ECTI Transactions on Computer and Information Technology
Journal homepage: https://ph01.tci-thaijo.org/index.php/ecticit/
Published by the ECTI Association, Thailand, ISSN: 2286-9131

Spatial-Frequency Redistribution-based Saliency Region Detection for Thai Text Localisation

Taravichet Titijaroonroj[1] Ungsumalee Suttapakti [2] and Walairach Nunsong[3]

ABSTRACT

Saliency region detection plays an important role in computer vision applications and related areas, such as human fixation, figure-ground separation, face detection, and image compression. Nevertheless, state-of-the-art methods can moderately detect the saliency regions of a psychological pattern dataset. Therefore, this paper proposes a spatial-frequency redistribution (SFR) method to improve the efficiency of the detection of the saliency regions. The proposed SFR method consists of two main procedures: (i) adaptive cosine image construction-and-implantation and (ii) saliency region detection using complete multiple-object-implanted images. The former procedure constructs an adaptive cosine image by using a cosine function based on an object structure and then individually implants it into each object detectable. This stage provides the first significant property, spatial redistribution, to the object implanted. The adaptive cosine image redistributes the objects for controllability. Then, the latter procedure transforms a complete multiple object-implanted image into the frequency domain. At this point, the second significant property, frequency redistribution, provides the simple technique for identifying and separating the target object from the unwanted objects and background. In this paper, these properties are referred to as spatial-frequency redistribution. This method was realized as a computer program, and then such program was tested with a psychological pattern dataset and a Thai text dataset. The experimental results, when compared with state-of-the-art methods, show that the proposed SFR method can achieve the clear detection of the attention region in the psychological pattern dataset. Moreover, the proposed method can achieve a F-value of 80.28% with a recall and precision of 76.32% and 85.23%, respectively, for Thai text localisation in the Thai text dataset.

Article information:

Keywords: Spatial-frequency Redistribution, Adaptive Cosine Image, Complete Multiple-object-implanted Image, Saliency Region Detection, Thai Text Localisation

Article history:
Received: September 4, 2021
Revised: December 20, 2021
Accepted: March 1, 2022
Published: March 31, 2022
(Online)

DOI: 10.37936/ecti-cit.2022161.245320

1. INTRODUCTION

Saliency region detection is a process for predicting an attractive area known as the visual attention of an input image. It is currently applied in many applications, such as human fixation [1], figure-ground separation [2], face detection [3, 4], and image compression [5]. Many papers have proposed saliency detection algorithms that can be categorized into two groups. In the first group, a top-down approach based on higher-order factors, namely semantic, structure, event, and action, is used to segment a whole image into an attention region [6, 7]. In the second group, a bottom-up approach based on low-level features, such as intensity, color, and orientation, is used to merge a set of patches into an attention region [8-13].

The bottom-up approach has been studied by many researchers, since scanning the attention area of the bottom-up approach is faster than that of the top-down approach [14]. In 1998, Itti et al. [11] introduced a model of saliency-based visual attention by using a linear combination of intensity, color, and orientation features. Itti's method splits a complex natural scene into a set of smaller patches so that it can

[1] Faculty of Information Technology, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand 10520, E-mail: taravichet@it.kmitl.ac.th.
[2] Faculty of Informatics, Burapha University, Chon Buri, Thailand 20131, E-mail: ungsumalee.su@informatics.buu.ac.th.
[3] Faculty of Science and Technology, Rajamangala University of Technology Srivijaya Nakhon Si Thammarat Campus, Nakhon Si Thammarat, Thailand 80110, E-mail: walairach.n@rmutsv.ac.th.
[3] Corresponding author: walairach.n@rmutsv.ac.th.

speed up the selection of the attention area. However, when psychological pattern images are used to test the efficiency of Itti's method, the results of detecting attention regions are not clear, as reported in [12, 13]. Perhaps, the use of low-level features alone cannot completely reflect the high-level features for attention region detection [4]. In 2007, Zhang et al. [1] demonstrated the application of a natural statistic for finding the saliency region in a whole image. Zhang's method defines the potential area based on learning through experience, which means that it requires a training dataset. In addition, this method is incorporated between the top-down information and the bottom-up information as a term of log-likelihood. Zhang's method is claimed to be more efficient due to the use of information from both sources. However, the performance of this method primarily results from the training data; therefore, it may possibly fail in the case of insufficient training data. In 2007, Hou et al. [12] presented the saliency region detection based on the Fourier domain. This method obtains the saliency region by using a concept of removing the spectrums that are frequent occurrences while keeping the spectrums that deviate from the norm. Consequently, the saliency region can be easily detected by subtracting the frequent spectrums from the original one. The subtracted spectrums, known as a spectral residue (SR), can represent the saliency region. In addition, Hou's method does not require prior knowledge from any objects. With this concept, the performance of Hou's method is fast and robust, but it is not effective in cases of large saliency, as investigated in [13]. Later, in 2012, the extension of the saliency region based on the frequency domain was again presented by Hou et al. [2]. It is an alternative method in the frequency domain known as sparse salient region (SAL) that attempts to highlight the salient region based on figure-ground separation. In this technique, the foreground of a gray-scale image is considered as the salient region in visual attention. It can be approximately isolated by using a sign of discrete cosine transform (DCT) coefficients. This coefficient sign is determined from the DCT phase spectrum, which means that the salient region can be detected by the DCT phase spectrum. In addition, it is claimed that using this method reduces the computational cost and that the salient region detection results and human perception are closely similar. However, in the frequency domain, the SR and SAL methods suppress the amplitude spectrum and enhance the phase spectrum, respectively. This is the reason that the performance of such methods is inadequate as reported in [13, 15]. In 2013, Li et al. attempted to tackle the problems in the SR method as mentioned by using scale-space analysis in conjunction with hypercomplex Fourier transform (HFT) for improving the efficiency of saliency region detection. This method can predict human fixation more accurately, but some failures still occur, as reported in [13], because this approach only uses the low-level features in the frequency domain. After that, the modified spectrum scale space [16], which is extended from Li's method based on the adaptive multiscale Gaussian filter, was presented in 2019 to suppress the Fourier spectrum of the non-saliency region. Thus far, most papers have used either spatial-domain features [1, 11] or frequency-domain features [2, 12, 13, 16-19]. However, the use of only one domain feature is insufficient to improve the detection efficiency.

As mentioned in the previous paragraph, the bottom-up approach of the saliency region detection has continually been of interest and investigated for almost two decades. Most of these methods work only moderately well with a psychological pattern dataset, since the higher-order structure cannot reflect the low-level features. This means that some low-level features are lost or distorted in each object. This phenomenon leads to the incorrect or inaccurate saliency region detection. In order to improve the efficiency of the saliency region detection algorithm, this paper mainly focuses on the bottom-up approach, but the object structure in the higher-order factors of the top-down approach is also used for efficiency improvement.

As recently mentioned, this paper proposes a spatial-frequency redistribution (SFR) method for saliency region detection. The proposed method consists of two main steps: the first step is constructing an adaptive cosine image generated by a cosine function in order to implant it into a target object. At this stage, the structural features of an object, such as curvature, edge, and ratio, are used to generate an individual cosine image for each individual object. As a result, the spatial information of the object-implanted image is redistributed to a controllable platform when it is transformed into the frequency domain. This concept provides a spatial-frequency redistribution that easily separates objects into two classes, which are wanted and unwanted objects. Also, the structural feature in the spatial domain used by the cosine function provides a simple method to control such a feature in the frequency domain. The second step is identifying the saliency region based on the Fourier spectrum. The correct and accurate saliency regions are identified by means of an accumulate matrix that is constructed from the object-implanted image. Certainly, in order to accurately detect the saliency region, the similar and different patterns are also designed to be incorporated with the accumulate matrix. With this concept, the novel contributions of the proposed method to improve the performance of saliency region detection are: (i) the use of the spatial-frequency redistribution for providing a controllable platform, and (ii) the definition of the similar and different patterns for more accurate identification. Also, when using the bottom-

up approach, natural or forest scene images are one of the most difficult tasks to clearly isolate objects from, as reported in [14]. In the same way, Thai text localisation with text characters in a natural scene image can be viewed as the saliency region in cases with a similar pattern that needs to be detected and segmented from the surrounding areas. For this reason, the SFR method is also designed for Thai text localisation.

The proposed method was evaluated with three experiments. The first two experiments tested with psychological pattern images and Thai text images in natural scenes. The experimental results show that the proposed method achieves success in the empirical response to the psychological pattern images and in predicting the text areas as well when compared with the state-of-the-art methods [1, 2, 11-13, 16-19]. The last experiment evaluated the computational time. Although its runtime is not the lowest, it is acceptable.

The remainder of this paper is organized as follows. The concept and implementation of the spatial-frequency redistribution (SFR) platform for saliency region detection are described and proposed in sections 2 and 3, respectively. The experimental results are reported and discussed in section 4. The limitations are explained in section 5. Finally, the conclusion is presented in section 6.

