MONTEAN Framework: A Magnificent Outstanding Native-Thai and Ecclesiastical Art Network

Pakpoom Mookdarsanit[1], Montean Rattanasiriwongwut[2]

[1]Faculty of Information Technology
King Mongkut's University of Technology North Bangkok
Bangkok, Thailand
e-mail: [1]pakpoom_m@hotmail.com
[2]Faculty of Information Technology
King Mongkut's University of Technology North Bangkok
Bangkok, Thailand
e-mail: [2]montean.r@it.kmutnb.ac.th

บทคัดย่อ—ประเทศไทยมีวัดและพุทธศาสนสถานหลายแห่งที่มีความงดงามทางสถาปัตยกรรมและวัฒนธรรมในสายตาของนักท่องเที่ยวทั่วโลก จากสถิติที่ผ่านมา พบว่าประเทศไทยมีพุทธศาสนสถานอยู่เป็นจำนวน 40,717 แห่ง กระจายอยู่ตามภูมิภาคต่าง ๆ ทั่วประเทศ อย่างไรก็ตาม การจดจำชื่อเต็มและการหาสถานที่ตั้งของพุทธศาสนสถานเหล่านี้เป็นเรื่องยากสำหรับนักท่องเที่ยวส่วนใหญ่ เพราะการไปท่องเที่ยวในสถานที่ที่ไม่รู้จักชื่อเต็มนั้น จะกลายเป็นเรื่องลำบากมากยิ่งขึ้นในการมาท่องเที่ยว ซึ่งอาจส่งผลกระทบให้ประเทศชาติสูญเสียรายได้สำหรับการท่องเที่ยว งานวิจัยนี้ผู้เขียนได้ออกแบบกรอบแนวคิดใหม่แบบระบบฐานความรู้สำหรับนักท่องเที่ยวในการค้นหาชื่อเต็มและสถานที่ตั้งของพุทธศาสนสถานในประเทศไทยจากรูปถ่าย โดยมีชื่อว่า "กรอบแนวคิดมณเฑียร" โดยระบบฐานความรู้จะทำการคำนวณเพื่อหาชื่อและสถานที่ตั้งโดยพิจารณาจากสถาปัตยกรรมของพุทธศาสนสถานที่ปรากฏอยู่ในรูปถ่าย มณเฑียรเป็นกรอบแนวคิดประยุกต์แรกในศาสตร์ของเทคโนโลยีสารสนเทศที่มีการบูรณาการระหว่างการคำนวณความเหมือนกันของภาพและพุทธศาสนสถานไทย ผลการทดลองแสดงให้เห็นว่า มณเฑียรมีค่าสมรรถนะโดยภาพรวมอยู่ในเกณฑ์ที่สูง โดยมีค่าความถูกต้องเท่ากับ 0.91, ค่าระลึกเท่ากับ 0.93 และค่าความแม่นยำเท่ากับ 0.89 ตามลำดับ

คำสำคัญ: ฐานความรู้, การค้นคืนภาพ, การรู้จำวัฒนธรรม

Abstract — Thai-ecclesiastical arts (also called "Thai-temple arts") are significantly attractive from many travelers. From the statistics, there are a total of 40,717 ecclesiastical temples that are located in Thailand. Many travelers could not go conveniently to these places because of the misremembering of their names and locations that made the country missed out a lot of income. In this paper, we design a novel framework as a knowledge-based system for travelers to find the name and location of an unknown Thai-ecclesiastical art from a single image, named "Magnificent Outstanding Native-Thai and Ecclesiastical Art Network (MONTEAN)". MONTEAN is the first groundwork in the field of information technology that integrally couples image matching and Thai-ecclesiastical arts together. From the experimental results, MONTEAN provided the high correctness in terms of accuracy, recall and precision as 0.91, 0.93 and 0.89, respectively.

Keywords- Knowledge-based System; Image Retrieval; Culture Recognition

I. INTRODUCTION

Nowadays, tourism has been rising in many sub-districts of Thailand. To stimulate the economic growth, Thai-government has designed the policy for invitation of travelers. The Tourism Authority of Thailand (TAT) also has conveyed the slogan "Amazing Thailand" with a campaign "Discover Thainess" to promote Thai tourism internationally for supporting Thailand to be one of the world's top tourist attractions. Since Thai-ecclesiastical arts (also called "Thai-temple arts") are huge much of interesting from travelers in views of culture, architecture and history. It is not surprise that tourism is one of the major economic factors in Thailand. As 2015, Thailand had a total of 9,881,091 travelers around the world [1].

The total country's income was estimated about 2.3 trillion baht of Thailand's GDP [2].

