SNRU Journal of Science and Technology 9(2) May – August (2017) 509 – 520

SNRU Journal of Science and Technology

Journal home page: snrujst.snru.ac.th

Ingredients estimation and recommendation of Thai-foods

Lawankorn Soimart[1,*], Pakpoom Mookdarsanit[2]

[1]Department of Business Computer, Faculty of Management Science, Chandrakasem Rajabhat University, Bangkok, 10900, Thailand

[2]Department of Computer Science, Faculty of Science, Chandrakasem Rajabhat University, Bangkok, 10900, Thailand

***Corresponding Author**: lawankorn.s@gmail.com

Received: 6 March 2017; **Revised:** 29 May 2017; **Accepted:** 30 May 2017; **Available online:** 1 August 2017

Paper selected from International Science, Social Sciences, Engineering and Energy Conference 2017 (I-SEEC 2017)

Abstract

Thai-foods are one of the most delicious tastes in the world. Many visitors around the world are fascinated in Thai-foods. In 2011, Thailand had the foods more than other countries from the list of “World’s 50 Most Delicious Foods” such as Tom-yam-goong (4th ranking), Pad-thai (5th ranking), Som-tam (6th ranking), Mas-sa-man (10th ranking), Green-curry (19th ranking), Thai-fried-rice (24th ranking) and Moo-nam-tok (36th ranking), respectively. However, some visitors cannot recognize these foods’ name. What’s more, the ingredients of food are difficult to be considered by the visitors’ eyes. To that end, this paper firstly introduces a novel ingredients estimation of Thai-foods as a Thai-food guidance of visitors that opens a new gate for Thai-foods with image processing and artificial intelligence. For the measurement, our system provided the high correctness of estimation in the criterions of accuracy, recall and precision which were summarized in term of their average values as 0.74, 0.64 and 0.78 respectively.

Keywords: Food recognition; Deep learning; Ingredients estimation of foods

1. Introduction

Thai-foods are one of the most delicious tastes in the world. Such in 2011, the 7 Thai-foods appeared on the list of the "World's 50 Most Delicious Foods (Readers' Pick)"— a worldwide voting of 35,000 people by CNN [1]. Thai-foods appeared more than other nations from the list. From Fig. 1, there were Tom-yam-goong (4th ranking), Pad-thai (5th ranking), Som-tam (6th ranking), Mas-sa-man (10th ranking), Green-curry (19th ranking), Thai-fried-rice (24th ranking) and Moo-nam-tok (36th ranking), respectively.

(a.) Tom-yam-goong (b.) Pad-thai (c.) Som-tam (d.) Mas-sa-man

(d.) Green-curry (e.) Thai-fried-rice (g.) Moo-nam-tok

Fig. 1 Thai-foods in the list of World's 50 Most Delicious Foods

Thai-government has designed the policy for supporting Thailand to be one of the world's top cuisine and tourism. The total of 9,881,091 visitors around the world came to Thailand in 2015 [2] which made the country's income about 2.30 trillion baht of Thai-GDP [3]. However, many visitors cannot recognize these foods' names with their ingredients. Some visitors do not eat some ingredients such as beef, pork or glutamate, etc. Some of them may be in diet. Hence, they do not buy the foods because of unknown ingredients which make the country lose off another income. Since Image processing and Artificial intelligence in computer science are available in other researches such as gender/age estimation of an image [4, 5], face detection [6, 7], place recognition [8 – 11], remote sensing [12 – 16], image forensics [17, 18] and biomedical images [19], etc. Likewise, it is possible to estimate the ingredients of the foods using only an image. This paper firstly introduces a novel method to estimate the ingredients of Thai-foods by image matching with the list of Thai-food images collected in the large-scale computer-brain [20]. For the computer-brain creation, all Thai-food images with their textual recommendations such as names, ingredients and other properties were crawled from the accessible data source [21, 22] over the internet [23, 24]. Then, all features of each image were extracted in term of a vector. After that, all similar images were grouped together. Finally, all images with their recommendations were modeled as the large-scale computer-brain by Deep learning (DP) [25 – 27]. For the ingredients estimation, a visitor can input an unknown Thai-food image to the system. The system will estimate the ingredients from an input image which considers from the most similar images from the computer-brain.

The organization of this paper consists of section 2 "Preliminary" as basic knowledge explanation. The system architecture is crucially described in section 3 and the measurement in section 4, respectively. The future work and our research direction are mentioned in section 5.

