

Analysis of Rock Mass Weathering Grade Using Image Processing Technique

Nursyafeeqa Mohamad Nasir, Md Yushalify Misro*

School of Mathematical Sciences, Universiti Sains Malaysia, 11800 Gelugor, Pulau Pinang, Malaysia

Abstract Image processing techniques refer to the process of converting an image into a digital format and then performing various operations on it to extract useful information. In this study, image processing technique has been used to categorize rock masses according to its weathering grades. The pixel values of the sample images were in the form of RGB color space before being converted to CIELAB color space. The conversion uses $D65/10^\circ$ as the illuminant. The a^* value in CIELAB color space represents the green-red opponent colors, with negative values for green and positive values for red. In contrast, the b^* value represents the blue-yellow opponents with negative values for blue and positive values for yellow. From the values of a^* and b^* of the samples, $K - Means$ clustering was used to classify the samples. This method will group the a^* and b^* values into seven clusters according to the closest distance between the values and the centroids. The proposed study can differentiate between the rock mass and rejected clusters containing plants and painted numbers on the rock. The painted number is placed in a rejected cluster due to the inability to determine the exact color of the rock, thereby impacting the data accuracy. The results have been discussed, and the rock masses have been categorized based on weathering grade. Several limitations have been identified, such as the presence of shadows in the sample images and the lack of arrangement of outcome images according to their a^* and b^* values. This research has also been validated and compared with previous studies. The JudGeo software utilized in prior research required human input to manually estimate suitable a^* and b^* values, whereas the proposed method automatically computes these values during color space conversion of the sample. Additionally, the proposed method can calculate the percentage of each cluster, facilitating the classification of rock mass into its respective weathering grade.

Keywords: Image Processing Technique, Rock Mass, $K - Means$ clustering, CIELAB, Weathering Grade.

Introduction

There are several types of rock that exists on Earth which are Igneous rocks, Metamorphic rocks and Sedimentary rocks [7]. One of the earliest reference of rock mass classification for the design of tunnel supports can be seen as described in [18]. Various rock mass categorization schemes have been used over the past decade in the weathering grade area. The foundation, slope stability, embankment collapse issues, and underground excavation are now mostly concerned with the challenges of engineering behavior in weathered rocks, which are common in humid tropical locations [1].

One of the criteria used to describe weathering grade and mineral composition is the color of the rock mass. Rock masses exposed to the elements such as sunlight and water steadily change over time, gradually covering their surface in weathering layers. Rock weathering is caused by physical disintegration, chemical deterioration, and biological factors [15]. Meanwhile, a new measurement system based on Structure from Motion photogrammetry had been proposed to evaluate tunnel faces, which aims to determine the tunnel faces discontinuity patterns and orientations for the tunnel face stability evaluation [20].

In 2020, a 2D image analysis method known as JudGeo software to identify weathering and discontinuity of the rock mass has been developed for users to calculate the potential of rock falls using weathering

*For correspondence:
yushalify@usm.my

Received: 21 Dec. 2023
Accepted: 14 April 2024

©Copyright Nasir. This article is distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use and redistribution provided that the original author and source are credited.

and discontinuity assessments. Depending on respective preferences, the analysis may produce various conclusions [19].

This study focuses on a geological area, which is the weathering process using CIELAB color space to differentiate weathered and unweathered rock masses. CIELAB is a color space that has become the industry standard for quantifying color. L^* specifies the brightness of the image, which is in the range of 0 to 100, where 0 denotes black and 100 denotes white. Additionally, a^* describes how many red or green tones are present in an image. Red corresponds to an a^* number that is significantly positive whereas green is significantly negative. The range of a^* is between -100 to 100 . The same range applied to b^* which indicates yellow for positive values and blue for negative values [17].

The color of an image may not match the RGB channels, which are frequently produced by color data capturing equipment. The faults can be objectively fixed by calibrating the system using industry-standard color charts. Color and weathering can be connected in a convincing way by using an image tool that computes the color channel into a calibrated color system, such as CIELAB. Moreover, CIELAB, as opposed to RGB, attempts to mimic human vision. CIELAB is useful for identifying tiny color changes despite having slightly uniform color axes.

In some of former studies, the authors discuss the method of transforming from one color space to another and implement it into applications. Chen *et al.* [4] introduced a new method for transforming from RGB to CIELAB color space based on Markov chain Monte Carlo in 2013. Meanwhile, in 2017, Bruce Lindbloom [11] presents matrices and formulations to convert RGB to XYZ color space and vice versa. Then, in order to match the information from the colorimeter data and Schmidt hammer rebound value, Lee *et al.* [10] analyzed natural slope images. By utilizing colorimeter data for calibration, the slope with high values for a^* and b^* from CIELAB color space was thought to be weathered. Yusoff *et al.* [21] used the Japanese Highway (JH) Rock Mass Classification to classify the rock mass tunnel face. Razali *et al.* [16] conducted a study that compared the efficiency and dependability of methods for measuring rock slope weathering that use the Munsell chart and CIELAB color space. There are several types of weathering classes. Therefore, the number of classes to categorize the weathering grade is based on the individual reference.

Before classifying a rock mass to its weathering grade, there is a technique to be completed which is by using clustering technique. Clustering is the unsupervised classification of patterns into groups [8]. $K - Means$ is a centroid-based clustering algorithm where the distance between each data point and centroid is calculated and assigned to a cluster. The assigning of each data point to the cluster is an iterative process based on similar features. There are some examples of clustering application which in 2018, Ahmar *et al.* [2] performed a study of data grouping to the current Provinces in Indonesia using $K - Means$ Clustering. Then, the template-based $K - Means$ and improved fuzzy $C - Means$ technique is the model recommended by Alam *et al.* [3] for automatically detecting human brain tumors in MRI images.

Kanjanawanishkul [9] proposed a study on an image-based grading method for Eri silkworm pupae according to their shape, color and size features focussing on RGB, HSV and CIELAB color spaces. Then, research on digital image processing, which includes methods like image acquisition, enhancement, segmentation, feature extraction, and classification, is covered by Chitradevi and Srimathi [5].

The motivation of this study is to propose a new improved method for rock mass weathering grade using computer-assisted image processing technologies, focusing on reducing the human bias in the manual characterization process. In viewing the sensitivity of JudGeo software, the software required several parameter values to classify and categorize the rock mass. These arbitrary values need multiple attempts to get the best output.

The objective of this study is to identify the color space of rock masses using improved computer-assisted image processing technology. Other than that, the rock mass will be categorized according to its weathering grade and compare the result of the proposed method with an existing study. This study focuses on the samples that were taken from the area of Kampung Wai, Perlis. The samples that will be used in the study do not cover all areas, but only several parts for the validation process.

Methodology

Every sample of rock mass was in a color image which contains sets of pixel numbers that will undergo two parts of image processing techniques. The result of these techniques will be used to analyze the weathering grade of the rock mass.

Conversion from RGB Color Space to CIELAB Color Space

In this section, conversion of the color space can be done using Equation 1 until Equation 9. D_{65} is an average daylight illuminant in the CIE standard illuminant and 10° is the viewing angle [6]. In RGB color space, there are 3 sets of matrices or channels that contain every value of the pixels. The first layer of the matrices is red, r , the second layer is green, g , and the last layer is blue, b .

$$R = \left(\frac{r}{255}\right), G = \left(\frac{g}{255}\right), B = \left(\frac{b}{255}\right). \tag{1}$$

If the value of R, G, B bigger than 0.04045 or above integer value of 10 in 8-bit color channels which is the RGB companding value, Equation 2 need to be done.

$$\bar{R} = \left(\frac{R + 0.055}{1.055}\right)^{2.4}, \bar{G} = \left(\frac{G + 0.055}{1.055}\right)^{2.4}, \bar{B} = \left(\frac{B + 0.055}{1.055}\right)^{2.4} \tag{2}$$

Or else,

$$\bar{R} = \frac{R}{12.92}, \bar{G} = \frac{G}{12.92}, \bar{B} = \frac{B}{12.92}. \tag{3}$$

Then,

$$\hat{R} = \bar{R} \times 100, \hat{G} = \bar{G} \times 100, \hat{B} = \bar{B} \times 100. \tag{4}$$

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 0.4124 & 0.3576 & 0.1805 \\ 0.2126 & 0.7152 & 0.0722 \\ 0.0193 & 0.1192 & 0.9005 \end{pmatrix} \begin{pmatrix} \hat{R} \\ \hat{G} \\ \hat{B} \end{pmatrix}. \tag{5}$$

The variables X, Y and Z will be divided by $D_{65}/10^\circ$ of standard illuminant as the following,

$$x = \left(\frac{X}{94.811}\right), y = \left(\frac{Y}{100}\right), z = \left(\frac{Z}{107.304}\right). \tag{6}$$

If the value of x, y and z bigger than 0.008856 which the break point to avoid rounding error given by the CIE standard [11], the variables will undergo the following algorithm,

$$\bar{x} = x^{1/3}, \bar{y} = y^{1/3}, \bar{z} = z^{1/3}, \tag{7}$$

Or else,

$$\begin{aligned} \bar{x} &= (7.787 \times x) + \left(\frac{16}{116}\right), \bar{y} = (7.787 \times y) + \left(\frac{16}{116}\right), \bar{z} \\ &= (7.787 \times z) + \left(\frac{16}{116}\right). \end{aligned} \tag{8}$$

Then,

$$L^* = (116 \times \bar{y}) - 16, a^* = 500 \times (\bar{x} - \bar{y}), b^* = 200 \times (\bar{y} - \bar{z}). \tag{9}$$

Hence, L^*, a^* and b^* are the CIELAB value of the specific pixel values from the RGB color space.