From the statistics, there are a total of 40,717 ecclesiastical temples that are located in Thailand [3]. Many beautiful images of those ecclesiastical places and their architectures are always shared on the social media. However, it is difficult for travelers to remember the names and locations of these ecclesiastical places. Hence, they cannot go to these places which make the country miss out a lot of income. Since a large-scale of Thai-ecclesiastical images with GPS-tags and other textual-data are shared on the social media. It is feasible to find the name and location of an unknown Thai-ecclesiastical art from a single image with the help of many images from social media. To give an illustration, Mookdarsanit et al. [4] have created the computer-brain that automatically tag the names and GPSs of unnamed world-heritage images from Europe and America.

In order to firstly apply in a knowledge-based system of Thai-ecclesiastical arts, we have designed a novel framework named "Magnificent Outstanding Native-Thai and Ecclesiastical Art Network (MONTEAN)" framework which enhanced from automatic tagging of images by one-to-many matching between the unknown image and many other Thai-ecclesiastical-art images with textual-tagging from the social media [5-6].

Figure 1. Many beautiful Thai-ecclesiastical arts

MONTEAN is the first groundwork in the field of information technology that integrally couples image matching and Thai-ecclesiastical arts together. For the supervised-model creation of MONTEAN, all Thai-ecclesiastical-art images with their names and GPSs were firstly crawled from social media. Then, all visual contents of each image were extracted. Finally, all images with textual information were modeled together using Deep learning. For finding the name and location of an unknown ecclesiastical temple, a traveler can input an image to the knowledge-based system [7]. And the system will show the name and location of the place which considers from the most similar image in the dataset. For the evaluation, our MONTEAN framework provided the recognition correctness in terms of accuracy, recall and precision which have averages at 0.91, 0.93 and 0.89, respectively.

Note that the Thai-meaning of word "MONTEAN" (Thai: มณเฑียร) usually refers to a great architectural place of the emperor or monk.

This paper is organized into 5 sections. Section 2 describes the preliminary. MONTEAN Framework is mathematically explained in section 3. Section 4 and 5 are experimental results and conclusion, respectively.

II. PRELIMINARY

A. Visual Content Extraction

Visual content extraction is a representation of overall visual contents in the image [8-9] in term of a vector. A visual content can be texture, color and shape which can determine the characteristics to distinguish one image from other ones [10-11]. Basically, interesting-points are dependent on the high contrast position such as edges or corners on the image. The visual contents at these interesting-points are mathematically described by visual content descriptors in term of a vector representation of an image. Later, the vector is input to the supervised learning model.

B. Deep Learning

Deep learning (DP) is the newest supervised model in artificial intelligence; was introduced by L. Deng and D. Yu in 2014 [12]. DP is a well-known approach in the algorithm titled AlphaGo that was developed by Google [13]. Google verified AlphaGo by providing the Go-matches between AlphaGo and some well-known Go-players. In 2016, AlphaGo [14] won Fun Hui in 2015 and Lee Sedol who are the world-class Go-players. Basically, the architecture of deep learning is based on Neural Networks which is one of a supervised learning (or named supervision). Supervised learning consists of a training set and test set [15-16]. Training set is a model which is formulated from the analysis of huge independent and dependent variables. As test set is a set of independent variables which input to the model for determining the results (or named dependent variables). Mathematical computation of DP is crucially explained in our proposed MONTEAN framework at the section 3. Recently, DP was applied in various applications, especially in computer vision. What's more, DP was used in other researches such as speech recognition [17-18], corpus of natural language processing [19-20], Glaucoma Image [21], gender estimation from an image [22] and Remote Sensing [23]. All things considered, DP also can be used in MONTEAN framework as a knowledge-based system to find the name and location of an unknown Thai-ecclesiastical-art from a single image.

III. MONTEAN FRAMEWORK

Our proposed MONTEAN framework opens a new branch about the relation between image matching and Thai-ecclesiastical arts which is originated from image retrieval. In this section, we describe deeply and user-friendly the architecture of MONTEAN framework as a knowledge-based system which can be categorized into 3 procedures: Preprocessing, Visual Content Extraction and Deep Learning, respectively.

A. Preprocessing

The framework is firstly initiated from many Thai-ecclesiastical-art images with GPS-tags and textual information which were crawled from Facebook as an example shown in “Fig. 2”.

