A comparative study of rice variety classification based on deep learning and hand-crafted features

Duc Phan Van Hoai[1], **Thongchai Surinwarangkoon**[2], **Vinh Truong Hoang**[3], **Huu-Thanh Duong**[4], and **Kittikhun Meethongjan**[5]

ABSTRACT

Rice is vital to people all around the world. The demand for an efficient method in rice seed variety classification is one of the most essential tasks for quality inspection. Currently, this task is done by technicians based on experience by investigating the similarity of colour, shape and texture of rice. Therefore, we propose to find an appropriate process to develop an automation system for rice recognition. In this paper, several hand-crafted descriptors and Convolutional Neural Networks (CNN) methods are evaluated and compared. The experiment is simulated on the VNRICE dataset on which our method shows a significant result. The highest accuracy obtained is 99.04% by using DenNet21 framework.

Keywords: Color Texture, Texture Classification, Deep Learning, Hand-crafted Feature, LBP, HOG, SIFT, GIST, Rice Seed Classification

1. INTRODUCTION

Rice is a staple and principal source of food consumed by nearly a half of the world's population. Products made with rice are essential to billions of people from all over the world according to the Food and Agriculture Census of the United Nations (FAO), the total rice production in the world increased from 570 million tonnes in 2002 to 720 million tons in 2012. Practically, it is necessary to prepare well all the steps to obtain high yielding rice in cultivation. Since some varieties are appropriate for specific applications, suitable selection is an important stage in the processing of rice grains. One of the most significant factors is the quality of rice seeds (i.e seeds of certain varieties should not be mixed with other seeds). Moreover, the quality of rice depends mainly on genetic features. In the growing and trading process, various varieties of rice can be combined and mixed, which will affect the quantity and quality of a rice crop. Therefore, an inspection process is necessary to make sure that all seeds belong to a certain rice variety. To recognize the unwanted seeds, it is necessary to ensure the purity of rice seeds of a certain variety of rice. Rice is Vietnam's most cultivated agricultural plant. Currently, the identification process of unwanted seeds is performed manually in Vietnamese companies by visual inspection by technicians with some expertise. This laborious process, which is both tedious and slow, can lead to a deterioration of purity in the seeds, thereby lowering the value of the seed. To solve this problem, an automated process based on a computer vision system is needed.

Computer vision is a rapidly growing scientific field. The advances in this field have brought our daily lives several benefits. Computer vision systems have been designed and used to solve problems in real life, namely detection of diseases in humans and plants, automatic manufacturing inspection, biometric recognition, etc. There are several publications about agriculture applying computer vision to defect analysis, automatic characterization and quality assessment of fruit, vegetables, and grain [1], [2], [3]. In the past, hand-crafted descriptors have been designed to extract appropriate features in computer vision. Identification of grain seed varieties is one of the major agricultural tasks. Several approaches to solving this problem have been proposed to identify the rice seed variety by the advancement of technology and engineering. The analysis and the detection fully connected. Several efforts have been made to improve the performance, including proposing a new regularization techniques [28], [29], new optimizations [30], [31], and designing new architectures [32], [33], [34], [35], [36], [37], [38], [39]. Kamilaris and Prenafeta-Boldú [40] give a complete survey of 40 works that apply deep learning in the domain of agriculture such as soil and vegetation analysis, crop monitoring, weed detection, green house monitoring, leaf disease detec-

Manuscript received on July 19, 2019 ; revised on October 12, 2019.
Final manuscript received on January 24, 2020.
[1,3,4] The authors are with Faculty of Computer Science, Ho Chi Minh City Open University, Vietnam., E-mail: 1551010028duc@ou.edu.vn, vinh.th@ou.edu.vn and thanh.dh@ou.edu.vn
[2] The author is with Faculty of Management Science, Suan Sunandha Rajabhat University, Thailand., E-mail: thongchai.su@ssru.ac.th
[5] The author is with Faculty of Science and Technology, Suan Sunandha Rajabhat University, Thailand., E-mail: kittikhun.me@ssru.ac.th
[2,5]Corresponding author: vinh.th@ou.edu.vn and kittikhun.me@ssru.ac.th
DOI 10.37936/ecti-cit.2020141.204170

tion and classification, and identification of weeds.

More recently, a hybrid method which combines CNN and hand-crafted features to boost the performance of learning systems has been proposed. For example, Local Binary Pattern Network (LBPNet) is proposed to efficiently extract and compare high-level over-complete features in a multilayer hierarchy. This method reflects the same topology of CNN architectures [41]. In [42], [43], the authors propose to feed CNN by the new encoded image via Neighbor-Center Difference Image (NCDI) for texture classification. Bello-Cerezo et al. [44] present a deep comparative evaluation of hand-crafted and CNN-based features for colour texture classification. The experimental results confirmed that hand-crafted descriptors give better performance than CNN-based features when there was little intra-class variability.

This paper shows a comparative study of the hand-crafted features extracted from well-known local image descriptors and the deep learning approach for rice seed classification. The remainder of this paper is structured as follows. Section 2 briefly introduces several local image descriptors in the literature. Next, the Convolutional Neural Network and its recent proposed models are introduced in section 3. Then, section 4 presents the dataset, experimental setup and comparative results. Finally, the conclusion is discussed in section 5.

