Application of data mining technique to complement photogrammetric roughness data

Dong-Hyun Kim[1], Chul-Ho Lee[2], Arumugam Balasubramaniam[3] and Ivan Gratchev[4]
[1, 2] Institute of research and development, Hulogue Inc., Seongnam-si, Gyeonggi-do, Republic of Korea
[3, 4]School of Engineering, Griffith University, Gold Coast, Australia
E-mail: d.kim@hulogue.com

ABSTRACT: This paper presents the applicability of data mining techniques to improve the accuracy of photogrammetric roughness data. Close range photogrammetry (CRP) has employed to create competent 3D slope models and can be used to highlight geological features on rock slopes. However, when a high degree of accuracy is required, this technology has also shown inconsistent values with respect to its accuracy because of various influencing factors induced by camera structures and survey environment. Thus, the uncertainty of this technology demands complex post image processing frequently. Data mining (DM) techniques are increasingly being used to identify valid and potentially useful correlations and patterns in existing data. In this study, a data mining technique is applied to reduce the noise of 3D images and to analyse multiple images effectively. Conclusions showed that a better performance was achieved using DM technique and this data mining approach was able to improve the accuracy of photogrammetric roughness data.

Keywords: Data mining, close range photogrammetry, image mining.

1. INTRODUCTION

Data mining (DM), which is also known as knowledge discovery from data (KDD) is the process of discovering interesting patterns and knowledge from large amount of data. As a fast-growing field, DM has been applied effectively to solve multiple problems connected to risk management, including error detection and quantification, detecting fraud etc. in our data-rich world. Higher level of this technique can be to build learning models to automatically extract knowledge from big and complex data. This technique has been progressively introduced to various industry sectors like marketing, manufacturing and construction.

In geotechnical engineering area, since early 1990s' Artificial Neural Network (ANN) has been increasingly employed to analyse various engineering issues such as constitutive modelling, bearing capacity of pile and slope stability. This technique was regarded as the most appropriate and effective tool among other data mining techniques as it is known as the most relevant to engineering disciplines. Recently, Tinoco et al., (2010) approached to predict compressive strength of jet grouting columns using a data mining technique. A data mining approach for the volcanic rock classification has been reported by Miranda et al. (2018).

Close distance photogrammetry has been increasingly adopted to create 3D models to investigate detailed aspects of rock slopes, as well as large-scale surveys in rock engineering and mining. Recent high-end equipment in the field of digital photographing can create high resolution images and this technology has encouraged the spread of close range photogrammetry to obtain surface features such as undulation and roughness which require higher degree of accuracy of 3D models. However, there are several known and unknown factors that can affect the accuracy of the results and the accuracy of photogrammetric models has insufficiently discussed. The authors have been studied on the accuracy of photogrammetric roughness data for many years (Kim et al., 2015; 2016). However, the relevant data have still been insufficiently explored and analysed by advanced data analysis techniques.

As the pixels of digital photographs embody large number of features, images represent a significant assembly of measurements. As a part of data mining, image mining as a specific term has been used for analysing various images of DNA microarrays, astronomical observations, satellite maps, medical images etc (Van der Walt et al., 2014). With regard to image data, data mining is associated with four broad problems: finding association, classification, sequential patterns and time series patterns (Ordonez and Omiecinski, 1998). The applications of image mining techniques have been reported in the literature (Soh, 1999; Uher et al., 2003; Thompson et al., 2005). For example, 3D images are used in a medical image analysis area, sets of 3D images of human brains obtained from MRI scanning have been analysed using a fully automatic method using a data mining program (Uher et al., 2013).

This study presents applicability of data mining techniques to predict the accuracy of photogrammetric roughness data. Data sets and attributes of joint roughness coefficient (JRC) and asperity heights obtained from a set of photogrammetric laboratory tests of a previous study are employed as input data. Also, the results of image mining approaches using 3D images corresponding to the data sets are discussed.

