International Journal of Innovative Computing, Information and Control
Volume **12**, Number **5**, October **2016**

ICIC International ©2016 ISSN 1349-4198
pp. 1645–1653

SCOTECT ALGORITHM: A NOVEL APPROACH FOR SOIL COLOR DETECTION PROCESS USING FIVE STEPS ALGORITHM

Fitri Utaminingrum and Ihwanudien Hasan Robbani

Faculty of Computer Science
Brawijaya University
Jl. Veteran No. 8, East Java, Malang 65145, Indonesia
f3_ningrum@yahoo.com; matno.robbani@gmail.com

Received March 2016; revised August 2016

Abstract. *Soil is a major component of the land. Soil color is often used as the initial impression when we view the soil. The color is also affected by the environment, deepness, mineral content, etc. Munsell Soil Color Chart is a book to classify soil colors, but so difficult to classify the color of soil using this book. In our paper, we research how soil color can be classified using the proposed algorithm, Soil Color Detection (Scotect) algorithm. Scotect is five steps algorithm to detection the color of the soil. The main process is to detect the color of the soil. The first step is creating the database, and we can use mode of RGB value to get representation data. Second, a median filter method is used to get the clearer image. Third, an image will be segmented by using K-means segmentation method. Furthermore, the segmented image will be filtered again by using median filter method. And the last process is matching each layer of image soil with color in the database using Euclidean distance. This research succeeded in finding the new way for detecting the color of the soil. We succeed in showing that the program can segment soil image. The most important is that this algorithm's output succeeded concluding that result of this program is 90.58% accurate to retrieve label of the testing data.*
Keywords: Soil color, Munsell Soil Color Chart, Color detection, Scotect algorithm

1. **Introduction.** Indonesia is an agricultural country, which means that most of the Indonesian people worked as farmers. Indonesia has a large area and good geographical conditions, and it is the main requirement for farming. Indonesian Data Center and Information System of Agriculture report that in 2012 the land area in Indonesia is about 39.594.536,91 Ha. Regarding to the differences in geographical condition, climate, weather, and many other factors, such lands have a different quality.

Soil is a major component of the land. Soil is a natural body comprised of solids (minerals and organic matter), liquid, and gases that occurs on the land surface, occupies space, and is characterized by one or both of the following: horizons, or layers, that are distinguishable from the initial material as a result of additions, losses, transfers, and transformations of energy and matter or the ability to support plants in the natural environment [1].

A soil horizon commonly is differentiated from the horizons adjacent to it partly by characteristics that can be seen or measured in the field, such as color, structure, texture, rupture-resistance class, and the presence or absence of carbonates [1]. Soil color is often used as the initial impression when we view the soil. The color is also affected by the environment: aerobic environments produce sweeping vistas of uniform or subtly changing color, and anaerobic (lacking oxygen), wet environments disrupt color flow with complex, often intriguing patterns and points of accent [2]. Deepness also affected the color of the soil, and colors usually become lighter, yellower, or redder [2].

Munsell Soil Color Chart is a book to classify soil colors, but so difficult to classify the color of soil using this book. Many colors variations make researcher difficult to choose the right color. To get the right result, it takes some time and a high accuracy. Researchers condition may also affect the result. Issues relating to the time, accuracy, and the effect of conditions now are not a serious problem. The machine is a reliable solution. The machine can complete some task faster than humans. The machine can give a decision with higher accuracy than humans. It is the reason why a tool must be created that can classify soil colors faster, reliable and accurately.

In our paper, we research how soil color can be determined by using the proposed algorithm, Soil Color Detection (Scotect) algorithm. Scotect is a five steps algorithm to detect the color of the soil, divided into pre-process and the main process. Pre-process consists of filtering and segmentation process. The main process is to detect the color of the soil. The object is a digital image. Euclidean distance is the algorithm used to classify the color of the soil.

Soil consists of various materials such as soil itself, gravels, roots, and garbages. It is important to eliminate the noise in the images before some subsequent processing, such as edge detection, image segmentation, and object recognition [3]. Moreover, the imperfection of pictures taken process will make a greater number of noise. In the soil taxonomy, we just look the soil to get the result of soil color. It is the reason why noise removal is the first and important process in this research.