2. HAND-CRAFTED FEATURE

Three well-known local image descriptors including Histogram of Oriented Gradients (HOG) [45], Scale Invariant Feature Transform (SIFT) [46], GIST [47], and Local Binary Pattern (LBP) [48] will be briefly discussed in the following subsections. The reason that we choose and apply these descriptors is because they are the most successful approaches for characterizing texture and objects.

2.1 Local Binary Pattern

LBP was first introduced by Ojala et al. [48]. It is a simple yet powerful local descriptor and captures the local structure and contrast information. Basic LBP value is extracted by thresholding circular neighbouring pixel values base on the central pixel value. Then the binary results are evaluated by a factor of 2^i and summed to obtain the LBP code. Mathematically, the computation of the LBP value at the central pixel and P neighbouring pixels on a circle of radius R is listed as:

$$\mathrm{LBP}_{P,R} = \sum\nolimits_{i=0}^{P-1} \xi(g_i - g_c) \times 2^i \tag{1}$$

where g_c is the central pixel value, $\{g_i\}_{i=0}^{P-1}$ are the grayscale values of its P neighboring pixels, and ξ is defined as:

$$\xi(u) = \begin{cases} 1 & \text{if } u \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

LBP is an efficient operator with high discriminative capabilities. However, it is sensitive to rotated and noisy images [49], and dimension of feature space is increased exponentially with the number of neighbors considered. From Ojala's work, a huge number of variants of the LBP operator have been introduced to enhance its robustness to noise and increase its discriminative power and applicability to solve different kinds of pattern recognition problems. Liu et al. [50] present a very detailed review of the LBP variant based on a gray scale image. We can mention several important and efficient LBP variants such as LBP uniform, Dominant Rotated LBP [51], Local Ternary Pattern (LTP) [49], Complete LBP [52], Median LBP [53], and Extend LBP [54].

2.2 Histograms of Oriented Gradient

The HOG descriptor counts the occurrences of gradient magnitude and orientation in localized portions of an image. It is commonly used for identifying and recognizing objects. HOG was presented by Dalal and Triggs [45] who first used it for detecting objects. The occurrences of gradient orientation based on the gradient angle and the gradient magnitude of local patches of an image are computed in order to extract HOG features. The authors in [45] calculate gradient angle and magnitude at each pixel in an 8×8-pixel patch. Next, 64 gradient feature vectors are divided into 9 angular bins 0-180°(20°each). Then, on each patch, the histogram of orientation is constructed by collecting the magnitudes of gradient. Finally, histograms of orientation from each patch are normalized and fused. The gradient magnitude G and angle L at each position (x, y) from an image I are computed as follows:

$$\Delta_x = |I(x-1, y) - I(x+1, y)| \tag{3}$$

$$\Delta_y = |I(x, y-1) - I(x, y+1)| \tag{4}$$

$$G(x, y) = \sqrt{\Delta_x^2 + \Delta_y^2} \tag{5}$$

$$L(x, y) = \tan^{-1}\left(\frac{\Delta_x}{\Delta_y}\right) \tag{6}$$

HOG features are used effectively in various computer vision tasks such as biometric recognition [55], [56], traffic recognition [57], hand gesture classification [58] and human detection [59].

2.3 GIST

GIST is an image descriptor proposed by Oliva and Torralba [47], it was firstly applied for scene recognition. GIST extracts gradient information from different parts of an image by combining the input image with several filters. In order to obtain GIST feature, the image is convolved with several Gabor filter. Then, Gabor response maps are divided into small regions. Next, the GIST features are obtained by averaging the value from each region of each Gabor response map. In the paper [47], Oliva and Torralba use 32 Gabor filters at four scales with eight different orientations. They divide each Gabor response map into 4×4. Therefore, $32 \times 16 = 512$ GIST features are extracted from each image. GIST has been successfully applied for indoor/outdoor scene recognition [60][61], traffic scene classification [62], and face recognition [63][64].

2.4 Scale Invariant Feature Transform

The SIFT descriptor was introduced by Lowe [46]. It was firstly applied to object recognition. A SIFT descriptor extracts key points of objects from a set of images and stores them in a feature database. Each key point of SIFT specifies the location, scale, and orientation. SIFT is robust against image scaling and rotation, and partially invariant to changes in illumination and 3D camera viewpoints. SIFT extraction process follows these steps. First, extrema points are detected from several Difference of Gaussian (DoG) of several scales. Then, key points are selected from candidate points based on their location and scale. Next, the dominant orientation of each key point is assigned to achieve rotation invariance. Finally, SIFT features are generated by concatenating several histograms of orientation, these histograms are extracted around the key point location with respect to its dominant orientation. SIFT has been used in many applications, such as object detection and tracking [65], [66], facial recognition [67], [68], and hand gesture recognition [69].

3. CONVOLUTION NEURAL NETWORKS

In deep learning, a CNN is a class of deep neural networks which is the most common method applied in computer vision applications. In this section, we briefly review various architectures, including AlexNet, VGG, Inception, ResNet, DenseNet, MobileNet, and NASNet [33], [34], [35], [36], [37], [38].

