Computers & Geosciences 103 (2017) 164–172

Contents lists available at ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

Research paper

Towards semi-automatic rock mass discontinuity orientation and set analysis from 3D point clouds

Jiateng Guo[a,b,*], Shanjun Liu[b], Peina Zhang[b], Lixin Wu[a,b], Wenhui Zhou[b], Yinan Yu[b]

[a] Key Laboratory of Ministry of Education on Safe Mining of Deep Metal Mines, Northeastern University, Shenyang 110819, China
[b] College of Resources and Civil Engineering, Northeastern University, Shenyang 110819, China

ARTICLE INFO

Keywords:
3D point cloud
Rock mass discontinuity
Point segmentation
Cluster analysis
Firefly algorithm (FA)
Fuzzy c-means (FCM)

ABSTRACT

Obtaining accurate information on rock mass discontinuities for deformation analysis and the evaluation of rock mass stability is important. Obtaining measurements for high and steep zones with the traditional compass method is difficult. Photogrammetry, three-dimensional (3D) laser scanning and other remote sensing methods have gradually become mainstream methods. In this study, a method that is based on a 3D point cloud is proposed to semi-automatically extract rock mass structural plane information. The original data are pre-treated prior to segmentation by removing outlier points. The next step is to segment the point cloud into different point subsets. Various parameters, such as the normal, dip/direction and dip, can be calculated for each point subset after obtaining the equation of the best fit plane for the relevant point subset. A cluster analysis (a point subset that satisfies some conditions and thus forms a cluster) is performed based on the normal vectors by introducing the firefly algorithm (FA) and the fuzzy c-means (FCM) algorithm. Finally, clusters that belong to the same discontinuity sets are merged and coloured for visualization purposes. A prototype system is developed based on this method to extract the points of the rock discontinuity from a 3D point cloud. A comparison with existing software shows that this method is feasible. This method can provide a reference for rock mechanics, 3D geological modelling and other related fields.

1. Introduction

A rock mass usually contains "planes of weakness". These planes of weakness occur at all scales and have a statistical distribution of spacing and orientation (Goodman, 1989). In rock engineering, these planes of weakness are referred to as "discontinuities" and include bedding planes, joints, fractures, and schistosity (Slob, 2010). The geomechanical behaviour (such as deformation) of a rock mass is determined by the overall fabric of discontinuities and intact rock. Thus, obtaining information on these discontinuities is important to understand the distribution of discontinuities within the rock mass and to analyse the deformation and instability of the rock mass (Gigli and Casagli, 2011; Slob, 2010). This information has been widely used in rock mechanics, mining engineering and slope engineering (Hoek and Bray, 1981; Hudson and Harrison, 2000). However, the traditional compass measurement method is a single point of contact measurement, which means that the information from the discontinuity is directly measured using tape and a compass for every joint of the discontinuity. When the position is easily measured and the number of discontinuities is small, this method is highly accurate and inexpensive. However, if the study area is vast, high and steep, this approach requires substantial manpower, materials, financial resources and time; this method can also be highly risky (Barton et al., 1974; Franklin et al., 1998; Slob et al., 2005). According to previous measurement experience, measuring 350 joint directions, processing the data, and charting requires a total of 3 individuals working for four days (Slob et al., 2005), which does not meet the requirements of modern geotechnical engineering projects with heavy workloads and tight time constraints. Therefore, a new method must be created to quickly extract rock mass discontinuities.

Remote sensing techniques can be used to acquire three-dimensional (3D) information from the terrain with high accuracy and high spatial resolution (Jaboyedoff et al., 2012) without direct contact. Light Detection and Ranging (LiDAR) and digital photogrammetry are two widely accepted techniques for discontinuity analysis (Riquelme et al., 2015). Using such techniques, one can obtain the point cloud of a rock mass with 3D information (3D coordinates) using non-contact measuring methods much more easily than with the traditional method. The goal of this study is to semi-automatically extract points of different joints based on a 3D point cloud of the study area. First, the point cloud

* Corresponding author at: College of Resources and Civil Engineering, Northeastern University, Shenyang 110819, China.
E-mail addresses: guojiateng@mail.neu.edu.cn (J. Guo), liushanjun@mail.neu.edu.cn (S. Liu), 18240496560@163.com (P. Zhang), awulixin@263.net (L. Wu), zhouwenhuimail@163.com (W. Zhou), 15140280920@163.com (Y. Yu).

http://dx.doi.org/10.1016/j.cageo.2017.03.017
Received 23 July 2016; Received in revised form 14 March 2017; Accepted 21 March 2017
Available online 24 March 2017
0098-3004/

is segmented into subsets. Then, the geometric parameters (dip direction and dip angle, i.e., the orientation of a discontinuity) are calculated through a series of calculations, such as LS (least-square), and a principal component analysis (PCA) of the points of each subset. Finally, the optimal class number can be automatically determined using the cluster validity and stereographic projection of the orientation. Clusters of the same class are merged and coloured after being classified with the FA (firefly algorithm) and FCM (fuzzy c-means, often also referred to as fuzzy k-means) algorithms.

The remainder of this study is organized as follows. Section 2 introduces related work in this area. In Section 3, the data and methods that are used in this study are introduced. The experimental results and analysis are discussed in Sections 4, 5 and 6. Section 7 presents the discussion and conclusions.

2. Related work

In recent decades, many scholars have studied non-contact measuring methods, such as close-range photogrammetry, laser scanning measurements, and the fusion of these two methods (Abellán et al., 2014; Chen et al., 2016; Gigli and Casagli, 2011; Jaboyedoff et al., 2012; Lato and Vöge, 2012; Oppikofer et al., 2009; Riquelme et al., 2014, 2015, 2016; Slob et al., 2007; Sturzenegger and Stead, 2009). The principles of these two methods are different. Point clouds can be directly obtained from laser scanning measurements. A series of photos could be acquired through close-range photogrammetry, and these photos could be converted into point clouds. Based on these point clouds, three types of procedures and software could extract the discontinuities, including the following: (a) commercial software packages, such as Split-FX (Slob et al., 2005), Coltop-3D (Jaboyedoff et al., 2007), JMX Analyst (available on http://www.3gsm.at/static/eng/home_eng.html), Sirovision (available on http://www.sirovision.com), PolyWorks (available on http://www.innovmetric.com), VirtuoZo (available on http://www.supresoft.com.cn/chinese/Download/Documentation/VirtuoZo/), ShapeMetrix 3D (available on http://www.3gsm.at/static/eng/home_eng.html) and LensPhoto (available on http://www.lensoft.com.cn/); (b) procedures for personal research, such as DiAna (Discontinuity Analysis) (Gigli and Casagli, 2011) and PlaneDetect (Vöge et al., 2013); and (c) open-source software, such as DSE (Discontinuity Set Extractor) (Riquelme et al., 2014).