In a proceeding by Macqueen in 1967, the author defined K -Means and its asymptotic behaviour [13].

Preliminaries. Let z_1, z_2, \dots be a random series of points in E_N , where each point is chosen using a set probability measure P , regardless of the points that came before it. Hence, for a given measurable set, $P[z_1 \in A] = p(A)$ and $P[z_{n+1} \in A \mid z_1, z_2, \dots, z_n] = p(A), n = 1, 2, \dots$, for A any measurable set in E_N . Relative to a given k -tuple $x = (x_1, x_2, \dots, x_k), x_i \in E_N, i = 1, 2, \dots, k$, define a minimum distance partition $S(x) = \{S_1(x), S_2(x), \dots, S_k(x)\}$ of E_N , by

$$\begin{aligned} S_1(x) &= T_1(x), S_2(x) = T_2(x)S'_1(x), \dots \\ S_k(x) &= T_k(x)S'_1(x)S'_2(x) \dots S'_{k-1}(x), \end{aligned} \tag{10}$$

where

$$T_i(x) = \{\xi: \xi \in E_N, |\xi - x_i| \leq |\xi - x_j|, j = 1, 2, \dots, k\}. \tag{11}$$

The points in E_N closest to x_i are in the set $S_i(x)$, with tied points being arbitrarily allocated to the set of lower indexes. Keep in mind that while using this tied-point convention, if $x_i = x_j$ and $i < j$ then $S_j(x) = \emptyset$. Sample K -Means $x^n = (x_1^n, x_2^n, \dots, x_k^n), x_i^n \in E_N, i = 1, \dots, k$, with associated integer weights $(w_1^n, w_2^n, \dots, w_k^n)$, are specified as follows: $x_i^1 = z_i, w_i^1 = 1, i = 1, 2, \dots, k$, and for $n = 1, 2, \dots$, if $z_{k+n} \in S_i^n, x_i^{n+1} = (x_i^n w_i^n + z_{k+n}) / (w_i^n + 1), w_i^{n+1} = w_i^n + 1$, and $x_j^{n+1} = x_j^n, w_j^{n+1} = w_j^n$ for $j \neq i$, where $S^n = \{S_1^n, S_2^n, \dots, S_k^n\}$ is the minimum distance partition relative to x^n .

Informally, the *K – Means* method is as follows: it starts with a set of *k* groups, each of which has a single random point. Then, it adds each new point to the group whose mean it is closest to. The mean of a group is modified to consideration for the addition of a point after it has been added. Because of this, the *K – Means* are actually the means of the groups its represent at each stage. The convenient assumptions was made which first, *p* is absolutely continuous with respect to Lebesgue measure on E_N , and second, $p(R) = 1$ for a closed and bounded convex set $R \subset E_N$, and $p(A) > 0$ for every open set $A \subset R$ in the analysis of the asymptotic behaviour of the *K – Means*. For the particular *k*-tuple $x = (x_1, x_2, \dots, x_k)$, this entity will be called a *k*-point. Let

$$W(x) = \sum_{i=1}^k \int_{S_i} |z - x_i|^2 dp(z), \tag{12}$$

$$V(x) = \sum_{i=1}^k \int_{S_i} |z - u_i(x)|^2 dp(z),$$

where $S = \{S_1, S_2, \dots, S_k\}$ is the partition with the least distance in relation to x , and $u_i(x) = \int_{S_i} z dp(z) / p(S_i)$ or $u_i(x) = x_i$, according to whether $p(S_i) > 0$ or $p(S_i) = 0$. If $x_i = u_i(x)$, $i = 1, 2, \dots, k$, where the *k*-point x is unbiased.

Theorem 1. The sequence of random variables $W(x^1), W(x^2), \dots$ converges a.s. and $W_\infty = \lim_{n \rightarrow \infty} W(x^n)$ is a.s. equal to $V(x)$ for a certain x in the class of *k*-points $x = (x_1, x_2, \dots, x_k)$ which are unbiased, and possess the attribute that $x_i \neq x_j$ if $i \neq j$.

The following theorem was derived in place of a good strong law of large numbers for *K – Means*.

Theorem 2. Let $u_i^n = u_i(x^n)$ and $p_i^n = p(S_i(x^n))$; then

$$\sum_{n=1}^m \left(\sum_{i=1}^k p_i^n |x_i^n - u_i^n| \right) / m \xrightarrow[\text{a.s.}]{} 0 \text{ as } m \rightarrow \infty. \tag{13}$$

Image Categorization using *K – Means* Clustering

The aim of this technique is to minimize the sum of distances between the data point and the centroid of the cluster. The explanation of the technique will be explained in simplest way in this section.

First step is to choose the number of clusters, *K*, which in this study using *K* = 7. This number will produce seven clusters based on the grouping of a^* and b^* values. Then, the centroid of each cluster needs to be initialized at random because the exact value is unknown. Now that the centroids of each cluster are known, the data points, x_n need to be assigned to their nearest cluster centroid, c_i . Euclidean Distance formula in Equation 14 will be used by each data point as the following,

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - c_i)^2}. \tag{14}$$

Next, the data point will be assigned to the cluster based on the minimum value. Then, the centroid need to be re-initialized by calculating the average of all data points using the Equation 15,

$$c_i = \frac{1}{|N_i|} \sum x_i. \tag{15}$$

The next step is the calculation of the distances using Equation 14 and re-assign the data points according to the nearest centroid is repeated. The iterative process continues until the data point in each cluster is fixed.

The segmentation shows the result according to the last iteration of previous steps, so, the image provided for each segment arranged by the color and not by the arrangement of the a^* and b^* values. Hence, the visualization of the weathering grade from the unweathered to weathered region will be determined using graphs and charts.

Data Visualization

From the result, the boxplot approach will be used to plot the data of the a^* and b^* values for each cluster. If there are any outliers while plotting the boxplot, the value of minimum and maximum will be used from the outliers and not from the minimum and maximum values of boxplot.

Then, the minimum and maximum values for each cluster are extracted to classify weathered and unweathered region. Radar charts are used in this section with y-axis represent $-b^*$ and b^* values while the x-axis will represent $-a^*$ and a^* values.

The range from unweathered to weathered regions was determined by referring to the work of Razali *et al.* [15]. This study was conducted in Lebuhraya Kuala Kubu Bharu–Teranum–Raub which located in Peninsular Malaysia. Since the proposed study are located in the same country which has the same tropical environment, it can be assumed the type of rock are the same hence the same classification of the rock mass can be used in the study. However, the classifications mentioned by Razali *et al.* [15] were only three grades which are unweathered, slightly weathered and highly weathered.

In the proposed study, the rock mass will be classified according to five weathering grades. Since the classes of weathering grade are individual preference, the value of the additional grades will be assumed as follows.

Table 1. Classes of Weathering Grade

Weathering Grade	a^*	b^*
Unweathered	< -3	< 12
Slightly weathered	$-2 - 9$	$13 - 15$
Moderately weathered	$10 - 18$	$16 - 30$
Highly weathered	$19 - 28$	$31 - 45$
Weathered	$29 >$	$46 >$

Since moderately weathered is in the middle of slightly and highly weathered in weathering grade, the value for moderately weathered is chosen based on the midpoint between slightly and highly weathered classes. Moreover, the weathered class is determined to have values that are the highest among all of the classes, hence, the range of the weathered class in weathering grade is higher than highly weathered. It is important to note that the value taken from Razali *et al.* [15] is the upper bound for each class except for the weathered class.

Next, the percentage of pixel number existing in each cluster is calculated and visualized in pie chart. The region with the highest percentage will represent the weathering grade of the sample.

Results and Discussion

In this section, there will be results for the proposed study and the discussion will be provided. Each sample goes through an image processing technique by using MATLAB R2021b which uses all their pixel numbers and the program will run their pixels of RGB one by one channel as discussed in previous section. The following are the results obtained from several samples provided by previous works.

