a\rho, at) = \frac{\iint_D f(x,y)\delta(\rho - ax\cos\theta - ay\sin\theta)dxdy}{\iint_D \delta(a\rho - ax\cos\theta - ay\sin\theta)dxdy} \tag{9}$$

where a is a constant factor and $t = (\cos\theta, \sin\theta)$ is a unit vector that is perpendicular to l. If $a = -1$, because of Eq. (10):

$$\delta(\rho - x\cos\theta - y\sin\theta) = \delta(-\rho + x\cos\theta + y\sin\theta) \tag{10}$$

it indicates that the modified Radon transform is an even function:

$$R_f'(-\rho, -t) = R_f'(\rho, t) \tag{11}$$

Due to the symmetry, the range of θ can be set as $(0°, 180°]$.

3.4.4. Edge effect and the bézier curve

When employing the density Radon transform, another issue arises: the edge effect, as marked in Fig. 6. The reason can be illustrated by Fig. 6:

when θ is neither horizontal nor vertical, there will be a sharp corner at each end of the measure line. When scanning at the corners, the density of the pixels may be very large or very small. Obviously, these don't mean that there is numerous or sparse fractures at this position of the rock mass. This is mainly caused by the narrow corner. In addition, the peaks identified by the arrow in Fig. 5 correspond to Fig. 6(a).

Therefore, a Bézier curve is used to solve this problem. Bézier curve was first described by Pierre Bézier in 1962 in the context of shape design of automobile bodies (Bézier, 1968), and has been commonly applied in many fields. A Bézier curve usually has several control points. However, the curve only passes through the first point and the last point, while the others are responsible for controlling the shape of the curve. A Bézier curve is defined as:

$$P(t) = \sum_{i=0}^{n} B_i^n(t)P_i, \quad t \in [0,1] \tag{12}$$

where P_i is the ith control point, $B_i^n(t)$ represents a Bernstein polynomial, and the Bernstein polynomial can be defined as follows:
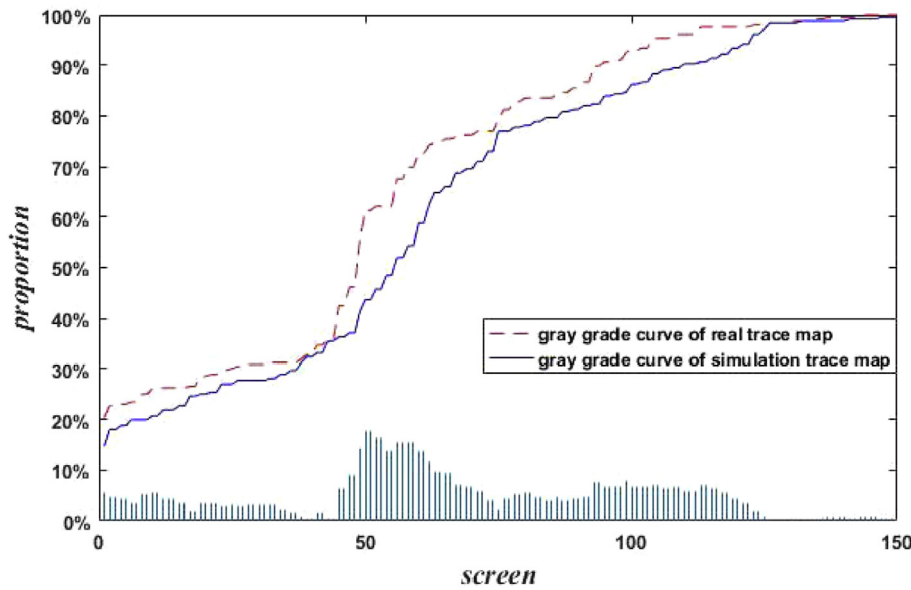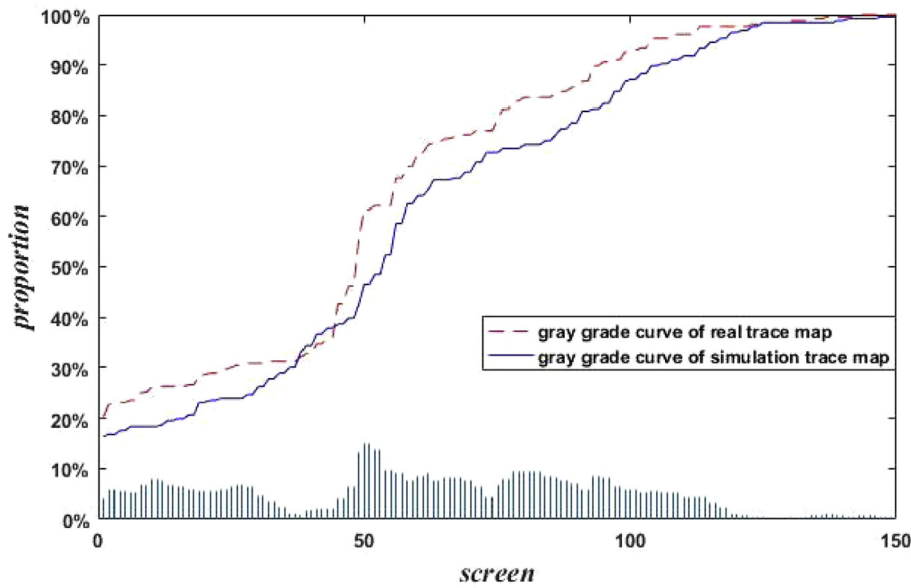
Fig. 16. Results of the gray grade curve analysis. (a) Gray grade curves of the real trace map and the first simulation trace map. (b) Gray grade curves of the real trace map and the second simulation trace map.