2. CONCEPT OF SPATIAL-FREQUENCY REDISTRIBUTION

In this section, the concept of the spatial-frequency redistribution (SFR) is proposed to improve the performance of saliency region detection, especially when applied to Thai text localisation. The SFR concept is based on an adaptive cosine image that is a type of period noise. The adaptive cosine image can easily identify the location of targets in the Fourier domain. In the application of object detection, a target implies a wanted object. Thus, the adaptive cosine image is very useful, both for separating wanted and unwanted objects and for grouping the same type of objects together. Furthermore, the adaptive cosine images not only can be generated for supporting a large number of different patterns, but they can be easily identified for supporting a variety of objects as well. The adaptive cosine image can be constructed by Eq. (1).

$$I_{cos}(x, y) = \alpha \times \cos(\varphi(x, y)) \quad (1)$$

where I_{cos}, α, φ, and (x,y) are the adaptive cosine image, the amplitude, the angular frequency, and the spatial domain coordinates, respectively. The angular frequency component can be expressed as Eq. (2).

$$\varphi(x, y) = \rho \times (x + \gamma \cdot y) \quad (2)$$

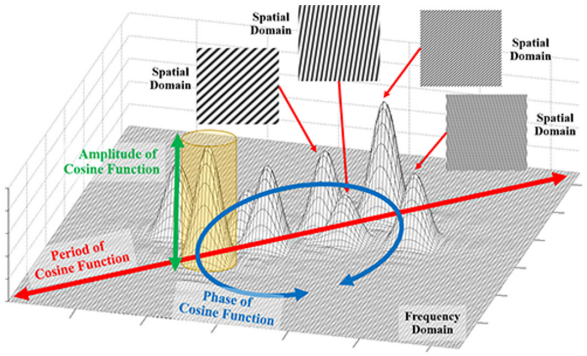


Fig.1: *Demonstration of a role of period, phase, and amplitude of a cosine function in frequency domain.*

where ρ and γ are the period and phase parameters, respectively. These parameters play an important role in controlling the location of a cosine spectrum in the frequency domain as illustrated in Fig. 1. Suppose that an adaptive cosine image generated by the cosine function is transformed into the Fourier domain; thus, its spectrum location is directly associated to the period and phase parameters. That is, the movement of this spectrum in the Fourier domain can be done by controlling these parameters. Another of the cosine function is as an amplitude component. The amplitude of the cosine function and its magnitude of the Fourier spectrum have a direct association; both increase or decrease simultaneously. Here, this fundamental characteristic property is called "spatial-frequency redistribution" and can be summarized as follows.

1. The period value can be used to define the attended location of the Fourier spectrum.
2. The phase value can be used to control the circular translation of the attended location of the Fourier spectrum.
3. The amplitude value directly relates to the height of the Fourier spectrum.

When this property is applied to the saliency region detection, it provides a simple and efficient method. To clearly understand these parameters in terms of the cosine function, they can be explained as follows.

2.1 Period of a Cosine Function

Period is one of the fundamental parameters in a cosine function that can generate different cosine images. In the spatial domain, the period is used to define the number of waves, as depicted in Figs. 2(a)-2(d). At the same time, it is also used to define the location of the high spectrum area in the Fourier domain as illustrated in Figs. 2(e)-2(h). When the period is set to a low value as shown in Fig. 2(a), its spectrum area is nearly located at the centroid of the Fourier domain as illustrated in Fig. 2(e). On the other hand, when the period value is constantly increasing from 0.4 to 1.0 as depicted in Figs. 2(b)-

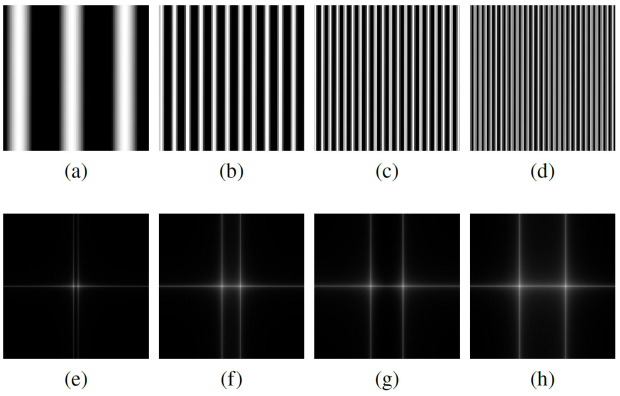


Fig.2: *Illustration of an adaptive cosine image in four different periods in the spatial domain: (a) ρ = 0.1, (b) ρ = 0.4, (c) ρ = 0.7, and (d) ρ = 1.0, and in the frequency domain: (e) ρ = 0.1, (f) ρ = 0.4, (g) ρ = 0.7, and (h) ρ = 1.0.*

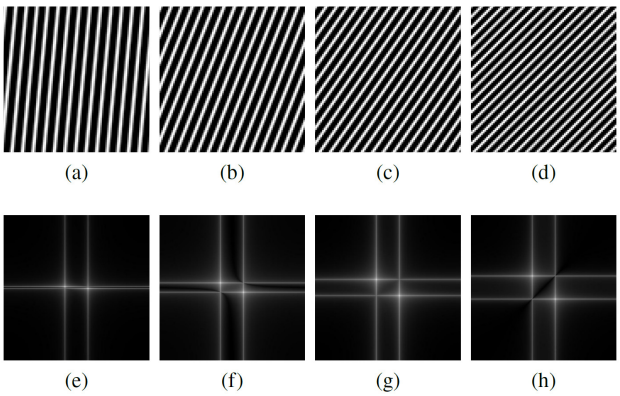


Fig.3: *Illustration of an adaptive cosine image in four different phases in the spatial domain: (a) γ = 0.1, (b) γ = 0.4, (c) γ = 0.7, and (d) γ = 1.0, and in the frequency domain: (e) γ = 0.1, (f) γ = 0.4, (g) γ = 0.7, and (h) γ = 1.0.*

2(d), the spectrum areas are also constantly further from the centroid of the Fourier domain, as illustrated in Figs. 2(f)-2(h). This becomes a powerful parameter in controlling the location of a high spectrum area in the Fourier domain.

2.2 Phase of a Cosine Function

Phase of a cosine function is a parameter that is used in controlling the rotation of the spatial frequency in the spatial domain and the circular translation of the high spectrum area in the frequency domain. In Fig. 3, the spatial-frequency is rotated clockwise if the phase is less than "0". Conversely, the spatial-frequency is rotated counterclockwise if the phase is greater than "0". Moreover, the rotation degree depends on the phase value; that is, both rotation degree and phase value increase or decrease together. We can say that the rotation degree relies on the phase value. For object detection, the phase parameter is used to move wanted and unwanted objects into different areas in the frequency domain. In other words, the location of unwanted and wanted objects is translated into circular direction in different areas in the Fourier domain.

2.3 Amplitude of a Cosine Function

Amplitude of an adaptive cosine image is defined as an α component. The relationship of the intensity in the adaptive cosine image and the amplitude in the cosine function is a linear association; that is, intensity and amplitude increase or decrease simultaneously. As illustrated in Fig. 4, the brightness of the adaptive cosine image in Fig. 4(a) is darker than that of the ones in Figs. 4(b)-4(d), because the amplitude of the cosine function in Fig. 4(a) is lower than that of the ones in Figs. 4(b)-4(d). Additionally, the amplitude in the spatial domain is also corresponds to the height of the spectrum area in the Fourier domain. As shown in Figs. 4(e)-4(h), the spectrum and amplitude rise or fall together. This property is very useful for controlling the increase or decrease of spectrums of the unwanted and wanted objects.

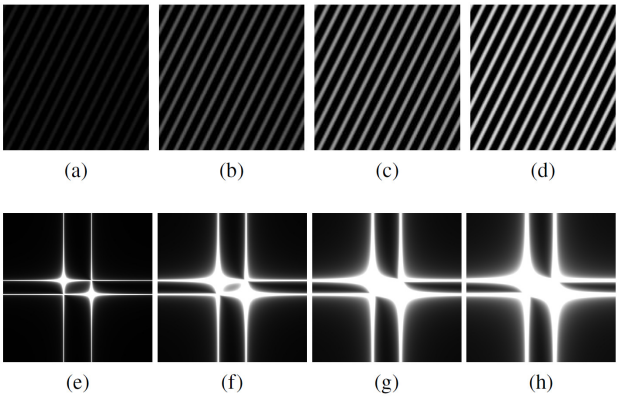


Fig.4: *Illustration of an adaptive cosine image in four different amplitudes in the spatial domain: (a) α = 0.10, (b) α = 0.35, (c) α = 0.60, (d) α = 0.85, and in the frequency domain: (e) α = 0.10, (f) α = 0.35, (g) α = 0.60, (h) α = 0.85.*

3. IMPLEMENTATION OF SPATIAL-FREQUENCY REDISTRIBUTION-BASED SALIENCY REGION DETECTION

As described in section 2, the concept of the spatial-frequency redistribution (SFR) method can be implemented with the following two main procedures: (i) adaptive cosine image construction-and-implantation and (ii) saliency region detection.