Sample 1

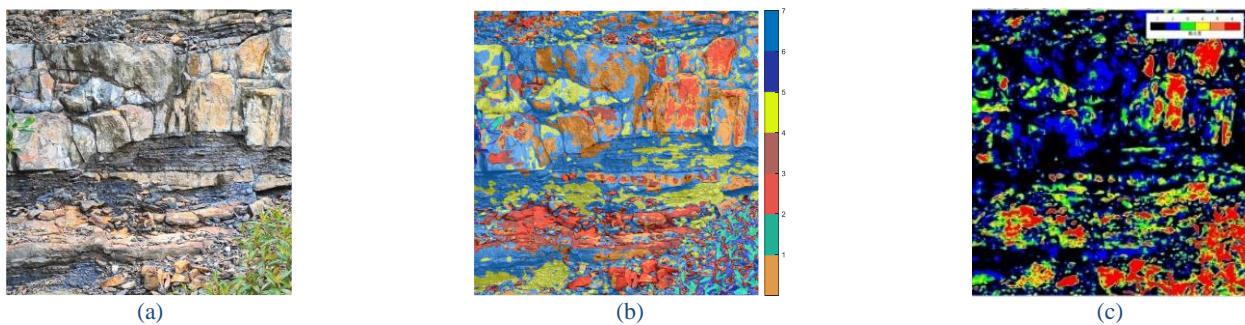


Figure 1. (a) Original Image (b) Output Image (c) Judgeo Software

Sample 1 image contain 938×938 pixels. As the aim of this study to identify the color space of rock mass by using computer-assisted processing technology, it can be observed from Figure 1(b) the output image can differentiate between plants and rock mass as compared to the image computed by using JudGeo Software. The software is not able to separate the plants from the rock mass through several attempt by applying different lower and upper bound and threshold values. The value of a^* and b^* of different objects can be differentiated based on its color.

From Figure 1 (a) and (c), the orange region can be considered weathered whereas the black surfaces are the unweathered region. So, the color bar from unweathered to weathered regions in Judgeo Software are arranged accordingly. However, the output image cannot determine which cluster is the weathered or unweathered. The image's color bar of the proposed method is arranged by the colored cluster provided from the *K - Means* clustering. Hence, the value of a^* and b^* are used to observe the arrangement from unweathered to weathered regions.

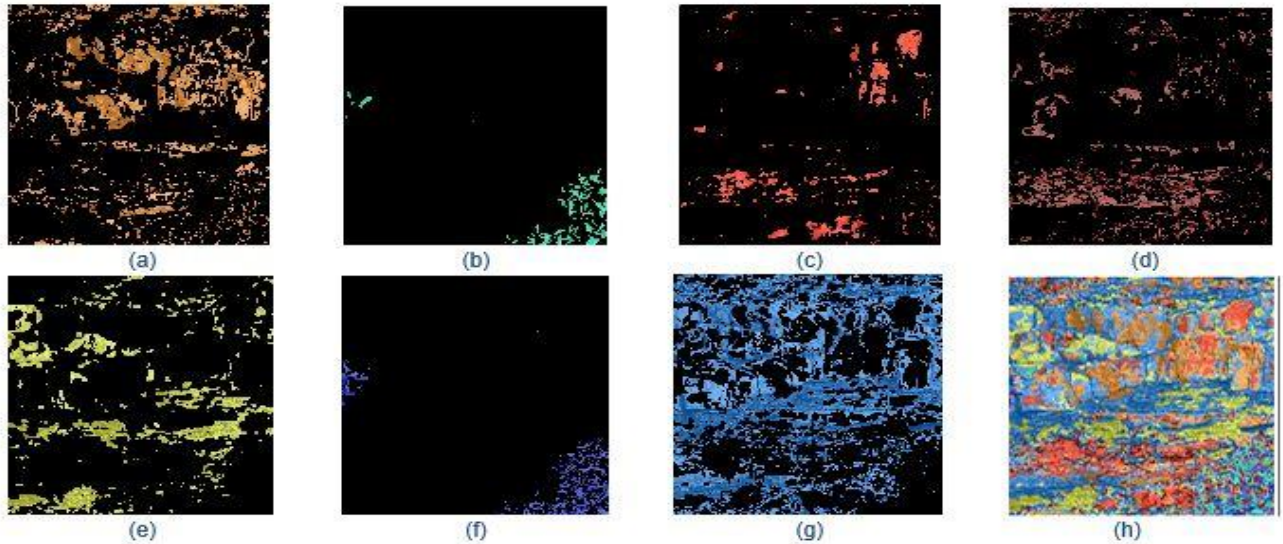


Figure 2. Clusters in Sample 1 (a) Cluster 1 (b) Cluster 2 (c) Cluster 3 (d) Cluster 4 (e) Cluster 5 (f) Cluster 6 (g) Cluster 7 (h) All Clusters

Figure 2 shows the clusters of output image separately after *K - Means* clustering. To grade Sample 1, the result of values a^* and b^* are visualized first using boxplot to see the position of each cluster. Note that Cluster 2 and Cluster 6 are the rejected clusters which in this case are the plants that are included in the image. Putting that aside, the image has five other clusters to be discussed which are Cluster 1,3,4,5 and 7.

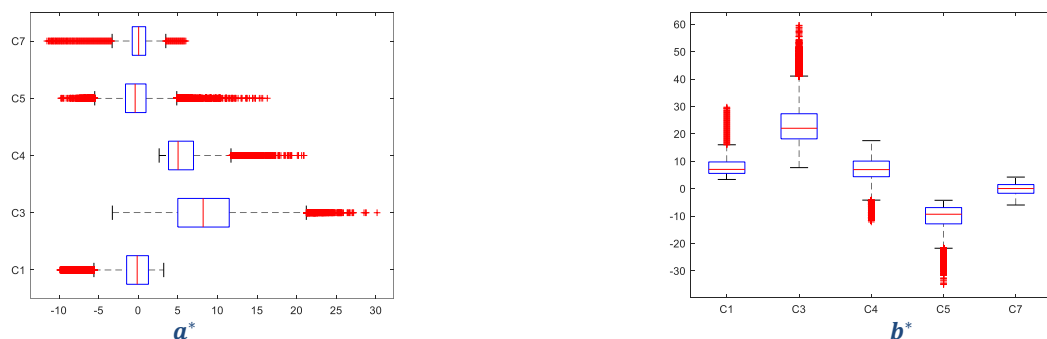


Figure 3. Boxplots for a^* and b^* values of Sample 1

From Figure 3, boxplot is used to show the range of every cluster's data such as the b^* values of Cluster 5 are the lowest among all the clusters but the lowest a^* value is in Cluster 7. However, for values a^* , there are several clusters seem to have low values. Besides, since two of the clusters are rejected clusters and five clusters are left to analyze, the accuracy of the output and previous result is low because image from Judgeo Software had six clusters of rock mass analyzed. Hence, in this study, 5 classes of weathering grades are used to classify the rock mass.

As mentioned before, the clusters are not arranged by the category of weathered to unweathered, thus, the values of minimum and maximum for each of the cluster will be used to visualized as in Figure 4 for better analysis and interpretation. Table 2 shows the extraction of the minimum and maximum value of a^* and b^* .

Table 2. Maximum and minimum values of a^* and b^* for clusters in Sample 1

Cluster	Max a^*	Min a^*	Max b^*	Min b^*
1	3.21	-9.92	29.83	3.33
3	30.15	-3.26	59.60	7.66
4	20.93	2.64	17.50	-12.14
5	16.29	-9.77	-4.25	-34.99
7	6.01	-11.52	4.21	-5.98