2. Preliminary

Feature Extraction

Feature extraction is a mathematical representation of the important features within the image [28] in term of vector. Features can be texture, color and shape [29] which are unique from other images [30]. Basically, important features are dependent on the high contrast position such as edges or corners on the image which are robust for illumination, scale, rotation, and shadowing. All features of an image are mathematically represented in term of a vector. Later, these vectors are input to the process for grouping of similar images.

Deep Learning

Deep learning is the latest calculator-brain in artificial intelligence; was officially proposed by Deng, et al. in 2014 [25]. Deep learning is a well-known approach in AlphaGo that was implemented by Google [26]. AlphaGo also was validated by providing the Go-matches between AlphaGo and some famous Go-players. Especially, AlphaGo [27] won Fun Hui in 2015 and Lee Sedol in 2016, respectively. Traditionally, the infrastructure of deep learning is based on the higher number of hidden nodes in Multi-layer perceptron which is one of a supervised learning. The supervised learning consists of a training set and test set [31]. Training set is a model which is constructed from analysis of huge independent and dependent variables [32]. Test set is a set of independent variables which input to the model for determining the results [33] (or named dependent variables).

3. Methodology of ingredients estimation and recommendation of Thai-foods

Our ingredients estimation of Thai-foods opens a new gate about the relation between Thai-food patterns and image processing with artificial intelligence which is a new field for researching as shown in Fig. 2. In this section, we crucially explain our system: RGB-to-grayscale conversion, Feature extraction and Deep learning: training, and testing, respectively.