3.1 AlexNet

AlexNet was proposed by Krizhevsky et al. [70]. It was the winner of the ImageNet Challenge in 2012 where it outperforms all of the other approaches by decreasing the top-5 error rate from 26% to 15.3%. Since AlexNet, all of the winners of later ImageNet Challenges have been CNN models. AlexNet consists of one $11{\times}11$, one $5{\times}5$, three $3{\times}3$ convolutional layers, three max pooling layers, and three fully connected layers. It is the first mode whichl uses Rectified Linear Unit (ReLU) instead of Tanh or Sigmoid function. The advantage of the ReLU is that it minimizes the vanishing gradient problem. Therefore, the model can be trained better and faster.

3.2 VGG

VGG was proposed by Simonyan and Zisserman in 2014. It achieved a top-5 accuracy of 92.3% on ImageNet Challenge [32]. VGG is an architecture based on AlexNet. To design VGG to be a deeper architecture, they only use 3×3 convolutional layers. By stacking several 3×3 convolutional layers, they acquire the same effective receptive field as a larger kernel-size convolutional layer. By doing so, the model is deeper and more complex. There are several VGG variants. Among these extensions, VGG16 and VGG19 are the most widely used.

3.3 Inception/GoogLeNet

Inception was proposed by Szegedy et al. [33]. It was the winner of ImageNet Challenge in 2014 which gave the top-5 accuracy of 93.3%. In this model, several inception modules are applied. Each inception module consists of convolutions of different sizes to capture details at different scales. In order to reduce the computation requirements of the inception module, they use 1×1 convolutional layer to reduce the depth of the input. Another salient point in this paper is that they use global average pooling to replace two fully connected layers after the last convolutional layer. By doing this, the total number of parameters of the model is greatly reduces. Inception has 12 times fewer parameters than AlexNet, and is much faster than VGG. Various enhanced Inception models have been proposed, including InceptionV2, InceptionV3, InceptionV4, Xception, Inception-ResNet, and Inception-ResNetV2.

3.4 ResNet

ResNet is proposed by He et al. [35] It it the winner of ImageNet Challenge in 2016 achieving a top-5 accuracy of 95.5%. As can be observed so far, increasing the depth of the model should increase the accuracy. However, when the model is deeper, the problem of vanishing gradient occurs. To tackle this problem, they devised a residual module which consists of two convolutional layers and a “skip connection”. The “skip connection” is created by adding the input to the output of the second convolutional layer. Therefore, the gradient can flow through the add gate, thus minimizing the vanishing gradient problem. They can build very deep models, so the computation requirement is very high. To design very deep models, they add a $1{\times}1$ convolutional layer to the beginning and the end of each residual module to reduce the depth of the input while not decreasing the performance of the model very much.

3.5 DenseNet

DenseNet is proposed by Huang et al. [37]. It is awarded the best paper in the Computer Vision and Pattern Recognition (CVPR 2017) conference. Inspired by the "skip connection" in residual module of ResNet model, they come up with dense block. Each dense block consists of several convolutional layers, where each convolutional layer connects to all subsequent convolutional layers. Instead of adding, they use concatenation of previous feature maps, which allows all subsequent layers can easily access the information from preceding feature maps.

Moreover, it allows the gradient to flow more easily through the model. To improve the computational efficiency of the model, they add a 1×1 convolution layer before each 3×3 convolution layers, furthermore, they add transition layers after each dense block to reduce the number of feature maps. DenseNet achieved state-of-the-art results on several datasets. On ImageNet, it outperforms ResNet when using 10-crop testing.

3.6 MobileNet

In 2017, MobileNet was proposed by Howard et al. [71]. This model was designed to be as small as possible, which is useful for mobile and embedded vision applications. To achieve that goal, they devised depth-wise separable convolution, which is a depth-wise convolution followed by a pointwise convolution. Depth-wise convolution applies a single filter to each input channel. Then the pointwise convolution combines the outputs. By doing so, they drastically reduce the number of parameters and computation time. To further reduce the size of the model, they introduce a width multiplier to control the number of input channels and a resolution multiplier to control the input resolution.

3.7 NasNet

In 2017, Zoph et al. [39] from AutoML of Google introduced a method to search for the best neural network architecture. They propose searching for building blocks on a small dataset and then to sample these building blocks to create end-to-end architectures. In the paper, they also introduced Scheduled Drop Path, which significantly improves the final performance of NASNets. Their proposed approach has achieved state-of-the-art results on ImageNet and COCO datasets.

4. EXPERIMENTAL RESULTS

4.1 Dataset and experimental setup

In order to evaluate and compare the hand-crafted features and deep learning, we propose to use the Vietnam rice seed varieties (VNRICE). This dataset consists of images of six common Vietnamese rice varieties (BC-15, Huong Thom-1, Nep-87, Q-5, Thien Uu-8, and Xi-23) from the north region. All images are acquired by a CMOS image sensor colour camera. Figure 1 illustrates example images from this dataset. Each column represents one class of rice seed variety.