2. OVERVIEW OF DATA MINING TECHNIQUES

Data mining applies statistical and logical methods to large data sets and allows people to locate and interpret data patterns, helping them better informed decisions. The analysis process, called Knowledge Discovery in Database (KDD), defines the procedure for transforming raw data into useful knowledge. The method can be used to categorize the data, or this can be used to create predictive models. As shown in Fig. 1, a general KDD analysis process consists of the following multiple steps: [1] Data selection, [2] Data cleaning, [3] Data transformation, [4] Pattern extraction and interpretation.

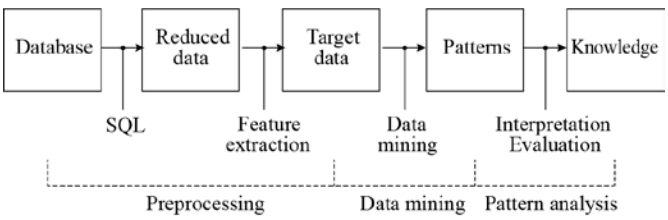

Figure 1 Process of data mining

In general process of data mining, firstly, we need to identify the objects of interests in the data and secondly, the quality of the data can be improved by techniques such as de-noising and cleaning. Third, the multivariate data can be converted into suitable forms for pattern recognition by pre-processing. Then, the patterns in the data can be identified using techniques such as classification, clustering and association rules. The overall data mining is interactive process. Data analysts involved in every step starting with an initial formation of data sets in the data collection process to validating the results obtained at each step. The process of this study is similar to the basic procedure of data mining shown in Figure 1.

3. METHODOLOGY

This study analyses rock surface roughness data obtained from a set of photogrammetry laboratory tests. The image data used in this study is parts of image sections extracted from the 3d surface models. JRC values and the corresponding measurements such as camera-to-object distances, pixel sizes and focal length of lenses are also employed in this study. To analyse the data in each process, this study uses a data mining program, RapidMiner studio (ver. 8.1) which is an open-source and flexible platform implemented in Java (RapidMiner Inc.). Also, in the pre-processing stage, the 3D images are filtered by smoothing to reduce the influence of the noise of raw images.

A digital image can be regarded as a group of discrete pixels, each of which has various ranges of colour and brightness information. Colour images with various geological features can be more complex than greyscale images. In this study, pixels converted from colour to 8-bit grey-scale are used as the input data. Thus, the processed values are in the range from 0 to 255. Common approach in data mining is to mine information from structured data with tabular form. However, image is an example of the unstructured data form. Thus, a feature extraction process which transforms an image into the tabular data is needed. In this study, Image mining extension (IMMI) operator is employed in RapidMiner for the feature extraction process.

3.1 Data cleaning

There are several operators which can enhance the quality of images in the IMMI operator. As a representative tool, median filtering is one of effective solutions to reduce the influences of the noise from the obtained photogrammetric roughness data. Poropat (2008) showed a clear effect of median filters from some simulations using roughness profiles (Fig. 2). To enhance the accuracy of the photogrammetric roughness data, the raw 3D images are processed with a median filter in this study.

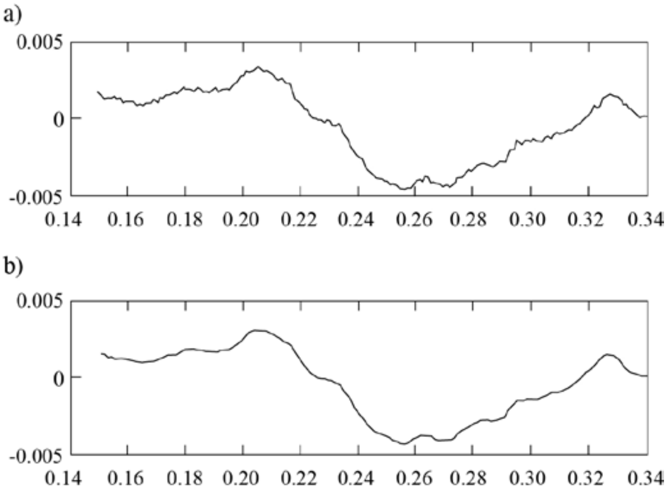

Figure 2 Details of the surface profile, corrupted with noise a), filtered with a 5 pixel median filter b) (Poropat, 2008)

3.2 Data training and extraction

In image mining, it is possible to connect each pixel intensity of an image to each input of process. In the case of 3D images, the pixels of the image can be given some meaning with their corresponding coordinates. This study attempts to correlate between the data of pixels and the photogrammetric roughness measurements on the corresponding locations. The first stage of this study is the learning stage. Pixel information of points along measured profiles on the different resolution images are compared by a learning procedure of the image mining and the results.