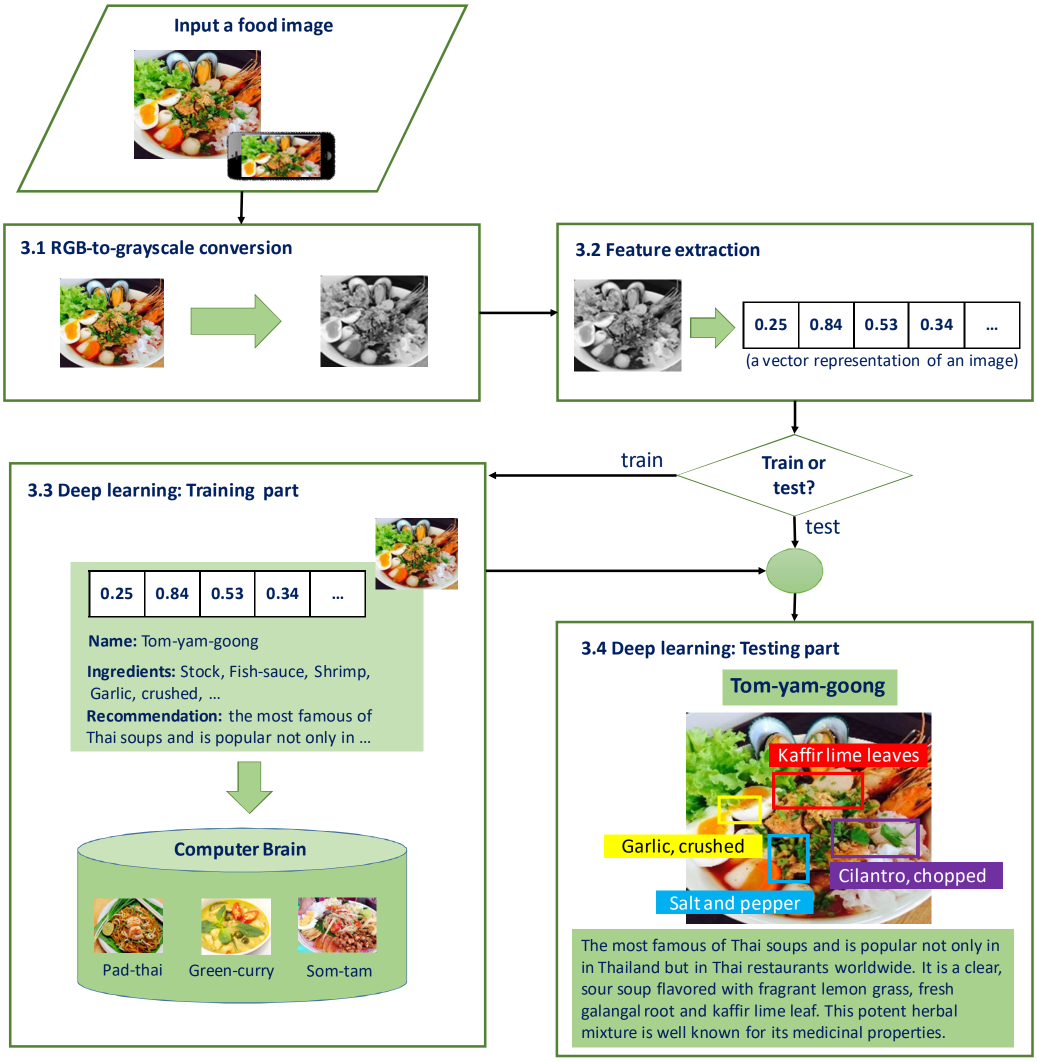

Fig. 2 System architecture of "ingredients estimation and recommendation of Thai-foods"

RGB-to-grayscale Conversion

Either an unknown Thai-food image or Thai-food image with its information is input to the system as shown in Fig. 3. Firstly, the image must be normalized the (color) dimensions. Gray-scale has the value between 0 – 255.

Fig. 3 An unknown Thai-food image (in case of Tom-yam-goong)

$$Gray = \sum_{1}^{n(pixel)} 0.29R + 0.58G + 0.14B \tag{1}$$

where Gray is the gray-scale value of a pixel within the image, R is the red-axis within the image, G is the green-axis within the image and B is the blue-axis within the image.

All pixels within the image are compressed from RGB to gray-scale using the equation (1) as shown in Fig. 4.

Fig. 4 The image in format of gray-scale

Feature Extraction

Before training or testing part, the gray-scale image must be in the form of a mathematical vector representation that automatically and recursively computed by the algorithm of computer (according to these 6 steps). The output of this procedure is a vector that represents all important features such as texture, color and shape of the image as shown in Fig. 5.

Fig. 5 A vector representation of an image

<u>Step 1</u>: Calculate the high-contrasts through the image in x and y axis. The high-contrast is a derivative distribution of intensity, by equation (2) and the transpose of (2) as (3).

$$D_H = \sum_{1}^{n(block)} [-1 \quad 0 \quad 1]_i \tag{2}$$

$$D_V = \sum_{1}^{n(block)} [-1 \quad 0 \quad 1]_i^T \tag{3}$$

where D_H and D_V is gradient value in horizontal and vertical axis

<u>Step 2</u>: Separate the image into 4 (or more) blocks as shown in Fig. 6.

Fig. 6 The image separated into blocks

Step 3 : Separate each block into 4 (or more) cells as shown in Fig. 7.

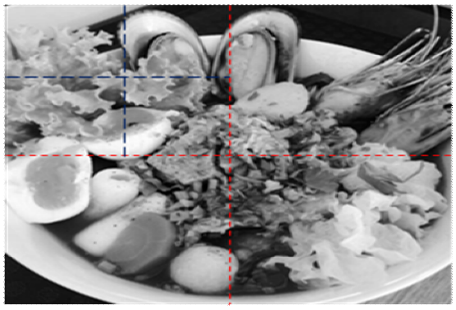

Fig. 7 Cells within a block

Step 4 : Calculate the magnitude of gradient in each block by (4)

$$M_G(x,y) = \sqrt{D_H(x,y) \bullet \cos\theta + D_V(x,y) \bullet \sin\theta} \tag{4}$$

where M_G is the magnitude of a gradient at the position (x, y)

Step 5 : Calculate the derivative of gradients by (5)

$$Q_G(x,y) = \frac{\partial D_V(x,y)}{\partial D_H(x,y)} \tag{5}$$

Step 6 : Extract the important features into a vector in (6) which parallel considers the step 4-5.

$$Features = \sum_{1}^{k^2} M_G(x,y) * O_G(x,y) \tag{6}$$

Deep learning: training part

Only the Thai-food image with its information (name, ingredients and other textual recommendation) is input to this procedure. All important features of an image are trained in a large number of hidden nodes in the deep learning which uses the least-squared error. Target value is the name, ingredients and textual recommendation of Thai-food images to construct a computer-brain as visualized in Fig. 8. For the mathematical definition, the parameters and target classes as $\{(x_1,t_1),\ldots,(x_K,t_K)\}$ where $x_K = [x_{k1},\ldots,x_{kR}]^T$, $t_k = [t_{k1},t_{k2}]$, $k = 1, 2$.