All hand-crafted feature extraction and evaluation processes were implemented using Matlab version 2015a and conducted on Google Cloud Computing by using the Keras framework. We built the system with 32 CPUs at 2.5 GHz, 32 GBs of memory 1 NVIDIA Tesla P100 GPU. In order to build the training and testing set for the dataset, all images were resized to 224×224, then 50% of the images from each of the six rice varieties were randomly selected to create the training set, the remainder of the images from each rice variety were used as the testing set. Characteristic images of this dataset are illustrated in table 1. The first column indicates the name of rice seed. The second and third columns give the number of images in the training and testing sets.

4.2 Experimental Results

LBP and its variants are extracted on each channel of an RGB image by using the central pixel and 8 neighbouring pixels on a circle of radius 1. SIFT features are extracted on training set with VLFeat 0.9.21, then K-means ($K = 1500$) is used to build Bag of Words from all SIFT features from the training set. After that, a histogram of each image is obtained by counting SIFT features that are near to each of the K centroids. GIST and HOG features are extracted with the same configuration as used the original papers. All hand-crafted features are then evaluated by using one-vs-one SVMs.

Table 1: *Summary of the rice seed database.*

Name	Training set count	Testing set count	Total image count
BC-15	917	917	1,834
Huong Thom 1	1,048	1,048	2,096
Nep-87	700	699	1,399
Q-5	962	962	1,924
Thien Uu-8	513	513	1,026
Xi-23	1,115	1,114	2,229

All CNN models use the ImageNet pre-trained weight provided by the Keras framework. These pre-trained models are then fine-tuned on the training set of the dataset with the same hyper-parameter. The batch size is set to 64. An Adam optimizer [30] is used with the initial learning rate of 10^{-4}. Finally, the learning rate is decreased by a factor of 10 when the validation set accuracy stops improving.

Table 2 has shown that the deep learning approaches outperform the others with a large margin of more than 10%. The highest accuracy achieved by hand-crafted descriptors in the experiment is

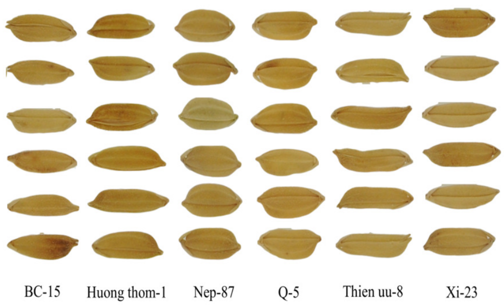

Fig.1: *Example images from six rice seed varieties.*

84.10%, which was obtained using the HOG descriptor. The best result of 99.05%, on the rice seed dataset achieved in the experiment is 99.05%, which is obtains by DenseNet with 121 layers.

Discussion on hand-crafted descriptors: LBP and its variants are mainly applied for face recognition and texture classification. Applying it for the rice seed classification task gives poor accuracy. The highest accuracy is 67.60% obtained by LBP Uniform. The SIFT descriptor was applied for several object classification tasks successfully. However, the accuracy obtained on the rice seed classification is only 69.31%. The GIST only captures global features from different regions of the image to apply to the scene classification task, but it has achieved a decent accuracy of 78.83%. The HOG descriptor is widely used for object classification task. It has been shown to provide a promising result on the rice seed dataset by achieving an accuracy of 84.10%.

Discussion on CNN models: All CNN models provided very good results on the rice seed classification task. Two of the best results are 98.84% and 99.05% from the ResNet 50 layers and DenseNet 121 layers respectively. Both models have several skip connections to let the gradient flow more strongly to the earlier layers, which allows the models to perform well on small datasets. MobileNet is lightweight, has very few parameters and requires very few computation resources. However, it achieved a reasonable accuracy of 97.66%, which is better than either the VGG16 or VGG19 models. The hybrid approach which combines hand-crafted descriptors and the CNN model does not perform as well as VGG. This method aims to change input of the CNN model and improve its efficiency for colour texture images. Since this set of rice seed images contain an object and are totally different, it does not outperform CNN-based methods. Similarly, we can see in the conclusion [44] that hand-crafted descriptors outperformed CNN-based features for colour texture classification. In the case of rice seed images, we found the opposite was true based on our experimental results.

Table 3 shows the number of parameters and computation time of each model. The computation time of each model was obtained by training the model on the training set one time using the configuration as shown in section 4.1. From Table 3, we can observe that DenseNet with 121 layers requires less memory and computational resources with a run time only 36 seconds. NasNetLarge takes more time to compute with more than 84 million parameters to adjust. We found the MobileNet and MobileNetV2 are more interesting since they can be employed for mobile applications with fewer parameters than other methods.