We can manually or semi-automatically extract information on discontinuities according to 3D geological models that are based on point clouds or that directly handle 3D point cloud data (Chen et al., 2016; García-Sellés et al., 2011; Gigli and Casagli, 2011; Lato and Vöge, 2012; Lato et al., 2009; Olariu et al., 2008; Slob et al., 2005; Sturzenegger and Stead, 2009; Van Knapen and Slob, 2009; Vöge et al., 2013). The manual method involves manually selecting the point of the discontinuity in the 3D display scene. This approach has been applied in both relevant commercial software (such as PolyWorks, VirtuoZo and ShapeMetrix 3D) and the literature (Maerz et al., 2013; Yang and Zhao, 2015). More knowledgeable and experienced operators could clearly obtain more accurate information through this manual method. The other method involves segmenting the point cloud to obtain small rectangular patches, which are manually clustered into different discontinuities (such as DiAna). Automatic/semi-automatic methods are divided into two categories. The first category is based on the principle of photogrammetry and a pattern recognition program and usually involves constructing a 3D surface model of discontinuities (such as PlaneDetect). This method has been extensively studied and has a high degree of automation. However, complex and irregular cases (such as cases that include folding or deep concavity) require more accurate and complex 3D geological modelling technology, which is often tedious, and the accuracy of the model thus cannot be guaranteed (Frank et al., 2007; Gigli and Casagli, 2011; Natali et al., 2013). Another category is the direct classification of point clouds (Riquelme et al., 2014), which can avoid the process of model reconstruction and directly obtain discontinuities from the original point cloud. This method has high efficiency and high accuracy when the number of points is not very large. When hundreds of thousands or millions of points are present, however, the calculation and classification of each point will take a long time (Slob et al., 2005).

In substance, the extraction of discontinuity sets based on a 3D point cloud is a point cloud classification problem, namely, to distinguish points that belong to different discontinuities. The classical classification method, which is known as the FCM algorithm, is simple and relatively rapid (Hammah and Curran, 1998). The classification results depend on the number of groups and the initial clustering centres, which are selected by the user, so this method is highly subjective, and the results may fall into local optima. The genetic algorithm (Cai et al., 2005), niche algorithm (Lu et al., 2007), particle swarm algorithm (Song et al., 2012) and other bionic algorithms have been introduced to overcome the random selection uncertainty of the initial clustering centres for the FCM algorithm, although the complexity of the encoding and decoding limits the efficiency of these algorithms. The FA (Yang, 2008) simulates the swarm behaviour of fireflies. The stochastic optimization algorithm has been widely used in many fields, such as multi-objective and industrial optimization problems, scheduling problems, and image processing. The FA has global optimization and high convergence speed and thus can be applied to determine the initial cluster centres for discontinuity extraction. Some scholars have attempted clustering analysis based on the FA for the dominative attitudes of rock mass discontinuities that are measured with compasses and tape (Song et al., 2015), and the results were acceptable. However, the FA has not yet been applied to the extraction of rock mass discontinuities based on 3D points. The FA is used to find the best initial cluster centres, and then the FCM algorithm is used to classify discontinuity clusters. Clusters that belong to the same class (the same discontinuity set) are merged.

3. Methodology

In this study, a method that combines the FA and FCM algorithms and is based on point cloud classification is proposed to semi-automatically extract the joint points from extensive 3D point cloud data. The optimal number (number of discontinuity sets) is determined by the clustering parameters. The extraction process of rock mass discontinuity sets is designed as follows: (Fig. 1).

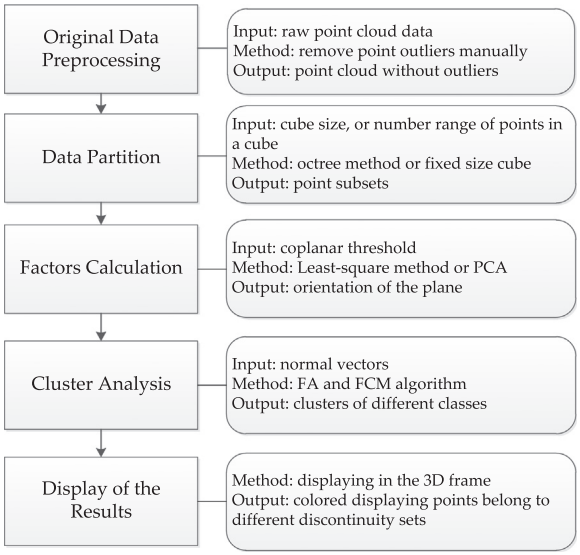

Fig. 1. Flowchart of the methodology.

1. Original data pre-processing.
2. Data partitioning: the imported data are segmented into point subsets according to the octree method or a fixed size cube.
3. Factor calculation: the orientations of the best fit plane of the subsets are computed by the LS method or PCA. A stereographic projection is applied during the analysis of the orientation data using Dips (available on https://www.rocscience.com/rocscience/products/dips/) to obtain the range of the number of discontinuity sets.
4. Cluster analysis: the FA is used to obtain the initial cluster centres, and the FCM algorithm is used to calculate the best classification result.
5. Merging and displaying of the results: clusters that belong to the same discontinuity are merged and displayed in the same colour in the 3D frame.

3.1. Description of the datasets

Three cases were used in this study: three field point clouds with different numbers of points and different numbers of discontinuity sets. Case study 1 was intended to provide a comparison between this method and the DSE software. The data from case study 2 were used for the parameter sensitivity analysis. The data from case study 3 were used to test the efficiency of the proposed method.

(1) Case study 1. The data from case study 1 (Fig. 2), which were collected in Padua, Italy, were provided by Siefko Slob and collected by Alessia Viero and Antonio Galgaro from the Department of Geosciences, University of Padova, Italy. These data contain natural fracture surfaces and vegetation points. A total of 148,298 points were collected. The size of the entire study area was 36.06 m by 47.87 m by 19.02 m. The data were stored in the XYZ format, with each line storing the x, y, and z coordinates and the RGB value of a point.

(2) Case study 2. The laser scanning data from case study 2 were obtained from https://www.researchgate.net/publication/289523409_raw_point_cloud_data_-_ascii_x_y_z_intensity_-_metadata (Slob's personal website at Research Gate). Photos from case study 2 are shown in Fig. 3. A portion of the slope (the area that was selected by the black rectangle; an enlarged display is shown on the right side) was used in this analysis. The photos were taken near the village of Torroja, Spain, along road number TP 7403. A total of 77,302 points were collected, and the resolution was approximately 6 mm by 6 mm. The data were scanned from a distance of approximately 15 m, and the entire study area was 1.38 m by 0.64 m by 1.11 m. The storage format was the same as in case study 1.

(3) Case study 3. The laser scanning data from case study 3 were obtained from http://geol.queensu.ca/faculty/harrap/RockBench/dataDownloads/index.html. A photo from case study 3 is shown in Fig. 4. The photos were taken near Ontario, Canada, along road number 15. The data were obtained using the Leica HDS6000 laser scanning equipment, and the point spacing was less than 1 cm. A total of 2,087,187 points were collected. The data were scanned from a distance of approximately 15 m, and the entire study area was 13.28 m by 4.21 m by 3.71 m. The storage format was the PLY format, which contains colour information on the points and their coordinates.