$$B_i^n(t) = C_n^i t^i (1-t)^{n-i}, \quad i = 0, 1, \ldots, n. \tag{13}$$

where C_n^i is the binomial coefficient:

$$C_n^i = \frac{n!}{(n-i)!i!} \tag{14}$$

A Bézier curve has many properties. One of them is that a Bézier curve is contained within the convex hull that produced by the control points, which can be used in this research. On one hand, an edge effect is caused by some points with exceptionally high or exceptionally low values. On the other hand, in the 3D graphic in Fig. 5, the cross sections of the surface perpendicular to the θ-axis can be regarded as curves that fluctuate strongly near the ends. The 3D graphic consists of lots of this kind of curves. Therefore, the edge effect can be weakened if the endmost fluctuations of the curves are reduced. To achieve this, for each of the curves, a Bézier curve is created to smooth its endmost fluctuations, as shown in Fig. 7.

As shown in Fig. 7, the blue curve [*P1, P2, …, P9*] represents an edge part of a cross section, and *P9* represents the last point of the edge. The green curve represents the Bézier curve created by the control points [*P1, P2, …, P9*]. On the Bézier curves, a new set of points, [*p1, p2, …, p9*], can be calculated, and *pi* has the same abscissa values as *Pi*. Then, the edge effect can be weakened by replacing [*P1, P2, …, P9*] with [*p1, p2, …, p9*]. In addition, the last point *pi* is always equal to the last point *Pi*, meaning that the edge effect caused by the last point will not be weakened. Therefore, it is better to reject the last point after the steps mentioned above.