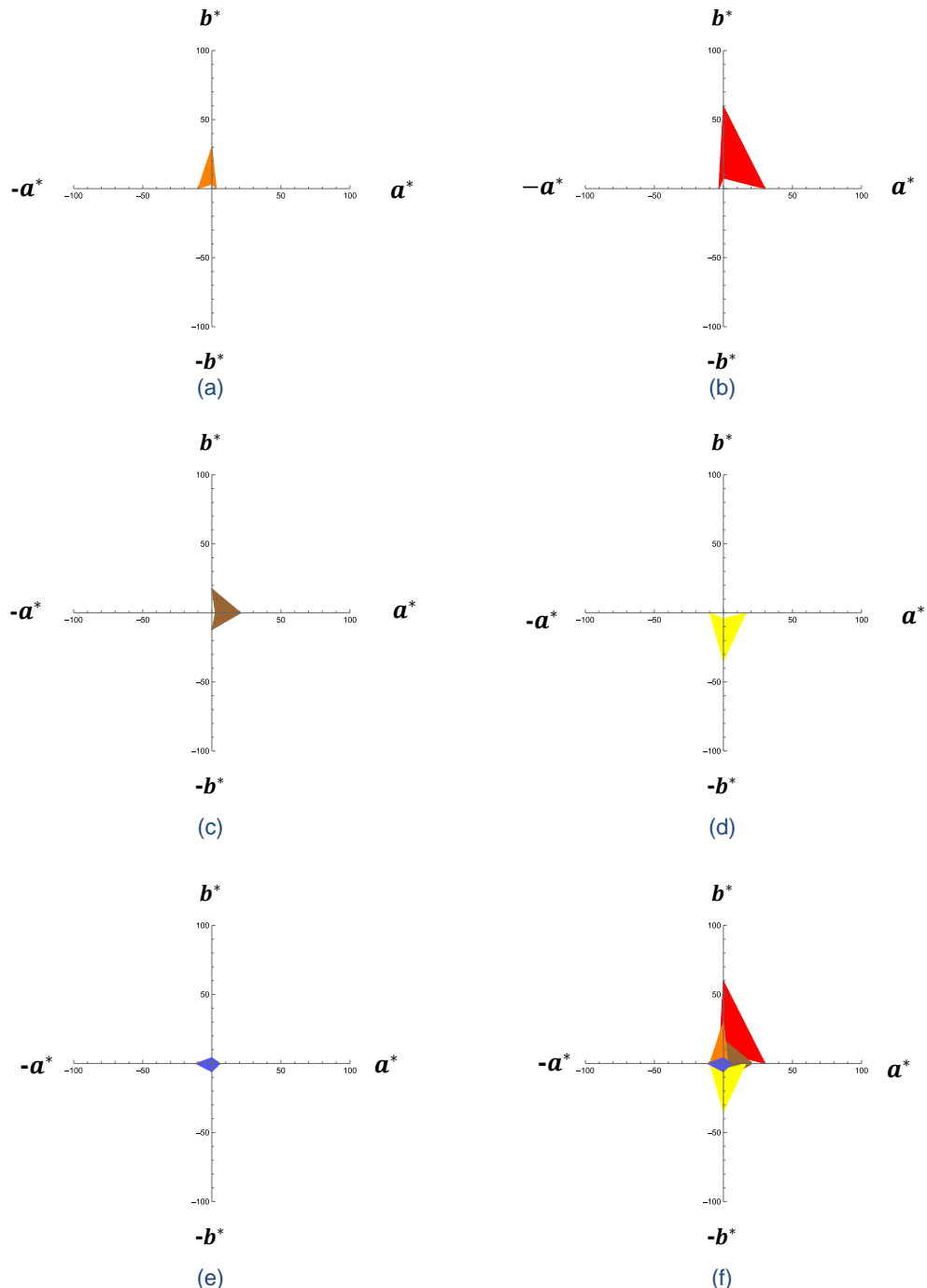


Figure 4. Radar Charts for Sample 1 (a) Cluster 1 (b) Cluster 3 (c) Cluster 4 (d) Cluster 5 (e) Cluster 7 (f) All Clusters

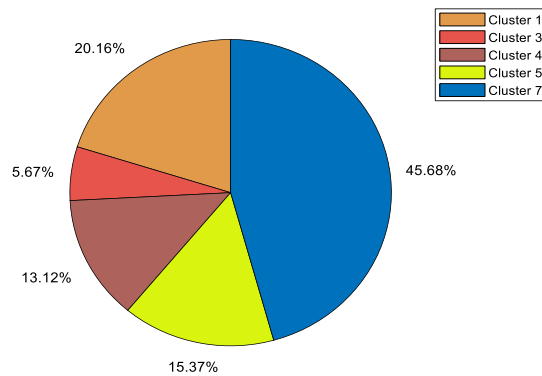


Figure 5. Pie chart for Sample 1

From the pie chart in Figure 5, Cluster 7 took the highest percentage which is 45.68%. Hence, this sample will be used to categorize the sample. Comparing the range of Cluster 7 from Table 2 and Table 1, the weathering grade for this sample is unweathered to slightly weathered.

Sample 2

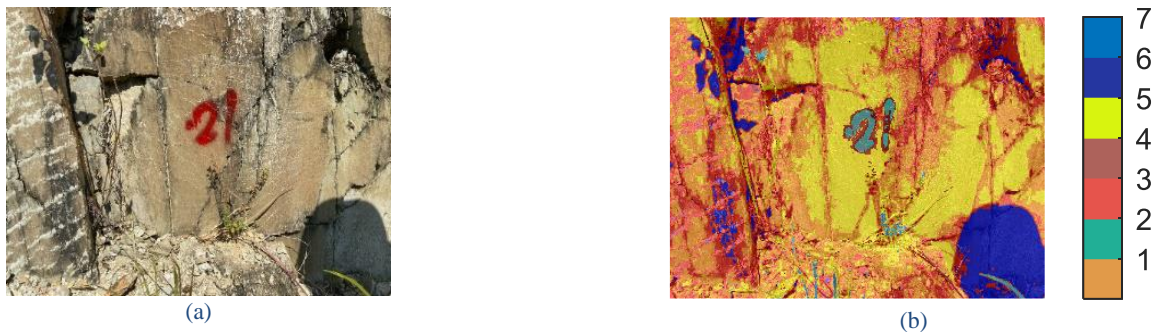


Figure 6. (a) Original Image (b) Output image.

This sample is different from previous sample which the surface of the rock mass is smaller than before. Sample 2 contains 960 × 1280 pixels. Figure 6 (b) shows the output image from the proposed method.

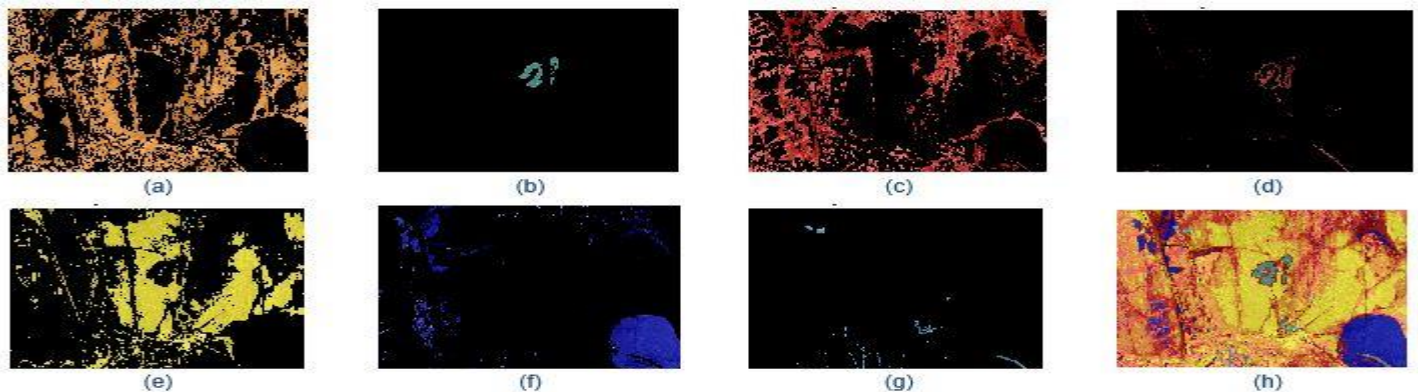


Figure 7. Clusters in Sample 2 (a) Cluster 1 (b) Cluster 2 (c) Cluster 3 (d) Cluster 4 (e) Cluster 5 (f) Cluster 6 (g) Cluster 7 (h) All Clusters

This study may be able to differentiate between plants and the rock mass. For Sample 2, three of the clusters are the rejected clusters which can be observed in Figure 7. Two of these clusters show the painted number on the rock mass which are Cluster 2 and 4, whereas Cluster 7 contains plants. So, there are four other clusters left to be discussed. However, by observing Cluster 6, there are shadows included. The surfaces with shadow should be considered as a limitation of this study. The visualization of values α^* and b^* of this sample are as follows.

From Figure 8, Cluster 6 has two outliers that are far from the interquartile range, so, it can be said that the other limitation of this study is the number of segmentation chosen.

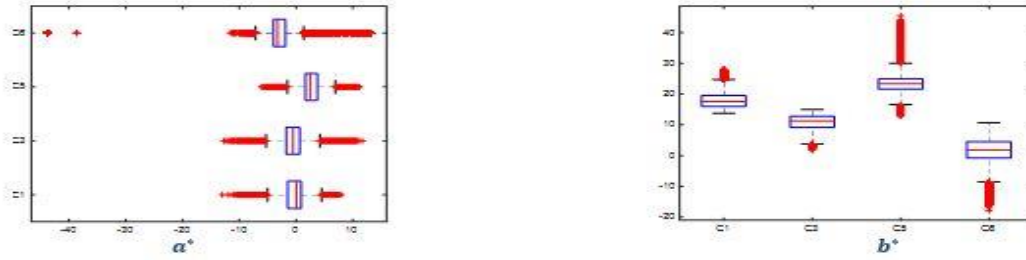


Figure 8. Boxplots for a^* and b^* values of Sample 2

Table 3. Maximum and minimum values of a^* and b^* for clusters in Sample 2

Cluster	Max a^*	Min a^*	Max b^*	Min b^*
1	7.88	-12.97	28.09	13.79
3	11.66	-12.67	14.95	1.87
5	11.16	-5.97	45.46	12.88
6	13.12	-43.77	10.69	-17.86

