Empirical Monocomponent Image Decomposition

UNGSUMALEE SUTTAPAKTI[ID]**, (Student Member, IEEE),**
KITSUCHART PASUPA, (Senior Member, IEEE),
AND KUNTPONG WORARATPANYA, (Member, IEEE)
Faculty of Information Technology, King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand

Corresponding author: Kuntpong Woraratpanya (kuntpong@it.kmitl.ac.th)

ABSTRACT Monocomponent image decomposition plays an important role in image analysis and related areas, such as image denoising, object detection, and texture segmentation. Existing image decomposition methods can extract monocomponents but their performances are insufficiently accurate because of interference and redundancy component problems caused by inaccurate spectrum segmentation. In this paper, an empirical monocomponent image decomposition (EMID) is proposed for fully recoverable monocomponents. The EMID method empirically decomposes an image into monocomponents based on energy concentration in Fourier support. This method is composed of two main processes: 1) energy concentration-based segmentation and 2) empirical image filter bank construction. In the former process, the base of a mountain-shaped energy concentration that can perfectly represent the monocomponent spectrum boundary is detected and identified. This process provides a more accurate spectrum segmentation which helps prevent the serious problems from interference and redundancy components. In the latter process, an empirical image filter bank is constructed in accordance with the actual monocomponent boundaries by means of ellipse and Gaussian functions and used to decompose an image into monocomponent images with fewer ringing artifacts. The experimental results show that the proposed EMID method achieves a better decomposition than the state-of-the-art methods in terms of the quality of monocomponent images that are evaluated by peak signal-to-noise ratio and structural similarity index. Furthermore, in a real world dataset, the EMID method is able to clearly detect text regions, thus significantly improving the efficiency of Thai text character localization in natural scene images.

INDEX TERMS Empirical wavelet transform, energy concentration-based segmentation, image decomposition, monocomponent image, monocomponent spectrum.

I. INTRODUCTION

Image decomposition is a process for separating an image into meaningful components that are essential for image analysis in various applications such as image denoising [1]–[8], object detection [9]–[14], and texture segmentation [15]–[20]. In image denoising, an efficient decomposition can separate the noisy part and the clear image part into distinct components such that meaningful components are fully recoverable. In object detection, image decomposition plays a role in separating object regions from the background and identifying them. Texture segmentation is also an important task in applications of computer vision that requires partitioning an image into a set of disjoint regions that have a certain homogeneous pattern characteristic. As mentioned, image decomposition is absolutely necessary for separating each component from a whole image with no interference from another component. This kind of non-interfered component is called a "monocomponent." For more comprehensive

characteristics of a monocomponent in the form of a signal, we refer to Fig. 1 that visually defines the monocomponent signal. A single "ridge" which corresponds to an elongated region of energy concentration is the monocomponent signal in a time-frequency domain [21]. The monocomponent is a key to the performance of applications based on image analysis. Consequently, many image decomposition methods have been investigated for better decomposition of fully recoverable components as reviewed in the following paragraphs.

For ease of explanation, the terms "component" and "subband" are interchangeable; hereafter, component and subband are used as the terms for the results of general signal/image decomposition and classical wavelet transform, respectively. Conventional image decomposition can be divided into non-adaptive and adaptive methods. In the non-adaptive method, 2D discrete wavelet transform (2D DWT) [22], a well-known method, was developed from a classical wavelet transform [23]. It is widely used to
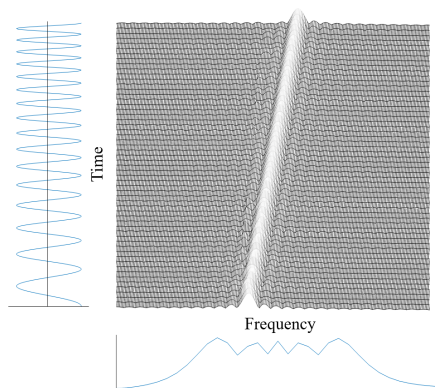
FIGURE 1. Monocomponent signal in a time-frequency representation.

decompose an image into multi-level subbands by using a prior function; that is, for the first decomposition level, the low- and high-pass filters are applied to the original image along the row direction and then both filtered images are downsampled by half. Along the column direction, this operation is repeated again. As a result, the original image is decomposed into four subbands: Low-Low, Low-High, High-Low, and High-High. For the next level, the Low-Low subband is repeatedly decomposed, similarly to the previous step, until the decomposition reaches a desired level, while the other three subbands are not repeatedly decomposed. This kind of decomposition is very useful for image analysis, but it has one drawback, i.e., in a multilevel decomposition, only the low-frequency subband (Low-Low) is selected for analysis, whereas the high-frequency subbands (Low-High, High-Low, and High-High) are disregarded for the next decomposition level. This easily leads to interference from other frequencies in each segmented subband [24]. In other words, it is hard to extract the monocomponents. Wavelet packet transform (WPT) proposed by Coifman *et al.* [25] was extended from the 2D DWT method by preserving both low- and high-frequency subbands. This method partitions an image into subbands in the form of a multi-level binary tree such that both low and high frequencies are preserved for the next level decomposition. However, it does not guarantee that the monocomponents will be successfully decomposed [24].