Next, the range of the edge should be determined. First, determine the dimension with the smaller size, *W*; as an example, the smaller dimension is the high, *H*. Then, carry out the density Radon transform. As the algorithm scans the image from one end to another, the length, *s,* will change, as shown in Fig. 8. When *s*=*H*, the range that has been scanned is the range of the edge. Clearly, the edges always come in pairs, and the
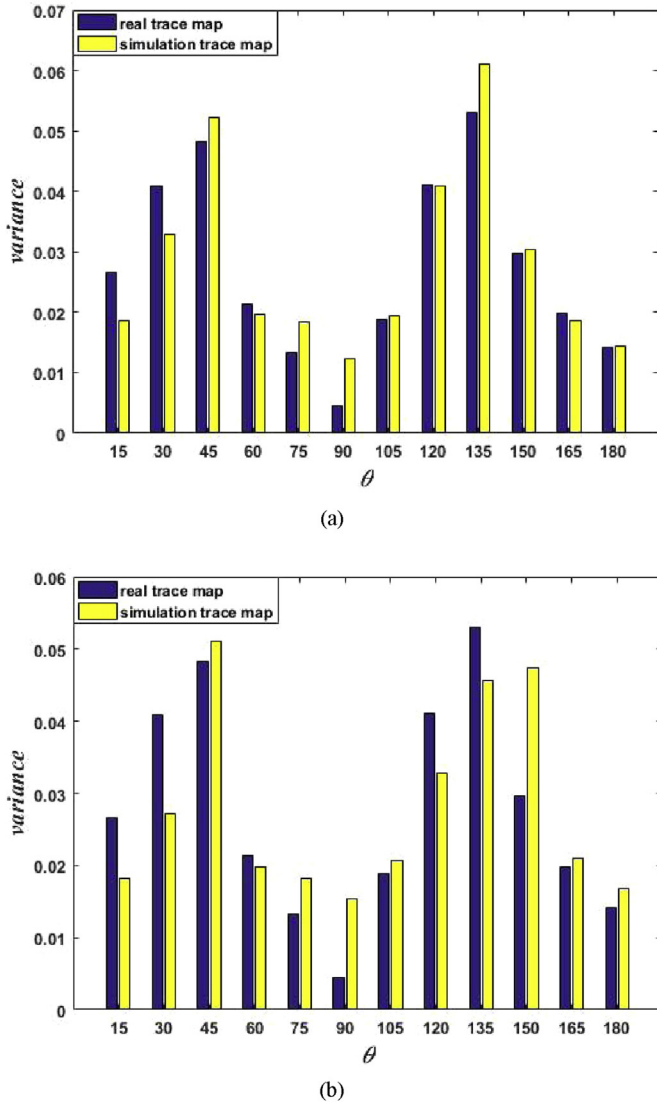
(a)

(b)

Fig. 17. Results of characteristic direction analysis. (a) Characteristic direction analysis of the real trace map and the first simulation trace map. (b) Characteristic direction analysis of the real trace map and the second simulation trace map.

Table 1
Results of the gray density distribution analysis.

θ	135°	45°	120°	30°	average
ωab	0.576	0.714	0.752	0.800	0.710
ωac	0.638	0.703	0.742	0.751	0.709

range of the other edge can be determined according to the symmetry principle.

As shown in Fig. 9, the edge effect has been largely reduced.

3.4.5. Characteristic direction detection

Considering that the dominant orientation of a joint group is a range, there is no need to change the θ degree by degree. Considering that in geology survey, a set of dominant orientations often vary in a range of 15°, and joints are analyzed along scanlines with 15° increments. (Bejari and Hamidi, 2013; Wasantha et al., 2015), we suggest changing θ every 15°. Then, the corresponding fluctuation degrees (variances) are determined at every θ. In total, there will be 12 variances. Generally, there will be 3-5 dominant joint sets in a project (Shanley and Mahtab, 1976; Zhan

et al. 2017). In this research we sort the variances from the largest to the smallest, and take the θs that corresponding to the four largest variances as the characteristic directions.