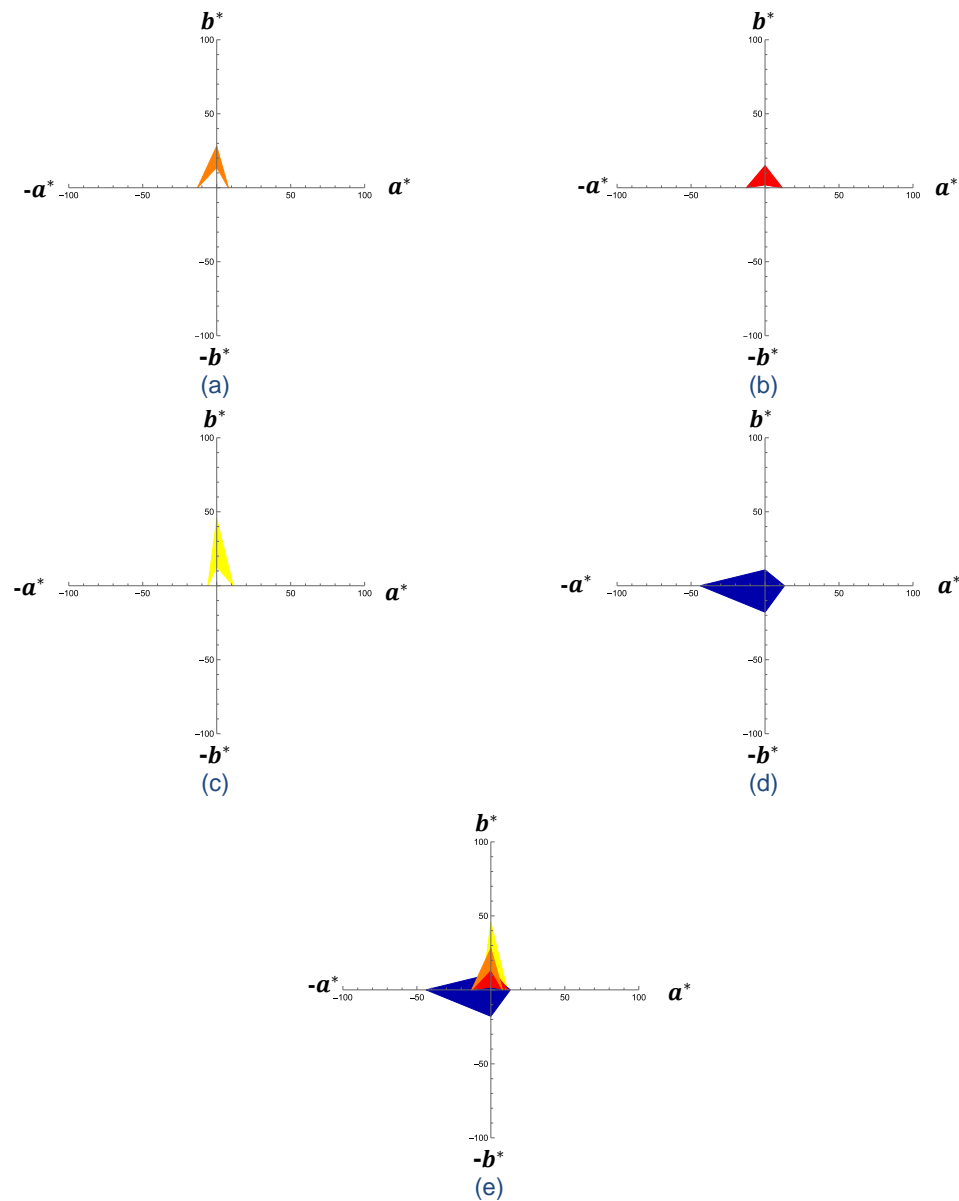


Figure 9. Radar Charts for Sample 2 (a) Cluster 1 (b) Cluster 4 (c) Cluster 5 (d) Cluster 6 (e) All Clusters

The a^* and b^* values from Table 3 are presented in radar charts in Figure 9. This sample has a value of a^* and b^* which are 9.43 and 27.37 respectively from the previous research by Lee *et al.* [10]. Moreover, the sample is described as slightly weathered. Based on the values mentioned above, it falls under Cluster 5.

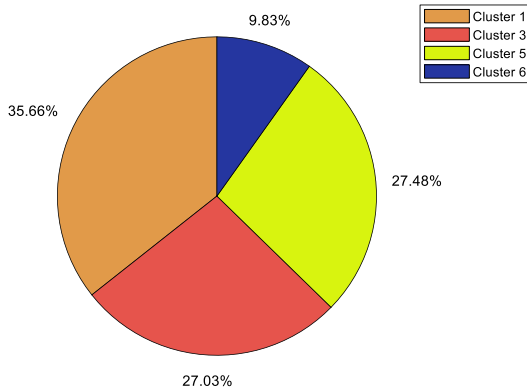


Figure 10. Pie chart for Sample 2

Based on the pie chart in Figure 10, Cluster 1 has the highest percentage which covers 35.66% of the surface. Then, followed by Cluster 5 with 27.48% and Cluster 3 with 27.03%. However, the result is slightly inaccurate because of the existence of the shadows. Comparing the value of Cluster 1 with Table 1, it can be observed that this cluster are categorized as unweathered to moderately weathered yet both a^* and b^* values exist in slightly weathered class.

Sample 3

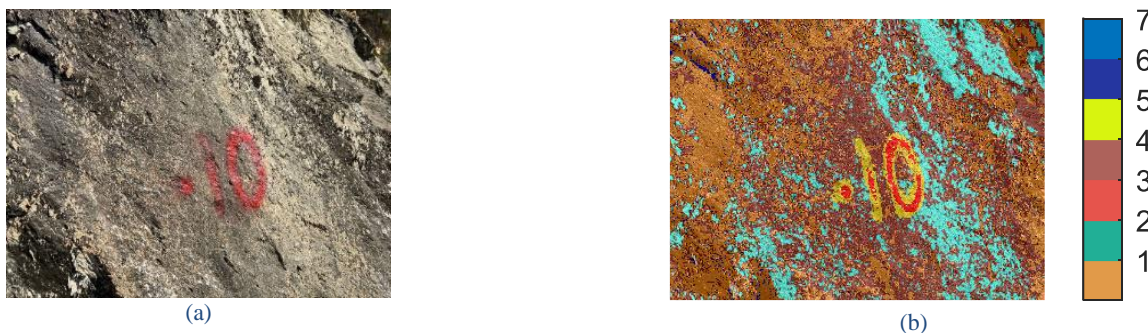


Figure 11. (a) Original Image (b) Output image

This sample contained 960×1280 pixels. Figure 11 shows the original sample and the output image. Compared to the samples before, the cluster is not fixed from one sample to another. For example, in Sample 2, the rejected cluster is Cluster 2, 4 and 7 whereas in this sample, the painted number is Cluster 3 and 5. So, note that the clusters in the following figure are not the same as samples before. Figure 12 shows the clusters in Figure 11 (b) separately.

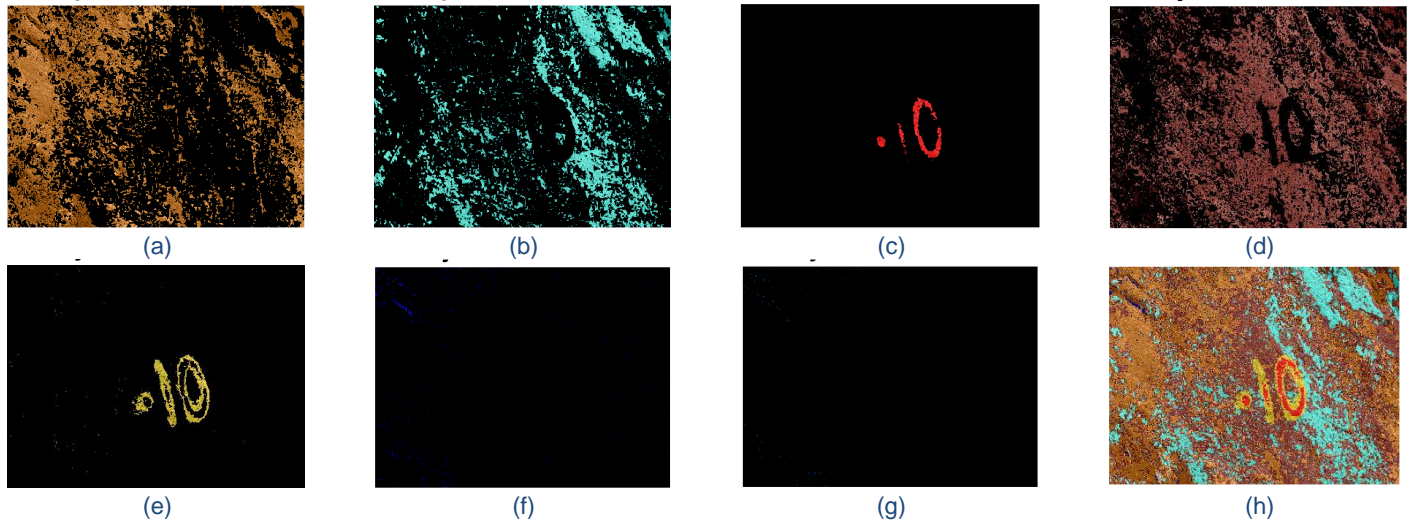


Figure 12. Clusters in Sample 3 (a) Cluster 1 (b) Cluster 2 (c) Cluster 3 (d) Cluster 4 (e) Cluster 5 (f) Cluster 6 (g) Cluster 7 (h) All Clusters