3.1 Adaptive Cosine Image Construction and Implantation

In this procedure, an adaptive cosine image is synthesized by the cosine function and is then implanted

into all objects to detect a target object or a wanted object. Hereafter, for ease of explanation, the term "synthesized cosine image" is used. The synthesized cosine image can be generated in a variety of patterns. Thus, each object can be implanted with an individual synthesized cosine image. This is the main reason that it is called the adaptive cosine image in this paper. With this concept shown in Fig. 1, it is not only useful to separate the wanted and unwanted objects into different areas in the Fourier domain, but also to group the same objects into the same areas. For a clear explanation of constructing the synthesized cosine images and implanting them into objects, a psychological pattern image as shown in Fig. 5(a) is used as a sample. Mathematically, let ς_i be an object-implanted image, and the complete multiple-object-implanted image can be expressed as Eq. (3).

$$\varsigma = \sum_{i=1}^{n} \varsigma_i \tag{3}$$

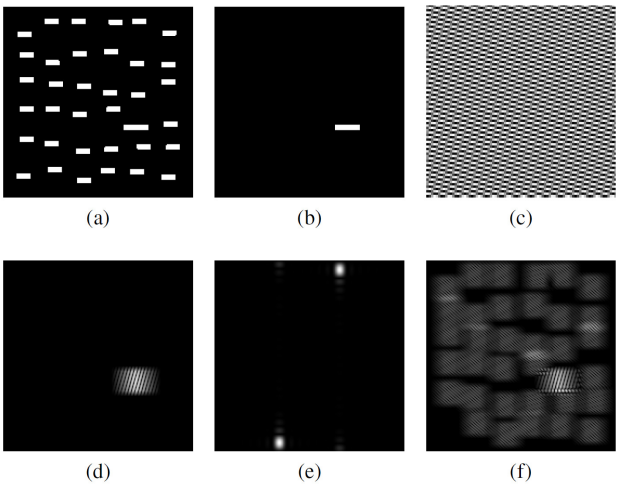


Fig.5: *Illustration of the stages of saliency region detection using a complete multiple-object-implanted image: (a) a complete multiple-object-implanted image in the Fourier domain, (b) an accumulated matrix, (c) a binary matrix or the candidate regions, (d) a complete weighted multiple-object-implanted image in the Fourier domain, (e) a saliency region in Fourier domain, and (f) a saliency region in the spatial domain.*

The adaptive cosine construction-and-implantation procedure can be described as follows.

Step 1: Detect each object of an input image by using the Moore-Neighbor tracing algorithm modified by Jacob's stopping criteria, which is employed to detect the wanted object and extract its structure features for generating an individual synthesized cosine image in each object, and then binarize the detected object by using the Otsu algorithm [20]. The outcome I_i of this step is a binary-object image as shown in Fig. 5(b).

Step 2: Construct an individual synthesized cosine image t_i for each binary object by using Eq. (4).

$$t_i(x, y) = \alpha \times cos(\rho(x + \gamma \cdot y)) \tag{4}$$

where α, ρ, and γ are the amplitude, a period, and a phase, respectively. These variables can be defined according to applications as described in section 4 and 5. This stage yields the output as shown in Fig. 5(c).

Step 3: Implant t_i into I_i by means of Eq. (5) to obtain the object-implanted image.

$$\varsigma_i = I_i \circ t_i \tag{5}$$

where $\circ$ is a Hadamard product operator. Figs. 5(d) and 5(e) illustrate the results of an object-implanted into the spatial and frequency domains, respectively.

Step 4: Repeat Steps 2 to 3 until all detected objects of the input image are implanted with individual synthesized cosine images.

As mentioned in section 1, the proposed method attempts to handle frequency-domain features through spatial-domain controllable features. This manipulation can be done as in the following description. Here, the wanted and unwanted objects of Fig. 5(a) are a longer rectangular shape and a regular shape, respectively. In this case, the wanted object can be viewed as a saliency region. To detect the wanted object and group the unwanted objects together, three components, namely the period, phase, and amplitude, of a cosine function play an important role in generating an individual cosine image for each individual object. This concept can be realized by the use of the structural features, clearly described in section 4, of each object in the spatial domain to construct the individual cosine image. Then, each object is implanted with its individual cosine image. We believe that the similar objects are implanted by the similar cosine image while the different objects are implanted by the different cosine images, thus making them simple to classify the different objects and to group the similar objects. Finally, a complete multiple-object-implanted image in the spatial domain is exhibited in Fig. 5(f).

3.2 Saliency Region Detection Using a Complete Multiple-object-implanted Image

This procedure is designed to identify the saliency region of a complete multiple-object-implanted image. At this stage, when such an image is transformed into the Fourier domain, its patterns can be categorized into two types: saliency and repeated pattern regions. The repeated pattern region is a pattern that is a more frequent occurrence in a whole image, wher-

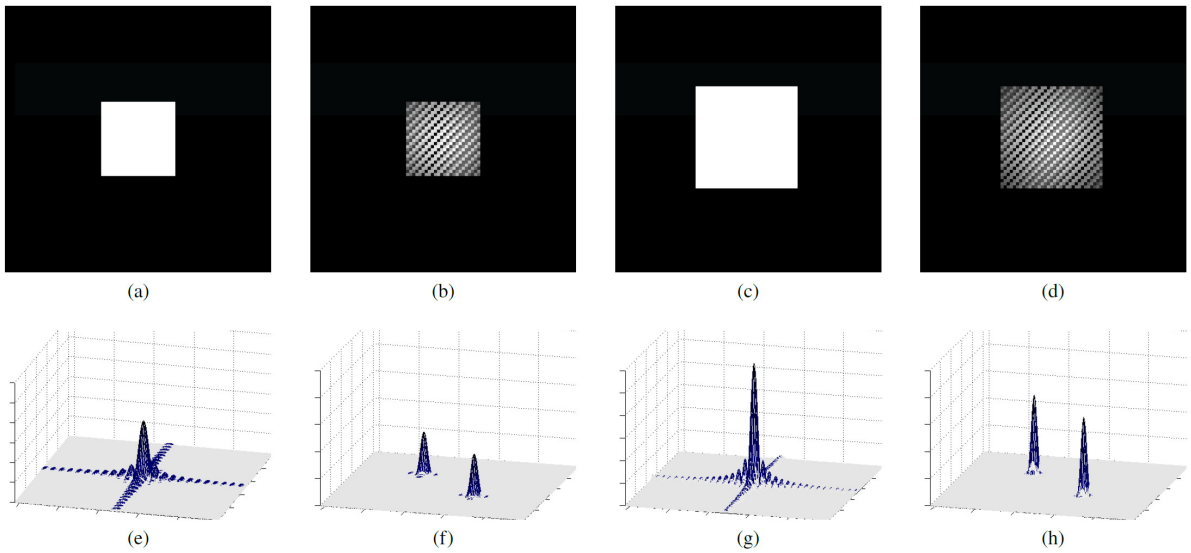


Fig.6: *Spatial-frequency redistribution: (a) a square object image with a size of 23×23 pixels, (b) a spatial redistribution of Fig. (a) implanted by an individual cosine image, (c) a square object image with a size of 32×32 pixels, (d) the spatial redistribution of Fig. (c) implanted by an individual cosine image, (e) the frequency spectrum of Fig. (a), (f) the frequency spectrum redistribution of Fig. (b), (g) the frequency spectrum of Fig. (c), and (h) the frequency spectrum redistribution of Fig. (d).*

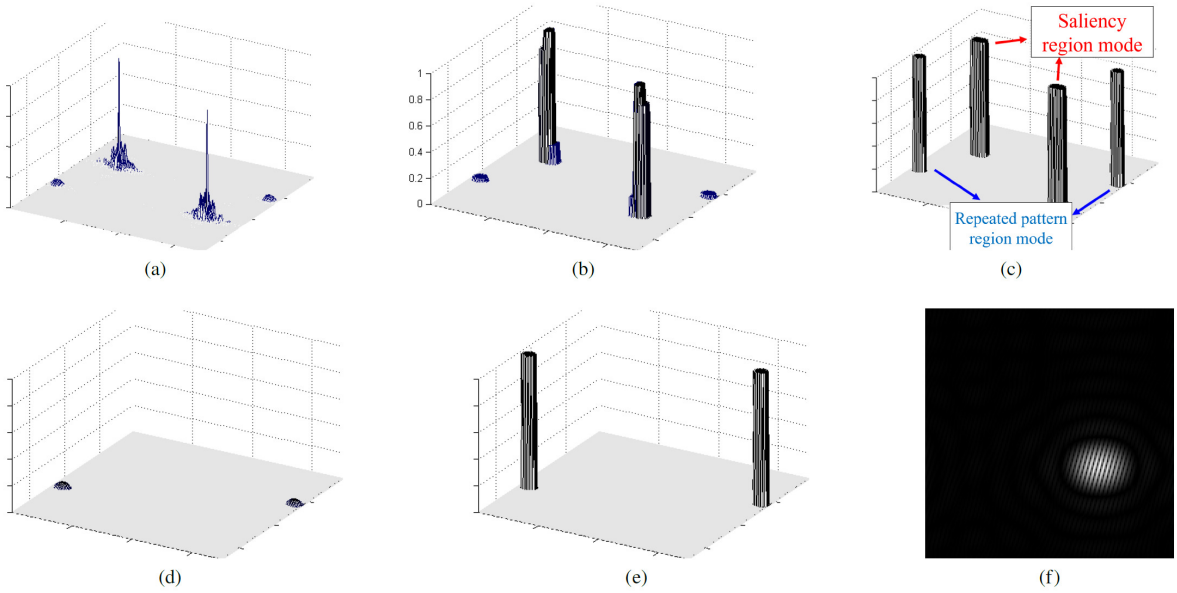


Fig.7: *Illustration of the stages of saliency region detection using a complete multiple-object-implanted image: (a) a complete multiple-object-implanted image in the Fourier domain, (b) an accumulated matrix, (c) a binary matrix or the candidate regions, (d) a complete weighted multiple-object-implanted image in the Fourier domain, (e) a saliency region in Fourier domain, and (f) a saliency region in the spatial domain.*

eas the saliency region is a pattern that is a less frequent occurrence in a whole image. Based on this definition, two modes of the saliency region detection using complete multiple-object-implanted image are proposed to support a variety of applications. The first mode is designed for identifying the saliency region of the object-implanted in a whole image, whereas the second mode is designed for identifying the repeated pattern region of the object-implanted in a whole image. However, identifying the saliency region in the frequency domain is sensitive to the magnitude of the spectrum of an object-implanted image. This phenomenon can be demonstrated by applying a Fourier transform to two different sized objects, 23×23 and 32×32 pixels, as shown in Figs. 6(a) and 6(c), respectively. The magnitude of spectrum of the smaller object in Fig. 6(e) is lower than that of the larger object in Fig. 6(g), even though the

shape of the two images is similar. It also occurs on the complete multiple-object-implanted image in different sizes that is illustrated in Figs. 6(b) and 6(d) in the Fourier domain and in Figs. 6(f) and 6(h) in the spatial domain. This leads to the incorrect and inaccurate identifying of the saliency region. Hence, to avoid and solve this problem, the object-implanted image is weighted by an accumulate matrix to obtain an accurate saliency region. Let ξ be a complete multiple-object-implanted image in the frequency domain as depicted in Fig. 7(a) and $FT(\cdot)$ be a discrete Fourier transform, ξ is computed by Eq. (6).

$$\xi = FT(\varsigma) \tag{6}$$

The procedure can be described as follows.