Table 2: *The classification results of hand-crafted features and deep learning on the VNRICE dataset. The value in bracket demonstrates the highest accuracy by using hand-crafted features while the value in brackets shows the best accuracy obtained among all methods.*

No	**Model name/ Descriptor**	**Accuracy** (%)
1	Four-Patch LBP [72]	49.59
2	LCP [73]	54.51
3	LBP [74]	54.88
4	LTP [49]	55.09
5	Color LBP [75]	58.16
6	LPQ [76]	58.27
7	Three-Patch LBP [72]	58.40
8	OCLBP [12]	58.44
9	LBP Neighbor intensity [77]	59.15
10	Median LBP [53]	61.83
11	LBP Num [78]	62.88
12	LBP Non-uniform [79]	63.43
13	LBP Uniform [80]	67.60
14	SIFT [46]	69.31
15	8-NCDI Histogram & LBP 8-NCDI [42]	78.70
16	GIST [47]	78.83
17	HOG [37]	{84.10}
18	NCDI-VGG [43]	92.65
19	NASNetMobile [39]	94.40
20	MobileNetV2 [81]	95.28
21	VGG19 [32]	96.88
22	VGG16 [32]	96.94
23	MobileNet [71]	97.66
24	Xception [38]	97.90
25	Inception-ResNet v2 [36]	97.93
26	InceptionV3 [33]	98.10
27	NasNetLarge	98.12
28	ResNet50 [35]	98.84
29	DenseNet121 [37]	99.04

5. CONCLUSION

In this study, we evaluated the performance of several hand-crafted descriptors and CNN models on the rice variety classification task. The experimental results show that deep learning approaches clearly give a higher accuracy. DenseNet with 121 layers achieved the best accuracy with 99.04%. LBP and its variants provided poor results. The highest accuracy was only 67.60% obtained by LBP Uniform. HOG is the most successful hand-crafted descriptor approach in the experiment and achieved an accuracy result of 84.10%.

The experiment results show that it is possible to develop an automation system for rice seed classification tasks. The purposes of this work were to evalu-

Table 3: *Number of parameters of each model and computation time to train each model per epoch.*

No	Model name	Computation time (seconds)	Number of parameters
1	InceptionV3	29	21,780,646
2	MobileNet	31	3,213,126
3	MobileNetV2	33	2,231,558
4	DenseNet121	36	6,960,006
5	ResNet50	37	23,546,886
6	VGG16	40	134,285,126
7	VGG19	46	139,594,822
8	NASNet Mobile	54	4,239,320
9	Inception-ResNetV2	57	54,285,414
10	Xception	64	20,819,246
11	NasNetLarge	188	84,744,348

ate the deep learning model for rice image recognition and implement a mobile application for end user. The future of this work is to design a suitable CNN model for this kind of rice seed image.

ACKNOWLEDGEMENT

This work was supported, in part, by Suan Sunandha Rajabhat University, Thailand and Ho Chi Minh City Open University under Grant No E.2019.06.1. The authors also would like to thank Dr. Hong Phan from the Vietnam National University of Agriculture for providing the Rice Seed Image dataset.

References

[1] J. F. S. Gomes and F. R. Leta, "Applications of computer vision techniques in the agriculture and food industry: A review," *European Food Research and Technologyŷ*, vol.235, pp.9891000, 2014.

[2] D. I. Patrício and R. Rieder, "Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review," *Computers and Electronics in Agriculture*, vol.153, pp.6981, 2018.

[3] A. C. Tyagi, "Towards a second green revolution," *Irrigation and Drainage*, vol.65, pp.388389, 2016.

[4] A. Humeau-Heurtier, "Texture feature extraction methods: A survey," in*IEEE Access*, vol.7, pp.89759000, 2019.

[5] A. Sinha, S. Banerji, and C. Liu, "Novel color Gabor-LBP-PHOG (GLP) descriptors for object and scene image classification," in *Proceedings of the Eighth Indian Conference on Computer Vision, Graphics and Image Processing, ACM*, pp.1-8, 2012.

[6] P. Vácha, M. Haindl, and T. Suk, "Colour and rotation invariant textural features based on Markov random fields," *Pattern Recognition Letters*, vol.32, pp.771779, April 2011.

[7] A. Bosch, A. Zisserman, and X. Mũnoz, "Scene classification via pLSA," in *European conference on computer vision*, Springer, pp.517530, 2006.

[8] A. Sinha, S. Banerji, and C. Liu, "New color GPHOG descriptors for object and scene image classification," *Machine Vision and Applications*, vol.25, pp.361375, 2014.

[9] R. Khan, J. van de Weijer, F. S. Khan, D. Muselet, C. Ducottet, and C. Barat, "Discriminative color descriptors," in *Proceedings of 23rd IEEE International Conference on Computer Vision and Pattern Recognition*, Portland, pp.28662873, 2013.

[10] K. Meethongjan, M. Dzuikifli, P. K. Ree, and M. Y. Nam, "Fusion affine moment invariants and wavelet packet features selection for face verification.," *journal of theoretical & applied information technology*, vol. 64, pp.606-615, 2014.

[11] T. Mäenpää, M. Pietikäainen, and J. Viertola, "Separating color and pattern information for color texture discrimination," In *Proceedings of the 16th IEEE International Conference on Pattern Recognition*, vol. 1, pp.668671, 2002.

[12] T. Mäenpää and M. Pietikäinen, "Classification with color and texture: jointly or separately?," *Pattern Recognition*, vol.37, pp.16291640, 2004.