<u>Step 1</u>: Identify the deep learning, the learning rate (η) as a small value convergence to 0, Least-error value as E_{th} (error threshold), and iteration of learning as L_{max} (maximum epoch) by defining a counter in finding and the counter for all inputs.

<u>Step 2</u>: Randomize the weight (W) value and bias (b) value.

<u>Step 3</u>: Input the training set as a series of $\{x_k, t_k\}$, input to the deep learning and also calculate the output and error of target class.

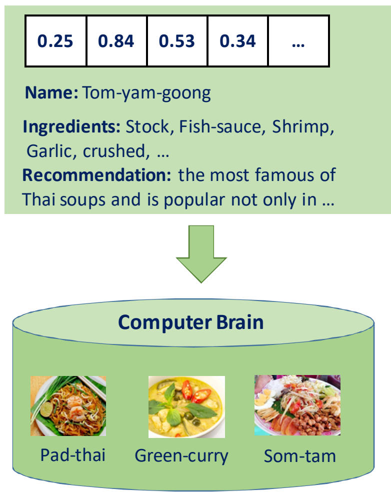

Fig. 8 A vector (with its name, ingredients and recommendation) trained to the computer brain by deep learning

<u>Step 4</u>: Update the weight value and the bias value of backward again from the output layer to the first hidden layer. Calculate the derivative of error. Update the new weight value and the bias value in the output layer, the second hidden layer and the first hidden layer.

<u>Step 5</u>: If $k < K$; $k = k + 1$ and go to step 3, else ; go to the step 6

<u>Step 6</u>: Calculate average square of error from (7).

$$MSE = \frac{1}{K}\sum_{k=1}^{K}\|e_k\|^2 = \frac{1}{K}\sum_{k=1}^{K}\|t_k - a_k\|^2 \tag{7}$$

In case of $MSE > E_{th}$ and no. of iterations $m < L_{max}$ then identify $k = 1, m = m + 1$ and start a new iteration of training at step 3

In case of $MSE \le E_{th}$ or $m \ge L_{max}$, then the training is finished.

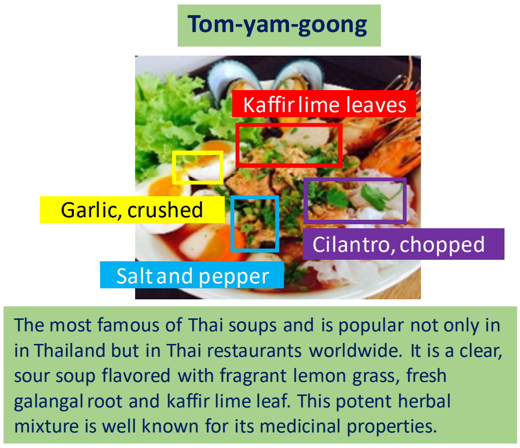

Fig. 9 Name, ingredients and textual recommendation tagged to the unknown image

Deep learning: testing part

The unknown Thai-food image is input to testing part for estimating its name, ingredients and textual-recommendation. Many images from the computer brain are used to compute the number of feature similarities with the unknown Thai-food. Deep learning automatically considers the highest similar number of features to estimate the name, ingredients and textual-recommendation. The unknown Thai-food image is finally tagged its information as shown in Fig. 9.

4. Experimental Measurement

In this experiment, all Thai-food images with their information were randomly crawled from the internet (for 458 images). The 80% of images were input to construct the computer-brain. The target classes were Tom-yam-goong, Pad-thai, Som-tam, Mas-sa-man, Green-curry, Thai-fried-rice, and Moo-nam-tok. For testing, we used another 20% of images; and to compute these criterions: Precision, Recall and Accuracy.

$$\Pr ecision = \frac{TP}{TP + FP} \tag{8}$$

$$\mathrm{Re}\,call = \frac{TP}{TP + FN} \tag{9}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

where TP = True Positive, FP = False Positive, TN = True Negative and FN = False Negative, respectively

The RGB to gray scale conversion, feature extraction and deep learning (either training or testing) were executed, respectively. This system was developed by Matlab. We used Precision, Recall and Accuracy in (8), (9) and (10) as criterions for the measurement of our system. The quantitative measurements from the names of unknown foods were calculated as the average of (8), (9) and (10) as shown in Table 1.