In the program `Rapidminer´, there are many operators for image transforms and manipulation. As a feature extraction algorithm, this study uses "local-level" feature extraction. This is suitable for extracting information such as pixel brightness integer values after image enhancement from each point in the image. Then, the image data can be transformed from image into a structured table form. The first operator used as a median filtering with a size of 3 pixels. In this stage, multiple image operator was also employed. This operator iterates the next analyses over all image files in a folder. This provided convenience for the reiterative procedure for dozens of images in this study.

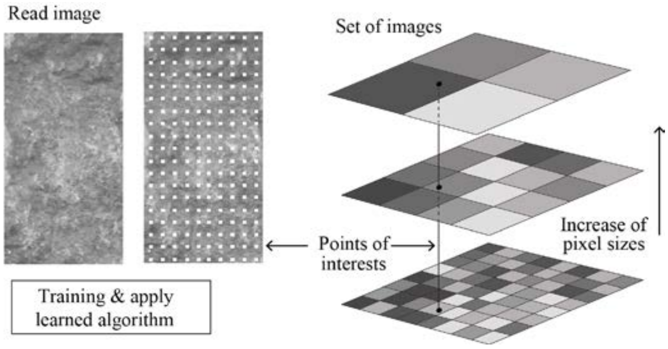

Figure 3 Concept of image mining using image processing operators with multiple images

Then, a learning process can be applied to reduce noise, recognize and identify a pattern of pixel information for multiple images. There are several learning algorithms in IMMI's operators such as support vector machines (SVM), decision tree and random forest. In this study, SVM with a Kernel method was selected for training stage based on the results of trials with other algorithms. It has been reported by Burget et al. (2012) that SVM with Kernel method was also successfully used for analysing a biomedical image. The designed process of Rapidminer in this study is shown in Figure 4. The results obtained from the learning process can be presented in tabular forms for further analysis steps. These structured data were used for statistical analyses using the differences of integer numbers between neighbouring points of interest. This data collection method was applied to the pixels at the given points of interest at the same locations on several sets of different resolution images.

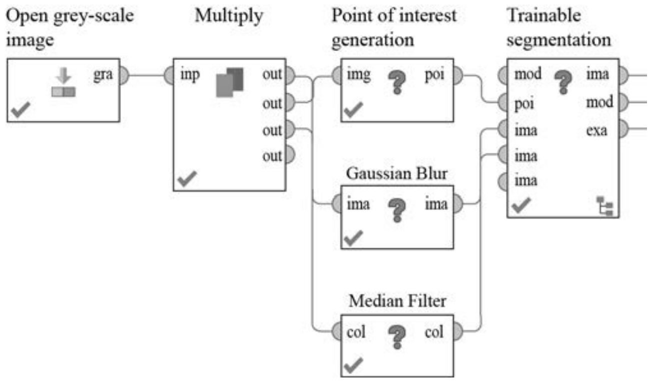

Figure 4 Designed process of image mining using IMMI

To measure the discrepancy of the bright integers, the root-mean-squared-error (RMSE) is employed in this study (See Eq. (1)). RMSE is defined to be the square root of the average of the squared discrepancies. For photogrammetry applications, RMSE has been widely used to identify the accuracy of data due to its high correlation between the predicted values and the observed values (ASPRS, 2014).

$$RMSE_{Int} = \mathrm{SQRT}\ \{\Sigma^{n}_{i=1}(Int_{o,i} - Int_{p,i})^2 / N\} \quad (1)$$

where Int_o is the brightness integers obtained from the highest resolution image; and Int_p is the brightness integers obtained from compared images.

