T-SNE COLOR AND SFTA TEXTURE FEATURES FOR AERIAL IMAGES PALM OIL PLANTATIONS AREA CLASSIFICATION

[1]AGUS HARJOKO, [2]SRI HARTATI, [3]BAKHTIAR ALLDINO ARDI SUMBODO, [4]AUFACLAV FRISKY ZATU KUSUMA

Department of Computer Science and Electronics, Universitas Gadjah Mada, Indonesia
E-mail: aharjoko@ugm.ac.id

Abstract- Indonesia has massive palm oil plantation area. Manual evaluation of the palm oil area would take a lot of effort and time. For that reason, an automated palm oil evaluation is proposed. This paper presents an early study and its results to automated palm oil evaluation. First, an aerial image of the palm oil plantation area is obtained using unmanned aerial vehicle (UAV). The images are then stitched altogether to obtain a single image with much greater area of coverage. Next, the image is divided into several patches, and features are extracted for each patch. This paper used newly joined features to represent a color and textures of the image. The texture features are extracted using Segmentation-based Fractal Texture Analysis (SFTA). The color features are extracted using T-SNE dimensional reduction in the RGB planes. The AdaBoost classifier is adopted to determine the class of these patches and classify it into three categories, i.e. palm oil plantation, non-palm oil plantation, and non-plantation area. Individual and join features performance is tested. The result shows that texture feature alone gives 94.23% classification accuracy, whereas color feature alone yields 45.06%. The join features improves the classification accuracy by 0.85% to give 95.08%.

Index Terms- palm oil, plantation, UAV, image processing, SFTA, T-SNE, AdaBoost

I. INTRODUCTION

Palm oil is one of the primary commodity in Indonesia. Ministry of Agriculture of the Republic of Indonesia said that the acreage of palm oil plantations in Indonesia is now about 8 million hectares, twice as many as in year 2000. This number is expected to rise to 13 million hectares in 2020. In 2008, the palm oil plantations was 7 million hectares, with a production of 19.2 million tons. With these vast and hard to reach areas, it is hard to evaluate all of the plantation areas manually.

In order to solve the problem of mapping and monitoring plantation region, Unmanned Aerial Vehicle (UAV) has been used by many researchers. UAV with RGB camera mounted on it was used to capture a wider area coverage. Aerial images have much information about a covered area. This information used by researchers to monitor a remote area [1]–[3] or wide vegetation area [4], [5]. Aerial image processing also popular for another reason such as urban aerial mapping [2] and disaster prevention [6]. In recent studies, further methods were applied to extract information in aerial images [1], [4], [7]. Machine learning methods have been used to extract the information in the image and classify it into different kind of regions. Many feature extraction methods were also applied to the image i.e. Local Binary Pattern [8] and Gray Level Co-Occurrence Matrix [7], [9] to obtain the representation of the images and then classifier like SVM [7] was used to determine the class of the image area. The study in [7] can classify the region very well. However, their model still can't differentiate an area with the same color, such as tea plantation area with a tree that grows in the region with the same color. In this paper, we propose a method to classify a palm oil plantation area into three categories i.e. palm oil tree, non-palm oil tree, and non-plantation area. A two stage approach is performed to achieve the results. These steps are feature extraction with new joined features and classification process, and it will be explained in section 3.

II. RELATED WORKS

Processing on aerial imagery taken by UAV have been done by many researchers. Harjoko et al.[7] used pictures from the UAV as a dataset to classify the tea plantation in several regions, Indonesia. This study aims to classify the tea plantation area into ready for harvest, should be revitalized and not a tea plantation area using images obtained from the UAV. In their study, joined features between Color Moment and GLCM are used as a features and Support Vector Machine as a method of classification. Their model reached 94.42% accuracy with 98.36% precision.

Gadzal et al. [5] conducted another study using NDVI video and level set methods. The study aims to monitor and estimates the vegetation area from an image taken by UAV. All the image that has been obtained by UAV were combined using a stitching method to get the whole image area. After that, the level set method is used to get the contour of the image and determine the limits of the vegetation area.