[13] B. Lurstwut and C. Pornpanomchai, "Image analysis based on color, shape and texture for rice seed (Oryza sativa L.) germination evaluation," *Agriculture and Natural Resources*, vol.51, pp.383389, 2017.

[14] X. Chen, Y. Xun, W. Li, and J. Zhang, "Combining discriminant analysis and neural networks for corn variety identification," *Computers and Electronics in Agriculture*, vol.71, pp.S48-S53, 2010.

[15] H.K. Mebatsion, J. Paliwal, and D.S. Jayas, "Automatic classification of non-touching cereal grains in digital images using limited morphological and color features," *Computers and Electronics in Agriculture*, vol.90, pp.99105, 2013.

[16] P. M. Szczypínski, A. Klepaczko, and P. Zapotoczny, "Identifying barley varieties by computer vision," *Computers and Electronics in Agriculture*, vol.110, pp.18, 2015.

[17] A. A. Chaugule and S. N. Mali, "Identification of paddy varieties based on novel seed angle features," *Computers and Electronics in Agriculture*, vol.123, pp.415422, 2016.

[18] T. Kuo, C. Chung, S. Chen, H. Lin, and Y. Kuo, "Identifying rice grains using image analysis and sparse representation-based classification," *Computers and Electronics in Agriculture*, vol.127, pp.716-725, 2016.

[19] H. li, Y. Qian, P. Cao, W. Yin, F. Dai, F. Hu, and Z. Yan, "Calculation method of surface shape feature of rice seed based on point

cloud," *Computers and Electronics in Agriculture*, vol.142, pp.416-423, 2017.

[20] F. Kurtulmuş and H. Ünal, "Discriminating rapeseed varieties using computer vision and machine learning," *Expert Systems with Applications*, vol.42, pp.1880-1891, 2015.

[21] H. Duong and V. T. Hoang, "Dimensionality reduction based on feature selection for rice varieties recognition," in *2019 4th International Conference on Information Technology (InCIT)*, pp.199202, 2019.

[22] R. Choudhary, J. Paliwal and D.S. Jayas, "Classification of cereal grains using wavelet, morphological, colour, and textural features of non-touching kernel images," *Biosystems Engineering*, vol.99, pp.330337, 2008.

[23] A. Douik and M. Abdellaoui, "Cereal grain classification by optimal features and intelligent classifiers," *International Journal of Computers, Communications & Control (IJCCC)*, vol.5, pp.506516, 2010.

[24] S. J. M. Rad, F. A. Tab and K. Mollazade, "Classification of rice varieties using optimal color and texture features and BP neural networks," in *2011 7th Iranian Conference on Machine Vision and Image Processing*, pp.15, 2011.

[25] P. T. T. Hong, T. T. T. Hai, L. T. Lan, V. T. Hoang, V. Hai and T. T. Nguyen, "Comparative study on vision-based rice seed varieties identification," in *2015 Seventh International Conference on Knowledge and Systems Engineering (KSE)*, Ho Chi Minh City, pp.377382, 2015.

[26] H. Vu, V. N. Duong and T. T. Nguyen, "Inspecting rice seed species purity on a large dataset using geometrical and morphological features," in *Proceedings of the Ninth International Symposium on Information and Communication Technology - SoICT 2018*, ACM Press, pp.321328, 2018.

[27] H. Zareiforoush, S. Minaei, M. R. Alizadeh, and A. Banakar, "Qualitative classification of milled rice grains using computer vision and metaheuristic techniques," *Journal of Food Science and Technology*, vol.53, pp.118131, 2016.

[28] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol.15, pp.19291958, 2014.

[29] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32Nd International Conference on International Conference on Machine Learning, ICML'15*, JMLR.org, vol.37, pp.448456, 2015.

[30] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," *International Conference on Learning Representations*, pp.15, 2014.

[31] L. Bottou, "Large-scale machine learning with stochastic gradient descent," in *Yves Lechevallier and Gilbert Saporta, editors, Proceedings of COMPSTAT2010*, Heidelberg, pp.177186, 2010.

[32] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv 1409.1556*, pp. 0106, 2014.

[33] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," *arXiv:1409.4842*, pp. 19, 2015.

[34] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 28182826, 2016.

[35] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.770778, 2016.

[36] Christian Szegedy, Sergey Ioffe, and Vincent Vanhoucke, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *AAAI*, 2016.

[37] G.Huang, Z. Liu, L. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.22612269, 2017.

[38] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.18001807, 2017.

[39] B. Zoph, V. Vasudevan, J. Shlens, and Q. V Le, "Learning transferable architectures for scalable image recognition," *arXiv:1707.07012* , pp. 86978710, 2018.

[40] A. Kamilaris and F. X Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Computers and Electronics in Agriculture*, vol.147, pp.7090, 2018.

[41] M. Xi, L. Chen, D. Polajnar, and W. Tong, "Local binary pattern network: A deep learning approach for face recognition," in *Image Processing (ICIP)*, pp.3224-3228, 2016.

[42] D. P. H. Hoai and V. T. Hoang, "Features based on enhanced neighbour center different image for color texture classification," in *2019 International Conference on Multimedia Analysis and Pattern Recognition (MAPR)*, pp.1-6, 2019.