Step 1: Calculate an accumulate matrix M based on the object-implanted image by using Eq. (7).

$$M = \sum_{i=1}^{n} F_{otsu}(FT(\varsigma_i)) \tag{7}$$

where ς_i and $F_{otsu}(\cdot)$ are the object-implanted image and the Otsu algorithm, respectively.

Step 2: Normalize the accumulate matrix M by using Eq. (8).

$$M'(u,v) = \frac{M(u,v) - m_{min}}{m_{max} - m_{min}} \tag{8}$$

where m_{max} and m_{min} are the maximum and minimum values of the accumulate matrix. The normalized accumulate matrix M' is shown in Fig. 7(b).

Step 3: Compute a binary matrix M'' by using Eq. (9).

$$M''(u,v) = \begin{cases} 1, & M'(u,v) \geq T_0 \\ 0, & M'(u,v) < T_0 \end{cases} \tag{9}$$

where T_0 is a thresholding for the binarization. At this step, the result is shown in Fig. 7(c).

Step 4: Detect a set of candidate regions $R = \{r_1, r_2, r_3, ..., r_n\}$ by using the Moore-Neighbor tracing algorithm from the binary matrix M''.

Step 5: Identify a saliency region for the first mode using Eq. (10).

$$\mathfrak{M} =_{\mathfrak{n}} AVG(\xi'(r_i)) \tag{10}$$

where $\mathfrak{M}$ is a saliency region in the first mode as shown in Fig. 7(c). $AVG(\cdot)$ is an average function, and ξ' is the complete weighted multiple-object-implanted image in the frequency domain in region r_i as shown in Fig. 7(d), which is computed by using Eq. (11).

$$\xi' = \xi \circ M' \tag{11}$$

However, the equation (10) is replaced by Eq. (12) when the repeated pattern region is required for detection in the second mode.

$$\mathfrak{M} =_{\mathfrak{n}} AVG(\xi'(r_i)) \tag{12}$$

Step 6: Identify a saliency region from the output region and the complete multiple-object-implanted image using Eq. (13).

$$\tilde{\varsigma} = IFT(\xi \circ \mathfrak{M}) \tag{13}$$

where $\tilde{\varsigma}$ and $IFT(\cdot)$ are the saliency region and the inverse Fourier transform, respectively.

For testing the response to a psychological pattern that is the less frequent occurrence of an object on the whole image, this procedure is therefore set to the first mode for saliency region detection as displayed in Fig. 7(e). When the original image in Fig. 5(a) is applied to the proposed method, the detectable result in Fig. 7(f) is the saliency region. Moreover, the Fourier spectrum of the complete multiple-object-implanted image is redistributed when compared to that of the original image. The result in Fig. 8 shows that the distribution of the original image is gathered in the centroid of the frequency domain while that of the complete multiple-object-implanted image is redistributed over the frequency domain. This is called "spatial-frequency redistribution" which can relocate the wanted and unwanted objects into different areas.

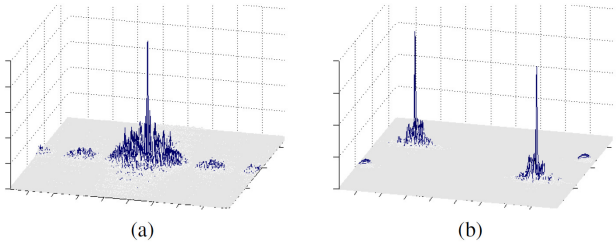


Fig.8: *Frequency redistribution: (a) Fourier spectrum of Fig. 5(a) and (b) Fourier spectrum of Fig. 5(f).*

4. EXPERIMENTS AND DISCUSSION

In order to evaluate the performance of the proposed method, three experiments were set up for each of the aspects. The first was designed to test an empirical response to 9 psychological pattern images for attention region detection. The second experiment was designed to measure a precision, a recall, and an F-value for 300 natural scene images for Thai text localisation. The last experiment aimed to evaluate the processing time. We implemented the tests and

Table 1: *Responses of the proposed SFR method when compared to the state-of-the art methods in testing with nine psychological pattern images.*

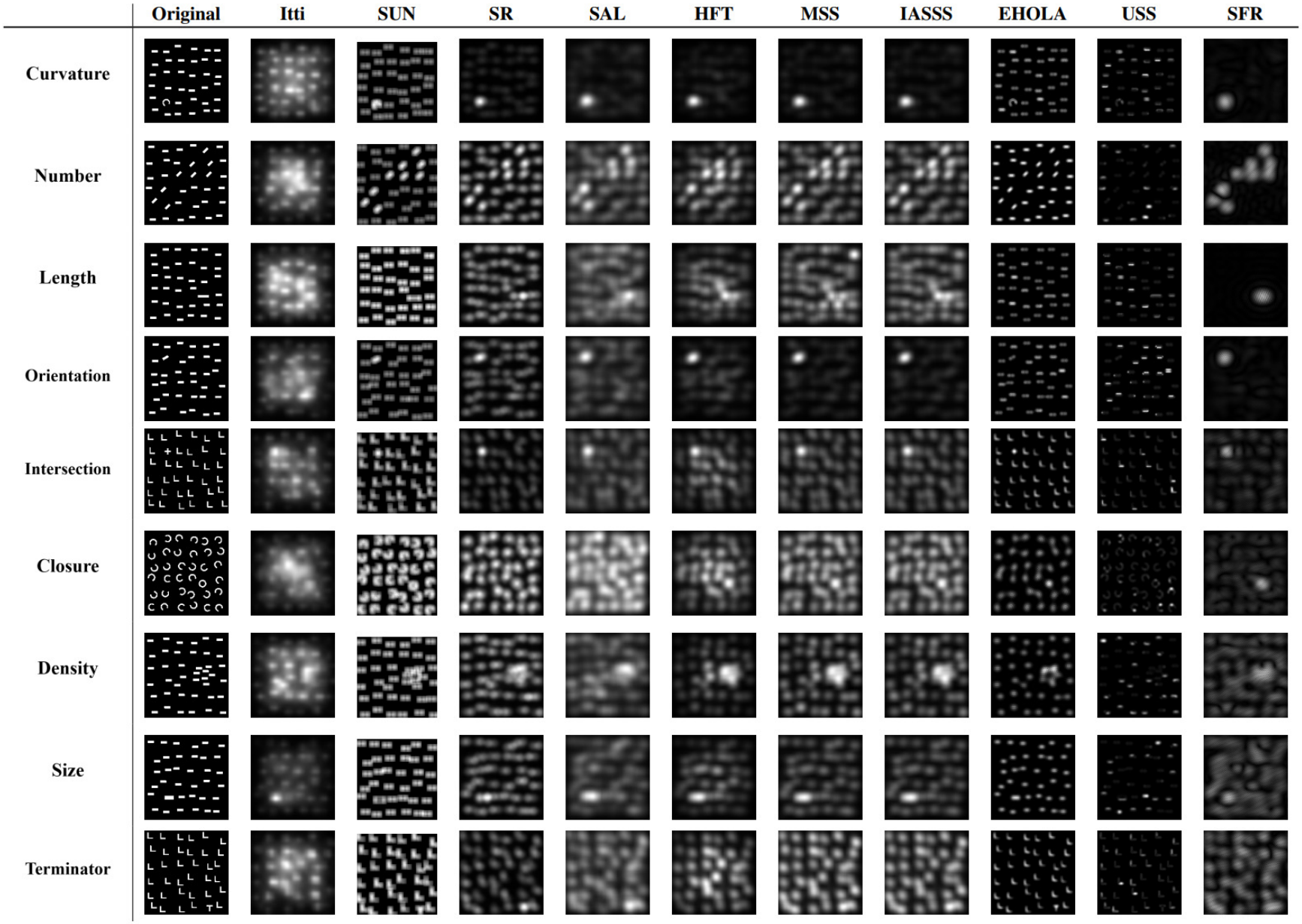


Table 2: *Comparison of the responses of all methods when tested by 9 psychological pattern images.*

	Itti	**SUN**	**SR**	**SAL**	**HFT**	**MSS**	**IASSS**	**EHOLA**	**USS**	**SFR**
Clear	0	0	1	1	1	1	1	0	0	4
Moderate	1	4	7	7	6	6	6	4	0	4
Unclear	8	5	1	1	2	2	2	5	9	1

ran them on a machine with Intel Core i5-830H @2.30GHZ with DDR4 8 GBytes with Windows 10 Home Single Language.