UAV research through the vegetation monitoring is also done by Mitchell et al. [1]. The study focused on the classification of the arid region of aerial imagery. In the study, data are clustered using the K-Means and ISODATA. The result from this study shows that K-means clustering method produced inconsistent results, and ISODATA generates data similar to the original data.

Finding the best representative feature in image analysis is one of the problems in classification research. Texture and color are two basic features in describing an image because it has strong links with the human perception. One way to extract the color feature is using Color Moments [10]. Color moments are scaling, and rotation invariant that characterize color distribution in an image. These features are quite robust in a light change condition, but cannot handle occlusion in an image [10]. For texture, Grey Level Co-Occurrence Matrix (GLCM) by Haralick [9] is widely used in many image classification. These features are joined features from many statistical occurrences in an image like Energy, Correlation, Contrast Level, and Homogeneity. Joined Color Moments and GLCM features are used in research conducted by Harjoko et al. [7]. In their research, they used this combined features to classify the area of tea plantation in several districts in Central Java, Indonesia.

Basically, an image with Red, Green, and Blue channel (RGB) has color properties in three channels that can be used as a feature. These resources can give information about an area or object which has a special color. Therefore, we tried to use dimensionality reduction method to extract color feature on three different channels in RGB. There are several ways to reduce the dimension of the vector, such as PCA, LDA, and T-SNE. Research performed by Maaten et al. shows that T-SNE performs better than PCA and LDA [11]. Based on this, T-SNE dimensionality reduction is performed in each channel separately and join the results into one vector. This final vector will be used as color features.

In another aspect, image texture is another representation which can be different from the spatial arrangement of color or intensities in an image. Image texture can be used to classify by using the information of perceived texture between two areas or objects in an image. GLCM and Gabor are the examples of this method which usually used to obtain the texture of the images [9], [12], [13]. In 2002, Costa et al. [13] in their research performed a new method to obtain the texture of the image, called Segmented based Fractal Texture Analysis (SFTA). Their study proved that SFTA was achieved a better result than GLCM and Gabor. Join the multi-feature is one of the best ways to improve the discrimination level of the feature. Many researchers join more than one features into one vector which represents each data. Most of them obtained better result by combining more than one features to represent each element in the data. This paper uses SFTA as a texture features and join it with a color feature.

III. PROPOSED METHOD

A. Dataset

In this experiment, we used RGB image of palm oil plantation taken from Kebun Inti II, South Sumatra province, using UAV. Each image was taken with a resolution of 3600 x 4000 pixels using UAV from around 350 meters above the ground. The images were stitched altogether to obtain one image covering much larger area as shown in Fig. 1(a). The research used 100x100 as the size of the patches (see Figure 1(b), 1(c), and 1(d)).

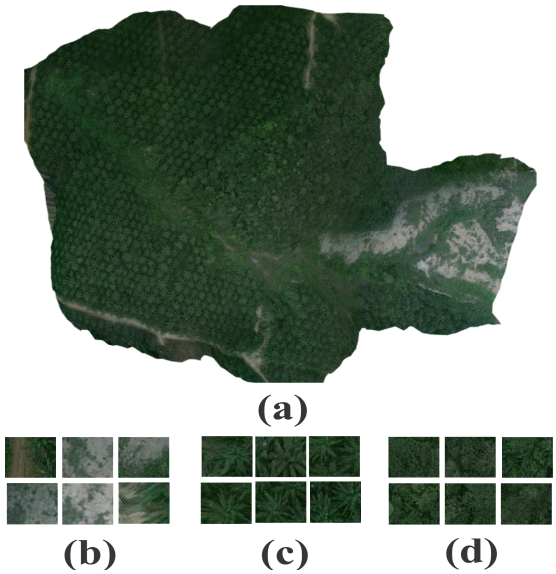


Fig. 1.(a) Dataset image (b) example of 100x100 non plantation area (c) example of 100x100 palm oil area (d) non palm oil area.