On the other hand, adaptive decomposition methods have been proposed to extract monocomponents from an image. Linderhed [26] presented a 2D empirical mode decomposition (2D EMD) in spatial domain to adaptively separate an image into a set of intrinsic mode functions (IMFs) based on spline interpolation. Likewise, Nunes *et al.* [27] introduced a bidimensional empirical mode decomposition (BEMD) to separate an image by means of radial basis function (RBF). These methods adaptively decompose an image by using the characteristics of the image itself as a decomposition criterion to obtain monocomponents [24]. Unfortunately, there is still some interference from adjacent monocomponents. This interference has a great impact on image quality for analysis [24]. Recently, Gilles *et al.* [6] proposed

a 2D empirical wavelet transform (2D EWT) that was the first adaptive image decomposition in frequency domain extended from the empirical wavelet transform [28] for signal analysis. Gilles's method adaptively decomposes an image by segmenting its spectrum in Fourier support. Then, it builds an adaptive wavelet filter according to its spectral segmentation to obtain monocomponents. The spectral segmentation of 2D EWT is based on local minimum detection that segments spectra at the local minima between two contiguous maximum spectra. Additionally, this method can be done by various types of segmentation such as 2D Tensor EWT, 2D Littlewood-Paley EWT, and 2D empirical curvelet transform. However, all of these 2D EWT methods make use of an average magnitude technique for segmentation. This generally leads to inaccurate spectral segmentation problems which can be categorized into interference and redundancy components. In the case of the interference component, the conventional 2D EWT methods cannot segment the spectra of different components into different monocomponents. In the case of the redundancy component, they frequently segment the spectrum of a single component into different components. Furthermore, ringing artifacts often appear in the components that are segmented by the conventional 2D EWT methods due to an inefficiently designed wavelet filter for decomposing the image. These problems have a great impact on the quality of monocomponents.

As mentioned in the previous paragraph, various adaptive image decomposition methods have been attempted to improve the quality of monocomponent images. However, those methods still cannot achieve high quality monocomponent images. Therefore, this paper proposes a new approach, called empirical monocomponent image decomposition (EMID), that applies energy concentration-based segmentation (described in Section III-B) for accurate decomposition of monocomponent spectra and for quality improvement of monocomponent images. The main idea of the proposed EMID method is that each centroid of energy concentration in Fourier support is identified as the first step. This centroid accurately represents the center of the monocomponent spectrum, thus clearly defining a monocomponent spectrum boundary. This can be done as follows. The proposed EMID method is composed of two main processes: (i) energy concentration-based segmentation and (ii) empirical image filter bank construction. In the first process, 2D local maximum point detection is performed on an empirical mean plane that is created by principal component analysis (PCA) to detect the central frequencies of candidate monocomponent spectra and also to eliminate a global trend problem as reported in [6]. Then a 2D local minimum boundary detection algorithm is used to detect a component boundary from each detected central frequency. After that, an actual monocomponent identification algorithm is used to identify actual monocomponent spectrum boundaries. In the second process, an empirical image filter bank is constructed in accordance with monocomponent spectrum boundaries by means of ellipse and Gaussian functions, and is used to

decompose an image into monocomponent images with fewer ringing artifacts. In this way, the proposed EMID method is able to achieve a high quality of monocomponent images in terms of image quality as measured by peak signal-to-noise ratio (PSNR) [29] and structural similarity index (SSIM) [30]. Furthermore, it can be successfully applied to text region detection for Thai text localization in natural scene images.

The remainder of this paper is organized as follows. 2D EWT is briefly explained and analyzed in Section II. Section III introduces the proposed EMID method. The experimental results are reported and discussed in Section IV, and finally, the conclusion is presented in Section V.

II. PROBLEM FORMULATION
In this section, the conventional empirical wavelet transform (EWT) method is briefly reviewed and its limitations are pointed out.

A. EMPIRICAL WAVELET TRANSFORM AND ITS EXTENSIONS
An assumption of image decomposition is that an image consists of N monocomponents, where N is a positive integer. Each monocomponent must be well segmented in Fourier spectrum [24]. In practice, the 2D empirical wavelet transform (2D EWT) introduced by Gilles *et al*. [6] is an adaptive image decomposition based on Fourier transform that is used to decompose an image into monocomponents. The efficiency of a conventional 2D EWT method depends on accurate Fourier spectrum segmentation and proper wavelet filter bank design. Furthermore, the 2D EWT was extended from 1D EWT [28], thus its spectrum segmentation and wavelet filter bank construction are based on the 1D EWT method. Both 1D and 2D EWT methods are described below.

1) 1D EMPIRICAL WAVELET TRANSFORM
1D empirical wavelet transform (1D EWT) was first introduced by Gilles [28]. It has two main signal decomposition steps: (i) Fourier spectrum segmentation and (ii) wavelet filter bank construction in accordance with segmented spectrum. In the first step, the Fourier spectrum is segmented based on local maximum detection. Local maximum points in the spectrum are detected and sorted in decreasing order. Then, a number of components are manually selected, and finally, spectrum boundaries are segmented by using the mean between two contiguous local maximum points. It is a simple technique. However, inaccurate segmentation usually occurs as shown in Fig. 2: an improper Fourier boundary is shown as the vertical solid line while the proper Fourier boundary should be at the vertical dashed line. Recently, the local minimum detection introduced by Gilles *et al*. [6] was proposed to solve the inaccurate spectrum segmentation problem. Fourier boundaries are segmented by using local minimum detection between two contiguous maximum spectra. Although this method is efficient in segmenting the Fourier spectrum as shown in Fig. 2 (the vertical dashed line), it is unsuccessful when the low frequency region contains
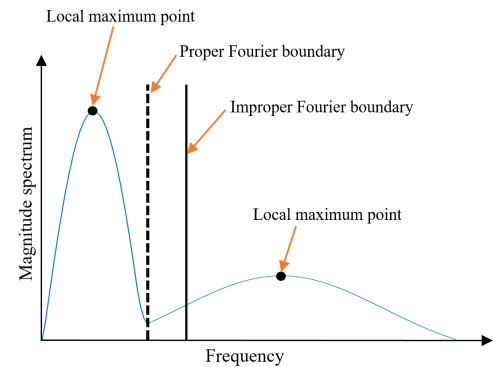