4.1 Testing with Psychological Pattern Images

The first experiment aimed to test the empirical response to 9 psychological pattern images including Curvature, Number, Length, Orientation, Intersection, Closure, Density, Size, and Terminator images as shown in Table 1, column name "Original". They were used to evaluate the basic ability for saliency region detection as reported in [2, 5, 11-13, 21, 22]. A pixel resolution of this dataset is 175×176. The proposed method is compared the effectiveness against the ability of five state-of-the-art methods, namely Itti [11], SUN [1], SR [12], SAL [2], HFT [13], MSS [16], IASSS [17], EHOLA [18], and USS [19]. Those methods are implemented as baselines. For fairness of comparison, the parameter setting of the baseline algorithms is based on their recommendation in [1, 2, 11-13, 16-19]. On the other hand, the parameters of the proposed method are defined by the structural feature in each object as shown in Eqs. (14), (15), and (16), respectively.

$$\rho = \frac{h}{w} \tag{14}$$

$$\gamma = \frac{l}{2 \times (h + w)} \tag{15}$$

$$\alpha = \frac{4\pi a}{l^2} \tag{16}$$

where a, h, w, and l are an area, a height, a width and a number of pixels in a boundary of objects, respectively. In addition, the procedure of the saliency region detection using complete multiple-object-implanted image is set to the first mode by using Eq. (10) and T_0 is set to "0.001". In pattern recognition, the various sized objects have a great impact on the performance in many applications, such as object detection, human detection, and character recognition. This becomes to be a key factor for structural feature selection. In order to make the proposed method robust to size invariances, the set of the structural features is used as follows. First, a ratio between height and width of an object is a powerful feature as recommended in [23]. It can be represented as a shape feature. In [23], they reported that this feature can isolate the wanted object from the unwanted object in a text detection application. As a result, it shows that the ratio between height and width called "HW-ratio" can represent the structure of an object. For this reason, it is used to serve as the period parameter in the cosine function. Second, the number of pixels on boundary edge is one kind of the structural features. Not only can it be represented as a texture feature, but also it can be referred to the density of pixels on the boundary edge. This feature can successfully separate the wanted and unwanted objects when applied to text detection in natural scene images as reported in [24], thus making it interest to serve as the phase in the cosine function. Last, the compactness of an object is a useful feature that can capture the information from an object boundary [25]. It is frequently used in superpixel segmentation. In other words, it can isolate the foreground from the background. Consequently, this feature is used to set the amplitude parameter in the cosine function. Here, we can conclude that those features are used to feed in the cosine function, since they can represent the structural features invariant to object size.

In this paper, an identification criterion for saliency region detection is defined as three levels: clear, moderate, and unclear. Based on Table 1, it is evident that the proposed method can clearly identify attention areas in four images: Orientation, Number, Length, and Curvature, and can moderately identify attention areas in other four images: Size, Intersection, Density, and Closure. However, it cannot clearly identify attention areas in the Terminator image. In the same way, SR, SAL, HFT, MSS, and IASSS methods achieve only in Curvature image. Itti's method can moderately respond to Size and Intersection images only, while SUN method can moderately respond to other four images: Orientation, Number, Density, and Curvature. For other five methods, the response results are not too much different; 7 images in moderate level for SR and SAL methods, and 6 images in moderate level for HFT, MSS, and IASSS methods. In unclear level, Itti's method is unsuccessful to detect an attention region in 8 images, whereas SUN method is unsuccessful in 5 images. SR and SAL methods have a response failure in only one image. Correspondingly, the HFT, MSS, and IASSS methods also have two response failures in Terminator and Closure images. Additionally, the EHOLA method can moderately respond to Intersection, Closure, Density, and Size images. It has five response failures in Orientation, Number, Length, Curvature, and Terminator images. Lastly, USS method cannot identify attention areas in 9 psychological pattern im-

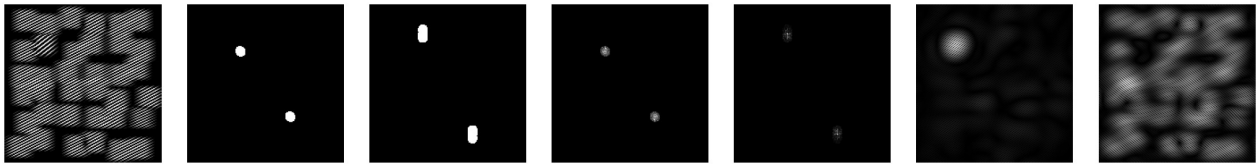


Fig.9: *Demonstration of a clear saliency region detection of the Orientation image: (a) a complete multiple object-implanted Orientation image in the spatial domain, (b) a candidate saliency region of the less frequent occurrence (first mode), (c) a candidate saliency region of the more frequent occurrence (second mode), (d) the spectrum of Fig. (b), (e) the spectrum of Fig. (c), a detectable saliency region using the first mode, and a detectable saliency region using the second mode.*

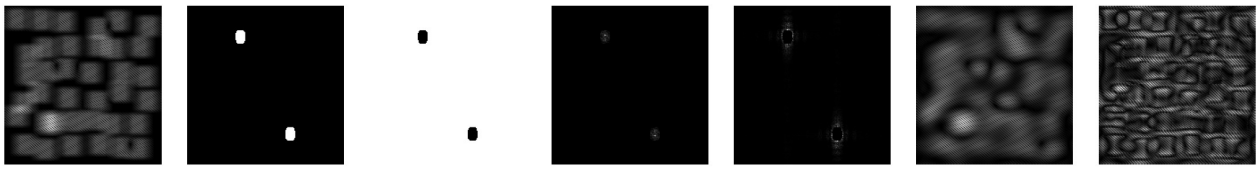


Fig.10: *Demonstration of a moderate saliency region detection of the Size image: (a) a complete multiple object-implanted Size image in the spatial domain, (b) a candidate saliency region of the less frequent occurrence (first mode), (c) a residual region of Fig. (b), (d) the spectrum of Fig. (b), (e) the spectrum of Fig. (c), (f) a detectable saliency region using the first mode, and (g) a detectable saliency region using a residual region.*

Fig.11: *Demonstration of a moderate saliency region detection of the Terminator image: (a) a complete multiple object-implanted Terminator image in the spatial domain, (b) a candidate saliency region of the less frequent occurrence (first mode), (c) a residual region of Fig. (b), (d) a spectrum of Fig. (b), (e) the spectrum of Fig. (c), (f) a detectable saliency region using the first mode, and (g) a detectable saliency region using a residual region.*

ages. Table 2 is summarized the achievement of each method in three-level criteria.

Furthermore, the achievement of the proposed method comes from the following reasons. In clear level, all object-implanted images of Orientation, Number, Length, and Curvature have much difference in spatial domain features as an example illustrated in Fig. 9(a). When such an image is transformed into Fourier domain, the candidate saliency regions are evidently clear-cut as shown in Figs. 9(b) and 9(c), thereby making them simply to identify the saliency regions as shown in Figs. 9(d) and 9(e). In this case, the saliency region of this image is an object that has the less frequent occurrence. The final saliency region is clearly detected as illustrated in Fig. 9(f). When regarded Fig. 9(g), the rest objects implanted with individual cosine images are also identified as demonstrated in Fig. 9(g). At this point we can say that the more clear implantation, the more simply identification. On the other hand, the baseline methods like SR, SAL, and HFT can achieve in Orientation image, since such methods use only the low-level features in frequency domain for detecting the saliency region.

In moderate level, as a sample illustrated in Fig. 10(a), all objects in Size image are implanted with individual similar cosine images. Consequently, the candidate saliency regions of the less and more frequent occurrences are almost occluded. Figs. 10(b) and 10(c) show a candidate saliency region of the less frequent occurrence and its residual region, respectively. When precisely consider their spectrums in Figs. 10(d) and 10(e), it is found that the spectrum sidelobe of Fig. 10(e) is radiated to surrounding area that excludes from a candidate saliency region of Fig. 10(b). Especially, the spectrum sidelobe of the more frequent-pattern occurrence is more radiated than that of the less frequent-pattern occurrence. As shown in Fig. 10(f), the intensity value of attention

area in the final saliency region is brighter than that of the residual area, since the spectrum of the less frequent-pattern occurrence is able to maintain more energy than that of the more frequent-pattern occurrence. When looking at Fig. 10(g), the brightness of a detectable saliency region spreads over a whole image. This is a reason why the saliency region of Size image is still detectable, even though both spectrum of saliency regions of the less and more frequent-pattern occurrences are largely occluded. We also can say that the adaptive cosine image implantation makes efficient frequency redistribution for object images and provides a strong discrimination for similar object-implanted images.

In unclear level, when all objects in Terminator image are implanted with individual similar cosine images as shown in Fig. 11(a). The spectrum and its sidelobe of the more and less frequent-pattern occurrences are almost identical. This makes a complete occluded spectrum between the more and less frequent-pattern occurrences. When a candidate saliency region in Fig. 11(b) is used and its spectrum shown in Fig. 11(d) is inverse-transformed into the spatial domain, as expected, all objects implanted with individual similar cosine images are detectable, but the saliency region is not detectable as resulted in Fig. 11(f). Moreover, Figs. 11(c) and 11(e) exhibit the residual region and spectrum sidelobe, respectively. When the residual region is employed, the brightness of the saliency region spreads over a whole image as displayed Fig. 11(g).