In this paper, T-SNE color and SFTA texture feature are used as a representative feature. These features are joined to obtain the final feature. For the next stage, features that will be utilized is the combined features F = {W, T}, which is a combination of a color feature W and texture T.

The Image was preprocessed using Histogram Equalization to improve its contrast and median filter to remove noise in the image. The image is then divided into a smaller-sized patch of 100x100 pixels. Color and texture feature extraction is carried out for each patch. Color feature is an essential feature that is often used as the main feature for image classification. Texture features are used because the texture displayed by palm oil plantation area is different from the other trees. The extraction method used in our research is the SFTA [13] to get the texture and t-SNE dimensionality reduction [14] to get a color feature in the three channels of RGB. Thus we obtained color features $W = \{w_1, w_2, w_3, ... w_n\}$ and texture features $T = \{t_1, t_2, t_3, ... t_n\}$, where n is the number of patches. The classification method used in this study is AdaBoost classifier [15].

B. Color Feature with T-SNE Dimensionality Reduction

Color feature extraction with T-SNE dimensionality reduction and texture features using SFTA is discussed in this section. Feature extraction is divided into two stages (see Fig. 2) to obtain the color and texture features.

Color feature is an essential feature that is often used as the main feature for image classification. The images which utilized in this study are the RGB image type. RGB image has three channels with color identity of red (R), green (G), and blue (B). For every images, there are R = { r_1, r_2, r_3, ... r_n }, G = { g_1, g_2, g_3, ... g_n } and B = { b_1, b_2, b_3, ... b_n }, where n is the number of patches.

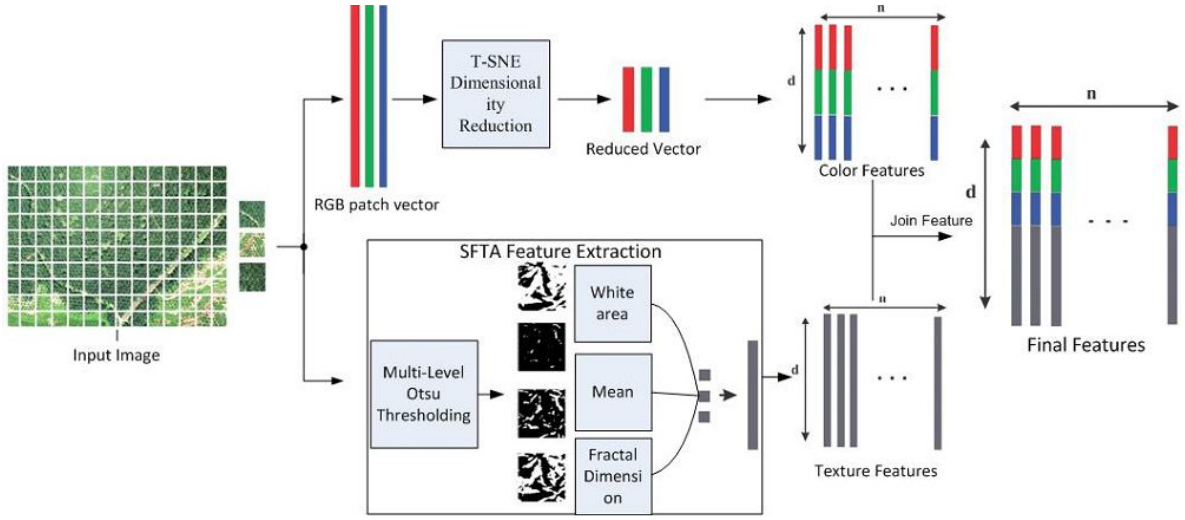


Fig. 2 Complete Process of feature extraction

Basically, T-Distributed Stochastic Neighbor Embedding (T-SNE) [14] minimize the difference between the two-dimensional distribution of data at high and low dimensional. For example, note the value of R = { r_1, r_2, r_3, ... r_n }, of the pixel values of the red entire high-dimensional patches, and d (r_i, r_j) measures the distance between the two points r_i and r_j. First of all, T-SNE determines joint probability p_{ij} which measures the similarity between points ri and r_j using probabilities using the following two functions:

$$p_{j|i} = \frac{\exp(-d(r_i, r_j)^2 / 2\sigma_i^2)}{\sum_{k \neq i} \exp(-d(r_i, r_k)^2 / 2\sigma_i^2)} \quad p_{i|i} = 0 \quad (1)$$