Table 1 Precision, Recall and Accuracy criteria

Accuracy	Recall	Precision
0.71	0.69	0.81
0.74	0.72	0.76
0.65	0.65	0.78
.	.	.
.	.	.
.	.	.
0.74	**0.64**	**0.78**

5. Conclusion

This paper introduces a novel ingredients estimation of Thai-foods which opens a new gate about the relation between Thai-food patterns and image processing with artificial intelligence for Thai-foods guidance of visitors. The system was categorized into RGB-to-grayscale conversion, Feature extraction and Deep learning (either training or testing). The system provides high correctness in criterions of accuracy, recall and precision which were measured from the food's name. In every minute, there are more than 100,000 uploaded-images around the world. Some of them are about foods with the description. The numbers of food images have been exponentially increased on the social media. With the concept of data mining, the more data produces the more correctness. Hence, all food images on the social media are large-scale (also called "Big data") that cover enough for all foods and dishes around the world. For future work, it is possible to further analyze nutrients; to compute the calories of Thai-foods from a single image. Imagine that how it is more convenient if the recipe or cookbook will be automatically shown to user from a food image. In coming soon, many street foods will be estimated their ingredients and calories using a mobile phone.

6. References

[1] Your pick: World's 50 best foods, http://travel.cnn.com/explorations/eat/readers-choice-worlds-50-most-delicious-foods-012321, 5 March 2017.

[2] Visitor Statistics, 1998 – 2016, http://tourism2.tourism.go.th/home/listcontent/11/221/276, 5 March 2017.

[3] R. Turner, Travel & Tourism, Economic Impact 2015, first ed., World Travel & Tourism Council, London, 2015.

[4] L. Soimart, P. Mookdarsanit, Gender Estimation of a Portrait: Asian Facial-significance Framework, The 6th International Conference on Sciences and Social Sciences, Rajabhat Maha Sarakham University, 23 September 2016, 122.

[5] R. Uricár, R. Timofte, J. Rothe, J. Matas, L.V. Gool, Structured Output SVM Prediction of Apparent Age, Gender and Smile from Deep Features, IEEE Conference on Computer Vision and Pattern Recognition Workshops, Las Vegas NV USA. 26 June 2016, 730 – 738.

[6] Y. Zheng, C. Zhu, K. Luu, C. Bhagavatula, T.H.N. Le, M. Savvides, Towards a deep learning framework for unconstrained face detection, IEEE 8th International Conference on Biometrics Theory, Applications and Systems, Niagara Falls NY USA. 6 September 2016, 1 – 8.

[7] K. Zhang, Z. Zhang, Z. Li, Y. Qiao, Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks, SPL. 23(10) (2016) 1499 – 1503.

[8] P. Mookdarsanit, M. Ketcham, Image Location Estimation of well-known Places from Multi-source based Information, The Eleventh International Symposium on Natural Language Processing, Ayutthaya Thailand. 10 February 2016, 75.

[9] X. Qian, Y. Zhao, J. Han, Image Location Estimation by Salient Region Matching, IP. 24(11) (2015) 4348 – 4358.

[10] P. Mookdarsanit, M. Rattanasiriwongwut, GPS Determination of Thai-temple Arts from a Single Photo, The 11th International Conference on Applied Computer Technology and Information Systems, Southeast Bangkok College. 25 January 2017, 42 – 47.

[11] P. Mookdarsanit, M. Rattanasiriwongwut, Location Estimation of a Photo: A Geo-signature MapReduce Workflow, EJ. (2017) (inpress).

[12] L. Soimart, M. Ketcham, An Efficient Algorithm for Earth Surface Interpretation from Satellite Imagery, EJ. 20(5) (2016) 215 – 228.

[13] L. Soimart, M. Ketcham, The Segmentation of Satellite Image Using Transport Mean-shift Algorithm, The 13th International Conference on IT Applications and Management, Naresuan University. 17 January 2015, 124 – 128.

[14] L. Soimart, M. Ketcham, Hybrid of Pixel-based and Region-based Segmentation for Geology Exploration from Multi-spectral Remote Sensing, The Eleventh International Symposium on Natural Language Processing, Ayutthaya Thailand. 10 February 2016, 74.