FIGURE 2. 1D Fourier spectrum segmentation based on local maximum detection.

a very high magnitude spectrum. Consequently, many local maximum points are detected in the same low frequency region. This problem, called a global trend problem, causes incorrect spectrum segmentation. In order to reduce this effect, Gilles *et al*. [6] proposed a mathematical morphology with logarithm to depress the magnitude of Fourier spectrum prior to segmentation. Both local maximum detection and local minimum detection are achieved by adaptive spectrum segmentation methods depending on the characteristics of the spectrum. Nevertheless, they need to predefine the maximum number of components for spectrum segmentation, i.e., manually select the maximum number of components. It is difficult to efficiently predefine the appropriate number of components for decomposing signals. Thus, Gilles and Heal [31] introduced a coarse histogram segmentation based on scale space for spectrum segmentation in order to automatically define a suitable number of components. The Fourier boundaries are determined by local minima of scale-space representation.

In the second signal decomposition step, detected Fourier boundaries are used to construct wavelet filter banks consisting of a low-pass filter with an empirical scaling function and band-pass filters with an empirical wavelet function. Then, the signal is decomposed into approximation and detail coefficients by using those low-pass and band-pass filters as described in [6].

2) 2D EMPIRICAL WAVELET TRANSFORM
Gilles *et al*. [6] introduced a 2D empirical wavelet transform (2D EWT) extended from 1D EWT. The 2D EWT segments average magnitude spectra based on local minimum detection and/or local maximum detection—a manually selected maximum number of components. The segmentation of 2D EWT is introduced in various methods. To explain each method, a synthesized image-I as established in Fig. 3(a) is used for demonstration. This synthesized image is generated from a combination of reference images shown in Figs. 3(b), 3(c), 3(d), and 3(e). When the synthesized image-I is transformed into frequency domain, monocomponent-1, -2, -3, and -4 of the reference images
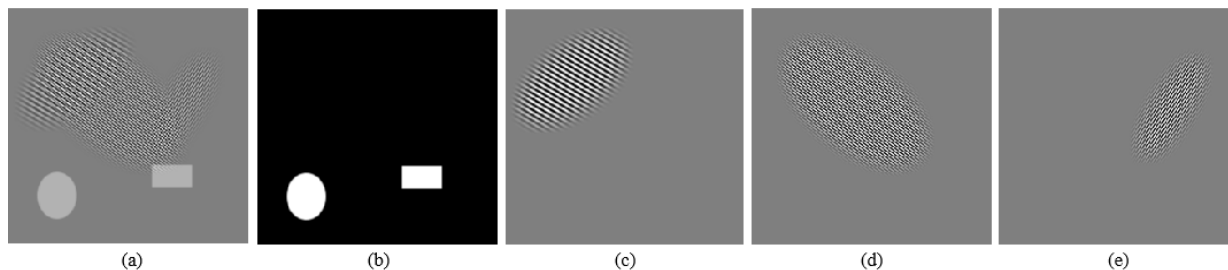
FIGURE 3. Synthesized image-I and its reference images in spatial domain. (a) Original synthesized image-I, (b) reference image-I, (c) reference image-II, (d) reference image-III, and (e) reference image-IV.
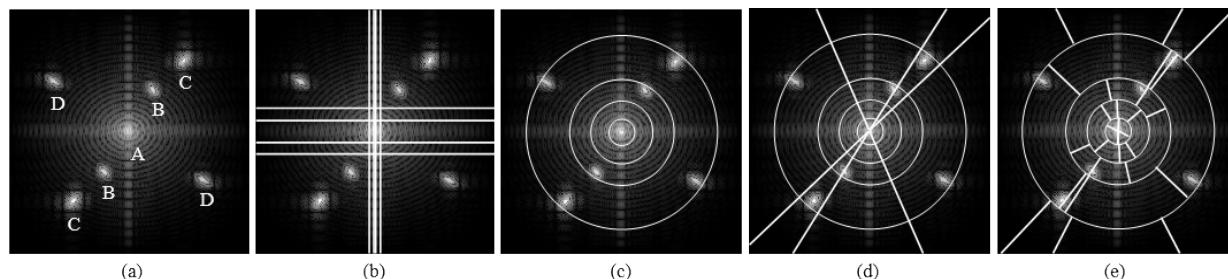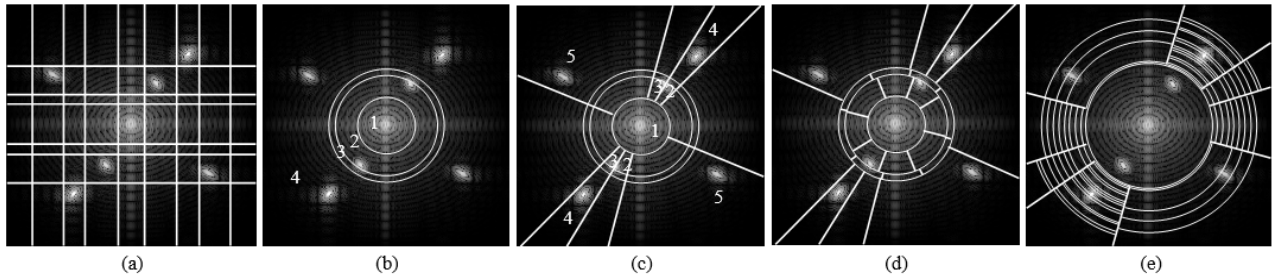

FIGURE 4. Detected Fourier boundaries of synthesized image-I using 2D EWTs in manual mode. (a) Fourier spectrum of Fig. 3(a), (b) 2D Tensor EWT, (c) 2D Littlewood-Paley EWT, (d) 2D Curvelet EWT-I, and (e) 2D Curvelet EWT-II.

clearly appear as illustrated in Fig. 4(a) labeled with letters A, B, C, and D, respectively. 2D EWT segmentations in a manually selected maximum number of components are of 2D Tensor EWT, 2D Littlewood-Paley EWT, 2D Curvelet EWT-I, and 2D Curvelet EWT-II as shown in Figs. 4(b), 4(c), 4(d), and 4(e), respectively. Each method is described below.