4. CASE STUDY

4.1 Photogrammetry laboratory data

The data used in this paper were collected from a previous laboratory work on accuracy of photogrammetry roughness data. The data were images of a rock sample with different resolutions as shown in Table 1. The photographs were taken in the range of 1.0 ~ 7.0 m, using three different focal length of lenses: F=24mm, 50mm and 85mm. The length and width of the targeted area on the rock surface were 20 cm and 8 cm, respectively. Details of the tests are presented in the previous literature by the authors (Kim et al., 2015). The spatial data of the 3D surface models were created using a photogrammetry code, Sirovision (CAE). JRC values were estimated along 8 measurement profiles on the rock sample.

This study selected high resolution images with a range of pixel sizes less than 1.0 mm as summarized in Table 1. The selected area of the sandstone surface was appeared to be reasonable in this study because in observations it was found that the range of brightness on the surface was quite simple as an indicator of the change of roughness based on its simple rock-forming. Figure 5 compares the greyscale images of the sample area with different resolutions. Using the targets on the images, 8 profiles were selected at the same locations. The data of pixels with 1 mm interval along the profiles were extracted by using an operator of IMMI. A total of 28,000 integer values were collected from 14 images and employed for the learning process.

Table 1 Summary of image data used in this study

FL (mm)	c-to-o distance (m)	Pixel size (mm)	No. images	No. data
24	1 ~ 3	0.5 ~ 1.0	3	6,000
50	1 ~ 5	0.1 ~ 1.0	5	10,000
85	1 ~ 7	0.1 ~ 1.0	6	12,000

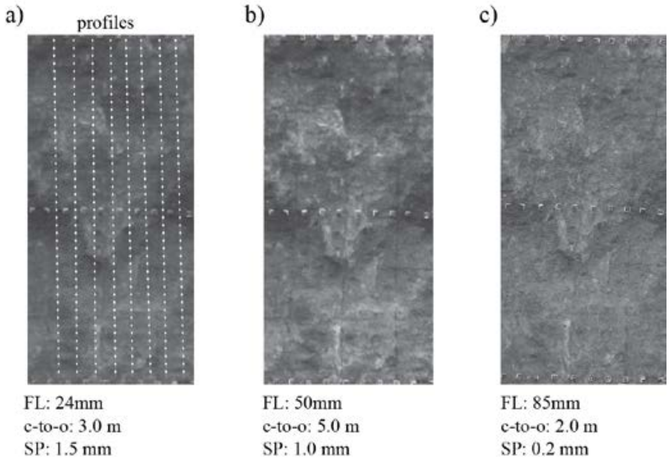

Figure 5 Comparison of images of the rock sample of pixel size = 1.5 mm a), 1.0 mm b) and 0.2 mm c)

4.2 Integer values of pixels

The designed learning process produced integer values of the smoothed images by a median filter in a tabular form. This image mining process enabled us to collect tens of thousands of data in a short period of time from hundreds of the selected points on multiple images. As presented in Figure 6, for the continuous integer of a profile on an image, absolute values of the difference between two consecutive integer values along the profile is regarded as a roughness parameter in this study. Sets of the differences of integer values between two consecutive pixels along the profiles of interest obtained from different resolution images obviously indicate downward trends as the camera moves farther from the rock sample, corresponding to the decrease of image resolution. This is directly dependent on the sizes of pixels of the created images.