There are many goals in designing noise removal methods: the visual result of the proposed method must have a smooth, clear texture and no artefacts in the filtering result; the important texture detail should not be lost; the filtering process is conducted especially on the noisy pixel, without engaging the important pixel that is indicated as original pixel; image boundaries should not be blurred or sharpened; preserve the integrity of edge area; and an approach to the filtering method is created to obtain a fast computation time [4]. In our research, we use median filter in the filtering process. Median filtering is a nonlinear process useful in reducing impulsive, or salt-and-pepper noise [5]. The median filter is best known to remove salt and pepper noise in image processing [6]. Median filters are widely thought to be better ones for removing impulse noise since they can also preserve the edges in a corrupted image [7]. Moreover, among all kinds of methods for impulse noise, the median filter is used widely because of its effective noise suppression capability and high computational efficiency [6].

The segmentation method is K-means clustering method. Clustering is a method which groups data into clusters, where objects within each cluster have high degree of similarity, but are dissimilar to the objects in other clusters [8]. Because of its simplicity and efficiency, clustering approaches were one of the first techniques used for the segmentation of (textured) natural images [9]. Image segmentation divides an image into a number of discrete regions such that the pixels have high similarity in each region and high contrast between regions [10]. K-means is a clustering algorithm, which partitions a data set into clusters according to some defined distance measure [9]. K-means algorithm is an unsupervised clustering algorithm that classifies the input data points into multiple classes based on their inherent distance from each other [9]. Referring to "*An Approach to Image Segmentation Using K-means Clustering Algorithm*" concludes that K-means for smaller values of k the algorithms give good results, but very coarse for larger values of k [9]. It is the reason why we use K-means for segmentation process, because soil layer number is below 10, and it means this research uses a small value of k variable.

Scotect algorithm is a novel approach to classify the color of the soil. Scotect algorithm can be applied to mobile devices or special device to detecting/classifying soil color. By Scotect algorithm determination of soil color can be performed more easily, quickly, and accurately. It will improve the work of soil science researchers. Future expected this will increase the agricultural products in the world.

2. **Soil Color.** Soil colour is often used as the initial impression when we view the soil. Since color is one of the most useful attributes to characterize and differentiate soils, which adequate determination has importance in many soil studies. Color can be used as a clue to mineral content of a soil. Iron minerals, by far, provide the most and the greatest variety of pigments in earth and soil as illustrated in Table 1 [2].

TABLE 1. Sample of coil color training data

No	Mineral	Formula	Munsell	Color
1	geothite	FeOOH	10YR 8/6	yellow
2	hematite	Fe2O3	5R 3/6	red
3	lepidocrocite	FeOOH	5YR 6/8	reddish-yellow
4	ferrihydrite	Fe(OH)3	2.5YR 3/6	dark red
5	glauconite	K(SixAl4-x)(Al,Fe,Mg)O10(OH)2	5Y 5/1	drk gray
6	iron sulfide	FeS	10YR 2/1	black
7	pyrite	FeS2	10YR 2/1	black (metallic)
8	jarosite	KFe3(OH)6(SO4)2	5Y 6/4	pale yellow
9	todorokite	MnO4	10YR 2/1	black
10	humus		10YR 2/1	black

Referring to Table 1, we conclude that it is important to get the right result in the soil color detection process. Routine determination of soil color in the field is usually accomplished by visually comparing a soil sample with the chips of standard color charts, of which the Munsell Soil Color Charts are the most familiar to pedologists [3]. Munsell Soil Color Chart describes the color of the soil in 3 features: hue, value, and chroma. For example, a reddish black soil may be noted as: hue value/chroma (10R 2.5/1).

Owing to several psychophysical and physical factors, substantial errors are involved in this visual, subjective method [11]. Basically, a human has a good vision, but the age, health, congenital abnormalities such as color blindness and some other factors may affect the human vision. Decreased ability of human vision will affect the decisions, and it makes the error value increase. It is the reason why a tool must be created that can classify soil colors faster, reliable and accurately.