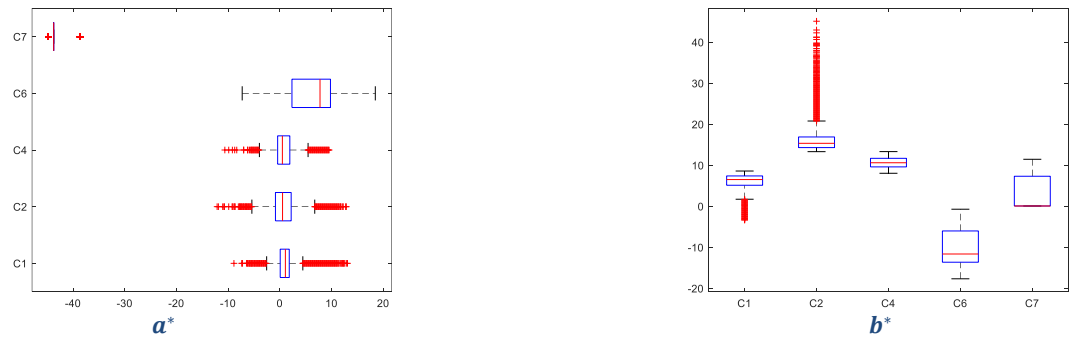


Figure 13. Boxplots for a^* and b^* values of Sample 3

Figure 13 shows that the visualization of values a^* and b^* in boxplots which a^* values of the first four clusters are near to each other, whereas the b^* values for each cluster are different from each cluster.

Table 4. Maximum and minimum values of a^* and b^* for clusters in Sample 3

Cluster	Max a^*	Min a^*	Max b^*	Min b^*
1	13.08	-8.88	8.61	-3.33
2	12.85	-12.14	45.12	13.33
4	9.55	-10.65	13.34	8.06
6	18.47	-7.29	-0.69	-17.65
7	-38.64	-44.79	11.47	0.09

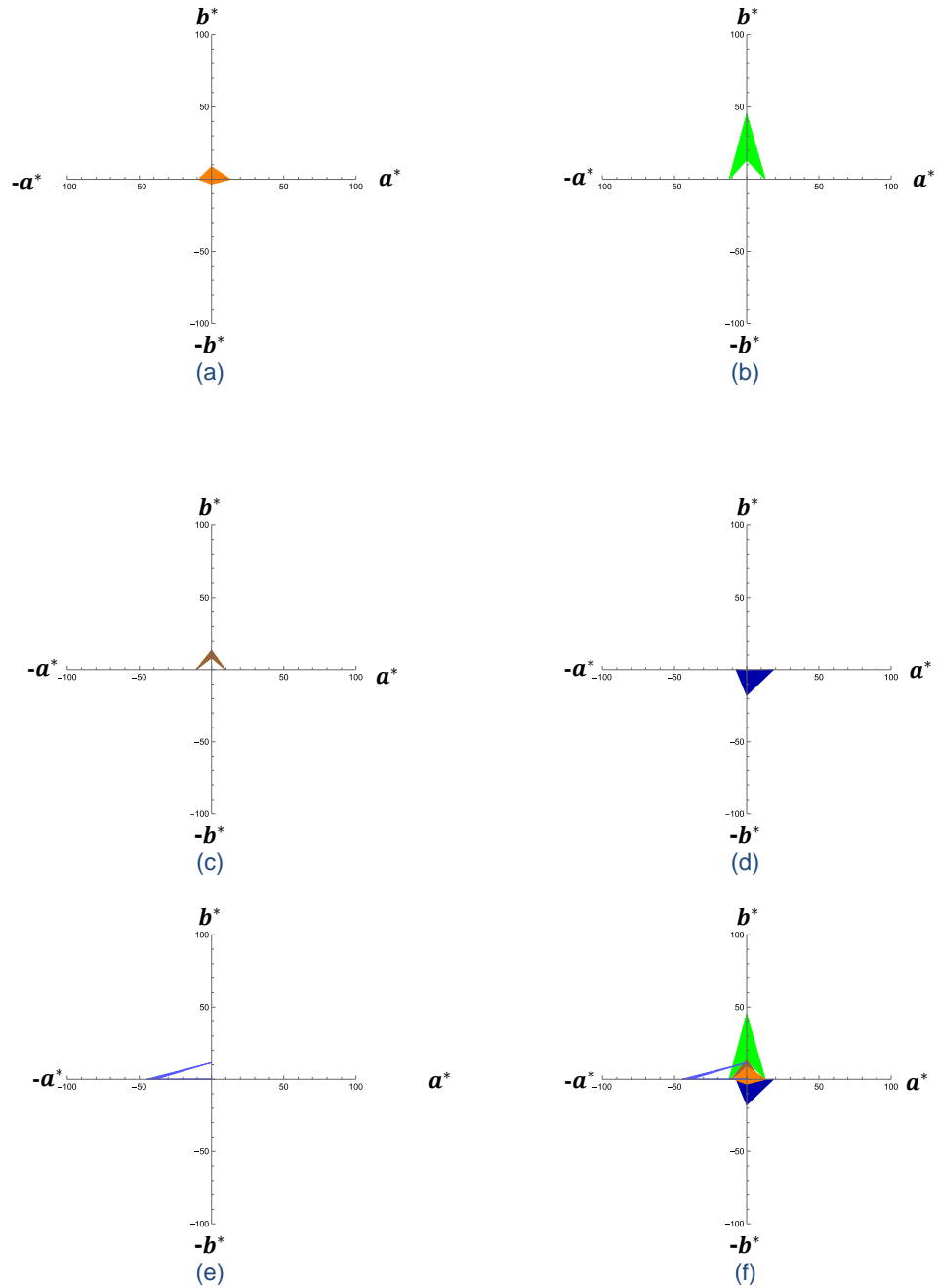


Figure 14. Radar Charts for Sample 3 (a) Cluster 1 (b) Cluster 2 (c) Cluster 4 (d) Cluster 6 (e) Cluster 7 (f) All Clusters

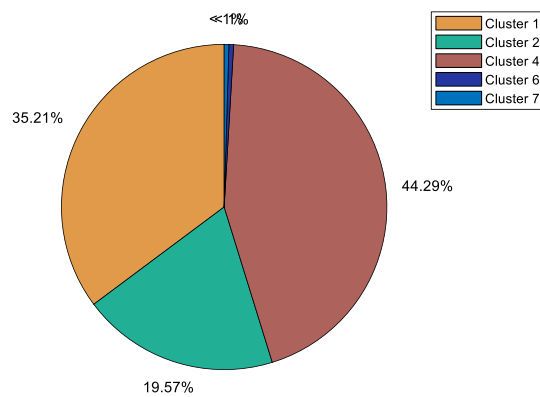


Figure 15. Pie chart for Sample 3

Figure 14 shows the visualization of the minimum and maximum a^* and b^* values which are stated in Table 4. Lee *et al.* [10] stated that this sample had a^* and b^* values which are 7.76 and 24.68 respectively and the sample are categorized as slightly weathered. The value obtain from previous study are in Cluster 2's range. Comparing the values of Cluster 2 in Table 4 and Table 1, both a^* and b^* values existed in the class of slightly to moderately weathered. On the other hand, the result from Figure 15 shows that Cluster 4 has the highest percentage among all the clusters with 44.29%. From the comparison of Cluster 4 and Table 1, this cluster categorized as unweathered to slightly weathered region. Hence, the weathering grade of this sample is slightly weathered.

Sample 4

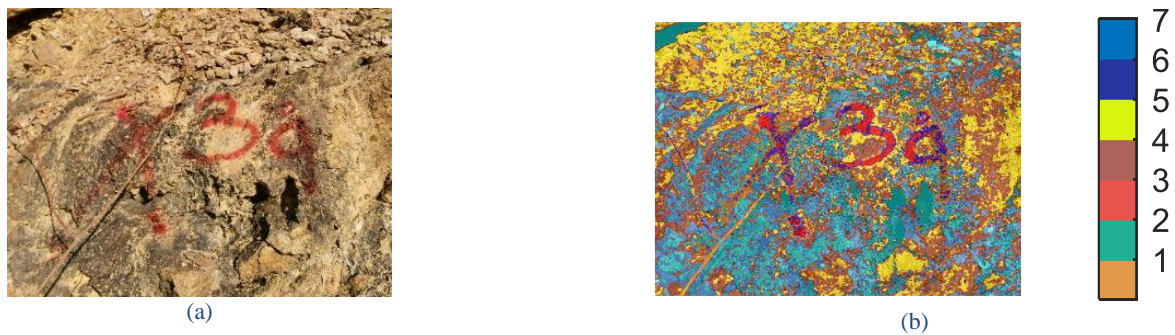


Figure 16. (a) Original Image (b) Output image

The last sample contains 3000×4000 pixels. Compared to the samples before, this is the largest value of the pixel count. Figure 17 shows the output image in Figure 16 (b) separately after $K - Means$ clustering.