Figure 2. A Thai-ecclesiasitcal-art image

First of all, the image must be reduced the (color) dimensions [24]. All pixels are converted from RGB image to gray-scale one using the equation (1). And the result is shown in “Fig.3”. Gray-scale image has the range of value between 0-255.

$$Gray = 0.3R + 0.6G + 0.1B \tag{1}$$

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

Figure 3. A grayscale image of Thai-ecclesiastical-art

B. Visual Content Extraction

Step 1: Compute all gradients through the grayscale image in x and y axis. The gradient is a slope distribution of intensity. The gradient value is divided into 3 values as {-1, 0, 1} by equation (2) and the transpose of (2) as (3).

$$D_H(x,y) = [-1 \quad 0 \quad 1] \tag{2}$$

$$D_V(x,y) = [-1 \quad 0 \quad 1]^T \tag{3}$$

where $D_H(x,y)$ is gradient value in horizontal axis, $D_Y(x,y)$ is gradient value in vertical axis and $[-1\ 0\ 1]^T$ is the transpose of $D_H(x,y)$, respectively

Step 2: Divide the grayscale image into *k* parts as shown in “Fig. 4” (*k*=4).

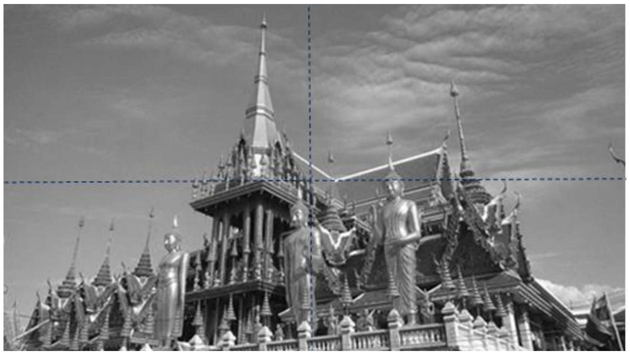

Figure 4. Parts of grayscale image

Step 3: Divide each part into *k* sub-parts as shown in “Fig. 5” (*k*=4). Note that a grayscale image has *k* parts and k^2 subparts (*k*=4 and k^2=16).

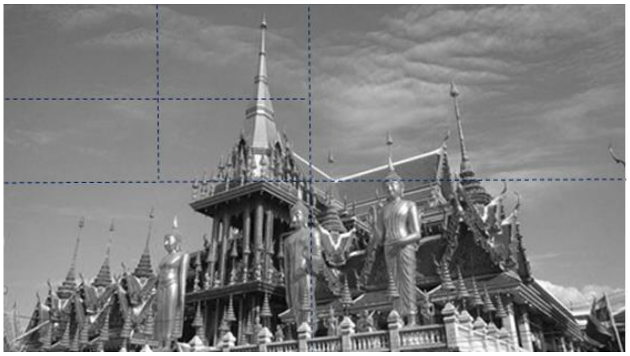

Figure 5. The *k* sub-parts within a part

Step 4: Compute the magnitude of gradient in each part by (4)

$$M_G(x,y) = \sqrt{D_H(x,y)^2 + D_V(x,y)^2} \tag{4}$$

where $M_G(x,y)$ is the magnitude of a gradient at the position (x, y), $D_H(x,y)$ is gradient value in horizontal axis and $D_Y(x,y)$ is gradient value in vertical axis, respectively

Step 5: Compute the slope of gradient in each sub-parts by (5)

$$O_G(x,y) = \tan^{-1}\left(\frac{D_H(x,y)}{D_v(x,y)}\right) \tag{5}$$

where $O_G(x,y)$ is the slope of gradient at the position (x,y), $D_H(x,y)$ is gradient value in horizontal axis and $D_Y(x,y)$ is gradient value in vertical axis, respectively

Step 6: Extract the visual contents into local descriptors in (6) which consider the step 4-5.

$$Descriptors = \int_{1}^{k^2} M_G(x,y) * O_G(x,y) \quad (6)$$

where *Descriptors* is a vector representation of Thai-ecclesiastical art images, $M_G(x,y)$ is the magnitude of a gradient at the position (x, y) and $O_G(x,y)$ is the slope of gradient at the position (x,y), respectively

C. Deep Learning

All descriptors in term of a vector (from the previous section) for all Thai-ecclesiastical art images are trained in a large number of hidden nodes in deep learning which uses the least-squared error. Target classes are the Thai-ecclesiastical places' names and GPSs which were downloaded from Facebook. The parameters and target classes are defined as {(x_1,t_1),…,(x_K,t_K)} where $x_K = [x_{k1},...,x_{kR}]^T$, $t_k = [t_{k1},t_{k2}]$, $k = 1,2$ (for the training set)

TABLE I. CONFIGURATION OF PARAMETERS IN DEEP LEARNING

Parameter	Value
Learning rate (η)	0.09
Error threshold (E_{th})	0.0020
Maximum epoch ($L_{\max}$)	81

Step 1: 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, as in TABLE I.