3. **Proposed Method.** This research will make the machine know how the color of the soil can be detected. Five steps soil color detection algorithm in our research is:

- Firstly, we will make a color database from Munsell Soil Color Chart image. The color database consists of three data features, and there are mode of R (red), G (green), and B (blue) in RGB value.
- Secondly, the image will be filtered to remove noise like stones, garbage, etc. Filtering process use median filter method.
- Thirdly, the filtered image will be segmented. From this process, soil layer boundary will be obtained. Segmentation process use K-means clustering method.
- Fourthly, the segmented image will be filtered again to remove over the noise. In this session, we reuse median filtering method.
- Fifthly, after we have images without noise, each segmentation will be matched with colors in the database. We use Euclidean distance to measure similarity of soil color with color in the database.

(a) Image cropping process (b) Sample of soil color database

FIGURE 1. Creating a database of soil color

(a) 10R 2.5/1 (b) R value (c) G value (d) B value

FIGURE 2. How to get red, green, and blue values

3.1. **Creating a database of soil color.** The color database is made from Munsell Soil Color Chart image. It was taken by iPhone 4 camera (5 MP). Full image is cropped to get one soil color characteristic. Database creation process is illustrated in Figure 1(a) and a sample of soil color database is shown in Figure 1(b). Munsell Soil Color Chart book standard has at least 276 color variants.

Soil color database consists of three data features, and there are mode of R (red), G (green), and B (blue) value. A mode is a number which appears most often in a set of numbers. Because it is an image that consists of only one color, the mode is the right formula to represent RGB values of the image. These are illustrated in Figure 3. Figure 2(a) is a sample of soil color classification in MSCC, and it is reddish black (10R 2.5/1). Suppose Figure 2(b) is R values, Figure 2(c) is G values, and Figure 2(d) is B values of Figure 2(a). From the sets of RGB values we know that R = 26, G = 18, and B = 16. It is mode value of each set data: R, G, and B set data.

3.2. **Median filter.** The median filter is a non-linear filter method. A major advantage of the median filter over linear filters is that the median filter can eliminate the effect of input noise values with extremely large magnitudes [12]. This filter uses a square window for filtering, and the size of a scan window is variable. Y is output of the median filter process, and given a pixel X_{ij}, the corresponding output using the median filter can be defined as Equation (1), where (s, t) represents the pixel position in a scan window and w represents a scan window [14].

$$Y_{ij} = median\{X_{i\pm s, j\pm r} \mid (s, t) \in w\} \tag{1}$$

3.3. **K-means.** The K-means algorithm (KM) partitions data into k sets data. The solution is a set of k center. The membership function is that each data point belongs to its nearest center, forming a Voronoi partition of the data [13]. KM is easy to understand and implement, making it a popular algorithm for clustering. Steps of the KM algorithm can be outlined as mentioned below [7]:

1. Initialize k number of clusters.
2. Choose k number of points randomly and make them initial centroids.
3. Select a data point from the collection, compare it with each centroid and if the data point is found to be similar with the centroid and then assign it into the cluster of that centroid.
4. When each data point has been assigned to one of the clusters, re-calculate the value of the centroids for each k number of clusters.
5. Repeat steps 2 to 3 until no data point moves from its previous cluster to some other cluster (termination criterion has been satisfied).

3.4. **Euclidean distance.** Euclidean distance is a method for detecting similarity between 2 objects. For 2 objects in the p dimensional area, a and b, if $a = (a_1, a_2, \ldots, a_p)$ and $b = (b_1, b_2, \ldots, b_p)$, with the output of the Euclidean distance process being d, Euclidean distance between a and b is described in Equation (2) [15].

$$d_{euc}(a, b) = \left[\sum_{j=1}^{p} (a_j - b_j)^2\right]^{\frac{1}{2}} \tag{2}$$

4. **Main Results.** The purpose of this research is to detect the color of each layer in the soil image. We will discuss step by step how we can get the result. The entire method was implemented in Intel(R) Core(TM) i3-2370M CPU 2.40GHz and RAM 2GB using MATLAB 7.10.0 (R2010a).