First, a 2D Tensor EWT segments an average magnitude spectrum in Fourier support in row and column directions independently. Suppose that a Fourier spectrum is segmented in the row direction, and the average magnitude spectrum is then calculated for each row. Next, the Fourier boundaries based on 1D EWT segmentation are detected from this average magnitude spectrum. Each boundary is used to build a unique filter bank. After that, the image is filtered along the row direction to obtain row components. For segmentation in the column direction, the Fourier spectrum is segmented in the same way as in the row direction. The 2D Tensor EWT is a simple technique to decompose image components, but it is unable to obtain the corner edges of the components as reported in [6]. In order to preserve corner edges, a 2D Littlewood-Paley EWT was introduced to decompose different components by means of ring-shaped Fourier spectrum segmentation. The Fourier spectrum segmentation is performed on a polar representation which is a two-dimensional coordinate system consisting of radius and angle. An average spectrum is segmented into different radius components. This method can decompose components and retain corner edges. However, it still cannot decompose an image into components with different angles in Fourier spectrum. For this reason,

a 2D Curvelet EWT is designed to segment different radii and different angles of an average Fourier spectrum on the polar representation. This method of segmentation can be of two types: 2D Curvelet EWT-I and 2D Curvelet EWT-II. In the first type, the Fourier spectrum is independently segmented in radius and in angle representations as shown in Fig. 4(d). In radius segmentation, the Fourier spectrum is divided into a series of ring-shaped segments; different rings can have different thicknesses. In angle segmentation, the Fourier spectrum is divided into a set of sectors with different angles. In the second type, the Fourier spectrum is initially segmented into different radii in ring-shaped segmentation, and then each radius is segmented into different angles based on its own radius average spectrum as shown in Fig. 4(e). The second type of segmentation differs from the first one in that the second type requires a specific sequence of segmentation: radius first and angle later.

Furthermore, a coarse histogram [31] can be used to automatically segment an average magnitude spectrum in the Fourier spectrum into different types: 2D Tensor EWT, 2D Littlewood-Paley EWT, 2D Curvelet EWT-I, 2D Curvelet EWT-II as depicted in Figs. 5(a), 5(b), 5(c), and 5(d), respectively. Moreover, 2D Curvelet EWT-III [32] as shown in Fig. 5(e) can segment the Fourier spectrum into different radii and angles that differ from 2D Curvelet EWT-I and 2D Curvelet EWT-II. The Fourier spectrum is primarily segmented into different angles, and then each angle is segmented into different radii based on its own angle average spectrum as displayed in Fig. 5(e). This series of 2D

FIGURE 5. Detected Fourier boundaries of synthesized image-I using 2D EWTs in automatic mode. (a) 2D Tensor EWT, (b) 2D Littlewood-Paley EWT, (c) 2D Curvelet EWT-I, (d) 2D Curvelet EWT-II, and (e) 2D Curvelet EWT-III.
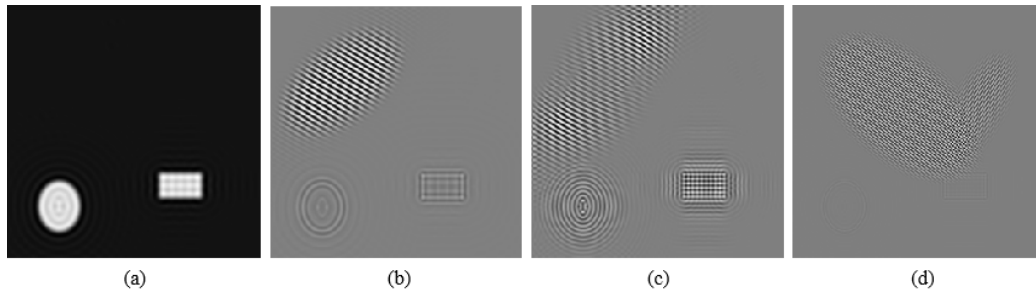
FIGURE 6. Interference components in spatial domain when Fig. 3(a) is decomposed by 2D Littlewood-Paley EWT. (a) Component C-I$_1$ of the Fourier boundary in Fig. 5(b) labeled with number 1, (b) component C-I$_2$ of the Fourier boundary in Fig. 5(b) labeled with number 2, (c) component C-I$_3$ of the Fourier boundary in Fig. 5(b) labeled with number 3, and (d) interference between monocomponent-3 and monocomponent-4 in component C-I$_4$ of the Fourier boundary in Fig. 5(b) labeled with number 4.

EWTs is specifically designed to increase the accuracy of spectrum segmentation. All of the segmented components are used to construct a set of wavelet filter banks for image decomposition.