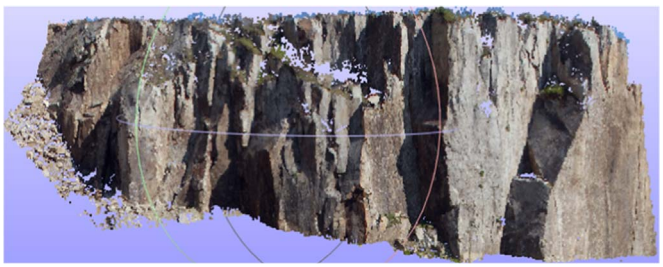

Fig. 2. Raw data for case study 1.

3.2. Data partitioning

The density of a 3D point cloud enables us to consider point subsets that meet the coplanarity conditions as a whole, which can improve the efficiency of point cloud processing. The advantage of point cloud segmentation is that the original point cloud can be directly analysed without surface reconstruction (Fig. 5).

3.2.1. Data pre-processing

The noise points in the cube both influence the orientation of the fitting plane and may lead to judgement errors regarding whether a cube is available. Therefore, data pre-processing was performed before the data partitioning to remove outlier points, such as points of vegetation, dust and insects. These points reduced the accuracy of the point extraction when determining the discontinuity and increased the execution time. Thus, these noise points had to be removed before the discontinuity extraction. These isolated outliers could be automatically removed with software (CloudCompare and Geomagic). However, the number of these points was relatively small, and the characteristics of these points were unclear (Lato et al., 2010). Thus, these noise points could not be automatically removed with a single key in the software, and some manual operations were required during data pre-processing.

3.2.2. Fixed cube segmentation

If the original point cloud is uniformly distributed, the 3D point cloud can be divided into equally sized cubes to avoid great differences in the number of points in each subset. Data partitioning is actually the gridding of a point cloud. A cube of fixed size (*Size0*) was used to segment the point cloud (Fig. 7b). The minimum bounding box of the point cloud, which had three lengths x_0, y_0, and z_0 in the x, y, and z directions, respectively, was used as the fundamental box. The fundamental box was then divided into N_0 cubes ($N_0 = N_x * N_y * N_z, N_x = \frac{x_0}{Size0}$, $N_y = \frac{y_0}{Size0}$, $N_z = \frac{z_0}{Size0}$). Cubes that contained fewer than 4 points or had no points were considered non-conformers and were removed. After that, the normal vector of the points in every cube were calculated.

The cube size was decided by the density of the point cloud and the roughness of the rock mass. Point clouds that were obtained by different laser scanners with different minimum angle steps or different angles of incidence for the laser beam (Lato et al., 2010) had different densities. Considering the difference in the roughness of the rock mass in different study areas, the cube size had to be manually set by the user. In accordance with the conclusion of the nearest neighbour points in the literature (Riquelme et al., 2014), the average number of points in the cube ranged between 15 and 30, and the cube size could be set between 4ρ and 7ρ when the point spacing ρ(the unit is millimetres) was known. If the minimum scale of the discontinuity (the smaller of the length and width of the minimum discontinuity) was S, the cube size was between $1/3S$ and S. Therefore, the user could obtain the best cube size range according to this feedback.

3.2.3. Octree partitioning

Segmenting a point cloud with a fixed cube is simple and convenient, but the size of the cube must be set in advance. The octree approach was used to segment the point cloud in this work to improve the segmentation and management of points.

The minimum bounding box of a point cloud was considered the “root node” of the octree, which was then segmented into 8 equally sized sub-cubes. Sub-nodes that met the constraint conditions were not split further. Otherwise, the corresponding sub-cubes continued to be divided into 8 equally sized sub-cubes until the entire space cube met the constraint conditions. Fig. 6 shows the octree partitioning principle.

Two parameters were set during this step: the minimum number (*nmin*) and maximum number (*nmax*) of the points in each cube. Then,

Fig. 3. Photos for case study 2 (Slob, 2010).

Fig. 4. Photo for case study 3 (Lato et al., 2012).

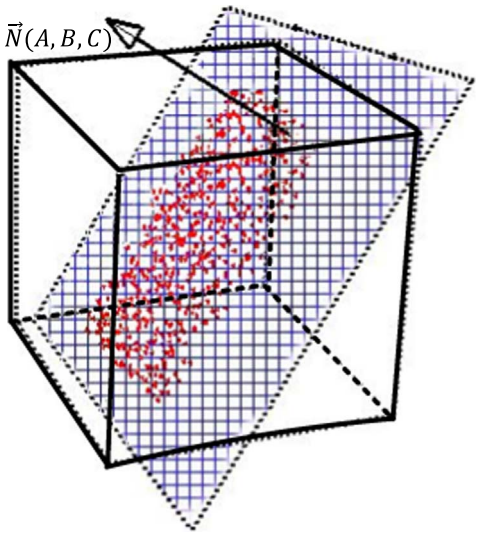

Fig. 5. Fitting plane (blue grid) of points within a cube (red points). The black arrow ($\vec{N}(A, B, C)$) is the normal direction of the plane. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the point cloud was divided by the octree principle; this division should satisfy that the number of points in one cube should be larger than *nmin* and less than *nmax*. Meanwhile, the coplanar test was conducted for the points in each cube; the average distance of all these points to the fitting plane should be less than a coplanarity threshold.

3.3. Calculation of normal vectors of point clusters

The LS can be used to calculate the fitting plane equation of point subsets (Feng et al., 2001). The LS method is feasible when the number of points is small, but this approach is slow and may not meet the requirements when the number of points is large, potentially reaching millions of points. Therefore, we introduced the PCA method to calculate normal vectors of the point subsets for a large number of point clouds.

PCA is based on dimensional reduction, namely, transforming multiple indicators into a small number of the most important indicators. Therefore, this method is very efficient and suitable for the calculation of a large amount of point cloud data (Jolliffe, 2002).

$$Ax+By+Cz+D = 0 \tag{1}$$

The normal vector of the fitting plane (A, B, C) can be obtained from the equation of the fitting plane (formula 1) of points within a cube. The normal vector is normalized to the unit normal vector $\vec{n}(l, m, n)$. Then, the orientation (dip/direction (α) and dip (β)) of the subset can be calculated according to formula (2) and formula (3):

$$\alpha = tan^{-1}\left(\frac{m}{l}\right)+Q \tag{2}$$

where if $l > 0$ and $m > 0$, Q=0°; if $l > 0$ and $m < 0$, Q=360°; otherwise, Q=180°.

$$\beta = tan^{-1}\left(\frac{n}{\sqrt{l^2 + m^2}}\right) \tag{3}$$

After the orientation data were obtained, the pole density contour map (Fig. 7c) was obtained using the stereographic projection software Dips. The number or number range of discontinuity sets was also obtained. The projected pole was the directional data from each fitting plane of points within a cube, namely, each cluster.