4.1. **Soil color database.** The image size is 472×472 pixels. It takes a lot of time to process 276 images with the size. This is a challenge, that how to get the data representation with the bigger size of the image and a greater number of samples. The important things are that we must know that the data training already represents the entire data. Table 2 shows some of the data training. R-value represents the entire value of R that exists in each data, as well as G and B values.

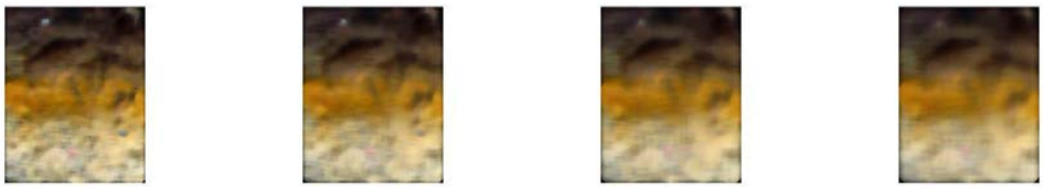
TABLE 2. Sample of coil color training data

No	Color Label	R	G	B
1	10R 2.5/1	26	18	16
2	10R 2.5/2	30	16	14
3	10R 3/1	36	28	28
4	10R 3/2	44	26	24
5	10R 3/3	46	22	18
6	10R 3/4	49	20	14
7	10R 3/6	55	11	6
8	10R 4/1	59	48	49
9	10R 4/2	69	45	40
10	10R 4/3	76	42	35

4.2. **First filter process (original image).** The first filter process is to remove noise from the original image. The filter method is the median filter. In this research, we vary the neighborhood value to get the best result. The neighborhood value is 3×3 to 17×17. Figure 3 shows the results of the first filter process with different sizes of neighborhood. The larger neighborhood makes the image blurry, shown in Figures 3(a) to 3(h), but it is good for the segmentation process. The same color will make the image easily to segment.

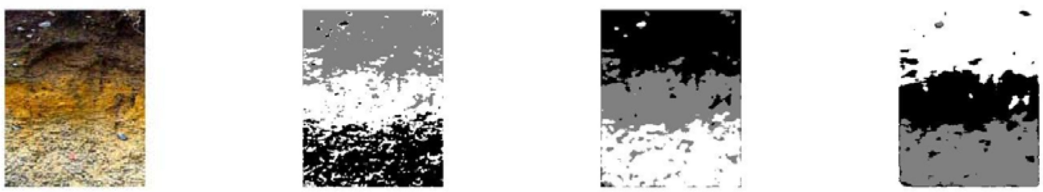
4.3. **Segmentation process.** In our research, the segmentation process uses K-means method. Sample will be segmented to three segmentations because the soil in the image has three layers. Figure 4 shows the results of the segmentation process with different sizes of neighborhood. We succeed in getting the boundary of soil layers, shown in Figure 5. This is due to the fact that soil has clear boundaries. The clearer boundary layer of