After calculating out the two sets of characteristic directions of the real trace map and simulated trace map, the characteristic direction similarity can be evaluated from two aspects: the overlap ratio and sort order. The overlap ratio, $\zeta1$, indicates how many directions exist in both sets of characteristic directions. The sort order, $\zeta2$, indicates that whether the directions that exist in both sets of characteristic directions have the same order in each of the two sets of characteristic directions. It contains three steps:

(1) Let [*rd1, rd2, rd3, rd4*] represent the characteristic directions of the real trace map and [*sd1, sd2, sd3, sd4*] represent the characteristic directions of the simulated trace map. *rdi* is the *i*th characteristic direction and *sdj* is the *j*th characteristic direction. Then, create a matrix from them:

$$M = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \\ m_{41} & m_{42} & m_{43} & m_{44} \end{pmatrix}, \quad m_{ij} = \begin{cases} 1, & rd_i = sd_j \\ 0, & rd_i \neq sd_j \end{cases} \quad (15)$$

(2) $\zeta1$ can be calculated as follows:

$$\zeta_1 = \frac{1}{4} \sum_{i=1}^{4} \sum_{j=1}^{4} m_{ij} \quad (16)$$

(3) Clearly, the number of mij that equal 1 is in the range of [0, 4]. Extract all the coordinates of mij that equal 1 by the following equation:

$$P = \left\{ (i, \ j) \big| i, \ j \in Z, \ 0 \leq i, \ j \leq 4, \ m_{ij} = 1 \right\} \quad (17)$$

Connect each of the coordinates to all the other coordinates to obtain a set of lines L (L=l1, l2, …). Count out the number of L with a positive slope and divide it by the total number of L. The calculation result is $\zeta2$.

Considering that the overlap ratio has a bigger impact on the similarity evaluation, the two-eight principle (Woolhouse et al., 2005) is used to take a weighted average of $\zeta1$ and $\zeta2$ to calculate a comprehensive ζ:

$$\zeta = 0.8\zeta_1 + 0.2\zeta_2 \quad (18)$$

For example, the two sets of characteristic directions are [135°, 15°, 60°, 160°] and [60°, 135°, 90°, 165°]. Therefore, $\zeta1$=3/4=0.75, $\zeta2$=2/3=0.67, and ζ=0.8×0.75+0.2×0.67=0.73.

3.5. Gray density distribution curve

If two trace maps are similar, the joints will be distributed in a similar way. The relative joint positions will also be similar. In this research, four cross sections of the 3D graphic created by density Radon transform are extracted. The four cross sections correspond to the four characteristic directions. Every cross section is a curve, which is called the gray density distribution curve. By comparing the similarity of the gray density distribution curves of the two images, the similarity of relative joint positions of the two images can be estimated.

3.5.1. Similarity of gray density distribution curves

At first, calculate the four characteristic directions, [*rd1, rd2, rd3, rd4*] and four corresponding gray density distribution curves, [*rc1, rc2, rc3, rc4*] of the real trace map. Then, extract the cross sections of the density Radon transform surface of the simulation trace map at [*rd1, rd2, rd3, rd4*] to produce four curves, [*sc1, sc2, sc3, sc4*]. For each *i*, if *rci* is similar to *sci*, it means that the simulated trace map has the same gray density distribution curves as the real trace map in the characteristic directions.

The similarity of the gray density distribution curves is calculated as follows:

$$\omega = \frac{1}{4} \sum_{i=1}^{4} \omega_i \qquad (19)$$

where ωi is the similarity between rci and sci.

3.5.2. Loop cosine similarity

In this paper, a loop cosine similarity is presented to determine whether two curves are similar based on cosine similarity. Cosine similarity is a method to calculate the similarity of two vectors by calculating the cosine value of them. (Dong et al., 2006; Ye, 2011). Therefore, the value of similarity is in the range of [-1, 1], and is determined by the orientation of the two vectors, independent of their magnitudes. If two vectors have the same orientation, the value is 1; if the two vectors are at 90°, the value is 0; if the two vectors have the opposite orientation, the value is -1. Given two vectors [$x1, x2, …, xn$] and [$y1, y2, …, yn$], the cosine similarity, $\cos(\theta)$, is represented as follows:

$$\cos(\theta) = \frac{\sum_{i=1}^{n}(x_i y_i)}{\sqrt{\sum_{i=1}^{n}(x_i)^2}\sqrt{\sum_{i=1}^{n}(y_i)^2}} \qquad (20)$$