Step 2: Randomize the weight (W) value and bias (b) value.

Step 3: 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.

Step 4: Adjust the weight value and the bias value of backward again from the output layer to the first hidden layer. Calculate the slope of error. Adjust the new weight value and the bias value in the output layer, the second hidden layer and the first hidden layer.

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

Step 6: Compute the average of error square from (7).

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

only if $MSE > E_{th}$ and no. of iterations $m < L_{\max}$ then identify $k = 1, m = m + 1$ and start a new iteration of training at step 3. only if $MSE \le E_{th}$ or $m \ge L_{\max}$, then the training is finished.

IV. EXPERIMENTAL RESULTS

All Thai-ecclesiastical-art images in this experiment were randomly crawled from Facebook which could be categorized into groups according to their regions: Northern, North-eastern, Western, Eastern, Central and Southern Thailand as in "Fig.6".

All images were input to the MONTEAN framework. The photo was converted into gray scale. Visual content extraction and deep learning were executed, respectively. For finding the name and location of an unknown Thai-ecclesiastical art, the traveler just input a single image to the knowledge-based system. The system finally shows the name and GPS of image.

Figure 6. Some Thai-ecclesiastical-art images in the experiment

Our MONTEAN framework was implemented by M-script in Matlab. The source code and dataset of Thai-ecclesiastical-art images can be requested for via the email (under in terms of use). The environmental setup was carried out under Intel Core™2, Quad CP Q8400 and 48 GB of RAM.

For the criteria evaluation, we used Precision, Recall and Accuracy in (8), (9) and (10) as benchmarks for our proposed MONTEAN framework.

Precision is referred to the fraction of changed target classes identified correctly.

Recall is referred to the fraction of changed target classes identified as unchanged.

Accuracy is referred to overall correctness of the proposed framework by the average of detecting changed target classes as changed and unchanged target classes as unchanged.

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

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

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

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

All experiments were computed as the average of these criteria as shown in TABLE II.

TABLE II. THE EXPERIMENTAL RESULTS

Accuracy	**Recall**	**Precision**
0.912	0.839	0.845
0.953	0.933	0.922
0.931	0.844	0.812
. . .	. . .	. . .
0.913	**0.934**	**0.891**

V. CONCLUSION

Million travelers are interested in Thai-ecclesiastical art. There are a number of 40,717 Thai-ecclesiastical places that are located in Thailand. However, the travelers cannot go to these places because of the misremembering of their names and locations that make the country missed out a lot of income. In conclusion, this paper proposes a novel framework named "Magnificent Outstanding Native-Thai and Ecclesiastical Art Network (MONTEAN)" which is the first groundwork in the field of information technology that integrally couples image matching and Thai-ecclesiastical arts together. From the results, MONTEAN provided the high correctness in terms of accuracy, recall and precision as 0.91, 0.93 and 0.89, respectively.

For future work (in view of basic research), the challenges are volume, velocity, veracity, value and variety of data [25-26]. Since there are large-scale information from billion social users around the world [27] which are the total samplings from world-wide people; and they are enough for development of the more correctness of Thai-ecclesiastical arts' names and locations.

For future work (in view of applied research), the annotation of Buddhist festivals such as the praying of 9-Scared Places (Thai: ไหว้พระ ๙ วัด) or Special Buddhist Days such as Visakah Puja Day (Thai: วันวิสาขบูชา) and Asalha Puja Day (Thai: วันอาสาฬหบูชา). Moreover, our proposed MONTEAN Framework can be applied to ontological structure of Thai-ecclesiastical places for knowledge-based system that can be categorized by Thai-regions (Northern, North-eastern, Western, Eastern, Central and Southern Thailand) or by Royal temples (Thai: พระอารามหลวง). The Royal temples consist of Special class (Thai: ชั้นราชวรมหาวิหาร), First class (Thai: ชั้นราชวรวิหาร), Second class (Thai: ชั้นวรมหาวิหาร) and Third class (Thai: ชั้นวรวิหารและสามัญ), respectively.

REFERENCES

[1] Department of Tourism Thailand, "Visitor Statistics 1998–2014," [Online]. Available: http://tourism2.tourism.go.th/home/list content/11/221/276. [Accessed: 18 March 2017].

[2] T. Rochelle, Travel & Tourism, Economic Impact 2015. World Travel & Tourism Council, London. 2015

[3] Dhammathai, "Theravada Buddhism Information Network," [Online]. Available: http://www.dhammathai.org/watthai/ watthai.php. [Accessed: 18 March 2017].