B. ANALYSIS OF 2D EMPIRICAL WAVELET TRANSFORM

This subsection aims to analyze problems of the conventional 2D EWT methods and point out key factors that degrade the quality of monocomponents. As described in Section II-A, the conventional 2D EWTs, including 2D Tensor EWT, 2D Littlewood-Paley EWT, 2D Curvelet EWT-I, 2D Curvelet EWT-II, and 2D Curvelet EWT-III, use 2D Fourier spectrum segmentation methods based on an average magnitude technique that leads to inaccurate spectrum segmentation. This inaccurate spectrum segmentation always creates two serious problems: interference and redundancy components. For the former problem, interference components occur when spectra of different components cannot be segmented into different components. In other words, spectra of different components are grouped into the same Fourier boundary. An example of interference components is illustrated in Fig. 6 when the 2D Littlewood-Paley EWT in an automatic mode is applied to the synthesized image-I. Detected Fourier boundaries in Fig. 5(b) labeled with numbers 1 to 4 are decomposed into component-1 to -4 (C-I$_1$ to C-I$_4$) in Figs. 6(a) to 6(d), respectively. The spectra of different components in Fourier spectrum labeled with letters C and D in Fig. 4(a) are detected within the same Fourier boundary shown in Fig. 5(b)

labeled with number 4. That is a cause of the interference between monocomponent-3 and monocomponent-4 as can be seen in Fig. 6(d). Interference components also occur from detected Fourier boundaries with 2D Tensor EWT and 2D Littlewood-Paley EWT in manual modes. For the latter problem, the redundancy components occur when the spectrum of a single component is segmented into different components. Fig. 7 shows a case of redundancy components. Figs. 7(a) to 7(e) are component-1 to -5 (C-I$_1$ to C-I$_5$) recovered from segmented Fourier boundaries of Fig. 5(c) labeled with numbers 1 to 5, respectively. The redundancy components of monocomponent-2 are explicitly shown in Figs. 7(b) and 7(c). The cause of the redundancy components comes from segmenting a spectrum of monocomponent-2 into different Fourier boundaries as illustrated in Fig. 5(c) labeled with numbers 2 and 3. The redundancy components still occur from detected Fourier boundaries with 2D Curvelet EWT-I, 2D Curvelet EWT-II in manual modes, 2D Curvelet EWT-II and 2D Curvelet EWT-III in automatic modes.

Moreover, the wavelet filter banks designed for the conventional 2D EWTs commonly decompose an image into different components that provide approximation and detail coefficients. Unfortunately, the use of wavelet filter banks in 2D EWT becomes another cause of ringing artifacts that appear as small ripples [33], [34]. Figs. 8(a) to 8(d) explicitly show that ringing artifacts appear around the sharp edges of each component. These artifacts have a direct impact on the quality of monocomponents.
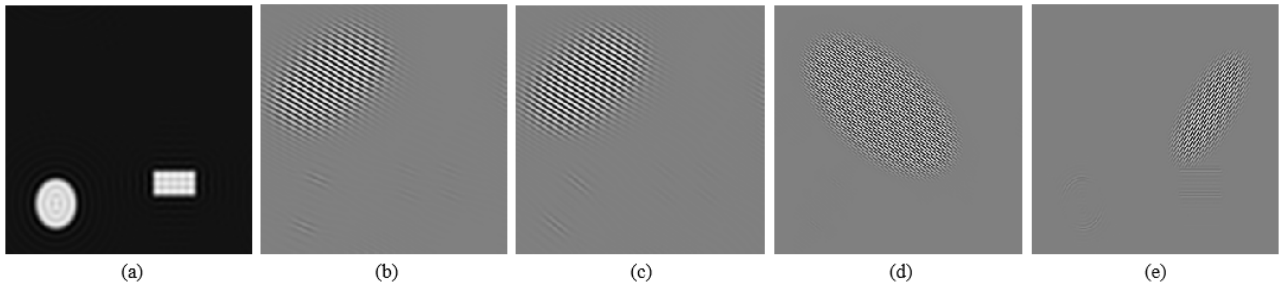
FIGURE 7. Redundancy components in spatial domain when Fig. 3(a) is decomposed by 2D Littlewood-Paley EWT. (a) Component C-I$_1$ of the Fourier boundary in Fig. 5(c) labeled with number 1, (b) and (c) redundancy monocomponent-2 of component C-I$_2$ and component C-I$_3$ of Fourier boundaries in Fig. 5(c) labeled with numbers 2 and 3, (d) component C-I$_4$ of the Fourier boundary in Fig. 5(c) labeled with number 4, and (e) component C-I$_5$ of the Fourier boundary in Fig. 5(c) labeled with number 5.
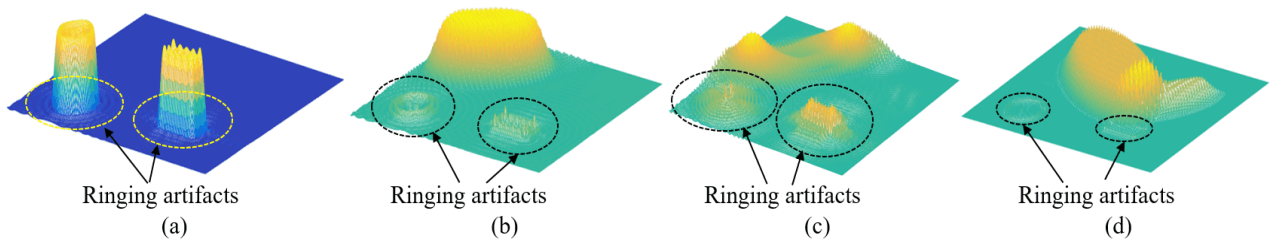
FIGURE 8. Ringing artifacts appearing in spatial domain when Fig. 3(a) is decomposed by 2D Littlewood-Paley EWT with wavelet filter banks. (a) Component C-I$_1$ of Fig. 6(a), (b) component C-I$_2$ of Fig. 6(b), (c) component C-I$_3$ of Fig. 6(c), and (d) component C-I$_4$ of Fig. 6(d).
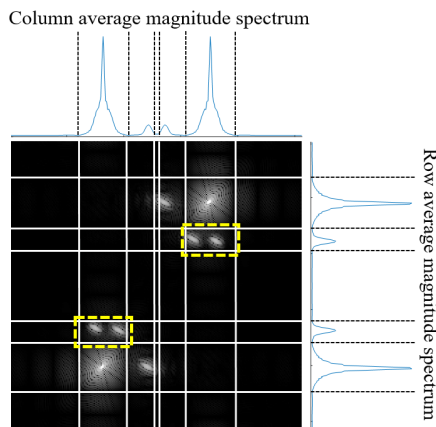
FIGURE 9. Inaccurate spectrum segmentation caused by using an average magnitude spectrum segmentation of 2D Tensor EWT, thus producing the interference components shown in dotted rectangles.