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2N} \quad (2)$$

where σ_i is a variance of the Gaussian kernel. In order to obtain the similarity in low dimensional, T-Student's kernel was used with a single degree of freedom instead of Gaussian Kernel, where y_i and y_j are two points from the different value of r and located in the lower dimension that has been embedded. The equation for calculating the similarity q_{ij} become:

$$q_{ij} = \frac{(1+\left\|y_i - y_j\right\|^2)^{-1}}{\sum_{k \neq l}(1+\left\|y_k - y_l\right\|^2)^{-1}} \quad q_{i|i} = 0 \quad (3)$$

Having obtained the two similarity values in high and low dimensions, then the location of y_i on low dimensional is updated by minimizing Kullback-Leibler divergence between the joint distribution in high-dimensional P and lower distribution Q using:

$$C(\varepsilon) = KL(P \,||\, Q) = \sum_{i \neq j} p_{ij} \log \frac{p_{ij}}{q_{ij}} \quad (4)$$

Because of the Kullback-Leiber divergence is not symmetric, the objective function will only focus on model p_{ij} (similar object) and the high value of q_{ij} (embedded neighbor point) and usually minimized by lowering the gradient:

$$\frac{\partial C}{\partial y_i} = 4\sum_{i \neq j}(p_{ij} - q_{ij}) q_{ij} Z(y_i - y_j) \quad (5)$$

where $Z = \sum_{k \neq l}(1+\left\|y_k + y_l\right\|^2)^{-1}$. To make it faster in finding the gradient $\partial C / \partial y_i$, this study used Barnes-Hut approximation [16]. The final value of the features is F_e={R_e,G_e,B_e}, where R_e, G_e, B_e are reduced vector of red, green, and blue channel respectively using T-SNE dimensionality reduction method. For the number of the dimension s in this study will be varied from 50 until 250 and we will use the configuration with the best result.

C. Texture Feature Extraction using Segmentation-based Fractal Texture Analysis

Texture feature is one of the popular characteristics, represent a form of a two-dimensional image of the image region. In this study we added this feature to distinguish the texture of the area of oil palm plantation and on palm oil tree, since the color in the image is similar.

Texture features are used because the texture displayed by palm oil plantation area is different from the other trees. To extract the texture features, the Segmentation-based Fractal Texture Analysis (SFTA) [13] feature extraction is used. The results of the Costa et al. study proved that SFTA feature extraction may work better than the Gray-Level Co-Occurrence Matrix (GLCM) [9] or Gabor [17] based on their recognition rate results and the speed of the classification process.

Tabel 1.Ten-Fold Cross Validation results using AdaBoost Classifier with nt=5 and s=50, where μ=mean

Iteration	10-fold cross validation results (%)										μ
100	86.11	85.51	86.80	86.01	86.89	86.11	86.01	86.01	86.20	86.11	86.18
200	85.41	85.51	90.97	85.31	86.89	85.41	86.01	86.01	86.89	86.11	86.45
300	91.67	85.51	90.97	86.01	97.24	93.75	93.00	86.01	86.89	86.11	89.64
400	92.67	93.10	91.27	86.01	96.55	93.75	95.10	95.80	90.34	86.11	92.07
500	93.47	93.10	92.56	86.01	96.55	93.75	95.10	95.80	90.34	96.52	93.32
600	93.47	93.10	93.27	94.40	96.55	93.75	95.10	95.80	90.34	96.52	94.23
700	94.59	93.10	94.77	94.40	96.55	93.75	95.10	95.80	90.34	96.52	94.49
800	96.63	93.10	96.28	94.40	96.55	94.44	95.10	95.80	90.34	96.52	94.85
900	96.65	93.10	96.27	94.40	96.55	96.77	95.10	95.80	90.34	96.52	**95.08**
1000	94.54	92.14	94.27	94.40	96.55	94.44	95.10	95.80	90.34	96.52	94.34