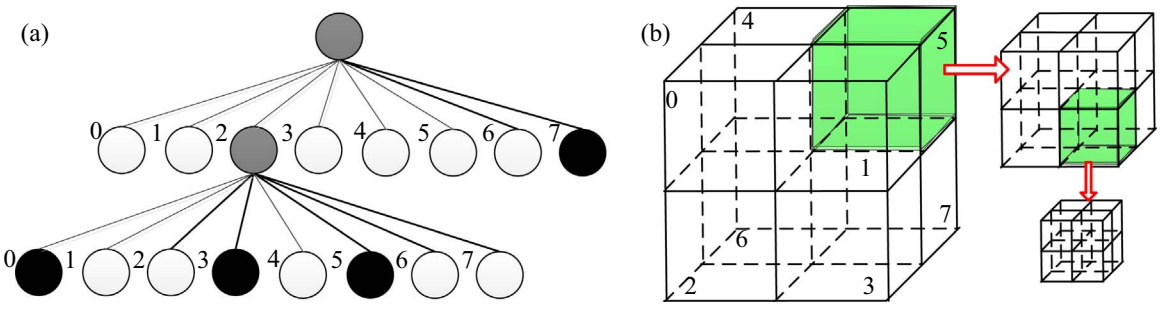

Fig. 6. Octree principle. (a) Structure of the octree. (b) Space partition based on the octree.

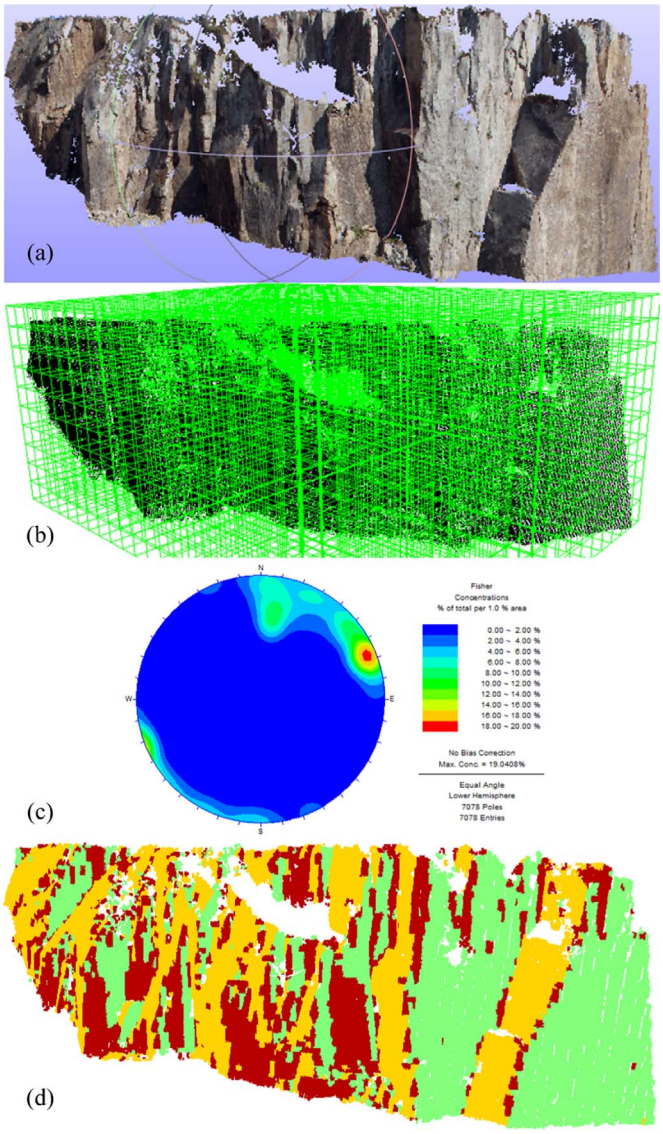

Fig. 7. Results of case study 1. (a) Point cloud after pre-processing. (b) Sketch of the segmentation of the point cloud. (c) Stereographic projection of the vector data. A total of 7084 poles were present. (d) Final classification figure.

3.4. Cluster analysis of directional data

Different discontinuities have different directions, so the cluster analysis method is based on the normal vector. A pole in a stereographic projection refers to the dip direction and dip angle (also called the orientation) of a cluster of points in the 3D space within a cube. Here, we clustered these poles mathematically while actually clustering these clusters of points. The classical clustering algorithm FCM is a local optimization algorithm, and the initial centres and number of discontinuity sets are specified by the user, which results in a lack of objectivity. If the initial centres are not reasonable, a satisfactory result cannot be obtained. The FA has global convergence, and the convergence speed is fast (Wong, 2012). Therefore, the FA was combined with the FCM algorithm in this study to obtain the global optimum centres and the best clustering results. We obtained a set of optimal initial cluster centres by FA; then, the directional data were classified according to the FCM algorithm. The number or number range of discontinuity sets was obtained according to a stereographic projection of the orientation data. If a number range was obtained, the cluster validity function was then introduced. Cluster validity parameters of different classification numbers were compared to find the optimal number of discontinuity sets.

3.4.1. Fuzzy c-means algorithms(FCM)

Therefore, the FCM algorithm is considered the optimal algorithm for the cluster analysis of directional data from rock mass discontinuities (Hammah and Curran, 1998). The dataset is divided into a certain number of classes according to the membership degree of each group. The membership degree ranges from zero to one; the greater the probability that the data belong to a collection, the closer the degree of membership of this collection to a value of 1 (Slob, 2010).

The dataset was $X=(X_1, X_2, \ldots, X_N)$, where N is the number of directional data points and the unit normal vector$X_i = (l_i, m_i, n_i)$. The purpose of clustering is to divide the directional data into C sets with C cluster centres $V=(V_1, V_2, \ldots, V_c)$, $V_i = (l_i, m_i, n_i)$, $1 \le i \le C$. The following objective function (J_m) is minimized (iteratively) during this process.

$$J_m = \sum_{j=1}^{N} \sum_{i=1}^{C} u_{ij}^2 d^2(X_j, V_i) \tag{4}$$

where $d^2(X_j, V_i)$ is the distance between pole X_j and cluster centre V_i. Because the clusters were classified according to the normal vector, the sine value of two normal vectors was used to measure the proximity of the two clusters, which we called the "distance". Thus, $d^2(X_j, V_i)$ could be represented as follows (Hammah and Curran, 2000):

$$d^2(X_j, V_i) = 1 - (X_j \bullet V_i)^2 \tag{5}$$

where u_{ij} is the membership degree between cluster j and centre i and $0 \le u_{ij} \le 1$, where i=1,2,...,C and j=1,2,...,N.

$$u_{ij} = \frac{1}{d^2(X_j, V_i)} \left[\sum_{k=1}^{C} \left(\frac{1}{d^2(X_j, V_k)} \right) \right]^{-1} \tag{6}$$

The update of the centres is

$$V_i = \sum_{j=1}^{N} u_{ij}^2 X_j / \sum_{j=1}^{N} u_{ij}^2 \tag{7}$$

3.4.2. Firefly algorithm (FA)

The principle of the FA is as follows. Each firefly represents a solution to the problem, namely, a group of cluster centres (Yang, 2008). The brightness of the firefly represents the superiority of the solution. To quantify the brightness of the firefly $I(s)$ (see formula (8)), the objective function (formula (4)) was used to express the specific relationship between the superiority of the solution (Song et al., 2012) and the data. The firefly with the maximum brightness was the optimal solution.

$$I(s) = \frac{1}{J_m + 1} \tag{8}$$

In nature, the brightest firefly will attract fireflies with less brightness. Based on this algorithm, the optimal solution of the problem can be obtained via iteration. The firefly F_i is attracted by firefly F_j and moves to firefly F_j. Changes in the position of the firefly F_i update this firefly; the position of F_i after moving can be determined as follows:

$$V_i = V_i + \eta(V_j - V_i) + \xi(rand - 0.5) \tag{9}$$

where V_i and V_j are the positions of F_i and F_j, respectively; η is the attraction between fireflies V_i and V_j; ξ is the step size; and $rand$ is a random number in [0,1].