As previously mentioned, the causes of the problems with conventional 2D EWTs come from two factors: (i) inaccurate spectrum segmentation and (ii) improper wavelet filter bank design. The first problem is caused by using an average spectrum segmentation that commonly produces interference and redundancy components. Fig. 9 is a clear example of the interference problem caused by applying row and column average magnitude spectrum segmentations of 2D Tensor EWT. In this case, the interference component is shown in a dashed rectangle. The second problem is caused by using improper wavelet filter banks that lead to ringing artifacts in decomposed monocomponent images. In order to overcome

these problems, this paper proposes a novel image decomposition method, called empirical monocomponent image decomposition (EMID), presented in Section III.

III. PROPOSED METHOD

This paper aims to decompose an image into monocomponent images for image analysis. The key concepts, implemented processes, and algorithms of the proposed method are covered in the following topics: definition of monocomponent image, energy concentration-based segmentation, and empirical image filter bank construction. In addition, mathematical notations used in this section and the next section are listed in the Appendix for a clear definition.

A. DEFINITION OF MONOCOMPONENT IMAGE

A monocomponent image can be defined as in *Definitions 1* and *2* below to expose its important characteristic—energy concentration.

Definition 1: Let O_s and F_s be a single flat object image and a single frequency image, respectively, in Fourier support. The convolution of the two images O_s and F_s yields energy concentration in the form of a single "mountain" shape. A single mountain energy concentration is a monocomponent spectrum and its inverse Fourier transform is a monocomponent image.

Definition 2: Let O_m and F_s be a multiple flat object image and a single frequency image, respectively, in Fourier support. The convolution of the two images O_m and F_s yields energy concentration in the form of a single "mountain" shape. A single mountain energy concentration
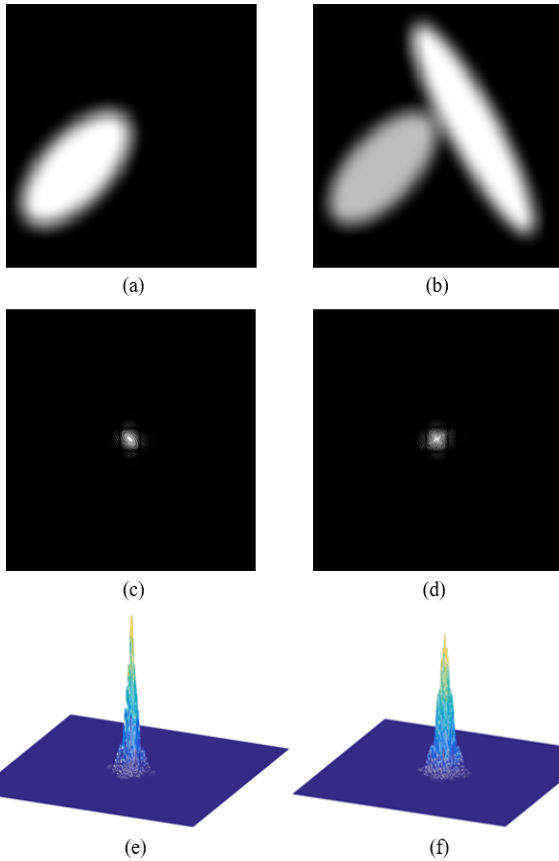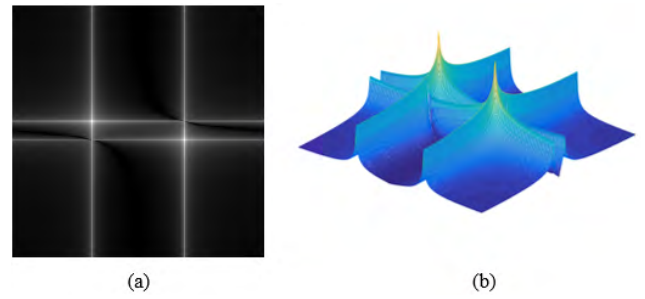
FIGURE 10. Synthesized flat object images. (a) A single flat object image in spatial domain, (b) a multiple flat object image in spatial domain, (c) a top view spectrum of Fig. 10(a), (d) a top view spectrum of Fig. 10(b), (e) a 3D view spectrum of Fig. 10(c), and (f) a 3D view spectrum of Fig. 10(d).

