The Soil Quality Analysis using K-Mean Technique and Model Color

Thidarat Pinthong
Dep. of Business Information Management
Phetchaburi Rajabhat University
Phetchaburi , Thailand.
Thidarat.pin@mail.pbru.ac.th

Mahasak Ketcham
Dep. of Management Information System
King Mongkut's University of Technology North Bangkok
Bangkok, Thailand
mahasak.k@it.kmutnb.ac.th

2022 International Conference on Cybernetics and Innovations (ICCI) | 978-1-6654-8619-4/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICCI54995.2022.9744162

***Abstract*—The objectives of paper are to design and develop procedures for soil quality analysis using image processing techniques and this research proposes soil quality analysis processes using K-mean grouping techniques and color modeling in soil quality analysis to help solve problems in soil quality analysis. Manage soil quality to be suitable for cultivation. by bringing soil data from the Department of Land Development and experts in soil management to assist in the development of advice, including transfer of knowledge in the development of information systems for soil analysis for the benefit of improving the quality of the soil in agriculture and the stability of the agricultural occupation of Thailand in the future, this research has 75% experimental results.**

Keywords— Soil Quality K-Mean, Model Color

I. INTRODUCTION

The problems of soil degradation are brought about by both natural conditions and misusing of land that does not match the academic principles. Examples of such problems are soil erosion and lack of organic matter. In addition, the problems arising from natural soil condition together with human action include saline soil, acidic soil, organic soil (peat soil), sandy soil and shallow soil. The problematic area for agricultural utilization in Thailand covers 108.87 million rai of soil erosion.

The most problematic area of soil erosion is in the northern region. That is, 98.70 million rai of soil is deficient in organic matter. The lack of organic matter in soil or 77% is in the northeastern region. Soil with problems for agricultural utilization or 209.84 million rai is mostly in the northern and northeastern regions. Saline soil, acidic soil and sandy soil are found in the northeastern area. For improper land use is approximately 36 million rai. The real problem is that farmers do not know the amount of minerals (NPK) in the soil, proper use of soil and soil quality improvement.

The problems mentioned lead to the expectation to develop an application for soil quality analysis using an imaging processing technique. This enables the goal for solving problems in managing soil quality that is suitable for cultivation. The soil data from Land Development Department and experts in soil management is applied to provide advice in development and to transfer of knowledge in the development of information system for soil analysis so that it can give the benefit to soil quality improvement in agriculture and the stability of agricultural profession of Thailand in the future.

II. THEORIES AND RELATED RESEARCH

A. Principles for determining soil quality

Office of Soil Resources Survey and Research from Land Development Department, Ministry of Agriculture and Cooperatives gives the definition for determining soil quality as follow [2]: there are 3 types of soil quality measurements. The details are as follows.

1. Measuring nutrients in soil by measuring NPK: nitrogen, phosphorus, and potassium which are essential nutrients for plants.

2. Measuring the acidity and alkalinity of soil is the nature of potting soil with neutral value or pH 7. When a pH is higher than 7, the soil is acidic. When a pH is lower than 7, the soil is alkaline. Acidic-alkaline soil is not suitable for growing crops [6].

3. Measuring the temperature in soil and soil microorganisms. The number of soil microorganisms depends on soil nutrition, moisture content and pH value. Therefore, soil temperature and soil microorganisms are important for soil quality.

III. METHODOLOGY

The study of "The Development of Application for Soil Quality Analysis Using Image Processing Technique techniques" employed the experiment to measure the efficiency of soil analysis accuracy as follows.

A. Planning and data preparation

This is a process of designing and developing a research conceptual framework from soil sampling collecting, soil transformation, data mining analysis, classification of soil minerals and result displaying. The overview is as follows.

1) Soil collection and land division (large land or different soil)

This is a process of collecting soil samples from farm areas and areas without crops using random technique. 100 soil samples were collected for the experiments.

978-1-6654-8619-4/22/$31.00 ©2022 IEEE

Fig. 1. 100 soil samples Sample from Land Development Department

2) The image photographing process

Soil samples were isolated and contaminants such as garbage, grass clippings, tree roots, and other foreign matters were removed. The soil samples were grounded by a 1-cm soil sifter to make them fine and suitable for mixing with nitrogen, phosphorus, potassium, NPK solution. The titrate powder and nitrogen, phosphorus, potassium NPK solution were mixed to chemically transform the soil.

Fig. 2. Color changes after titrations and cotton wool was used to reduce color variations.