The firefly attraction η between fireflies is related to the environmental medium and distance (r) between fireflies:

$$\eta = \eta_0 * exp(-\gamma r^2) \tag{10}$$

where η_0 is the greatest attraction, namely, the attraction when $r=0$, and γ is related to the light absorption factor for the medium.

The distance between the fireflies is set as the sum of the distances between the corresponding centres:

$$r^2 = d^2(X_1, X_1') + \quad d^2(X_2, X_2') + \ldots + d^2(X_C, X_C') \tag{11}$$

where C is the number of centres.

When using the FA in the classification of clusters, the first step is to initialize the parameters, namely, the number of fireflies N, maximum attraction η_0, light absorption factor for the medium γ, step size ξ, number of clusters C, and maximum number of iterations $Tmax$. The implementation of the FA involves an initial solution through iteration to obtain the optimal solution. Therefore, the parameters (such as the maximum attractiveness degree, the light absorption factor and the step length) merely affect the number of iterations to some extent but barely affect the final results. In theory, the greater the number of fireflies and number of maximum iterations, the more reliable the results become. However, this iterative process is convergent, so when the loop is stable, the increase in these two parameters has little effect on the results. Therefore, we can refer to the parameter settings of the algorithm in the existing literature. The maximum attractiveness degree, the light absorption factor and the step length were set to 1, 0.8, and 0.5, respectively. The number of fireflies and the maximum number of iterations were set to 30 and 15, respectively.

Algorithm 1. Firefly Algorithm

```
Algorithm 1:
For t=0 to Tmax
  For i=0 to n
  For j=0 to n
  If I(V_j) > I(V_i), then
  Compute η_ij using formula 10
  Update V_i using formula 9
  Update U_i using formula 6
  Update I(V_i) using formula 8
  End if
  End for
  End for
  The fireflies are sorted by their brightness in ascending order. The
  brightest firefly is the optimal solution.
End for
```

First, the initial position of the fireflies $V_1, V_2, \ldots, V_N$ is obtained. Then, the membership degree matrix U_i and brightness $I(V_i)$ of each firefly can be calculated according to formula (6) and formula (8), respectively. The fireflies are sorted by their brightness in ascending order and updated through an iterative process. The pseudo code is shown in algorithm 1.

The final optimal solution results in the best centres. The membership matrix can be calculated according to formula (6). Then, the cluster analysis and merging of clusters in the same group are performed.

3.4.3. Parameters of cluster validity

The optimal feature number is automatically decided by an introduced cluster validity function. The clustering validity evaluation method provided the evaluation results of the clustering results in the form of clustering parameters based on an evaluation of the density and dispersion degree of the clustering results (Hammah and Curran, 1998). The corresponding clustering parameters of different group numbers were calculated and compared with each other to determine the optimal classification group number. The clustering validity function is mainly divided into two categories: one category considers the basic geometric structure of the dataset in the classification, and the other category introduces fuzzy factors (Tang and Yang, 2009). The Xie-Beni validity index was discussed and adopted by Hammah and Curran (2000) and Van Knapen and Slob (2009) and performed very well in their experiments. Adopting more functions into the operation would result in more objective results (Slob, 2010), so both types of functions were used to ensure the reliability of the results. The representative functions are the partition coefficient V_{pc} (Bedzek, 1973) and the Xie-Beni validity index V_{xb} (Xie and Beni, 1991). Large V_{pc} values (i.e., close to 1) indicate good classification results with few errors, whereas small V_{xb} values (i.e., close to 0) indicate a low correlation between the different clusters and a large clustering effect. The calculation formulas are as follows:

$$\mathrm{V}_{pc} = \frac{1}{N}\sum_{j=1}^{N}\sum_{i=1}^{C} u_{ij}^2 \tag{12}$$

$$V_{xb} = \frac{\sum_{j=1}^{N}\sum_{i=1}^{C} u_{ij}^2 r^2(P_j, V_i)}{N\cdot(min_{i\neq k} d^2(V_i, V_k))} \tag{13}$$

where N is the number of orientation data points, C is the number of classes, $u_{ij}$is the membership degree between cluster P_j and centre V_i (formula (6)), and $r^2(P_j, V_i)$ is the distance between cluster P_j and centre V_i (formula (11)).

Based on this workflow, a prototype system was developed based on the QT framework, and three experiments were performed on a ThinkPad laptop (1.7_GHz(R) Intel Core i5-4210U, 8 GB of RAM memory, and an Intel HD graphics card).

4. Results from case study 1: experimental analysis

This method involved the deletion of point outliers (vegetation points, isolated points, and noise points) and an analysis of the influence of the aforementioned process. Section 4.1 presents a comparison of the proposed methods and DSE software.

4.1. Comparison experiments for case study 1

Fig. 7a shows the point cloud after pre-processing (removal of outlier points). The segmenting of the point cloud and the cube size was set to 0.400 m (Fig. 7b). Point sets that satisfied the conditions formed clusters, and the coplanarity threshold value (σ_t) was set to 0.035 m. Then, the orientations of each point cloud subset were calculated. The stereographic projection of the orientation data with Dips is shown in Fig. 7c. The classification range is 2–3. The cluster analysis results are shown in Fig. 7d, and the optimal classification group number is 3 categories. A total of 17680 fewer points are present in Fig. 7d than in Fig. 7a, which represents untreated points, including points along the edge cubes and other non-coplanar point subsets.

Riquelme (Riquelme et al., 2014) proposed a method to automatically identify discontinuity sets. In contrast, we used the proposed methods and DSE in this work to operate on data from case study 1. Method 1 used fixed cubes during segmentation and the LS during vector calculation. The results are shown in Fig. 7. Method 2 used the octree approach during segmentation and PCA during vector calculation. The results are shown in Fig. 8. According to a comparison of Fig. 7d, Fig. 8b and Fig. 9, which were the results when using the DSE software, the classification results were basically the same. Table 1 compares the detailed information on the results of these methods. Our method requires improvement regarding some finer details.

5. Results from case study 2: sensitivity analysis

Sensitivity tests were conducted in this study, including the determination of suitable parameters to obtain better results.