(a) Resource (b) 3×3 (c) 5×5 (d) 7×7

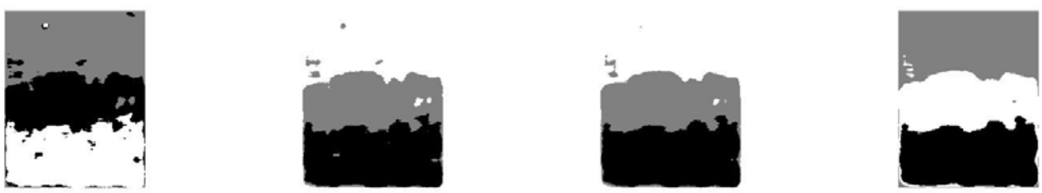


(e) 9×9 (f) 11×11 (g) 13×13 (h) 15×15

Figure 3. First filter process result, neighborhood value variations

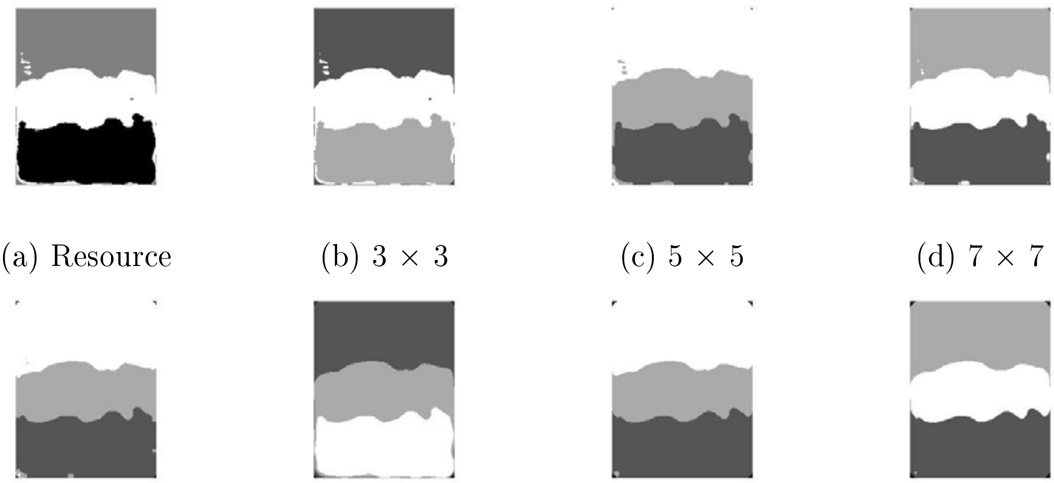


(a) Resource (b) 3×3 (c) 5×5 (d) 7×7



(e) 9×9 (f) 11×11 (g) 13×13 (h) 15×15

Figure 4. Segmentation process result



(a) Resource (b) 3 × 3 (c) 5 × 5 (d) 7 × 7

(e) 9 × 9 (f) 11 × 11 (g) 13 × 13 (h) 15 × 15

FIGURE 5. Second filter process result

soil will make the segmentation process more easily, and easy to get the better result. If the boundary layer of the soil is not clear, it will be worse.

4.4. **Second filter process (get clear boundary).** The second filter process is to get clear boundary of soil layers. The filter method is a median filter. Like the first filter process, in this process we vary the neighborhood value to get the best result. The neighborhood value is 3 × 3 to 17 × 17. Figure 5 shows the results of the second filter process with different sizes of neighborhood. Larger neighborhood creates the clearer boundary. This process depends on the previous process, and if the segmentation process results are good, then this process will get the good result. This research highly depends on the source image. The original image must have a good quality, and have a clear boundary layer of soil. It is shown in Figure 5.

4.5. **Color matching (main feature).** The color matching process is to get the color of each soil layer. Each layer of soil will match using the color in the database to get the result. The matching method is Euclidean distance. In our research, we use 276 data testing. The data testing image is taken by the same camera with the database image. Distance of testing data and result is about 0 and 12. If the distance is getting closer, the data is increasingly similar. Result of this program is 90.58% accuracy to retrieve label of the data. Table 3 shows sample of testing data and the result distance.

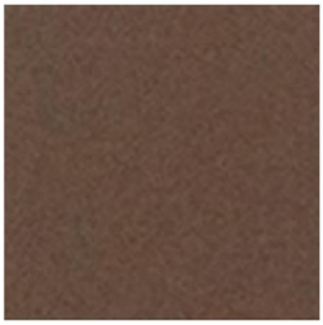
Figure 6 is a comparison between the original image and result from the image. Original image is taken by different cameras. From this comparison, we can conclude that the quality of image determines the result. Images were taken by different cameras, and then the result is wrong. The best results are obtained from images taken with the same camera.