[15] R. Hang, Q. Liu, H. Song, Y. Sun, F. Zhu, H. Pei, Graph Regularized Nonlinear Ridge Regression for Remote Sensing Data Analysis, STARS. 10(1) (2017) 277 – 285.

[16] Q. Zou, L. Ni, T. Zhang, Q. Wang, Deep Learning Based Feature Selection for Remote Sensing Scene Classification, GRSL. 12(11) (2015) 2321 – 2325.

[17] A. Dada Warbhe, R.V. Dharaskar, V.M. Thakare, Digital image forensics: An affine transform robust copy-paste tampering detection, IEEE 10th International Conference on Intelligent Systems and Control, Coimbatore India. 7 January 2016, 1 – 5.

[18] P. Mookdarsanit, L. Soimart, M. Ketcham, N. Hnoohom, Detecting Image Forgery Using XOR and Determinant of Pixels for Image Forensics, IEEE 11th International Conference on Signal-Image Technology & Internet-Based Systems, Bangkok. 26 November 2015, 613 – 616.

[19] B. Dong, C. Sun, H.F. Zhang, Optical Detection of Ultrasound in Photoacoustic Imaging, BME. 64(1) (2017) 4 – 15.

[20] L. Soimart, P. Pongcharoen, Multi-row Machine Layout Design using Aritificial Bee Colony, International Conference on Economics Development & Research, Bangkok. 2011, 103 – 108.

[21] J. Wang, B. Wang, A hybrid main memory applied in virtualization environments, IEEE 1st International Conference on Computer Communication and the Internet, Wuhan China. 15 October 2016, 413 – 417.

[22] L. Soimart, P. Mookdarsanit, Multi-factor Authentication Protocol for Information Accessibility in Flash Drive, The 9th International Conference on Applied Computer Technology and Information Systems, Nakhon Pathom Rajabhat University. 19 February 2016, 10 – 13.

[23] S. Imai, S. Patterson, C. A. Varela, Cost-Efficient High-Performance Internet-Scale Data Analytics over Multi-cloud Environments, The 15th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, Shenzhen. 7 May 2015, 793 – 796.

[24] P. Mookdarsanit, S. Gertphol, Light-weight operation of a failover system for Cloud computing, IEEE 5th International Conference on Knowledge and Smart Technology, Burapha University. 31 January 2013, 42 – 46.

[25] L. Deng, D. Yu, Deep Learning: Methods and Applications, FTiSP. 7(3 – 4) (2014) 197 – 387.

[26] F.Y. Wang, J.J. Zhang, X. Zheng, X. Wang, Y. Yuan, X. Dai, J. Zhang, L. Yang, Where does AlphaGo go: from church-turing thesis to AlphaGo thesis and beyond, J-AS. 3(2) (2016) 113 – 120.

[27] D. Silver, A. Huang, C.J. Maddison, A. Guez, L. Sifre, G.V.D. Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, D. Hassabis, Mastering the game of Go with deep neural networks and tree search, Nature, Int Wkly J Sci. 529(7578) 484 – 489.

[28] A.A. Olaode, G. Naghdy, C.A. Todd, Unsupervised Classification of Images: A Review, IJIP. 8(5) (2014) 325 – 342.

[29] A.A. Olaode, G. Naghdy, C.A. Todd, Bag-of-Visual Words Codebook Development for the Semantic Content Based Annotation of Images, IEEE 11th International Conference on Signal-Image Technology & Internet-Based Systems, Bangkok. 26 November 2015, 7 – 14.

[30] A.A. Olaode, G. Naghdy, C.A. Todd, Efficient region of interest detection using blind image division, Signal Processing Symposium, Debe Poland. 10 June 2015, 1 – 6.

[31] L. Soimart, P. Mookdarsanit, An Admission Recommendation of High-school Students using Apriori Algorithm, The 6th International Conference on Sciences and Social Sciences, Rajabhat Maha Sarakham University. 23 September 2016, 121.

[32] D. Chen, B.K.W. Mak, Multitask Learning of Deep Neural Networks for Low-Resource Speech Recognition, ASLP. 23(7) (2015) 1172 – 1183.

[33] J. Tao, S. Ghaffarzadegan, L. Chen, K. Zechner, Exploring deep learning architectures for automatically grading non-native spontaneous speech, IEEE International Conference on Acoustics, Speech and Signal Processing, Shanghai China. 25 March 2016, 6140 – 6144.