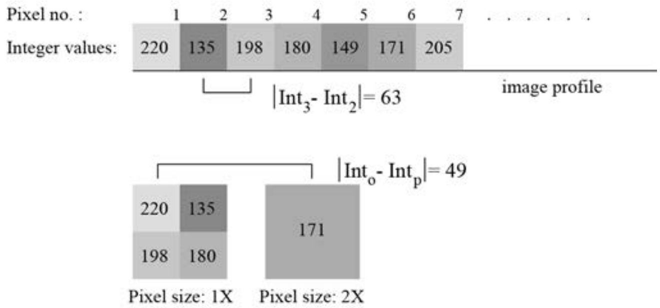

Figure 6 Schema of image profile

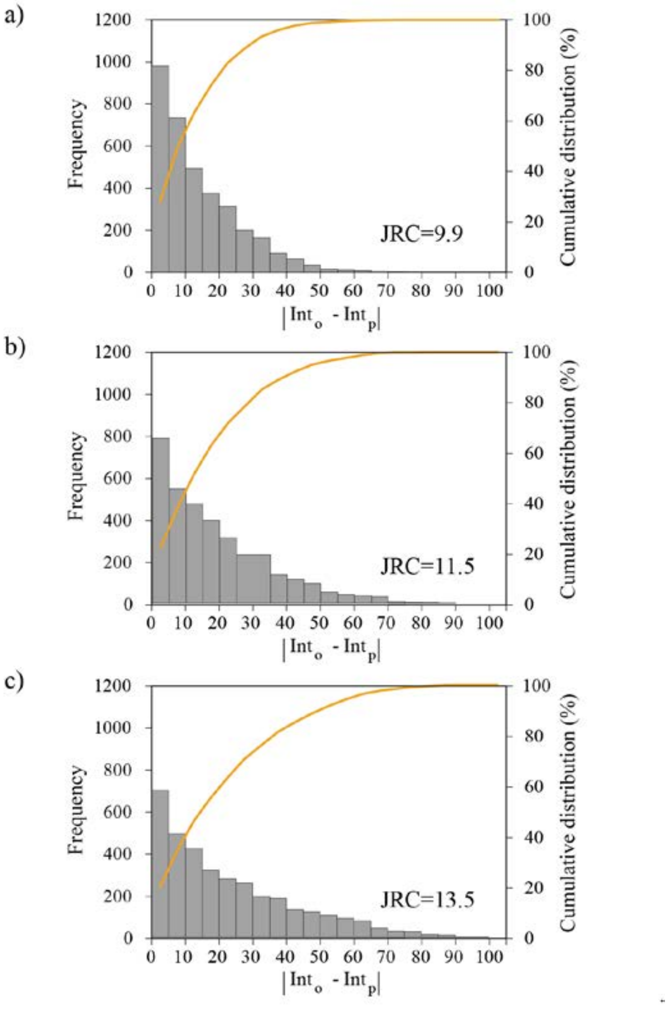

Figure 7 Data distributions of integer values

The histograms of all training showed an interesting trend. Using all resolution images, Figure 7 indicates the differences of integer values obtained from comparison between highest resolution (pixel size = 0.1 mm, FL=85mm) and target images. The data formed an exponential distribution for each profile. Mean values of the integer differences was ranged from 13.7 to 15.7 for all image profiles. Also, a high range of standard deviation which ranges from 11.6 to 21.2 was calculated from the set of data values. The ranges of statistics were quite similar to each other in the different resolution images.

In Figure 7, the histograms and the cumulative distribution curves distributed by their corresponding JRC values also indicate interesting trends to evaluate the accuracy of photogrammetric JRC values. Overall, bigger JRC values formed wider dispersion of data. The range of integer values for smaller JRC values is composed of smaller $|Int_o\text{-}Int_p|$ values. The threshold values which composes 50% of data are 10 (JRC=9.9), and 14 (JRC=11.5) and 17 (JRC=13.5). This may be attributed to the fact that the standard deviations increase with the JRC values. The inspection of cumulative distribution plots indicates that the integer values obtained from different resolution images can be used to verify photogrammetric JRC values.

4.3 Integer values and JRCs

Using the $RMSE_{INT}$ values obtained as described in Section 3.2 and using the equation (1), statistics analysed a relationship between $RMSE_{INT}$ and JRC values as presented in Figure 8. The ranges of $RMSE_{INT}$ obviously increases as the corresponding JRC values are raised. This is an interesting finding because the integer values are obtained from 2D images which exclude any information of coordinates of the rock surface.

In the authors' previous study, the data distributions of normalized JRC values comparing photogrammetric JRC values and manual measurement were fairly scattered showing the ranges of the coefficient of determination, R^2 from 32% to 48%. The uncertainty of this technology to estimate JRC values need assistants to improve its accuracy to quantify roughness data of rock surface. The integer values obtained from different resolution greyscale images can complement the shortcoming of photogrammetry.