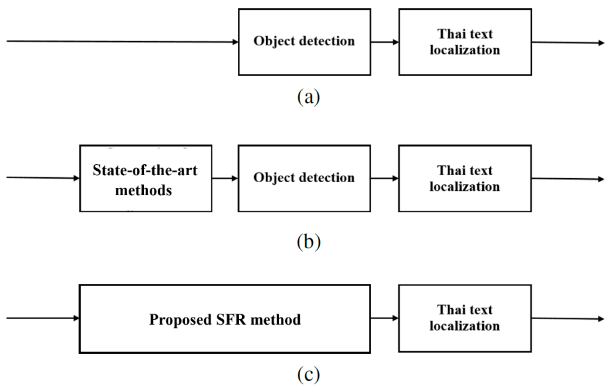


Fig.12: *Three experimental schemes designed for evaluating the performance of Thai text localisation: (a) conventional Thai text localisation scheme, (b) the use of state-of-the-art methods as a preprocessing stage, and (c) the use of the proposed SFR method as a preprocessing stage.*

4.2 Thai Text Localisation

The second experiment aims to evaluate the effectiveness of applying the proposed method to detect Thai text in natural scene dataset. As schematically presented in Fig. 12, the experimental framework is designed for two main purposes: (i) to evaluate an impact on the use of saliency region detection as a preprocessing stage of Thai text localisation scheme, and (ii) to compare to the efficiency of state-of-the-art [1,2,11-13] and proposed methods. The dataset comprised of 300 images with a pixel resolution of 480×360. It is divided into three groups: (i) GROUP#1, (ii) GROUP#2, and (iii) GROUP#3. Examples of this dataset in each group are shown in Fig. 13. This dataset is categorised in background complexity as reported in [26]. Text and background in signs of GROUP#1 dataset have a uniform color. The scenes of the GROUP#2 dataset have complex backgrounds. In the GROUP#3 dataset, scenes have illumination effects. The recall, precision, and F-value as described in [23],[26] are metrics used to evaluate the performance in text character localisation of each method. In this paper, modified connected component analysis (MCCA) is used as Thai text detection [23]. Based on Fig. 12, all methods play a role in identifying the candidate text region before feeding it into the Thai text localisation method. The parameters of the proposed method are defined by Eqs. (17), (19), and (20), respectively.

$$\rho = f(\frac{h}{w}) \tag{17}$$

$$f(s) = \begin{cases} \frac{1}{s}, & s > 1 \\ s, & s \leq 1 \end{cases} \tag{18}$$

$$\gamma = \frac{l}{2 \times (h + w)} \tag{19}$$

$$\alpha = \frac{\rho + \gamma}{2} \tag{20}$$

where ρ, α, and γ are period, amplitude, and phase parameters, respectively. $f(\cdot)$ is the structural object function for separating the unwanted object and text character into different areas and for grouping same type of objects together. Based on the constraint in [23], the ratio of height and width of a text character is nearly equal to "1" whereas the ratio of an unwanted object is nearly equal to "0" or "∞". Therefore, the structural object function and period are unavoidably defined to relocate the location of text character and unwanted object into different areas in the frequency domain. On the other hand, the amplitude, α, is defined by averaging between period and phase parameters instead of a roundness metric. Additionally, the procedure of the saliency region detection using complete multiple-object-implanted image is set to the second mode by using Eq. (12) to detect the saliency region in case of the repeated pattern and T_0 is set to "0.001".

As depicted in Table 3, the Itti's, SUN, SR, SAL, HFT and proposed methods achieve an average precision value of 81.97%, 81.19%, 83.34%, 81.03%, 83.06% and 81.91%, respectively, when compared to

Table 3: *Comparison of the performance of all methods tested by a Thai text dataset categorized into three groups.*

Method	GROUP#1			GROUP#2			GROUP#3			Average		
	Recall	Precision	F-value	Recall	Precision	F-value	Recall	Precision	F-value	Recall	Precision	F-value
MCCA	82.42	82.21	82.32	76.04	74.91	75.47	52.16	70.69	60.03	70.21	75.94	72.61
Itti	74.20	88.98	80.92	43.59	80.07	56.45	35.15	76.87	48.24	50.98	81.97	61.87
SUN	82.38	89.39	85.74	64.13	79.42	70.96	48.94	74.75	59.15	65.15	81.19	71.95
SR	81.99	89.51	85.59	61.27	84.60	71.07	40.93	75.91	53.18	61.40	83.34	69.95
SAL	70.74	88.32	78.56	56.03	79.01	65.56	44.68	75.76	56.21	57.15	81.03	66.78
HFT	79.51	90.02	84.44	56.82	82.86	67.41	46.74	76.31	57.97	61.02	83.06	69.94
SFR	80.40	88.53	84.27	71.89	79.97	**75.71**	50.07	**77.22**	**60.75**	67.45	81.91	**73.58**

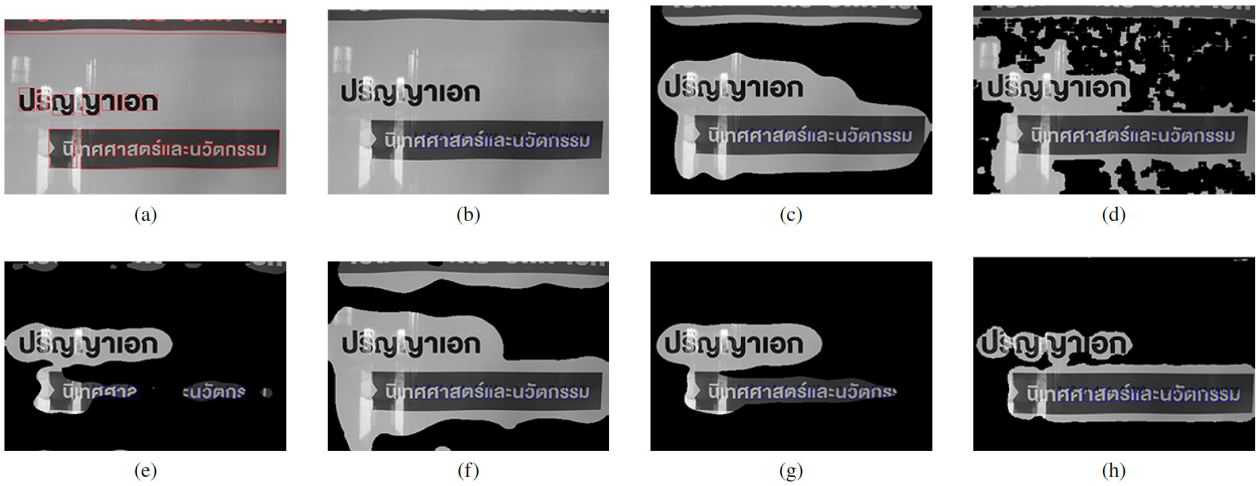
Table 4: *Comparison of the performance of all methods tested by Thai text dataset categorized into three groups.*

Method	GROUP#1			GROUP#2			GROUP#3			Average		
	Recall	Precision	F-value	Recall	Precision	F-value	Recall	Precision	F-value	Recall	Precision	F-value
MCCA	86.16	86.08	86.12	80.23	75.89	78.00	56.30	68.51	61.81	74.23	76.83	75.31
Itti	75.07	95.39	84.02	49.03	83.22	61.71	41.17	77.16	53.69	55.09	85.26	66.47
SUN	90.19	93.39	91.76	70.19	83.86	76.42	53.59	74.74	62.42	71.32	84.00	76.87
SR	88.91	93.98	91.38	76.06	86.71	81.04	51.06	76.22	61.16	72.01	85.64	77.86
SAL	75.66	94.95	84.22	62.47	84.62	71.88	52.27	76.13	61.98	63.47	85.23	72.69
HFT	84.89	94.74	89.54	64.40	86.60	73.87	53.02	78.00	63.13	67.44	86.45	75.51
SFR	89.79	94.23	**91.96**	**81.16**	83.95	**82.53**	**58.02**	77.50	**66.36**	**76.32**	85.23	**80.28**

Fig.13: *Examples of Thai text scene dataset of (a) GROUP#1, (b) GROUP#2, and (c) GROUP#3.*

that of the MCCA method. However, a trade-off between precision and recall occurs; that is, the recall values of Itti's, SUN, SR, SAL, HFT and proposed methods are 19.23%, 5.06%, 8.81%, 13.06%, 9.19%, and 2.76%, respectively, lower than that of the MCCA method. This leads to the performance of state-of-the-art methods in overall (F-value) is decreased except the proposed method that is 1% higher than the MCCA method. In this experiment, although all methods can correctly detect candidate text regions on natural scene image, the MCCA method cannot retrieve Thai text character from those regions due to the constraint in MCCA itself. This constraint effect appears as displayed in Fig. 14. Therefore, in order to set up a fairness of experiments, the 150 selected images from that Thai text character dataset are used to evaluate the performance of all methods.