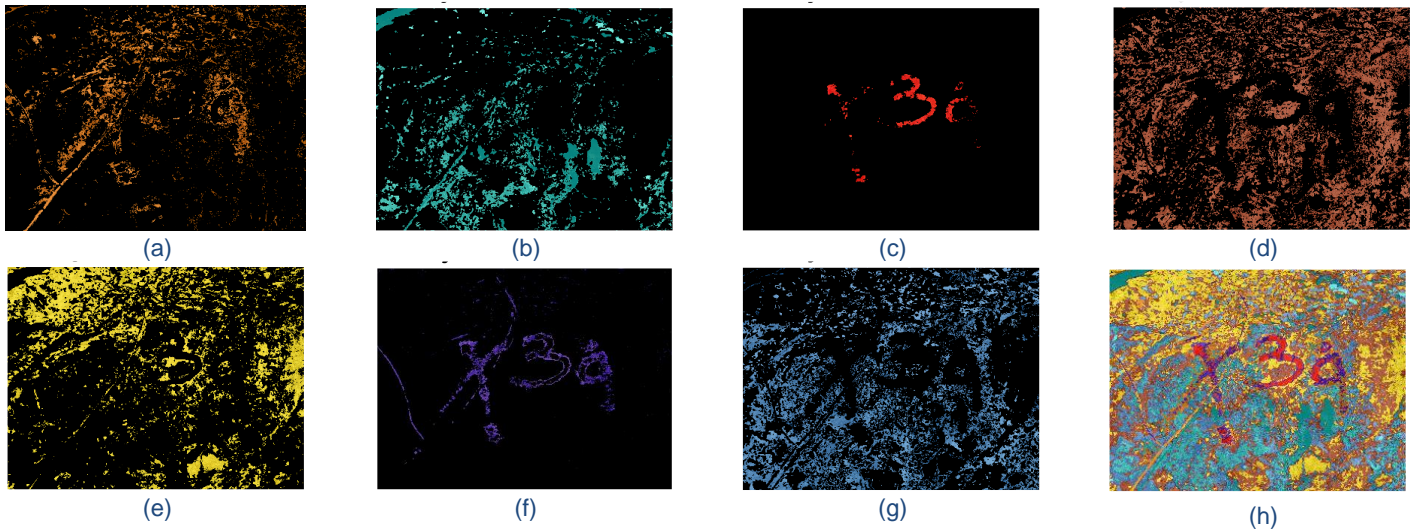


Figure 17. Clusters in Sample 4 (a) Cluster 1 (b) Cluster 2 (c) Cluster 3 (d) Cluster 4 (e) Cluster 5 (f) Cluster 6 (g) Cluster 7 (h) All Clusters

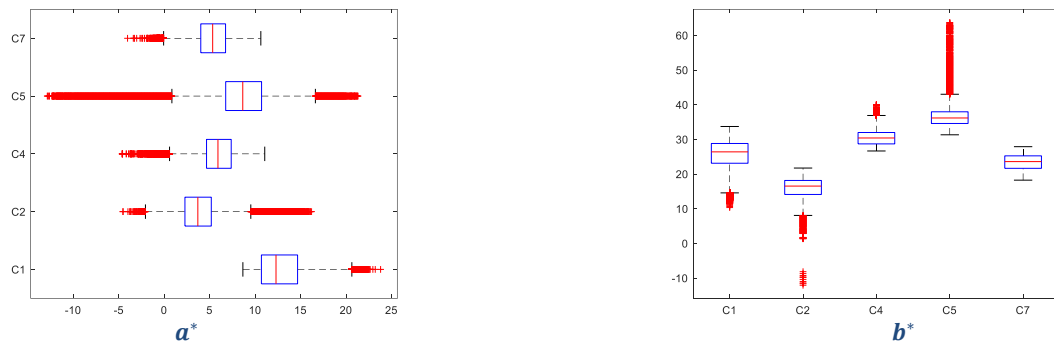


Figure 18. Boxplots for a^* and b^* values of Sample 4

Note that Cluster 3 and 6 are the rejected clusters. Figure 18 shows the visualization of the results for better observations and discussions. Figure 18 below shows a^* and b^* values presented in boxplot.

Table 5. Maximum and minimum values of a^* and b^* for clusters in Sample 4

Cluster	Max a^*	Min a^*	Max b^*	Min b^*
1	23.80	8.65	33.71	10.33
2	16.21	-4.54	21.72	-12.06
4	11.06	-4.60	39.95	26.65
5	21.32	-12.85	63.68	31.31
7	10.63	-4.02	27.89	18.23

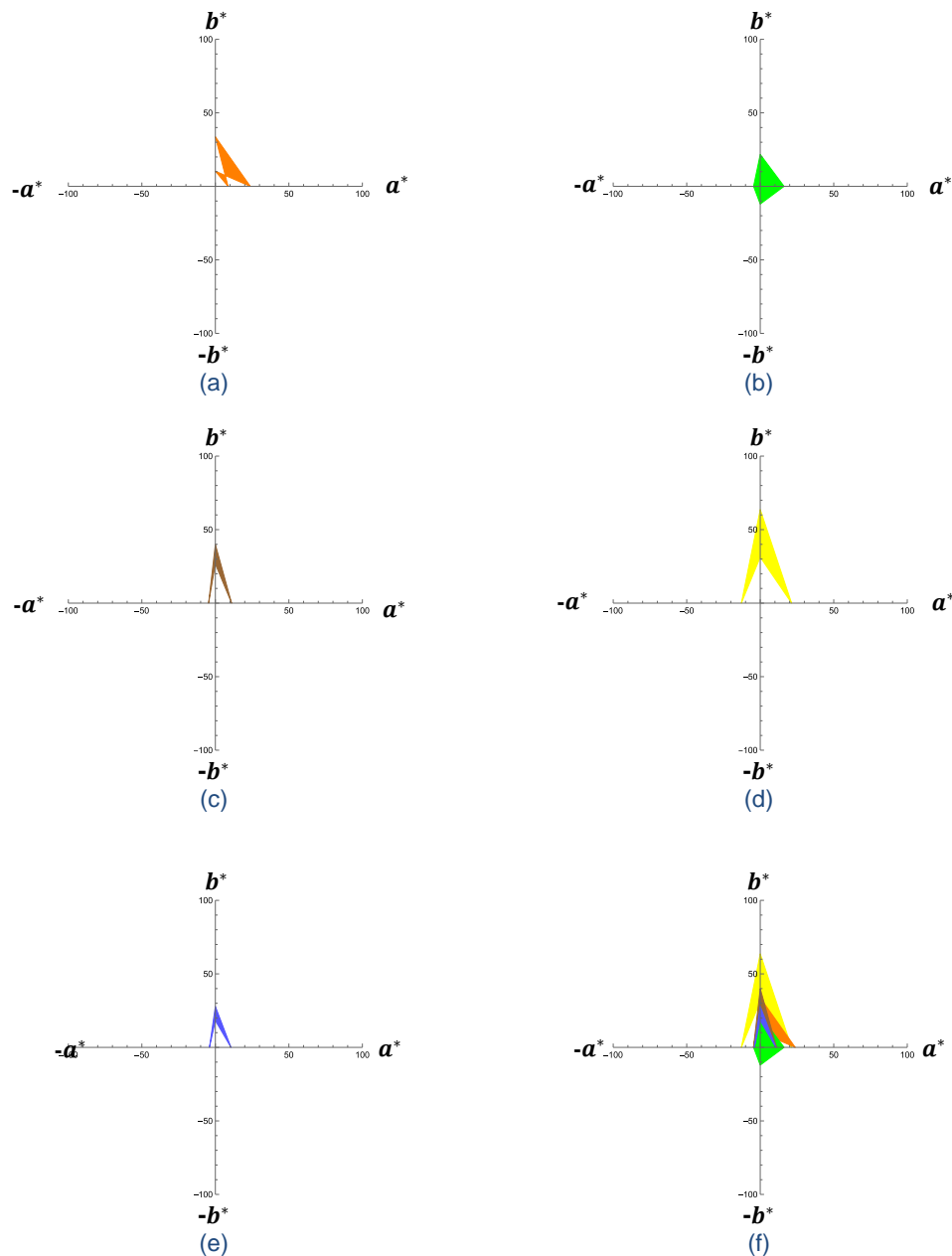


Figure 19. Radar Charts for Sample 4 (a) Cluster 1 (b) Cluster 2 (c) Cluster 4 (d) Cluster 5 (e) Cluster 7 (f) All Clusters

From Table 5, the radar charts in Figure 19 will give a better explanation. In the previous research by Lee *et al.* [10], it is stated that this sample had a^* value which is 28.27 and b^* value which is 51.23. Moreover, the sample is categorized as moderately weathered to highly weathered surface. However, value a^* is not in the range of this purpose study but it can be assumed to be in the closest cluster. Hence, those values from previous study are in the range of Cluster 5. Cluster 5 classified as highly weathered since the values in Table 5 are in a^* and b^* range of highly weathered class in Table 1. As an explanation purposes, even though the values of minimum and maximum for Cluster 5 belongs to several classes in Table 1, the final classification taken by considering the value of a^* and b^* are both included in respected weathering grade.