5.1. Influence of cube size

The data from case study 2 were used to perform a size sensitivity

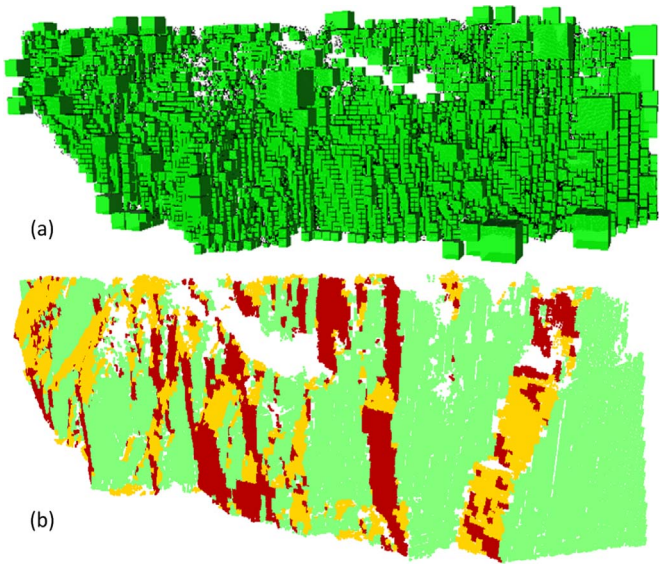

Fig. 8. Results of case study 1 when using Method 2. (a) Segmentation of the point cloud for case study 1 based on the octree approach. (b) Final classification figure of case study 1.

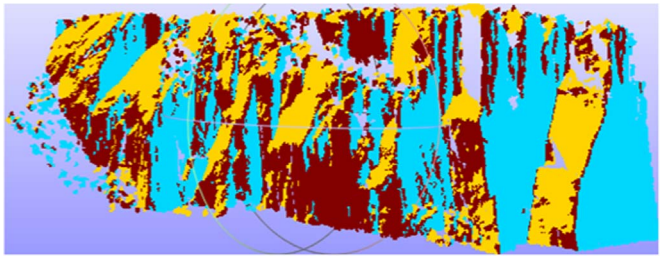

Fig. 9. Classification results from case study 1 when using DSE.

analysis experiment. Different cube sizes, which influenced the results (Fig. 10), could be selected. Small cube sizes indicate smaller features. However, an overly small cube can only contain a few points, which means that the normal vector is then determined by only a few points and is readily changed. Large cube size improves the cluster validity parameters within a certain range (Fig. 11). The cluster validity parameters tended to be stable when the cube size was larger than 0.04 m (Fig. 11).

5.2. Coplanarity threshold

Coplanar detection was performed to ensure that the points in the subset belonged to one plane. Clusters that did not meet the coplanar conditions may have contained boundary points and vertices. The σ value of points in a cluster concerning the best fit plane was then compared with σ_t. If $\sigma < \sigma_t$, a cluster formed. The chosen value of $\sigma_t$is important because this value affects both the number of available points and the number of available cubes (Fig. 12). Changes in the cluster validity parameters were relatively smooth. The curves for the unused points and unused cubes changed more slowly when the threshold value was between 0.025 m and 0.045 m.

6. Analysis of results from case study 3

Case study 3 (more than two million points) was introduced to further verify whether this method still had high efficiency for data with a large number of points. The results showed that the processing of case study 3 took a total of 1 min and 25 s. The point cloud is shown in Fig. 13(a), and the partitioning of the point cloud based on the octree approach is shown in Fig. 13(b). The maximum and minimum numbers of points within a cube were set to 3000 and 10. Point subsets that met the coplanarity conditions formed clusters. As shown in Fig. 13(c), the classification range was 3–4. The final results of the cluster analysis are shown in Fig. 13(d), and the best classification group number was 3. In total, 2442 fewer points were left untreated, including points in clusters at the boundary and other non-coplanar clusters.

7. Discussion and conclusions

A new method for extracting joint points from 3D point clouds was proposed in this study, and a prototype system was developed to conduct the experiments. Data with different number of points were processed with different methods. The results of case study 1, which used three different methods (the proposed methods and DSE), were compared in Section 4, and acceptable results were obtained in less than two minutes. Furthermore, when the data contained more than two million points, proposed method 2 could still process these data within two minutes. The case studies in this work showed that the proposed methods could extract points of different discontinuity sets with high efficiency from a 3D point cloud, which is helpful to understand surface discontinuities in large-scale geological bodies.

In this method, the FA with global optimality was used for the rock mass discontinuity extraction process based on the 3D point cloud. Meanwhile, not all points had to be calculated, which avoided redundant computations. This method was only based on raw point clouds without any complex triangulation and the establishment of a 3D model of the geological body. The quality of the results when using different methods (the proposed methods and DSE) was also compared in Section 4, and the proposed methods could efficiently obtain more acceptable results for data with both ten thousand points and two million points.

Considering the limits of point cloud segmentation, some deficiencies had to be handled: (1) points near the boundary could not be accurately classified because the cube at the boundaries contained at least two discontinuities,; and (2) accurate boundary lines could not be obtained. Addressing these problems will be the focus of the authors' future work.

Acknowledgments

We acknowledge and thank the Editor-in-Chief Professor Gregoire Mariethoz and the three reviewers, namely, Professor Siefko Slob, Professor Adrián Riquelme, and an anonymous reviewer, for their thoughtful and constructive comments on this manuscipt. The laser scanner raw data from case study 1 were privately provided by Siefko Slob, and the data for case study 2 were obtained from his personal website at Research Gate (https://www.researchgate.net/profile/Siefko_Slob); the help of Siefko Slob is kindly appreciated. In

Table 1
Comparison of the proposed methods and DSE.

	All points	Removed points	Untreated points	Points in Set 1	Points in Set 2	Points in Set 3	Process time
DSE	148,298	0	19,462	39,879	42,267	45,695	9 min 45 s
Method 1	148,298	11,202	9663	34,229	32,850	63,539	1 min 30 s
Method 2	148,298	11,202	7575	33,049	53,590	42,882	1 min 25 s

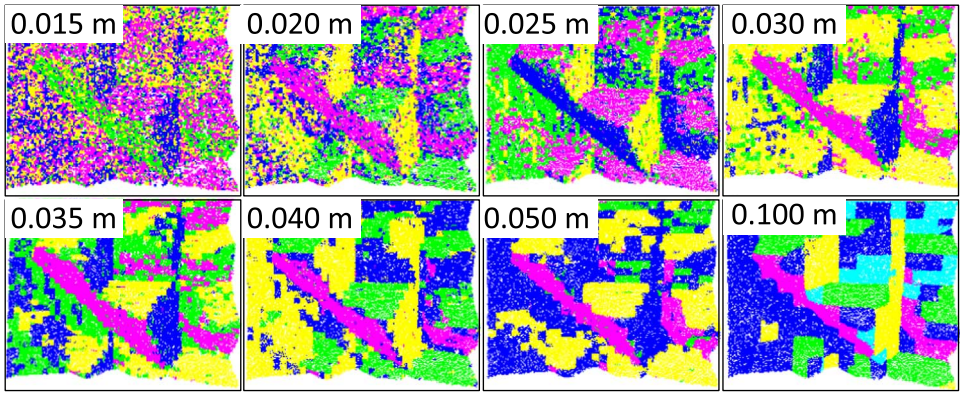

Fig. 10. Results of case study 2 with cube sizes from 0.015 m to 0.1 m.