[4] P. Mookdarsanit and M. Ketcham, "Image Location Estimation of well-known Places from Multi-source based Information," The 11th International Symposium on Natural Language Processing, 2016, pp. 75.

[5] X. Qian, Y. Zhao and J. Han, "Image Location Estimation by Salient Region Matching," in IEEE Transactions on Image Processing, vol. 24, no. 11, pp. 4348-4358, Nov. 2015.

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

[7] L. Ploywattanawong, "The Wisdom Cultural Heritage Information Systems of Suphan Buri," International Journal of Applied Computer Technology and Information Systems (IJACTIS), vol.6, no.1, pp.7-10, 2016.

[8] L. Soimart and M. Ketcham, "The Segmentation of Satellite Image Using Transport Mean-shift Algorithm," 13th International Conference on IT Applications and Management (ITAM-13), 2015, pp. 124-128.

[9] L. Soimart and M. Ketcham, “Hybrid of Pixel-based and Region-based Segmentation for Geology Exploration from Multi-spectral Remote Sensing,” The 11th International Symposium on Natural Language Processing, 2016, pp. 74.

[10] L. Soimart and P. Mookdarsanit, “Ingredients Estimation and Recommendation of Thai-foods,” The 8th International Science, Social Science, Engineering and Energgy Conference (selected to SNRUJST publication), March 2017, pp. 43.

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

[12] L. Deng and D. Yu, “Deep Learning: Methods and Applications,” Foundations and Trends® in Signal Processing, vol. 7, no. 3-4, pp. 197-387, 2014.

[13] F. Y. Wang et al., "Where does AlphaGo go: from church-turing thesis to AlphaGo thesis and beyond," IEEE/CAA Journal of Automatica Sinica, vol. 3, no. 2, pp. 113-120, April 2016.

[14] D. Silver et al., “Mastering the game of Go with deep neural networks and tree search,” Nature International Weekly Journal of Science, vol. 529, no. 7578, pp. 484-489, 2016.

[15] L. Soimart and P. Mookdarsanit, “An Admission Recommendation of High-school Students using Apriori Algorithm,” The 6th International Conference on Sciences and Social Sciences, 2016.

[16] L. Soimart and P. Pongchareon, “Multi-row Machine Layout Design using Aritificial Bee Colony,” The International Conference on Economics and Business Information, Bangkok, Thailand, 2011.

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

[18] Z. H. Ling et al., "Deep Learning for Acoustic Modeling in Parametric Speech Generation: A systematic review of existing techniques and future trends," IEEE Signal Processing Magazine, vol. 32, no. 3, pp. 35-52, May 2015.

[19] O. Koller, H. Ney and R. Bowden, "Deep Learning of Mouth Shapes for Sign Language," 2015 IEEE International Conference on Computer Vision Workshop (ICCVW), 2015, pp. 477-483.

[20] Y. He, S. Xiang, C. Kang, J. Wang and C. Pan, "Cross-Modal Retrieval via Deep and Bidirectional Representation Learning," IEEE Transactions on Multimedia, vol. 18, no. 7, pp. 1363-1377, July 2016.

[21] R. Srivastava, J. Cheng, D. W. K. Wong and J. Liu, "Using deep learning for robustness to parapapillary atrophy in optic disc segmentation," 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), New York, NY, 2015, pp. 768-771.

[22] L. Soimart and P. Mookdarsanit, “Gender Estimation of a Portrait: Asian Facial-significance Framework,” The 6th International Conference on Sciences and Social Sciences, 2016.

[23] Q. Zou, L. Ni, T. Zhang and Q. Wang, "Deep Learning Based Feature Selection for Remote Sensing Scene Classification," IEEE Geoscience and Remote Sensing Letters, vol. 12, no. 11, pp. 2321-2325, Nov. 2015.

[24] L. Soimart and M. Ketcham, “An Efficient Algorithm for Earth Surface Interpretation from Satellite Imagery,” Engineering Journal, vol.20, no.5, pp.215-228, October, 2016.

[25] P. Mookdarsanit and M. Rattanasiriwongwut, “Location Estimation of a Photo: A Geo-signature MapReduce Workflow,” Engineering Journal, in press.

[26] L. Soimart and P. Mookdarsanit, “Multi-factor Authentication Protocol for Information Accessibility in Flash Drive,” The 9th Applied Computer Technology and Information Systems, Nakhon Pathom, February 2016, pp. 10-13.

[27] P. Mookdarsanit and S. Gertphol, "Light-weight operation of a failover system for Cloud computing," 2013 IEEE 5th International Conference on Knowledge and Smart Technology, Chonburi, Thailand, 2013, pp. 42-46.