FIGURE 11. One single frequency image in Fourier support plotted on a log scale. (a) Top view and (b) 3D view.
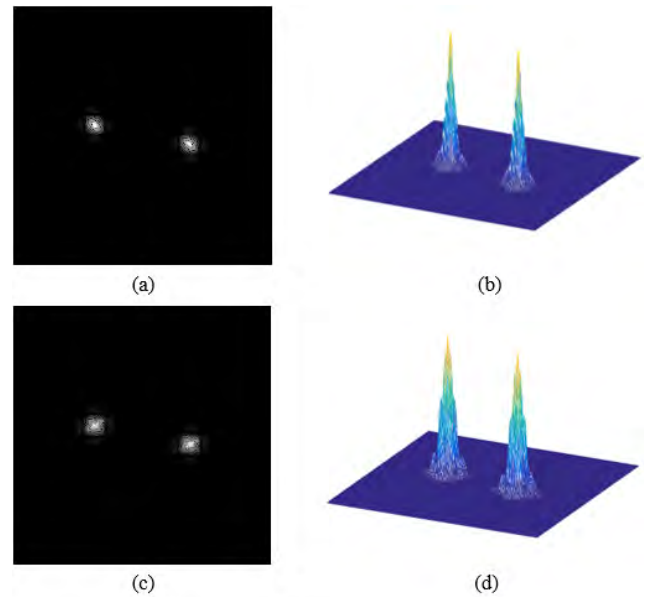
FIGURE 12. Monocomponent spectra in Fourier support. (a) A top view monocomponent spectrum of *Definition 1*, (b) a 3D view monocomponent spectrum of *Definition 1*, (c) a top view monocomponent spectrum of *Definition 2*, and (d) a 3D view monocomponent spectrum of *Definition 2*.
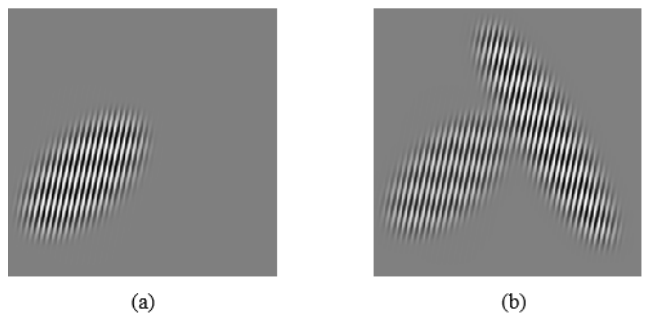
FIGURE 13. Monocomponent images in spatial domain representation of (a) *Definition 1* and (b) *Definition 2*.

is a monocomponent spectrum and its inverse Fourier transform is a monocomponent image.

From *Definitions 1* and *2*, a visual definition can be expressed by providing a single flat object O_s and a set of flat objects O_m in spatial domain as shown in Figs. 10(a) and 10(b), respectively. Both O_s and O_m images are transformed into Fourier domain as exhibited in Figs. 10(c) and 10(d). The 3D view spectra of Figs. 10(c) and 10(d) are graphically plotted on a log-scale as shown in Figs. 10(e) and 10(f) to expose a ''mountain'' shape of energy concentration. Also, a single frequency image in Fourier support is clearly provided and graphically plotted on a log-scale as illustrated in Figs. 11(a) for top view and 11(b) for 3D view.

In *Definition 1*, a monocomponent spectrum determined from the convolution of Fig. 10(c) and Fig. 11(a) is illustrated in Fig. 12(a). A mountain shape of such a monocomponent spectrum is clearly shown in Fig. 12(b). In the same way, the top and 3D view monocomponent spectra of *Definition 2* are still similar to the previous ones from *Definition 1* as demonstrated in Figs. 12(c) and 12(d), respectively. Figs. 13(a) and 13(b) are inverse Fourier transforms

of Figs. 12(a) and 12(c) that show monocomponent images of *Definitions 1* and *2*, respectively. Both definitions demonstrate that a mountain peak location of a single frequency is a principal location of energy concentration.

According to *Definitions 1* and *2* that clearly illustrate the important characteristics of monocomponent spectra and images, the proposed EMID method uses energy concentration-based segmentation as a key process in order to segment monocomponent spectra based on mountain shapes in a 2D Fourier support representation. Furthermore, a mountain peak location of the energy concentration always represents the central frequency of the monocomponent spectrum. This makes it completely different from the state-of-the-art methods, 2D EWTs described in Section II-B, which use a 1D average magnitude spectrum segmentation in decomposing monocomponent spectra. More details of the EMID method are given below.

B. ENERGY CONCENTRATION-BASED SEGMENTATION
Energy concentration-based segmentation is a process that segments more accurate N-monocomponent spectrum boundaries in Fourier support. It consists of three subprocesses: (i) 2D local maximum point detection implemented with PCA that creates an empirical mean plane for eliminating a global trend problem and for detecting the more accurate central frequencies of candidate monocomponent spectra in Fourier support; (ii) a 2D local minimum boundary detection algorithm used for solving an inaccurate spectrum segmentation problem and for defining monocomponent spectrum boundaries; and (iii) an actual monocomponent identification algorithm used for identifying actual monocomponent spectrum boundaries. Each subprocess is described as follows:
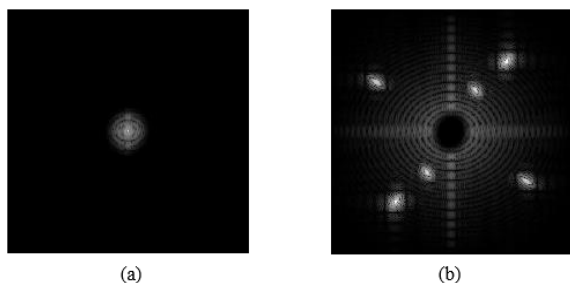
FIGURE 14. Separated Fourier support of synthesized image-I. (a) Approximation and (b) detail regions.