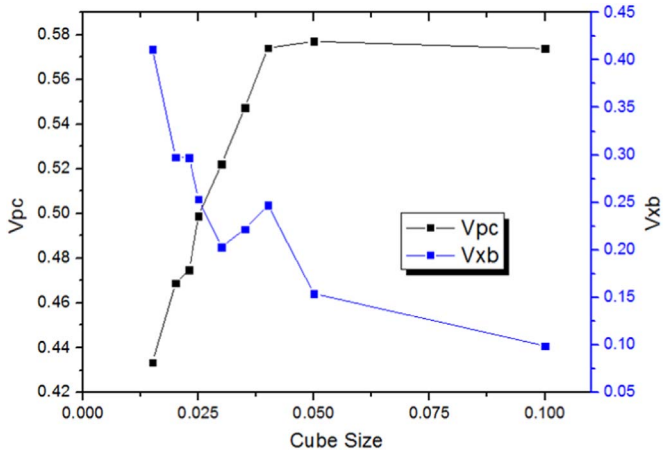

Fig. 11. Changes in the cluster validity parameters for case study 2 to different cube sizes (horizontal coordinates; unit: m). The partition coefficient V_{pc} (black line with black points) and the Xie-Beni validity index V_{xb} (blue line with blue points) are shown. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

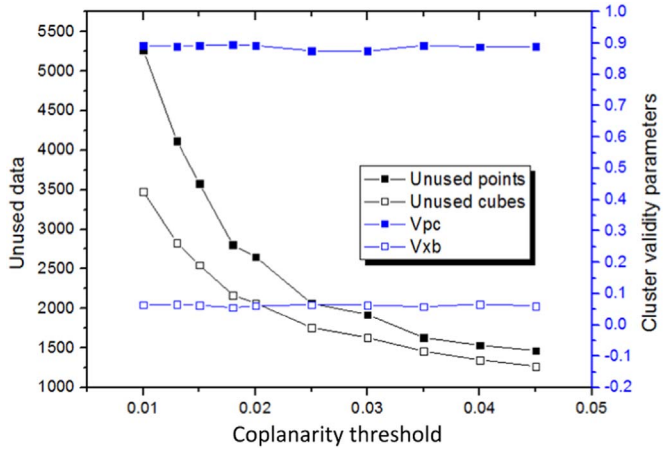

Fig. 12. Graph of the changes in the cluster validity parameters with different coplanar thresholds (horizontal coordinates; unit: m).

addition, data for case study 1 were collected by Alessia Viero and Antonio Galgaro from the Department of Geosciences, University of Padova, Italy. We also express gratitude to them. This research was jointly supported by the National Natural Science Foundation (41671404), the Liaoning Province Natural Science Foundation (2015020581) and the China Geological Survey Projects (DD20160355).

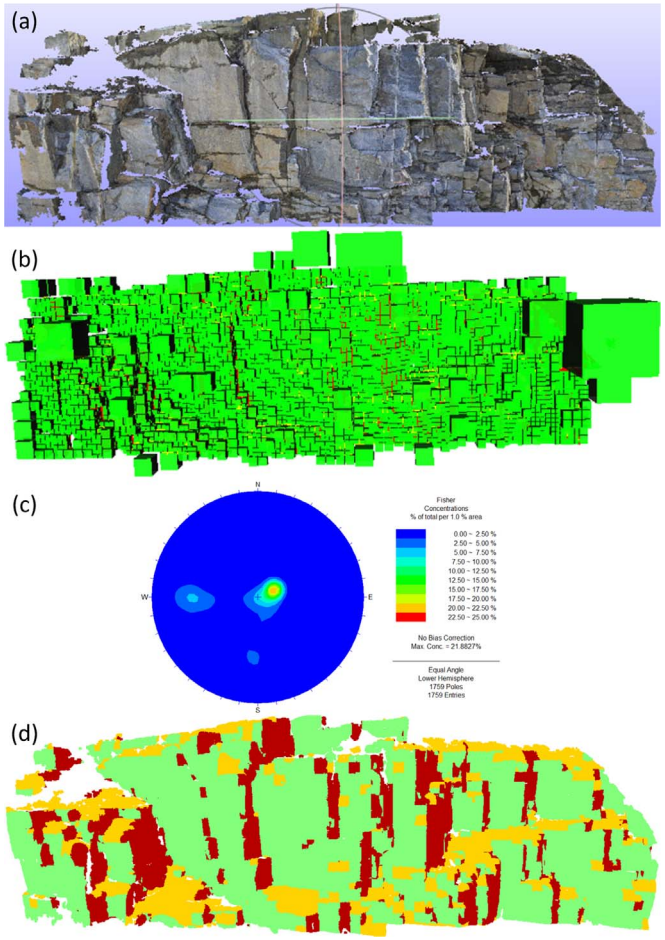

Fig. 13. Results of case study 3 based on Method 2. (a) Point cloud of case study 3. (b) Segmentation of the point cloud based on the octree approach. (c) Stereographic projection figure of case study 3. (d) Final classification figure of case study 3.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.cageo.2017.03.017.

References

Abellán, A., Oppikofer, T., Jaboyedoff, M., Rosser, N.J., Lim, M., Lato, M.J., 2014. Terrestrial laser scanning of rock slope instabilities. Earth Surf. Process. Landf. 39 (1), 80–97.

Barton, N., Lien, R., Lunde, J., 1974. Engineering classification of rock masses for the design of tunnel support. Rock Mech. 6 (4), 189–236. http://dx.doi.org/10.1007/BF01239496.