This method has its limitations. For example, A and B represent two different curves, their peaks and valleys reflect the distribution of the traces, as shown in Fig. 10. The cosine similarity between them is very low, meaning that the traces drawn on the two images are completely different. However, it can also be found that there are some similar segments between A and B, especially when connect the head and end of each of the two curves respectively, they will be two same sawtooth circles. It means that although the two curves are absolutely similar, the relative positions of their peaks and valleys are similar, meaning that the relative joints positions of the two images are similar.

Based on the above, the loop cosine similarity method is proposed, the pseudocode is as follows.

Loop cosine similarity

A=[a1, a2, a3, …, an]; B=[b1, b2, b3, …, bn];//A and B are two curves

Similarity=1×n vector;//*Similarity* is a 1×n dimension vector

for i=1 to n

Btmp=[bi, bi+1, bi+2, …, bn, b1, b2, …, bi-1]

Similarity (i)=CosineSimilarity (A, Btmp)//Calculate the cosine similarity between A and Btmp

end

LoopCosineSimilarity=Maximum (Similarity).

3.6. Comprehensive similarity

After the η, ξ, ζ and ω have been calculated in steps 3.2-3.5, the comprehensive similarity can be calculated using a weighted average.

$$S = c_1\eta + c_2\xi + c_3\zeta + c_4\omega \qquad (21)$$

However, it is difficult to exactly determine all the weights and we use a heuristic evaluation method, brain-storming method (Isaksen, 1998), to estimate these weights: (1) five experts are asked to act as the evaluators and present their comments on the weights, (2) by synthesizing their comments, the [$c1, c2, c3, c4$] is set as [0.2, 0.2, 0.3, 0.3] because the calculation methods of η and ξ are relatively rough and their weights should be smaller than the other two metrics.

4. Results and discussions

4.1. Calculation results

In this section, an example is provided of the performance of the algorithm. As shown in Fig. 11, (a) is the real trace map obtained from the geological survey, (b) is the first simulated trace map obtained from the first 3D model, and (c) is the second simulation trace map obtained from the second 3D model (in the following, a, b and c represent the real trace map, the first simulated trace map and the second simulated trace map, respectively). Fig. 12 shows the real adit, the adit section of the first 3D model and the adit section of the second 3D model.

After being preprocessed, the original images are converted into the images shown in Fig. 13. The total gray analysis results are shown in Fig. 14 and Fig. 15. The total gray of a, b, can c can be indicated by the red area in Fig. 13, the blue area in Fig. 15(a), and the blue area in Fig. 15(b). The calculations show that the total gray of a is 6527, the total gray of b is 6 932 and the total gray of c is 7 544. Therefore, $\eta_{ab} = 0.835$ and $\eta_{ac} = 0.838$.

Fig. 16 shows the gray grade curve analysis results. Among them, (a) shows the gray grade curves of a and b and (b) shows the gray grade curves of a and c. The iterative calculations indicate that $\xi_{ab} = 0.863$ (the gray grade similarity between a and b) and $\xi_{ac} = 0.904$ (the gray grade similarity between a and c).

For the characteristic direction analysis, Fig. 17(a) shows the comparison of a and b, and Fig. 17(b) shows the comparison of a and c. The four characteristic directions of a are [135°, 45°, 120°, 30°], the characteristic directions of b are [135°, 45°, 120°, 30°], and the characteristic directions of c are [45°, 150°, 135°, 120°]. According to section 3.4.5, the characteristic direction similarity between a and b is $\zeta_{ab} = 0.8 \times 1 + 0.2 \times 1 = 1$, and the characteristic direction similarity between a and c is $\zeta_{ac} = 0.8 \times 0.75 + 0.2 \times 0.667 = 0.733$.