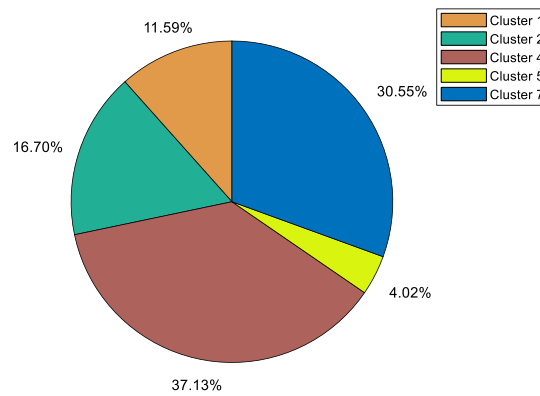


Figure 20. Pie chart for Sample 4

The pie chart in Figure 20 shows the highest percentage of Sample 4, which is Cluster 4 that covers 37.13%. Comparing the values of Cluster 4 in Table 1, this cluster categorized as moderately weathered. Hence, the weathering grade for Sample 4 is moderately weathered.

In summary, all of the samples are categorized according to their weathering grade based on the highest percentage from the pie charts. The comparisons are not exactly accurate but the results are close to the previous research by Lee *et al.* [10].

Although there are issues encountered, this study successfully identified the rejected cluster which contain plants and painted numbers that are included in the image compared to the previous research. Since a^* and b^* values in the boxplots are shown separately, the radar charts help to visualize both values from the boxplot simultaneously. However, there are outliers which are far from the interquartile range and that may have affected the radar charts. On the other hand, since the boxplots and radar charts cannot display the quantity in the range of values for each cluster, the pie charts show the percentage of a^* and b^* for each cluster to complete the categorizing process. The highest percentage of the pie charts does not represent the weathered region of the sample, it represents the quantity of pixel numbers contained in each cluster.

Conclusions and Future Work

In summary, one of the limitations found in the proposed study is the differences between each sample in term of the clustering. For example, the number of rejected cluster can be two or three clusters for each sample. Next, the shadows that appeared during the capture of the sample image may affected the result. Although the values of a^* and b^* are not in the range for Sample 4, the explanation and the pie chart helped to categorize the rock mass to its weathering grade that are similar to the previous study. Upon observing the box plots for all samples, it is evident that there are numerous outliers, necessitating greater attention to their removal for improved discussion.

In the future, the process of the proposed method can be automated by integrating deep learning or machine learning. Additionally, research using the proposed method can be conducted on several types of rocks to compare the analysis of the a^* and b^* values. Moreover, the study's sample images could be the focal point for image testing using an established approach such as the support vector machine (SVM) technique.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgment

This research was supported by the Ministry of Higher Education Malaysia through the Fundamental Grant Scheme (FRGS/1/2023/STG06/USM/03/4) and the School of Mathematical Sciences, Universiti Sains Malaysia. The authors are very grateful to the anonymous referees for their valuable suggestions.

References

- [1] Abad, S. V. A. N. K, Mohamad, E. T., Komoo, I. & Kalatehjari, R. (2015). Assessment of weathering effects rock mass structure. *Jurnal Teknologi*, 72(1), 2180-3722.
- [2] Ahmar, A. S., Napitupulu, D., Rahim, R., Hidayat, R., Sonatha, Y. & Azmi, M. (2018). Using K-Means clustering to cluster provinces in Indonesia. *Journal of Physics: Conference Series*, 1028.
- [3] Alam, M. S., Rahman, M. M., Hossain, M. A., Islam, M. K., Ahmed, K. M., Ahmed, K. T., Singh, B. C. & Miah, M. S. (2019). Automatic human brain tumor detection in MRI image using template-based K-Means and improved fuzzy C means clustering algorithm. *Big Data and Cognitive Computing*, 3(27).
- [4] Chen, Y., Liu, D. & Liang, J. (2013). A new method for RGB to CIELAB color space transformation based on Markov chain Monte Carlo. *Proc of SPIE*, 8920, 89200G
- [5] Chitradevi, B. & Srimathi, P. (2014). An overview on image processing techniques. *International Journal of Innovative Research in Computer and Communication Engineering*, 2(11), 6466-6472.
- [6] Color Math and Programming Code Examples. (2023). [Online]. [Accessed 26th May 2023]. Available from the World Wide Web: <http://www.easyrgb.com/en/math.php>.
- [7] Hobart, M. K. (2023). Rocks: Igneous, Metamorphic and Sedimentary [Online]. [Accessed 18th August 2023]. Available from the World Wide Web: <https://geology.com/rocks/>.
- [8] Jain, A. K., Murty, M. N. & Flynn, P. J. (1999). Data clustering: A review. *ACM Computing Surveys*, 31(3), 264-323.
- [9] Kanjanawanishkul, K. (2022). An image-based Eri Silkworm Pupa grading method using shape, color, and size. *International Journal of Intelligent Engineering and Systems*, 15(2), 563-571.
- [10] Lee, K. Y. A., Razali, M., Ismail, M. A. M., Yusoff, I. N., Nagendran, S. K., Zainal, Z., Tobe, H., Date, K. & Yokota, Y. (2023). Preliminary analysis of rock mass weathering grade using image analysis of CIELAB color space with the validation of Schmidt hammer: A case study. *Physics and Chemistry of the Earth*, 129, 103291.
- [11] Lindbloom, B. J. (2017). RGB/XYZ Matrices. [Online]. [Accessed 21th August 2023]. Available from the World Wide Web: http://www.brucelindbloom.com/index.html?Eqn_RGB_XYZ_Matrix.html.
- [12] Manickavasagan, A., Al-Mezeni, N. K. & Al-Shekaili, H. N. (2014). RGB color imaging technique for grading of dates. *Scientia Horticulturae*, 175, 87-94.
- [13] MacQueen, J. B. (1967). Some method for classification and analysis of multivariate observations. *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability 1967*. University of California Press, 281-297.
- [14] Patil, B. G. & Jain, S. N. (2014). Cancer cells detection using digital image processing methods. *International Journal of Latest Trends in Engineering and Technology*, 3(4), 45-49.
- [15] Razali, M., Jamal, M. H., Ismail, M. A. M., Yusoff, I. N., Nagendran, S. K., Zainal, Z., Zainuddin, N. E., Tobe, H., Miyoshi, T., Date, K. & Yokota, Y. (2021). Quantitative weathering assessment of rock slope using CIELAB color space and image analysis technique. Research Square.
- [16] Razali, M., Ismail, M. A. M., Yusoff, I. N., Tobe, H., Miyoshi, T., Date K. & Yokota, Y. (2022). Analysis of rock mass surface colour using CIELAB and Munsell Chart. *IOP Conference Series: Earth and Environmental Science*, 1091, 012048.
- [17] Service PT, CIELAB System. (2013). Defining and communicating color: The CIELAB system (2013). In Sappi Fine Pap North Am, 1-8.
- [18] Terzaghi, K. (1946). Rock defects and loads on tunnel supports. *Rock Tunneling with Steel Supports*, 1, 17-99.
- [19] Tobe, H., Miyajima, Y., Shirasagi, S. & Yamamoto, T., (2020). A new method for predicting of rock fall from the tunnel face. *Rock Mechanics for Natural Resources and Infrastructure Development- Proceedings of the 14th International Congress on Rock Mechanics and Rock Engineering. ISRM 2019*, 1761-1768.
- [20] Yusoff, I. N., Ismail, M. M. A., Tobe, H., Miyoshi, T., Date, K. & Yokota, Y. (2023a). Detection of discontinuity pattern and orientation for tunnel faces by using 2D and 3D image analysis techniques. *Physics and Chemistry of the Earth*, 129, 103330.
- [21] Yusoff, I. N., Ismail, M. M. A., Tobe, H., Date, K. & Yokota, Y. (2023b). Quantitative granitic weathering assessment for rock mass classification optimization of tunnel face using image analysis technique. *Ain Shams Engineering Journal*, 14, 101814.