Bedzek, J.C., 1973. Cluster validity with fuzzy sets. J. Cyber. 3 (3), 58–73.
Cai, M.F., Wang, P., Zhao, K., Zhang, D.K., 2005. Fuzzy C-means cluster analysis based on genetic algorithm for automatic identification of joint sets. Chin. J. Rock Mech. Eng. 24 (3), 371–376, (in Chinese).
Chen, J., Zhu, H., Li, X., 2016. Automatic extraction of discontinuity orientation from rock mass surface 3D point cloud. Comput. Geosci. 95, 18–31.
Feng, Q.J., Jing, L.J., Stephansson, O.J., Vejde, S.J., 2001. A new approach for geological surveying of exposed rock faces. In: Proceedings of ISRM International Symposium–2nd Asian Rock Mechanics Symposium, International Society for Rock Mechanics, Lisbon.
Frank, T., Tertois, A.L., Mallet, J.L., 2007. 3D-reconstruction of complex geological interfaces from irregularly distributed and noisy point data. Comput. Geosci. 33, 932–943.
Franklin, J.A., Maerz, N.H., Bennett, C.P., 1998. Rock mass characterization using photo analysis. Int. J. Min. Geol. Eng. 6, 97–112.
García-Sellés, D., Falivene, O., Arbués, P., Gratacos, O., Tavani, S., Muñoz, J.A., 2011. Supervised identification and reconstruction of near-planar geological surfaces from terrestrial laser scanning. Comput. Geosci. 37 (10), 1584–1594. http://dx.doi.org/10.1016/j.cageo.2011.03.007.
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. http://dx.doi.org/10.1016/j.ijrmms.2010.11.009.
Goodman, R.E., 1989. Introduction to Rock Mechanics second ed.. Wiley, New York.
Hammah, R.E., Curran, J.H., 1998. Fuzzy cluster algorithm for the automatic identification of joint sets. Int. J. Rock Mech. Min. Sci. 35 (7), 889–905. http://dx.doi.org/10.1016/S0148-9062(98)00011-4.
Hammah, R.E., Curran, J.H., 2000. Validity measures for the fuzzy cluster analysis of orientations. IEEE Trans. Pattern Anal. Mach. Intell. 22 (12), 1467–1472.
Hoek, E., Bray, J.D., 1981. Rock Slope Engineering. CRC Press, Boca Raton.
Hudson, J.A., Harrison, J.P., 2000. Engineering Rock Mechanics-An Introduction to the Principles. Elsevier, Burlington.
Jaboyedoff, M., Metzger, R., Oppikofer, T., Couture, R., Derron, M.H., Locat, J., Turmel, D., 2007. New insight techniques to analyze rock-slope relief using DEM and 3D-imaging cloud points: COLTOP-3D software. In: Eberhardt, E., Stead, D., Morrison, T. (Eds.), Rock Mechanics: Meeting Society's Challenges and Demands. Taylor & Francis, London, 61–68.
Jaboyedoff, M., Oppikofer, T., Abellán, A., Derron, M., Loye, A., Metzger, R., Pedrazzini, A., 2012. Use of LIDAR in landslide investigations: a review. Nat. Hazards 61 (1), 5–28. http://dx.doi.org/10.1007/s11069-010-9634-2.
Jolliffe, I., 2002. Principal Component Analysis. Springer, New York.
Lato, M.J., Vöge, M., 2012. Automated mapping of rock discontinuities in 3D lidar and photogrammetry models. Int. J. Rock Mech. Min. Sci. 54, 150–158. http://dx.doi.org/10.1016/j.ijrmms.2012.06.003.
Lato, M.J., Diederichs, M.S., Hutchinson, D.J., 2010. Bias correction for view-limited Lidar scanning of rock outcrops for structural characterization. Rock Mech. Rock Eng. 43 (5), 615–628. http://dx.doi.org/10.1007/s00603-010-0086-5.
Lato, M.J., Diederichs, M.S., Hutchinson, D.J., Harrap, R., 2009. Optimization of LiDAR scanning and processing for automated structural evaluation of discontinuities in rockmasses. Int. J. Rock Mech. Min. Sci. 46 (1), 194–199. http://dx.doi.org/10.1016/j.ijrmms.2008.04.007.
Lu, B., Ding, X., L., Wu, A., Q., 2007. Study on method of orientation data partitioning of randomly distributed discontinuities of rocks. Yanshilixue Yu Gongcheng Xuebao/Chin. J. Rock Mech. Eng. 26 (9), 1809–1816, (in Chinese).
Maerz, N.H., Youssef, A.M., Otoo, J.N., Kassebaum, T.J., Duan, Y., 2013. A simple method for measuring discontinuity orientations from terrestrial LIDAR data. Environ. Eng. Geosci. 19 (2), 185–194.
Natali, M., Lidal, E.M., Parulek, J., Viola, I., Patel, D., 2013. Modeling terrains and subsurface geology. In: Eurographics (STARs), pp. 155–173.
Olariu, M.I., Ferguson, J.F., Aiken, C.L.V., Xu, X., 2008. Outcrop fracture characterization using terrestrial laser scanners: deep-water jackfork sandstone at Big Rock Quarry, Arkansas. Geosphere 4 (1), 247–259. http://dx.doi.org/10.1130/GES00139.1.
Oppikofer, T., Jaboyedoff, M., Blikra, L., Derron, M.-H., Metzger, R., 2009. Characterization and monitoring of the Åknes rockslide using terrestrial laser scanning. Nat. Haz. Earth Syst. Sci. 9 (3), 1003–1019. http://dx.doi.org/10.5194/nhess-9-1003-2009.
Riquelme, A.J., Abellán, A., Tomás, R., 2015. Discontinuity spacing analysis in rock masses using 3D point clouds. Eng. Geol. 195, 185–195.
Riquelme, A.J., Tomás, R., Abellán, A., 2016. Characterization of rock slopes through slope mass rating using 3D point clouds. Int. J. Rock Mech. Min. Sci. 84, 165–176.
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. http://dx.doi.org/10.1016/j.cageo.2014.03.014.
Slob, S., 2010. Automated Rock Mass Characterisation using 3-D Terrestrial Laser Scanning (Ph.D. Dissertation, TU Delft). Delft University of Technology, Delft, Netherlands, 119.
Slob, S., Van Knapen, B., Hack, R., Turner, K., Kemeny, J., 2005. Method for automated discontinuity analysis of rock slopes with three-dimensional laser scanning. Transp. Res. Rec. 1913, 187–194.
Slob, S., Hack, H.R.G.K., Feng, Q., Röshoff, K., Turner, A.K., 2007. Fracture mapping using 3D laser scanning techniques. In: Ribeiro e Sousa, L., Olalla, C., Grossmann, N. F. (Eds.), Proceedings of the 11th Congress of the International Society for Rock Mechanics. Vol. 3, Taylor & Francis, London, pp. 299–302.
Song, J.L., Liu, J.P., Zhang, J.X., 2012. Study on dynamic clustering method based on fuzzy clustering of structural interface of rock mass. Yangtze River 9, 17, (in Chinese).
Song, T.J., Chen, J.P., Zhang, W., Song, S.Y., 2015. Clustering analysis of dominative attitudes of rock mass structural plane based on firefly algorithm. J. Northeast Univ. Nat. Sci. 2, 028, (in Chinese).
Sturzenegger, M., Stead, D., 2009. Quantifying discontinuity orientation and persistence on high mountain rock slopes and large landslides using terrestrial remote sensing techniques. Nat. Haz. Earth Syst. Sci. 9 (2), 267–287. http://dx.doi.org/10.5194/nhess-9-267-2009.
Tang, M.H., Yang, Y., 2009. Research and development of fuzzy clustering validity. computer. Eng. Sci. 9, 41, (in Chinese).
Van Knapen, B., Slob, S., 2009. Identification and characterisation of rock mass discontinuity sets using 3D laser scanning. Engineering Geology for Tomorrow's Cities 22. Engineering Geology Special Publications, Geological Society, London.
Vöge, M., Lato, M.J., Diederichs, M.S., 2013. Automated rock mass discontinuity mapping from 3-dimensional surface data. Eng. Geol. 164, 155–162. http://dx.doi.org/10.1016/j.enggeo.2013.07.008.
Wong, J.S., 2012. Existence of positive solutions for second order multi–point boundary value problems. Differ. Eqn. Appl. 4 (2), 197–219.
Xie, X.L., Beni, G., 1991. A validity measure for fuzzy clustering. IEEE Trans. Pattern Anal. Mach. Intell. 13 (8), 841–847. http://dx.doi.org/10.1109/34.85677.
Yang, W.Z., Zhao, P., 2015. Research on extraction method of rock structural plane based on three-dimensional laser point cloud data. Site Investig. Sci. Technol. 3, 22–25, (in Chinese).
Yang, X., 2008. Firefly algorithm. In: Yang, X. (Ed.), Nature-Inspired Metaheuristic Algorithms. Luniver Press, Frome, UK, 79–90.