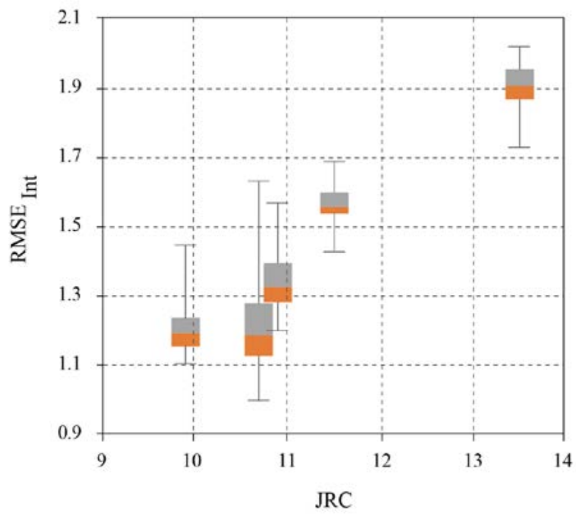

Figure 8 Correlation between $RMSE_{Int}$ and JRC values

5. CONCLUSION

This paper presented an applicability of image mining technique to analyse multiple images of rock surface for rock surface investigation. A learning process has been designed to perform the image mining process using a data mining program, Rapidminer. Using an image data set collected from a previous photogrammetry laboratory tests, following conclusions can be drawn:

- Image mining techniques can be effectively used for analysing multiple images including a large volume of image data. In this study, IMMI operators of Rapidminer was employed.
- Differences of integer values of pixels are used to identify rock joint roughness coefficient (JRC) in two different types in this study; [1] $|Int_{i+1}\text{-}Int_i|$ and [2] $|Int_o\text{-}Int_p|$. The cumulative distribution of $|Int_o\text{-}Int_p|$ indicated a relationship between JRC values and the integer values.
- Using $RMSE_{INT}$, a range of JRC values can be estimated and this method can complement when photogrammetry is employed to quantify JRC values.

6. REFERENCES

American Society for Photogrammetry & Remote Sensing (2014) "ASPRS positional accuracy standards for digital geospatial data". PE & RS, 81, 3, ppA1-A26.

Burget, R., Vaclav, U. and Masek, J. (2012) "Trainable segmentation based on local-level and segment-level feature extraction". IEEE International symposium on biomedical imaging: from nano to macro, pp1-8.

Han, J., Kamber, M., Pei, J. (2012) "Data mining concepts and techniques". 3rd edition, Elsevier Inc., p.703.

Hofmann, M. and Klinkenberg, R. (2014) "RapidMiner dataming use cases and business analytics applications". Taylor & Francis Group, LLC, p.518.

Kim, D-H., Poropat, G., Gratchev, I., Balasubramaniam, A. S. (2015) "Improvement of photogrammetric JRC data distributions based on parabolic error models". International Journal of Rock Mechanics & Mining Sciences, 80, pp19-30.

Kim, D-H., Poropat, G., Gratchev, I., Balasubramaniam, A. S. (2016) "Assessment of the accuracy of close distance photogrammetric JRC data". Rock Mechanics and Rock Engineering, 49, pp4285-4301.

Miranda, T., Sousa, L. R., Gomes, A. T., Tinoco, J. and Ferreira, C. (2018) "Geomechanical characterization of volcanic rocks using empirical systems and data mining techniques". Journal of rock mechanics and geotechnical engineering, 10, pp138-150.

Ordonez, C. and Omiecinski, E. (1998) "Image mining: a new approach for data mining". Georgia Institute of Technology technical report, GIT-CC-98-12.

Poropat, G. V. (2008) "Remote characterisation of surface roughness of rock discontinuities". In: Proceedings of the 1st Sourthern hemisphere international rock mechanics symposium, Perth, Australia, pp 447-458.

Soh, L-K. (1999) "Segmentation of satellite imagery of natural scenes using data mining". IEEE Transactions on geoscience and remote sensing, 37, 2, pp1086-1099.

Thompson, D. R., Smith, T. and Wettergreen, D. (2005) "Data mining during rover traverse: from images to geologic signatures". In: Proceedings of 8th International Symposium on Artificial Intelligence, Robotics and Automation in Space, Germany.

Uher, V., Burget, R., Masek, J. and Dutta M. K. (2003) "3D brain tissue selection and segmentation from MRI". In: Proceedings of 36th International conference on telecommunications and signal processing (TSP), Rome, Italy.