In the previous step, the characteristic directions of the real trace map have been found. According to these directions, the gray density distribution similarity can be calculated and is shown in Table 1, where ω_{ab} is the similarity between a and b and ω_{ac} is the similarity between a and c. The final values of ω_{ab} and ω_{ac} are 0.710 and 0.709, respectively.

From the above analysis, all the η, ξ, ζ and ω have been calculated. Then, the comprehensive similarity between a and b, S_{ab}, and the comprehensive similarity between a and c, S_{ac}, can be calculated as follows:

$$S_{ab} = 0.2\eta_{ab} + 0.2\xi_{ab} + 0.3\zeta_{ab} + 0.3\omega_{ab} = 0.853;$$
$$S_{ac} = 0.2\eta_{ac} + 0.2\xi_{ac} + 0.3\zeta_{ac} + 0.3\omega_{ac} = 0.781$$

This means that b is more similar than c to a and is closer to the reality. Therefore, it is better to take the first 3D model as the results.

4.2. Discussions

Through the experiment described above, the algorithm shows its utility as a graphical validation method. The two simulated trace maps are both similar to the real one. The total gray analysis and gray grade curve analysis shows that the quantities of joints in the two simulation trace maps are very close to the real trace map, both from macro and micro perspectives. The largest difference is in the characteristic direction analysis. It shows that the first simulation trace map has higher similarity with the real trace map. In the gray density distribution analysis, the two trace maps almost have the same similarity as the real trace map. Finally, the comprehensive similarity analysis indicated that the first simulation trace map has a higher similarity from the comprehensive perspective. Note that the comprehensive similarity does not show the absolute similarity between the simulation trace map and the real trace map, and it is just a tool to determine which simulation trace map is more suitable to use.

5. Conclusions

When building a geological model, the most important factors are whether the model matches reality and whether the model is valid. Therefore, it is necessary to determine an effective validation method. In this paper, an algorithm that can estimate the similarity between real trace maps and simulated trace maps is described to improve the graphical validation. At the beginning of this algorithm, a preprocessing step is adopted to standardize the two trace maps, and make them more recognizable. Then a total gray comparison, which is commonly used in many other image processing fields, was carried out to analyze the two trace maps from a macro perspective. Next, a new concept named the gray grade curve was proposed to detect the local features of the trace maps. Then, a modified Radon transform was presented based on Radon transform to detect the characteristic directions of the two images. Besides, we proposed how to use Bézier curve to reduce the edge effect. After that, the loop cosine similarity was put forward based on cosine similarity to analyze the gray density distribution similarity. Finally, a comprehensive evaluation equation was put forward based on the above analysis. To demonstrate the performance of the algorithm, an example was provided. The results show that the algorithm can determine which simulated trace map is more similar to the real trace map.

Although the algorithm is feasible, there are several subjective decisions, including the designing of screens in chapter 3.3 and the weights in Eqs. (15), (18) and (21). The assignments of them are just empirical, summarized from repeated experiments. In the next work, these sections will be further studied. In addition, this method cannot determine the availability of a 3D model directly. It only helps to select the optimal model from repeated simulating. The availability of a 3D model should be analyzed by considering numerical validation and graphic validation comprehensively.

Acknowledgements

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.cageo.2018.03.002.

References

Aydin, O., Caers, J., 2013. Image transforms for determining fit-for-purpose complexity of geostatistical models in flow modeling. Comput. Geosci. 17 (2), 417–429.

Bejari, H., Hamidi, J.K., 2013. Simultaneous effects of joint spacing and orientation on TBM cutting efficiency in jointed rock masses. Rock Mech. Rock Eng. 46 (4), 897–907.

Bézier, P.E., 1968. How Renault Uses Numerical Control for Car Body Design and Tooling (No. 680010). SAE Technical Paper).

