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

Pakpoom Mookdarsanit¹, Montean Rattanasiriwongwut²

¹Faculty of Information Technology
King Mongkut's University of Technology North Bangkok
Bangkok, Thailand
e-mail: ¹pakpoom_m@hotmail.com

²Faculty of Information Technology
King Mongkut’s University of Technology North Bangkok
Bangkok, Thailand
e-mail: ²montean.r@it.kmutnb.ac.th

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].

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".

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.

\[ \text{Gray} = 0.3R + 0.6G + 0.1B \]  \hspace{1cm} (1)

where \( \text{Gray} \) 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.

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).

\[ D_H(x, y) = [-1 \ 0 \ 1] \]  \hspace{1cm} (2)

where \( D_H(x, y) \) is gradient value in horizontal axis, \( D_V(x, y) \) 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” (\( k=4 \)).

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 \( k^2=16 \)).

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

\[ M_G(x, y) = \sqrt{D_H(x, y)^2 + D_V(x, y)^2} \]  \hspace{1cm} (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 \( D_V(x,y) \) is gradient value in vertical axis, respectively.

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

\[ O_c(x, y) = \tan^{-1}\left(\frac{D_H(x, y)}{D_V(x, y)}\right) \]  \hspace{1cm} (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.

$$\text{Descriptors} = \int M_G(x,y)*O_G(x,y) \quad (6)$$

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 $\{(x_1,t_1),\ldots,(x_k,t_k)\}$ where $x_k = [x_{k1},\ldots,x_{kk}]^T$, $I_k = [I_{k1},I_{k2}]$, $k = 1,2$ (for the training set).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning rate ($\eta$)</td>
<td>0.09</td>
</tr>
<tr>
<td>Error threshold ($E_{\text{th}}$)</td>
<td>0.0020</td>
</tr>
<tr>
<td>Maximum epoch ($L_{\text{max}}$)</td>
<td>81</td>
</tr>
</tbody>
</table>

TABLE I. CONFIGURATION OF PARAMETERS IN DEEP LEARNING

Step 1: Identify the deep learning, the learning rate ($\eta$) as a small value convergence to 0, Least-error value as $E_{\text{th}}$ (error threshold), and iteration of learning as $L_{\text{max}}$ (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).

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

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

\[
\text{Precision} = \frac{TP}{TP + FP} \tag{8}
\]

\[
\text{Recall} = \frac{TP}{TP + FN} \tag{9}
\]

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}
\]

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>
<thead>
<tr>
<th>Accuracy</th>
<th>Recall</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.912</td>
<td>0.839</td>
<td>0.845</td>
</tr>
<tr>
<td>0.953</td>
<td>0.933</td>
<td>0.922</td>
</tr>
<tr>
<td>0.931</td>
<td>0.844</td>
<td>0.812</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>0.913</td>
<td>0.934</td>
<td>0.891</td>
</tr>
</tbody>
</table>

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: ๙ทัศนาทัพ) 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: พระอารามหลวง), First class (Thai: พระอารามวิหาร), Second class (Thai: พระอารามวิหาร) and Third class (Thai: พระอารามแลว) respectively.

REFERENCES