5. **Conclusions.** This research succeeded in finding the new way for detecting the color of the soil. We succeed in showing that the program can segment soil image. The most important is that this algorithm's output succeeded concluding that result of this program is 90.58% accurate to retrieve label of the testing data. The reason shows that we can detect the color of soil by using Scotect algorithm. There are several things to note in this study. First, image for data testing and data training must be taken by the same

Table 3. Color matching result

No	Image Label	Result	Distance
1	10R 2.5/1	10R 2.5/1	1.7321
2	10R 2.5/2	10R 2.5/2	1
3	10R 3/1	10R 3/1	0
4	10R 3/2	10R 3/2	1.4142
5	10R 3/3	10R 3/3	1
6	10R 3/4	10R 3/4	1.7321
7	10R 3/6	10R 3/6	1
8	10R 4/1	10R 4/1	0
9	10R 4/2	10R 4/2	1.7321
10	10R 4/3	10R 4/3	0
11	10R 4/4	10R 4/4	1.7321
12	10R 4/6	10R 4/6	1

(a) Original image 7.5YR 4/3



(b) Result image 5YR 5/2

Figure 6. Color detection of image – 7.5YR 4/3

camera. Second, the results depend on the initial image. Third, it takes a lot of time to process more and bigger data in the database cretaion process. And the last, in the filtering process, larger neighborhood makes the image blurry.

REFERENCES

[1] USDA, *Soil Taxonomy: A Basic System of Soil Classification for Making and Interpreting Soil Surveys*, U.S. Government Printing Office, Washington DC, 1999.

[2] USDA, *The Color of Soil*, United States Department of Agriculture, 2000.

[3] Y.-L. Kuo and C.-W. Tai, A simple and efficient median filter for removing high-density impulse noise in images, *International Journal of Fuzzy System*, vol.17, no.1, pp.67-75, 2015.

[4] F. Utaminingrum, K. Uchimura and G. Koutaki, Mixed Gaussian and impulse noise removal based on kernel observation and edge direction, *International Journal of Innovative Computing, Information and Control*, vol.11, no.5, pp.1509-1523, 2015.

[5] M. Nagu and N. V. Shanker, Image de-noising by using median filter and Weiner filter, *International Journal of Innovative Research in Computer and Communication Engineering*, vol.2, no.9, pp.5641-5649, 2014.

[6] R. Zhu and Y. Wang, Application of improved median filter on image processing, *Journal of Computers*, vol.7, no.4, pp.838-842, 2012.

[7] M.-H. Hsieh, F.-C. Cheng, M.-C. Shie and S.-J. Ruan, Fast and efficient median filter for removing 1-99% levels of salt-and-pepper noise in images, *Engineering Applications of Artificial Intelligence*, vol.26, pp.1333-1338, 2013.

[8] M. K. I. Rahmani, N. Pal and K. Arora, Clustering of image data using K-means and fuzzy K-means, *International Journal of Advanced Computer Science and Applications*, vol.5, no.7, pp.160-163, 2014.

[9] C. Chandhok, S. Chaturvedi and A. A. Khurshid, An approach to image segmentation using K-means clustering algorithm, *International Journal of Information Technology*, vol.1, no.1, pp.11-17, 2012.

[10] V. Jumb, M. Sohani and A. Shrivas, Color image segmentation using K-means clustering and Otsus adaptive thresholding, *International Journal of Innovative Technology and Exploring Engineering*, vol.3, no.9, pp.72-76, 2014.

[11] J. Torrent and V. Barron, Laboratory measurement of soil color: Theory and practice, *Soil Science Society of America*, vol.31, pp.21-33, 1993.

[12] A. Rajasekaran and S. P, Image denoising using median filter with edge detection using canny operator, *International Journal of Science and Research*, vol.3, no.2, pp.30-34, 2014.

[13] G. Hamerly and C. Elkan, Alternatives to the K-means algorithm that find better clusterings, *Conference on Information and Knowledge Management*, Virginia, 2002.

[14] R. C. Gonzales and R. E. Woods, *Digital Image Processing*, Prentice-Hall, Inc., New Jersey, 2012.

[15] M. M. Deza and E. Deza, *Encyclopedia of Distance*, Springer, Berlin Heidelberg, 2009.