TABLE I. TABLE TYPE STYLES

No.	Soil set	Type of soil set	Image	X Pixels	Y Pixels
1	5th soil set	River sediment		960	720
2	11th soil set	Soil from lowland area		960	720
3	18th Soil set	Soil from upland area		960	720
4	34th soil set	Soil from rainy area		960	720
5	62nd soil set	Soil from high slope area		960	720

B. The Image enhancement.

It is a preparatory step starting from image analysis to determination of RGB values including red, green, and blue which were used to classify color groups and link with agricultural information and geographic information.

1) The unit of color division in this research was defined as Pixel 1 unit : 1. Then, the area of ROI was divided into non-overlapping block with the size of 8 × 8 pixel. X was an original image. In the ROI region, a grayscale image with the size of N 1 × N2 pixel can be substituted by the following equation (3-3).

$$X = \{X_{(i,j)};\ 0 \le i < N_1, 0 \le j < N_2\}$$

When $X(i, j) \in \{0, \ldots, 2L - 1\}$ is the darkness of pixel (i , j) and L is the amount of memory in bits used in each pixel. The total number of regions associated with the original image is $\frac{N_1}{8} \times \frac{N_2}{8}$ regions or is equal 64×64 regions when $N_1 = N_2 = 512$. This is illustrated in Figure 3-4.

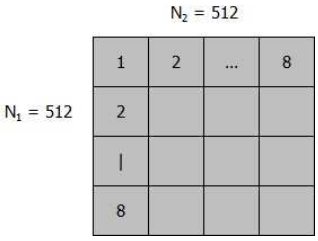

Fig. 3. Non-overlapping block of the original image

2) The RGB color average of the image was determined to analyze the color range for appropriate class of dataset by finding the color average value from all pixels of the image using the internal color average and the color value of RGB

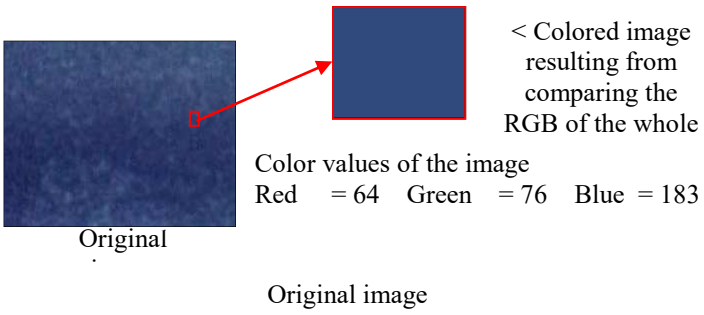

Fig. 4. Determining RGB color average from the image

3) RGB color value classification. The RGB color values were classified into different classes to determine GRB color groups so that the data format can be used to support soil quality improvement and the suitability of each type using K-

Mean method. The precise data group classification was determined by different color ranges as follows.

TABLE II. TABLE FOR DETERMINATION OF THE NUMBER OF DATASETS

Type of data	Data format	Details	Imported datasets before data mining
Nitrogen	5 Class	N0-N4	300 images
Phosphorus	5 Class	P0-P4	300 images
Potassium	5 Class	K0-K4	300 images
pH	12 Class	pH3.0 -8.5	300 images
NPK	15 Class	N0-N4, P0-P4, K0-K4	300 images

NPK and pH ranges in data class were determined by using the data from Land Development Department to determine NPK and pH value for determination of data class range.

TABLE III. DATASETS WITH NPK VALUES

Type of data	N	p	k
NPK,1	0	5	50
NPK,2	10	10	200
NPK,3	20	20	400
NPK,4	40	50	600

TABLE IV. DATASETS WITH PH VALUES

Class pH	pH	Class pH	pH
pH1	3.0	pH7	6.0
pH2	3.5	pH8	6.5
pH3	4.0	pH9	7.0
pH4	4.5	pH10	7.5
pH5	5.0	pH11	8.0
pH6	5.5	pH12	8.5

4) Data mining was created by K-Mean.

The RGB data with unclassified class or unknown data was used to determine data distance in each class using Euclidian Distance. The following equation was applied.

$$d(X_i, X_j) = \sqrt{\sum_{K}^{N} (X_i - X_j)^2}$$

When

d (xi , xj) distance between sample xi and sample xj

n = total number of sample properties

K= data location in xi

C. Collecting information for soil quality management

This process is to analyze NPK value to identify the relation with agricultural data to formulate solutions for soil quality improvement and pH value by using information from Land Development Department. The process is as follows.

1) Agricultural information is information on solving problems of soil surface, soil management suitability for growing cash crops. The information used in soil management is as follows.

- Location information where soil samples were collected was obtained from identification via display communication devices.
- Fertilizer formula information for improving soil quality that is suitable for cash crops.
- Graphically displayed data of the soil set with a high-angle image that can be converted into an aerial photograph.
- Information of soil problem management such as acidity and alkalinity management, soil improvement [5], [8], [9] was obtained from Ministry of Agriculture, Rice Department.