Texture feature extraction using SFTA consists of two main stages. The first stage is Two-Threshold Binary Decomposition (TTBD) to determine the threshold and change the input into several binary images. The second stage is SFTA extraction algorithm to process the binary image processing results of the first stage. The TTBD process takes a grayscale image I(x, y) along with a number of thresholds n_t which are set by the user as its input. Next, TTBD generates a number of binary images using two threshold values from T. To get the value of the threshold, Multilevel Otsu Thresholding [18] is used. Otsu multi-level algorithm works to find the threshold that minimizes the variance of the input. Recursively, Otsu algorithm is applied to each image until the desired binary images is obtained using the number of threshold n_t. One of them becomes the lower limit and the other become upper limit. The segmentation process become:

$$I_b(x,y) = \begin{cases} 1 \text{if } t_\ell < I(x,y) \le t_u \\ o, \text{otherwise} \end{cases} \quad (6)$$

where t_ℓ and t_u is the lower and upper limit threshold from n_t.

SFTA texture features consist of three elements. First is the number of white value from the whole image, second is the mean value, and the third is the fractal dimension. For the third feature, edge detection is needed using:

$$\Delta(x,y) = \begin{cases} 1 \text{ if } \exists (x',y') \in N_8[(x,y)]: \\ \quad I_b(x',y') = 0 \wedge \\ \quad I_b(x,y) = 1 \\ 0, otherwise\ . \end{cases} \quad (7)$$

where $N_8[(x,y)]$ are neighbor pixels that connected with (x,y). Therefore, $\Delta(x,y)$ will have the value of 1 if the (x,y) point is 1 and have at least one neighbors that have the value of 0. Otherrwise, all the value of the pixel will be set to 0.

D. Classification

Classification is the process of determining a class of an object based on the value of a given feature. There are so many techniques used for classification as Support Vector Machine, K-Nearest Neighbors, and Neural Network. However, in this study we used Adaboost [15] for the classification. The Adaboost classifier was proven to increase the value of "weak learner". The AdaBoost refers to a method for training the classifier by strengthening its learner value. Using increased learning value, every data will be separated easier than using original value. The AdaBoost method used in this paper is shown in Algorithm 1. After classifying the data, we evaluate our classification system based on the accuracy of predicted labels for the test files.

$$\frac{\text{number of correct classified (TP+TN)}}{\text{all samples}} \text{x100} \quad (8)$$

IV. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, palm oil plantation classification results will be shown. In this experiment, we used RGB image with resolution of 3600 x 4000 pixels taken from UAV with around 350 meters from the ground (see Figure 1(a)) and used 100x100 as the size of the patches (see Figure 1(b), 1(c), and 1(d)). The image was preprocessed using histogram equalization and median filter to improve contrast and remove noise. Then, we tried to find the best parameter in Texture and Color Feature. We varied the parameter of the number of threshold n in SFTA and number of reduced dimension k in color feature using T-SNE.

In the next step, we tried to classify all the patches. We used ten-fold cross-validation using AdaBoost classifier. A one-tenth from the complete data become the test data, and the rest become the training data. This data will be repeated ten times using the different training data until all the data are used. The mean of all the results will be used as the final recognition rate. This experiment used 127 patches area of a

non-plantation area, 638 patches of palm oil plantation area, and 181 patches area of a tree inside plantation area. We manually determine the ground truth as the basis for the performance measurement step.

In the first experiment, we varied the number of threshold in SFTA feature extraction from 5 until 25 in multiple of 5. This experiment was intended to show the relation between the number of threshold parameters (n) in SFTA with the classification performance of the method. Figure 3 shows the result of 10 fold cross validation results where accuracy is calculated to measure the performance of plantation classification using texture feature only. These features proved that for this dataset SFTA obtained the best result when n=20 and after that, there is a slight decline in the performance results. In this first experiment, we used 1000 as the number of iteration on AdaBoost classifier parameter.