[43] D. P. H. Hoai and V. T. Hoang, "Lbp-based edge information for color texture classification," *Industrial Networks and Intelligent Systems*, pp.232-239, 2019.

[44] R. Bello-Cerezo, F. Bianconi, F. Di Maria, P. Napoletano and F. Smeraldi, "Comparative evaluation of hand-crafted image descriptors vs. off-the-shelf CNN-based features for colour texture classification under ideal and realistic conditions," *Applied Sciences*, vol.9, pp.738, 2019.

[45] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol.1, pp.886893, 2005.

[46] D. G. Lowe, "Distinctive image features from scale-invariant key-points," *International Journal of Computer Vision*, vol.60, pp.91110, 2004.

[47] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *International Journal of Computer Vision*, vol.45, pp.145-175, 2001.

[48] T. Ojala, T. Maenpaa, M. Pietikainen, J. Viertola, J. Kyllonen, and S. Huovinen, "Outex - new framework for empirical evaluation of texture analysis algorithms," in *Object recognition supported by user interaction for service robots*, Quebec City, Que., Canada, vol.1, pp.701706, 2002.

[49] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," in *IEEE Transactions on Image Processing*, vol.19, pp.16351650, 2010.

[50] L. Liu, P. Fieguth, Y. Guo, X. Wang, and M. Pietikäinen, "Local binary features for texture classification: Taxonomy and experimental study," *Pattern Recognition*, vol.62, pp.135160, 2017.

[51] S. Liao, M. W. K. Law and A. C. S. Chung, "Dominant Local Binary Patterns for texture classification," *IEEE Transactions on Image Processing*, vol.18, pp.11071118, 2009.

[52] Z. Guo, D. Zhang, and D. Zhang, "A completed modelling of local binary pattern operator for texture classification," *IEEE Transactions on Image Processing*, vol.19, pp.16571663, 2010.

[53] L. Liu, S. Lao, P. W. Fieguth, Y. Guo, X. Wang, and M. Pietikainen, "Median robust extended Local Binary Pattern for texture classification," *IEEE Transactions on Image Processing*, vol.25, pp.13681381, 2016.

[54] L. Liu, L. Zhao, Y. Long, G. Kuang and P. Fieguth, "Extended local binary patterns for texture classification," *Image and Vision Computing*, vol.30, pp.8699, 2012.

[55] O. Deniz, G. Bueno, J. Salido, and F. De la Torre, "Face recognition using histograms of oriented gradients," *Pattern Recognition Letters*, vol.32, pp.15981603, 2011.

[56] H. T. M. Nhat and V. T. Hoang, "Feature fusion by using LBP, HOG, GIST descriptors and Canonical Correlation Analysis for face recognition," in *2019 26th International Conference on Telecommunications (ICT)*, pp.371375, 2019.

[57] T. Surinwarangkoon, S. Nitsuwat and E. J. Moore, "A traffic sign detection and recognition system," *International Journal of Circuits, Systems and Signal Processing*, vol.7, pp.58-65, 2013.

[58] W. T. Freeman and M. Roth, "Orientation histograms for hand gesture recognition," *Technical Report TR94-03, MERL Mitsubishi Electric Research Laboratories*, Cambridge, MA 02139, December 1994.

[59] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, pp.886893, 2005.

[60] A. Swadzba and S. Wachsmuth, "Indoor scene classification using combined 3d and gist features," in *Ron Kimmel, Reinhard Klette, and Akihiro Sugimoto, editors, Computer Vision ACCV 2010*, Springer Berlin Heidelberg, vol. 6493, pp.201215, 2010.

[61] W. Tahir, A. Majeed, and T. Rehman, "Indoor/outdoor image classification using gist image features and neural network classifiers," in *2015 12th International Conference on High-capacity Optical Networks and Enabling/Emerging Technologies (HONET)*, pp.15, 2015.

[62] I. Sikiríc, K. Brkíc, and S. Segvíc, "Classifying traffic scenes using the gist image descriptor," in *Croatian Computer Vision Workshop*, 2013.

[63] Vinay A., Gagana B., Vinay.S. Shekhar, Anil B., K.N. Balasubramanya Murthy, and S. Natarajan , "A double filtered GIST descriptor for face recognition," *Procedia Computer Science*, vol.79, pp.533542, 2016.

[64] A. Vinay, B. Gagana, Vinay S. Shekhar, Vasudha S. Shekar, K. N. Balasubramanya Murthy and S. Natarajan, "Face recognition using the novel fuzzy-gist mechanism," in *D. S. Guru, T. Vasudev, H.K. Chethan, and Y.H. Sharath Kumar, editors, Proceedings of International Conference on Cognition and Recognition*, Singapore, Springer Singapore, pp.397406, 2018.

[65] D.G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol.2, pp.1150-1157, 1999.

[66] S. Ha and Y. Moon, "Multiple object tracking using SIFT features and location matching," *International Journal of Smart Home*, vol.5, pp.17-25, 2011.