Cacas, M.C., Ledoux, E., Marsily, G.D., Tillie, B., Barbreau, A., Durand, E., Feuga, B., Peaudecerf, P., 1990. Modeling fracture flow with a stochastic discrete fracture network: calibration and validation: 1. The flow model. Water Resour. Res. 26 (3), 479–489.

Cao, T., Xiao, A., Wu, L., Mao, L., 2017. Automatic fracture detection based on Terrestrial Laser Scanning data: a new method and case study. Comput. Geosci. 106, 209–216.

Dong, Y., Sun, Z., Jia, H., 2006. A cosine similarity-based negative selection algorithm for time series novelty detection. Mech. Syst. Signal Process. 20, 1461–1472.

Ferrero, A.M., Umili, G., 2011. Comparison of methods for estimating fracture size and intensity applied to Aiguille Marbrée (Mont Blanc). Int. J. Rock Mech. Min. Sci. 48 (8), 1262–1270.

Fitton, N., Cox, S., 1998. Optimising the application of the Hough transform for automatic feature extraction from geoscientific images. Comput. Geosci. 24, 933–951.

Gigli, G., Casagli, N., 2011. Semi-automatic extraction of rock mass structural data from high resolution LIDAR point clouds. Int. J. Rock Mech. Min. Sci. 48 (2), 187–198.

Guo, L., Li, X., Zhou, Y., Zhang, Y., 2015a. Generation and verification of three-dimensional network of fractured rock masses stochastic discontinuities based on digitalization. Res. J. Environ. Earth Sci. 73, 7075–7088.

Guo, L., Li, X., Zhou, Y., Zhang, Y., 2015b. Simulation of random 3D discontinuities network based on digitalization and its validation test. Chin. J. Rock Mech. Eng. 34, 2854–2861.

Han, X., Chen, J., Wang, Q., Li, Y., Zhang, W., Yu, T., 2016. A 3D fracture network model for the undisturbed rock mass at the Songta dam site based on small samples. Rock Mech. Rock Eng. 49, 611–619.

Herbert, A., 1996. Modelling approaches for discrete fracture network flow analysis. Dev. Geotech. Eng. 79, 213–229.

Isaksen, S.G., 1998. A review of brainstorming research: six critical issues for inquiry. In: Monograph #302, Creativity Research Unit. Creative Problem Solving Group - Buffalo, Buffalo, NY.

Kemeny, J., Post, R., 2003. Estimating three-dimensional rock discontinuity orientation from digital images of fracture traces. Comput. Geosci. 29, 65–77.

Lambert, C., Thoeni, K., Giacomini, A., Casagrande, D., Sloan, S., 2012. Rockfall hazard analysis from discrete fracture network modelling with finite persistence discontinuities. Rock Mech. Rock Eng. 45, 871–884.

Lemy, F., Hadjigeorgiou, J., 2003. Discontinuity trace map construction using photographs of rock exposures. Int. J. Rock Mech. Min. Sci. 40, 903–917.

Li, M., Zhang, Y., Zhou, S., Yan, F., 2017a. Refined modeling and identification of complex rock blocks and block-groups based on an enhanced DFN model. Tunn. Undergr. Space Technol. 62, 23–34.

Li, X., Chen, J., Zhu, H., 2016. A new method for automated discontinuity trace mapping on rock mass 3D surface model. Comput. Geosci. 89, 118–131.

Li, Y., Chen, J., Shang, Y., 2017b. Connectivity of three-dimensional fracture networks: a case study from a dam site in Southwest China. Rock Mech. Rock Eng. 50, 241–249.

Mendoza-Torres, F., Díaz-Viera, M., Erdely, A., 2017. Bernstein copula modeling for 2D discrete fracture network simulations. J. Petrol. Sci. Eng. 156, 710–720.