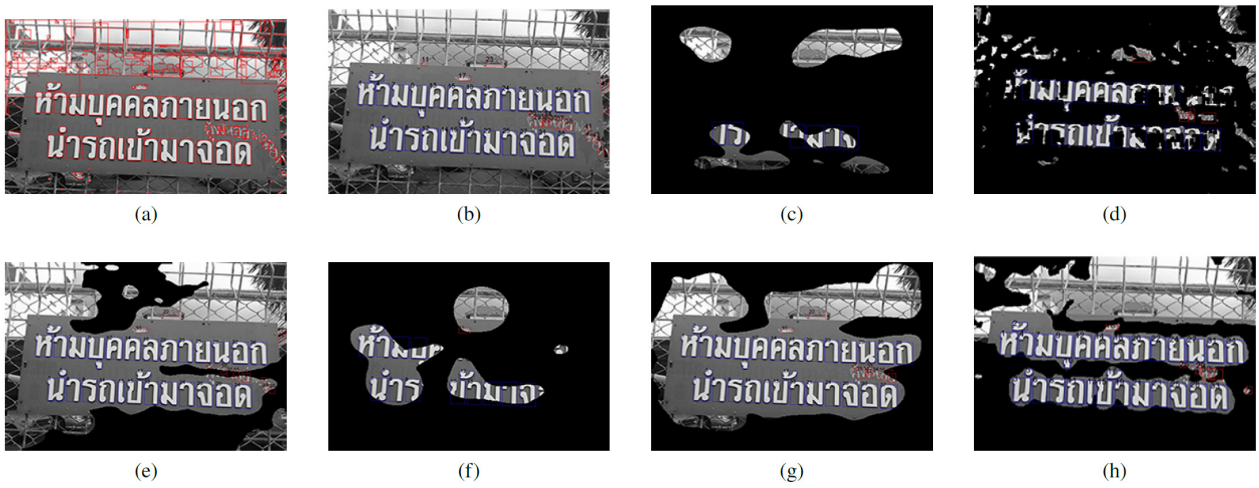
The experimental results in Table 4 shows that the precision values of the Itti's, SUN, SR, SAL, HFT, and proposed methods are 8.43%, 7.17%, 8.81%, 8.40%, 9.62%, and 8.40%, respectively, higher than that of the MCCA method, since the candidate text region detected by all methods is more accurate as empirically illustrated in Fig. 15. This improvement mainly comes from the reduction of a number of unwanted objects. However, the recall values of the Itti's, SUN, SR, SAL, and HFT methods are 19.14%, 2.91%, 2.22%, 10.76% and 6.79%, respectively, lower than that of the MCCA method. On the other hand, the recall value of the proposed method is 2.09% higher than that of the MCCA method. In this case, it is proved that the proposed method can clearly identify the text region and also can reduce a large number of unwanted objects as demonstrated in Fig. 16. Moreover, the F-values achieved by SR, SUN, HFT, and the proposed methods are 0.20%, 1.56%, 2.55%, and 4.97%, respectively, higher than that by the baseline methods except Itti and SAL that are 8.84% and 2.26% lower than the MCCA method. This success mainly comes from: (i) spatial-frequency redistribution and (ii) the definition of similar and different patterns. It is evident that spectrum of unwanted object and text character is redistributed into different areas when the proposed method is served as the preprocessing stage. This leads to the more accurate text region identification. It is evident that the performance of Thai text localisation in terms of



(a) (b) (c) (d)

(e) (f) (g) (h)

Fig.14: *Example of MCCA's constraint effect in which it cannot retrieve text characters from the text regions: (a) objects detectable, (b) MCCA, (c) Itti's method, (d) SUN, (e) SR, (f) SAL, (g) HFT, and (h) SFR.*



(a) (b) (c) (d)

(e) (f) (g) (h)

Fig.15: *Example of the experimental results to show that how the precision value achieved by all methods: (a) objects detectable, (b) MCCA, (c) Itti's method, (d) SUN, (e) SR, (f) SAL, (g) HFT, and (h) SFR.*

recall, precision, and F-value is significantly improved when the proposed method is applied.

4.3 Computational Time

This section compares the computational time of the proposed method to baseline methods [1, 2, 11-13, 16-19]. We experimented with five image sizes including 64×64, 128×128, 256×256, 512×512, and 1024×1024. The result is presented as a line graph in Fig. 17. This graph consists of two axis including time (seconds) in the y-axis and image size (pixels) in the x-axis. The computational time of each method is denoted with a different line-and-color style.

There are two types of computational time trends including (i) relatively stable and (ii) steadily increasing. In the first type, there are four methods including EHOLA, USS, IASSS, and HFT that have trends which are relatively stable on a log scale. All methods of this type handle a range of image sizes and they still take less than 10 seconds to process. This is an advantage of those methods although its computational time average is higher than one second. In the second type, the trend of the SUN, MSS, SAL, SR, Itti, and SFR steadily increased. For image sizes lower than 256 ×256, the computational time is nearly stable. Runtime increased after the image size became larger than 256×256. This means that they cannot maintain computational time stability. The performance depends on the given image size. Furthermore, the SUN method got the highest computational time line while the Itti method got the lowest line, which is less than one second, to execute a single given image.

In other words, we can divide the time in the y-axis into three ranges–(i) $0 < t \leq 1$, (ii) $1 < t \leq 10$, and (iii) $t > 10$. The first group consists of three methods including the SAL, SR, and Itti methods. The computational time are lower than one second for all image sizes. In the second group, there are five

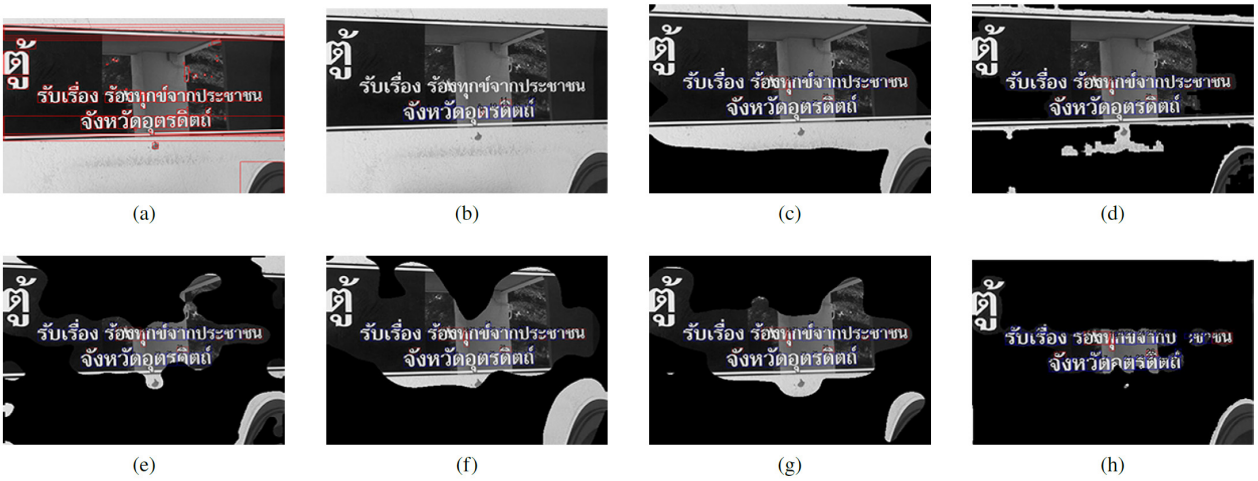


Fig.16: *Example of the experimental results to show that how the recall value achieved by all methods: (a) objects detectable, (b) MCCA, (c) Itti's method, (d) SUN, (e) SR, (f) SAL, (g) HFT, and (h) SFR.*

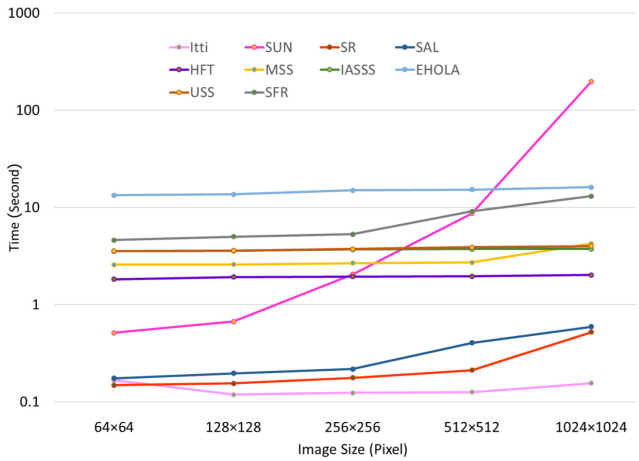


Fig.17: *Comparison of the execution time of all methods when tested by the given images at five different size.*

methods in this range–IASSS, USS, MSS, HFT, and SFR. Their processing time is in a range of [1 10]. They require setup time for hypercomplex Fourier transform and filtering generation, thus making the time longer than one second. In the last group, the EHOLA method takes more than ten seconds to process the given image because it spends time generating the six filters before applying them to each channel image. The EHOLA method uses the LAB color domain composing the L, A, and B channels. Therefore, the six filters are applied to the three channels separately. This is the main reason why the EHOLA takes more than ten seconds.

Based on the SFR method, the trend of the computational cost is steadily increasing. Although the processing time is higher than ten seconds when the image size is larger than 512×512, it is in a range of $1 < t \leq 10$ if the image size is smaller than 512×512. The processing time of the SFR method is in the same range as the IASSS, USS, MSS, and HFT methods and it is still lower than the EHOLA method. The proposed SFR method spends processing time on the object detection in Sect. 3.1 at Step 1. The Moore-Neighbor tracing algorithm, which is used to detect the object in our work, processes the image pixel-by-pixel. This is more time-consuming. The processing time is higher if the given image size is larger, whereas the processing time is lower if the image size is smaller. However, it is not a contribution to our work. It can be replaced by more efficient object detection to reduce the processing time in the future as explained in Sect. 5.