[67] C. Geng and X. Jiang, "Face recognition using sift features," in *2009 16th IEEE International Conference on Image Processing (ICIP)*, pp. 3313-3316, 2009.

[68] J. Krizaj, V. Struc, and N. Pavesic, "Adaptation of SIFT features for robust face recognition," in *Aurélio Campilho and Mohamed Kamel, editors, Image Analysis and Recognition*, Springer Berlin Heidelberg, vol. 6111, pp.394-404, 2010.

[69] H. Mahmud, K. Hasan, and M. A. Mottalib, "Hand gesture recognition using SIFT features on depth image," *Proceedings of the the Ninth International Conference on Advances in Computer-*

Human Interactions (ACHI), pp. 7, 2016.

[70] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol.60, pp.8490, 2017.

[71] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications., *arXiv:1704.04861*, pp.19, 2017.

[72] L. Wolf, T. Hassner, and Y. Taigman, "Descriptor Based Methods in the Wild," in *Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition*, Marseille, France, Erik Learned-Miller and Andras Ferencz and Frédéric Jurie, pp.1-14, 2008.

[73] Y. Guo, G. Zhao, and M. Pietikäinen, "Texture Classification using a Linear Configuration Model based Descriptor," in *Proceedings of the British Machine Vision Conference 2011*, British Machine Vision Association, Dundee, pp. 119.1119.10, 2011.

[74] T. Ojala, M. Pietikäinen, and T. Mäenpää, "A generalized local binary pattern operator for multiresolution gray scale and rotation invariant texture classification," in *Proceedings of the Second International Conference on Advances in Pattern Recognition*, Springer-Verlag, pp. 397406, 2001.

[75] M. Pietikäinen, T. Mäenpää, and J. Viertola, "Color texture classification with color histograms and local binary patterns," in *Workshop on Texture Analysis in Machine Vision*, pp. 109-112, 2002.

[76] V. Ojansivu and J. Heikkilä, "Blur Insensitive Texture Classification Using Local Phase Quantization," in *David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum, Abderrahim Elmoataz, Olivier Lezoray, Fathallah Nouboud, and Driss Mammass, editors, Image and Signal Processing*, Springer Berlin Heidelberg, Berlin, Heidelberg, vol. 5099, pp.236243, 2008.

[77] R. Nava, G. Cristóbal, and B. Escalante-Ramírez, "A comprehensive study of texture analysis based on local binary patterns," *Proceedings of SPIE - The International Society for Optical Engineering* , Brussels, Belgium, pp.84360E, 2012.

[78] R. Nava, G. Cristóbal, and B. Escalante-Ramírez, "Acomprehensive study of texture analysis based on local binary patterns," in *Optics, Photonics, and Digital Technologies for Multimedia Applications II*, vol. 8436, pp. 84360E., 2012.

[79] Y. Ma, "Number Local binary pattern: An Extended Local Binary Pattern," in *2011 International Conference on Wavelet Analysis and Pattern Recognition*, Guilin, China, pp.272-275, 2011.

[80] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution grayscale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.24, pp.971-987, 2002.

[81] M. B. Sandler, A. G. Howard, M. Zhu, A. Zhmoginov, and L. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4510-4520, 2018.

Duc Phan Van Hoai is graduate student of Ho Chi Minh City Open University, Vietnam. He obtained his B.Sc. degree in Computer Science in 2019. His research interests include optical character recognition and image analysis.

Thongchai Surinwarangkoon was born in Suratthani, Thailand in 1972. He obtained B.Sc. degree in mathematics from Chiang Mai University, Thailand in 1995. He received M.Sc. degree in management of information technology from Walailak University, Thailand in 2005 and Ph.D. in information technology from King Mongkut's University of Technology North Bangkok, Thailand in 2014. He is now a Lecturer in Department of Business Computer, Suan Sunandha Rajabhat University, Bangkok, Thailand. His research interests include digital image processing, machine learning, artificial intelligence, business intelligence, and mobile application development for business.

Vinh Truong Hoang received his master's degree from the University of Montpellier in 2009 and his PhD degree in computer science from the University of the Littoral Opal Coast, France. He is currently an assistant professor and Head of Image Processing and Computer Graphics Department at the Ho Chi Minh City Open University, Vietnam. His research interests include image analysis and feature selection.

Huu-Thanh Duong received his master's degree from the Ho Chi Minh City University of Science, Vietnam. He is currently lecturer at the Ho Chi Minh City Open University. His research interests include machine learning and text mining.

Kittikhun Meethongjan is a full lecturer in the Computer Science Program, Head of Apply Science Department, Faculty of Science and Technology, Suan Sunandha Rajabhat University (SSRU), Thailand. He received his B.Sc. Degree in Computer Science from SSRU in 1990, a master's degree in computer science from King Mongkut's University of Technology Thonburi (KMUTT) in 2000 and a PhD degree in Computer Graphic from the University of Technology Malaysia in 2013. His research interest is Computer Graphic, Image Processing, Artificial Intelligent, Biometrics, Pattern Recognition, Soft Computing techniques and application. Currently, he is an advising of Master and PhD student in Forensic Science of SSRU.