Merrien-Soukatchoff, V., Korini, T., Thoraval, A., 2012. Use of an integrated discrete fracture network code for stochastic stability analyses of fractured rock masses. Rock Mech. Rock Eng. 45, 159–181.

Monga, V., Evans, B.L., 2006. Perceptual image hashing via feature points: performance evaluation and tradeoffs. IEEE Trans. Image Process. 15 (11), 3452–3465.

Ni, P., Wang, S., Wang, C., Zhang, S., 2017. Estimation of REV size for fractured rock mass based on damage coefficient. Rock Mech. Rock Eng. 50, 555–570.

Otsu, N., 1979. A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man Cybern. 9, 62–66.

Park, R.G., 2013. Foundation of Structural Geology, third ed. Routledge, London, New York. 201pp.

Riquelme, A.J., Abellán, A., Tomás, R., Jaboyedoff, M., 2014. A new approach for semi-automatic rock mass joints recognition from 3D point clouds. Comput. Geosci. 68, 38–52.

Rossmanith, H.P., 2014. Rock fracture mechanics. Springer, 484pp.

Schultz, R., Gu, Y.J., 2013. Flexible, inversion-based Matlab implementation of the Radon transform. Comput. Geosci. 52, 437–442.

Shanley, R.J., Mahtab, M.A., 1976. Delineation and analysis of clusters in orientation data. Math. Geol. 8 (1), 9–23.

Tahir, M.A., Bouridane, A., Kurugollu, F., Amira, A., 2003. An FPGA based coprocessor for calculating Grey level co-occurrence matrix. In: Circuits and Systems, 2003 IEEE 46th Midwest Symposium on, vol. 2. IEEE, Cairo, pp. 868–871.

Tegen, I., Lacis, A.A., 1996. Modeling of particle size distribution and its influence on the radiative properties of mineral dust aerosol. J. Geophys. Res.: Atmosphere 101 (D14), 19237–19244.

Toft, P.A., Sørensen, J.A., 1996. The Radon Transform: Theory and Implementation. Technical University of DenmarkDanmarks Tekniske Universitet, Department of Informatics and Mathematical ModelingInstitut for Informatik og Matematisk Modellering. København, 312pp.

Umili, G., Ferrero, A., Einstein, H.H., 2013. A new method for automatic discontinuity traces sampling on rock mass 3D model. Comput. Geosci. 51, 182–192.

Wang, S., Chen, J., Shi, B., 2004. Verification of random 3-D fractures network model of rock mass. J. Eng. Geol. 12 (2), 177–181.

Wasantha, P.L.P., Ranjith, P.G., Zhang, Q.B., Xu, T., 2015. Do joint geometrical properties influence the fracturing behaviour of jointed rock? An investigation through joint orientation. Geomech. Geophys.Geo-Energy Geo-Resour. 1, 3–14.

Woolhouse, M.E.J., Shaw, D.J., Matthews, L., Liu, W.C., Mellor, D.J., Thomas, M.R., 2005. Epidemiological implications of the contact network structure for cattle farms and the 20–80 rule. Biol. Lett. 1 (3), 350–352.

Xu, T., Pruess, K., 2001. Modeling multiphase non-isothermal fluid flow and reactive geochemical transport in variably saturated fractured rocks: 1. Meth.Am. J. Sci. 301 (1), 16–33.

Ye, J., 2011. Cosine similarity measures for intuitionistic fuzzy sets and their applications. Math. Comput. Model. 53, 91–97.

Zeeb, C., Gomez-Rivas, E., Bons, P.D., Virgo, S., Blum, P., 2013. Fracture network evaluation program (FraNEP): a software for analyzing 2D fracture trace-line maps. Comput. Geosci. 60, 11–22.

Zhan, J., Chen, J., Xu, P., Zhang, W., Han, X., Zhou, X., 2017. Automatic identification of rock fracture sets using finite mixture models. Math. Geosci. 49 (8), 1021–1056.