5. LIMITATIONS

In this paper, we aim to apply the concept of the saliency region detection to the Thai text Localisation task. Based on our analysis, we believe that the characteristic of the Thai characters is closely similar to that of the repeated pattern in the frequency domain. Therefore, a spatial-frequency redistribution (SFR) method is designed to mainly support the Thai text Localisation task by using the saliency region detection concept. However, our performance depends on three factors as follows:

Firstly, the SFR method requires an algorithm that can detect the objects in an input image. This is not the focus of this paper. Nevertheless, to show the performance of the proposed SFR method, the Moore-Neighbor tracing algorithm modified by Jacob's stopping criteria was used based on our previous study on Thai text localisation as discussed in [26]. For other applications, the detection algorithm can be changed. It is not limited to the Moore-Neighbor tracing algorithm if there is a method that is more appropriate.

Secondly, the amplitude, period, and phase parameters for generating the adaptive cosine image should be defined from the structural features of the object

in the spatial domain. Although our performance has increased, it may decrease when applied to another application that our recommended feature cannot discriminate. This means that the method to define the parameters is important for the SFR method. Therefore, the solution to define our parameters can be changed when applied to another application. For the second experiment of this paper, the amplitude, period, and phase parameters were computed from the Thai text features recommended in [23, 26].

Lastly, there are two modes in the SFR method that can identify the saliency region and/or the repeated pattern region depending on the application. For example, in the first experiment, the psychological pattern images were used to show the performance when an area that is different from another area is required. The SFR method was performed by identifying the saliency region for this case. In the second experiment, the characteristic of the Thai text is closely corresponding to the concept of the repeated pattern region. Therefore, to detect Thai text, the SFR method was performed by identifying the repeated pattern region. Consequently, users can choose the appropriate mode by themselves in step 5 of section 3.2.

As mentioned above, these three factors directly impacted our performance. Users can choose or modify them accordingly, depending on the application used.

6. CONCLUSION

In this paper, a spatial-frequency redistribution (SFR) method is proposed for improving saliency region detection and Thai text localisation. In the saliency region detection, the proposed SFR method includes two steps: The first step generates an adaptive cosine image based on a structural target object and then implants it into an object detectable. The first contribution is the spatial redistribution that occurs in this step. When an object in the spatial domain is implanted by the adaptive cosine image, its intensity, as spatial domain information, is redistributed into an easy-to-control platform. In the second step, when a complete multiple-object-implanted image is transformed into the Fourier domain, its spectrum is also redistributed. This phenomenon is called a frequency redistribution. This property provides a simple method to identify the target object from the unwanted objects and background. The second contribution occurs in the second step, which is that the similar and different patterns are designed to be incorporated with the accumulate matrix to detect the saliency or repeated pattern region. Based on the experiments, the proposed SFR method can improve the performance when tested with a response to psychological pattern images and can effectively identify the attention area in Thai text images with no prior knowledge.

In future work, we plan to use the deep learning technique to estimate the amplitude, period, and phase parameters in each object automatically. Moreover, we will develop a solution to extend the adaptive cosine image to support the complex objects.

7. COMPLIANCE WITH ETHICAL STANDARDS

The author declares that there are no conflicts of interest. This article does not contain any studies involving human participants performed by any of the authors.

References

[1] L. Zhang, M.H. Tong, T.K. Marks, H. Shan, and G.W. Cottrell, "SUN: A Bayesian framework for saliency using natural statistics," *Journal of Vision*, vol.8, no.7, pp.1–20, Dec. 2008.

[2] X. Hou, J. Harel, and C. Koch, "Image signature: Highlighting sparse salient regions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.34, no.1, pp.194–201, Jan. 2012.

[3] D. Walther and C. Koch, "Modeling attention to salient proto-objects," *Neural Networks*, vol.19, no.9, pp.1395–1407, Nov. 2006.

[4] M. Cerf, J. Harel, W. Einhaeuser, and C. Koch, "Predicting human gaze using low-level saliency combined with face detection," *Advances in Neural Information Processing Systems 20*, pp.241–248, 2007.

[5] C. Guo and L. Zhang, "A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression," *IEEE Transactions on Image Processing*, vol.19, no.1, pp.185–198, Jan. 2010.

[6] W. Einhauser, M. Spain, and P. Perona, "Objects predict fixations better than early saliency," *Journal of Vision*, vol.8, no.14, p.1–26, 2008.

[7] L. Elazary and L. Itti, "Interesting objects are visually salient," *Journal of Vision*, vol.8, no.3, pp.1–15, 2008.

[8] L. Itti and C. Koch, "A saliency-based search mechanism for overt and covert shifts of visual attention," *Vision Research*, vol.40, no.10–12, pp.1489–1506, Jun. 2000.

[9] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," *Proceedings of Neural Information Processing Systems*, pp.545–552, 2006.

[10] J.K. Tsotsos and N.D.B. Bruce, "Saliency based on information maximization," *Advances in Neural Information Processing Systems 18*, MIT Press, pp.155–162, 2006.

[11] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.20, no.11, pp.1254–1259, Nov. 1998.

[12] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pp.1–8, Jun. 2007.

[13] J. Li, M.D. Levine, X. An, X. Xu, and H. He, "Visual saliency based on scale-space analysis in the frequency domain," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.35, no.4, pp.996–1010, Apr. 2013.

[14] A. Kimura, R. Yonetani, and T. Hirayama, "Computational models of human visual attention and their implementations: A survey," *IEICE Transactions on Information and Systems*, vol.E96-D, no.3, pp.562–578, Mar. 2013.

[15] Y. Zhang, Y. Li, and S. Safavi-Naeini, "A spectrum-based saliency detection algorithm for millimeter wave insar imaging with sparse sensing," *IEICE Transactions on Information and Systems*, vol.E100-D, no.2, pp.388–391, Feb. 2017.

[16] J. Jaemsiri, T. Titijaroonroj, J. Rungrattanaubol, "Modified scale-space analysis in frequency domain based on adaptive multiscale Gaussian filter for saliency detection," *2019 International Joint Conference on Computer Science and Software Engineering*, pp.212–217, Jul. 2019.

[17] T. Titijaroonroj, J. Jaemsiri and J. Rungrattanaubol, "Improved adaptive spectrum scale-space in frequency domain for saliency detection," *2020 12th International Conference on Knowledge and Smart Technology (KST)*, pp.97–102, 2020.

[18] D. Kakanopas and K. Woraratpanya, "An efficient Hola filter for saliency detection," *2020 12th International Conference on Information Technology and Electrical Engineering (ICITEE)*, pp. 338–343, 2020.

[19] D. Kakanopas and K. Woraratpanya, "A unified framework for saliency segmentation," *2021 17th International Conference on Computing and Information Technology (IC2IT)*, pp. 191–200, 2021.

[20] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol.9, no.1, pp.62–66, Jan. 1979.

[21] C. Guo, Q. Ma, and L. Zhang, "Spatio-temporal saliency detection using phase spectrum of quaternion Fourier transform," *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pp.1–8, Jun. 2008.

[22] A.M. Treisman and G. Gelade, "A feature-integration theory of attention," *Cognitive Psychology*, vol.12, no.1, pp.97–136, Jan. 1980.

[23] K. Woraratpanya, P. Boonchukusol, Y. Kuroki, and Y. Kato, "Improved Thai text detection from natural scenes," *2013 International Conference on Information Technology and Electrical Engineering*, pp.137–142, Oct. 2013.

[24] R.C. Chang, "Intelligent text detection and extraction from natural scene images," *2011 15th North-East Asia Symposium on Nano, Information Technology and Reliability*, pp.23–28, Oct. 2011.

[25] A. Schick, M. Fischer, and R. Stiefelhagen, "Measuring and evaluating the compactness of superpixels," *Proceedings of the 21st International Conference on Pattern Recognition*, pp.930–934, Nov. 2012.

[26] K. Woraratpanya, K. Pasupa, U. Suttapakti, P. Boonchukusol, T. Titijaroonroj, R. Hokking, Y. Kuroki, and Y. Kato, "Text-background decomposition for Thai text localization and recognition in natural scenes," *2014 6th International Conference on Information Technology and Electrical Engineering*, pp.1–6, Oct. 2014.

Taravichet Titijaroonroj is currently an assistant professor in the Faculty of Information Technology, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand. He received the B.S., M.S., and Ph.D. degrees in Information Technology from King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand in 2013, 2014, and 2018, respectively. His research interests include machine learning, deep learning, image transformation, pattern recognition, and medical image analysis.

Ungsumalee Suttapakti is currently a lecturer in the Faculty of Informatics, Burapha University, Thailand. She received the B.Sc. degree in Food Technology from the Suranaree University of Technology, Nakhon Ratchasima, Thailand, in 2006, and the M.Sc. degree in Information Technology from the King Mongkut's University of Technology North Bangkok, Bangkok, Thailand, in 2010. She obtained the Ph.D. degree in Information Technology from the King Mongkut's Institute of Technology Ladkrabang, Thailand, in 2018. Her research interests include signal analysis, image processing and decomposition, computer vision, pattern recognition, and machine learning.

Walairach Nunsong is an assistant professor in the Faculty of Science and Technology, Rajamangala University of Technology Srivijaya Nakhon Si Thammarat Campus, Nakhon Si Thammarat, Thailand. She received the Bachelor of B.B.A.(Business Computer) from Rajamangala Institute of Technology, the M.Sc. degree in Information Technology from King Mongkut's University of Technology North Bangkok, and the Ph.D. degree in Information Technology from King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand in 1999, 2002, and 2018, respectively. The research interests include machine learning, deep learning, pattern recognition, and image processing.