The second experiment was intended to show the performance of the color feature in our dataset. We used the same parameter of the classifier with the previous experiment. The number of dimension in T-SNE dimensionality reduction become the varied parameter in this second experiment. We believed that the number dimension in dimensionality reduction become the main factor of the performance. Search for the correct number of dimension can improve the discrimination level of feature. This experiment tried to vary the number of dimension from 50 until 250 in the multiple of 50.

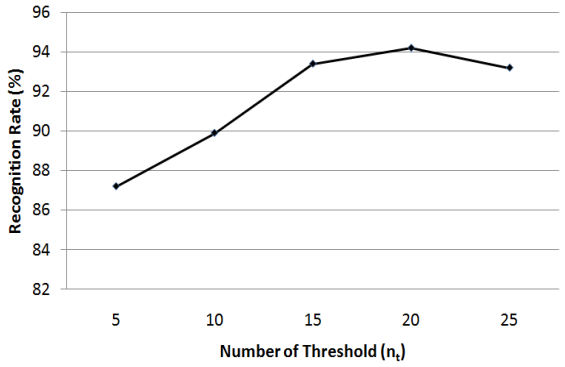


Fig. 3 Mean results of 10-fold Cross-validation with AdaBoost Classifier using Texture Feature.

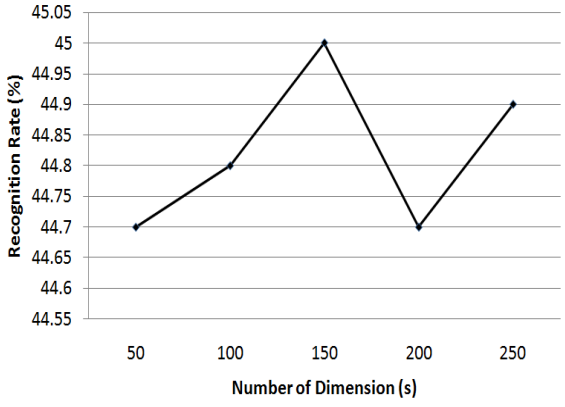


Fig. 4 Mean results of 10-fold Cross-validation with AdaBoost Classifier using Color feature.

From Fig. 4, we can see from the chart that the result of using color feature only in this dataset is not good. These results show that even though the number of thresholds has been varied, it cannot change the result. The classification rate using only color feature is around 45% in all different thresholds. This performance is caused by the color of the dataset is the same for the entire images, except for the non-plantation area. The system can work very well in separating the plantation area and the non-plantation area, but it does not work to classify the oil palm area and the forest inside the plantation area.

In order to improve the performance, joined feature is used in the system. In the third experiment, joined feature from texture and color feature are used together. We used the best result in each feature as a parameter, i.e. n=20 for the number of threshold in SFTA and 150 as the dimension in T-SNE color feature. We varied the iteration number in AdaBoost parameter from 100 until 1000. Table 1 shows all the classification results with 10-fold cross-validation and the mean in each category. Overall, these results prove that this system can map the area correctly with the mean accuracy of 95.08% in 900 iterations (see Fig. 3). These Table also prove that joined feature gives better performance than Color or Texture feature only. However, there are some misclassified patches, especially palm oil and forest area. This case occurred because of the forest area both color and texture are similar. For future work, the system needs to be improved to perform better by adding more training data such as variation of an unhealthy plant.

CONCLUSION

In this paper, we proposed a palm oil area classification using join T-SNE color and SFTA fractal analysis texture feature. The first experiment shows texture feature work quite well with 94.23% recognition rate. The second experiment shows that the color feature alone doesn't separate palm oil trees and other trees well. However, color feature separates plantation and non-plantation area quite well. In the third experiment, texture and color feature are joined into one feature and proved that this feature could improve the accuracy of the classification to 95.08%. In the future, we will classify the palm oil plantation area into a number of categories based on the palm oil condition, such as ages, health etc. We also hope that this system can be extended to map different plantation area e.g. tea, coffee, and paddy field.