All data was analyzed to determine the suitability of soil improvement to create value and to develop soil quality improvement. The data from Land Development Department and the data from agricultural research [11][12] defined 4 cash crops including 1. photosensitive rice, 2. non-photosensitized rice, 3. cassava and 4. sugarcane and they were used in the preliminary analysis.

2) The NPK data was linked. The NPK and pH value were verified to retrieve appropriate information on soil problem solutions. Appropriate data conditions were determined according to the data class.

IV. EXPERIMENTAL RESULTS

A. Results of System Development

The results of information system development for soil quality analysis by using K-Mean Technique and Model Color. The efficiency of use and analysis of soil quality was tested.

1) Soil quality analysis experiments

TABLE V. SUMMARY OF NITROGEN, PHOSPHORUS, AND POTASSIUM (NPK) ANALYSIS

Time	Image	Accuracy	Average
1	100 images	74	74%
2	100 images	75	75%
3	100 images	76	76%
Total	300 images	273	75%

Form Table shows the color detections. The three experiments of 100 images each was found to be highly accurate in the analysis of nitrogen, phosphorus, and potassium (NPK) value. Based on image processing using the application of soil quality analysis, the rate was at 75 percent. This is in line with the assumption at an accuracy level above 70%.

V. CONCLUSION

The researcher focused on designing and developing the application for soil quality management that is suitable for agriculture. The system for detecting nitrogen, phosphorus, and potassium in soil using color theory from soil images that have undergone the titration was developed. The potential performance of the application is comparable with a computer system's performance. The application applies image processing theory and K-Mean data mining.

According to the experiments, it was found that the application was able to detect the colors, the NPK and the pH value of the experimental soil sets compared with the actual data obtained from the laboratory of Land Development Department. In soil analysis, it was found that the accuracy of the NPK analysis was at 75 percent of the total experiments.

REFERENCES.

[1] W.Ekburanawat.(2011) "ICD-10 Coding Guide for Occupational Diseases". Samma Vocational Foundation.

[2] N.Chumuang,N.Meesang, M.Ketcham, W.Yimyam, J.Chalermdit, N.Wittayakhom and P.Pramkeaw. "An Efficiency Random Forest Algorithm for Classification of Patients with Kidney Dysfunction." 2020 15th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP). IEEE, 2020.

[3] M.Ketcham,W. Yimyam, and N. Chumuang. "Segmentation of overlapping Isan Dhamma character on palm leaf manuscript's with neural network." Recent Advances in Information and Communication Technology 2016. Springer, Cham, 2016. 55-65.

[4] K. Yeh, N. Lo, L.Chen, and P.Lin, (2015). "A Fraud Detectio System for Real-time Messaging Communication on Android Facebook Messenger" In 2015 IEEE 4th Global Conference on Consumer Electronics (GCCE) . 361-363.

[5] J. Brixey., et al. (2017). "A Facebook chatbot for Sexual Health Information on HIV/AIDS" In Proceedings of the SIGDIAL 2017 Conference. 370–373.

[6] D.R. Vukovic., et al. (2016). "Facebook Messenger Bots and Their Application for Business" In 24th Telecommunications forum TELFOR 2016.

[7] A. Vegesna., et al. (2018). "Ontology based Chatbot (For E-commerce Website)" In 2018 International Journal of Computer Applications (0975 – 8887). 51-55.

[8] LV. Zhe-peng., et al. (2016). "Effect of Instant Messenger Use on Purchase Decision of Consumers" In 2016 National Natural Science Foundation of China.

[9] F. Daniel., et al. (2018). "Toward Truly Personal Chatbots On the Development of Custom Conversational Assistants" In 2018 ACM/IEEE 1st International Workshop on Software Engineering for Cognitive Services. 31-36.

[10] A.M. Rahman., et al. (2017). "Programming challenges of Chatbot: Current and Future Prospective" In 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC). 75-78.

[11] N.Chumuang, S.Hiranchan,M. Ketcham, W.Yimyam, P.Pramkeaw, and S.Tangwannawit. "Developed Credit Card Fraud Detection Alert Systems via Notification of LINE Application." 2020 15th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP). IEEE, 2020.

[12] W.Yimyam, and M. Ketcham. "The Grading Multiple Choice Tests System via Mobile Phone using Image Processing Technique." International Journal of Emerging Technologies in Learning (iJET) 13.10 (2018): 260-269.

[13] T.Pinthong, W. Yimyam, N. Chumuang and M. Ketcham. "Face Recognition System for Financial Identity Theft Protection." 2020 15th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP). IEEE, 2020.