ACKNOWLEDGMENT

This research was partly funded by the Directorate General of Higher Education, Ministry of Research and Higher Education of the Republic of Indonesia. The researchers also would like to thank PT. Pina Gautama, Palembang, South Sumatra for their support on this research.

REFERENCES

[1] J. J. Mitchell, N. F. Glenn, M. O. Anderson, R. C. Hruska, and A. H. Charlie, "Unmanned aerial vehicle (UAV) hyperspectral remote sensing for dryland vegetation monitoring hyperspectral image and signal sensing," 4th Work. Hyperspectral Image Signal Process., no. Ua V, 2012.

[2] J. F. Nel, "Post-processing of UAV-captured images for enhanced mapping by image stitching," in 2015 IEEE 5th International Conference on Consumer Electronics - Berlin (ICCE-Berlin), 2015, pp. 500–504.

[3] K. N. Tahar, A. Ahmad, W. A. A. W. M. Akib, and W. M. N. W. Mohd, "Aerial mapping using autonomous fixed-wing unmanned aerial vehicle," 2012 IEEE 8th Int. Colloq. Signal Process. its Appl., pp. 164–168, 2012.

[4] R. L. Thomas Arnold, Martin De Biasio, Andreas Fritz, "UAV-based measurement of vegetation indices for environmental monitoring," 2013, p. 3, 2013.

[5] M. Ghazal, Y. Al Khalil, and H. Hajjdiab, "UAV-based remote sensing for vegetation cover estimation using NDVI imagery and level sets method," in 2015 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), 2015, pp. 332–337.

[6] T. Chou, M. Yeh, Y. Chen, and Y. Chen, "Disaster monitoring and management by the unmanned aerial vehicle technology," ISPRS TC VII Symp., vol. XXXVIII, pp. 137–142, 2010.

[7] A. Harjoko, A. Prahara, S. Hartati, and B. A. A. Sumbodo, "Tea plantation evaluation using support vector machine on UAV aerial image," in Proceedings of the Universal Academic Cluster International Autumn Conferences, 2015, pp. 41–51.

[8] A. L. Sumalan, D. A. N. Popescu, and L. Ichim, "Flooded areas detection based on LBP from UAV images," in Recent Advances on Systems, Signals, Control, Communications and Computers, 2015, pp. 186–191.

[9] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," IEEE Trans. Syst. Man. Cybern., vol. 3, no. 6, pp. 610–621, 1973.

[10] M.-K. Hu, "Visual pattern recognition by moment invariants," IRE Trans. Inf. Theory, vol. 8, pp. 179–187, 1962.

[11] L. Van Der Maaten, "Accelerating t-SNE using tree-based algorithms," J. Mach. Learn. Res., vol. 15, pp. 1–21, 2014.

[12] B. Sujatha and T. Santhanam, "A novel approach integrating geometic and gabor wavelet approaches to improvise visual lipreading," Int. J. Soft Comput., 2010.

[13] A. F. Costa, G. Humpire-Mamani, and A. J. M. Traina, "An efficient algorithm for fractal analysis of textures," in Brazilian Symposium of Computer Graphic and Image Processing, 2012, pp. 39–46.

[14] L. Van Der Maaten, G. Hinton, and H. G. van der Maaten L, "Visualizing data using t-SNE," J. Mach. Learn. Res., vol. 9, pp. 2579–2605, 2008.

[15] Y. Freund and R. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," Comput. Learn. theory, vol. 55, pp. 119–139, 1995.

[16] W. Dehnen, "A hierarchical (N) force calculation algorithm," J. Comput. Phys., vol. 179, no. 1, pp. 27–42, 2002.

[17] B. S. Manjunath and W. Y. Ma, "Texture features for browsing and retrieval of image data," IEEE Trans. Pattern Anal. Mach. Intell., vol. 18, no. 8, pp. 837–842, 1996.

[18] P. S. Liao, T. S. Chen, and P. C. Chung, "A fast algorithm for multilevel thresholding," J. Inf. Sci. Eng., vol. 17, no. 5, pp. 713–727, 2001.

★★★