1) 2D LOCAL MAXIMUM POINT DETECTION
2D local maximum point detection is employed to detect central frequencies of candidate monocomponent spectra of an image in Fourier support. Let F be a 2D Fourier spectrum of Fig. 4(a) transformed from the synthesized image-I in Fig. 3(a). Generally speaking, F in Fourier support can simply be divided into two regions: approximation and detail. The approximation and detail regions are located at the central and surrounding regions of Fourier support as depicted in Figs. 14(a) and 14(b). Based on *Definitions 1* and *2*, each energy concentration location is detected by a local maximum point. However, in a real situation, a global trend problem always occurs when a very high magnitude spectrum occurs in the approximation region, a low frequency region, as mentioned in Section II-A. Consequently, many local maximum

points are detected only in the approximation region as shown in Fig. 15(a) for a top view and 15(b) for a 3D cross-section view. This leads to a failure of detecting monocomponent spectra in the detail region that contains valuable information such as texture and strong edges. Technically, in order to achieve accurately detected monocomponent spectra in the detail region, the global trend problem is first solved by excluding an approximation region from Fourier support as shown in Fig. 14(b). Then an empirical mean plane of the detail region is formed by means of PCA [35] and used as a new reference plane instead of the original Fourier support plane for projecting spectra. As shown in Fig. 15(c), a solution of the proper detected local maximum points is illustrated by red dot marks in the detail region. With this technique, the central frequencies of candidate monocomponent spectra in the detail region are accurately obtained from the projected spectra that are above the empirical mean plane. An algorithm of 2D local maximum detection can be described as having the following stages.

In the first stage, the main purpose is segmenting an approximation region. Fortunately, the central frequency of the approximation region is typically located at the centroid of F as shown in Fig. 16(a). Thus, the approximation region segmentation is easily done by identifying the centroid location of F. Then, a monocomponent spectrum boundary is detected in the region surrounding the central frequency as exhibited in Fig. 16(b). Subsequently, the complete monocomponent spectrum boundary is generated from the monocomponent spectrum boundary of Fig. 16(b). Fig. 16(c) shows the approximation region segmented by this stage. The determination of a monocomponent spectrum boundary and of a complete monocomponent spectrum boundary will be described in Sections III-B-2) and III-C, respectively.

In the second stage, all candidate central frequencies of monocomponent spectra of detail regions are targets. In order to achieve this goal, the segmented approximation region derived from the first stage is subtracted from F. Resulting from this step is the detail region F', which is used to calculate the empirical mean plane. Fortunately, one of the important properties of Fourier transform is the symmetry property [36]. This property is applied in this stage for creating the empirical mean plane. To do this, the detail region F' is divided into four symmetrical quadrants: an upper-right q_1, a lower-right q_2, a lower-left q_3, and an upper-left q_4 as shown in Fig. 17(a). The quadrants q_1 and q_2 are used to construct empirical mean planes for each. With the symmetry property of Fourier transform, the empirical mean planes of quadrants q_3 and q_4 are simply constructed from those of quadrants q_1 and q_2, respectively. To create each empirical mean plane, u-coordinate, v-coordinate, and F' of each quadrant are arranged in a matrix form $\mathbf{X}$ as

$$\mathbf{X} = \begin{bmatrix} u_1 & v_1 & F'_1 \\ \vdots & \vdots & \vdots \\ u_h & v_h & F'_h \end{bmatrix}_{h \times w}, \tag{1}$$
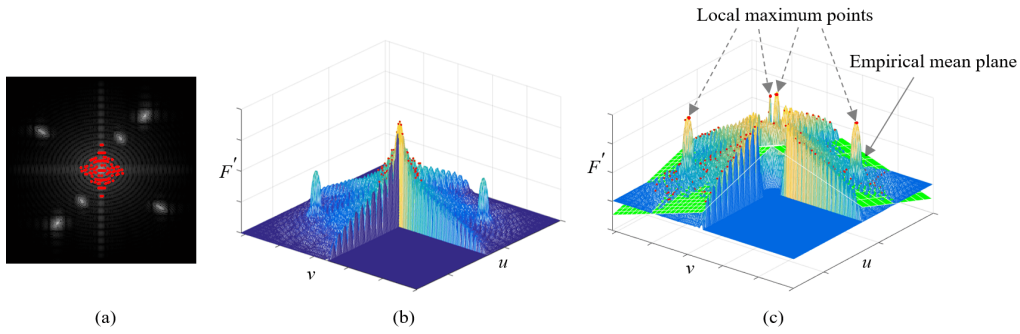
FIGURE 15. Global trend problem and its solution. (a) A top-view plot of many improper local maximum points (red dot marks) detected at the approximation region, (b) a 3D cross-section plot of Fig. 15(a) to show a cause of improper detected local maximum points (red dot marks), and (c) a 3D cross-section plot without the approximation region of Fig. 15(a) to show a solution of the proper detected local maximum points by means of an empirical mean plane.
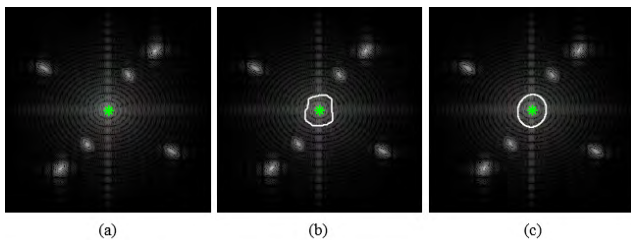
FIGURE 16. Detected approximation region. (a) A central frequency, (b) a monocomponent spectrum boundary, and (c) a complete monocomponent spectrum boundary.

where h and w are the number of rows and columns of each quadrant, respectively.

Then a square symmetric covariance matrix $\mathbf{A}$ is calculated by

$$\mathbf{A} = \frac{1}{w-1}(\mathbf{X} - \mathbf{\Gamma})^T (\mathbf{X} - \mathbf{\Gamma}), \qquad (2)$$

where $\mathbf{\Gamma}$ is an $h \times w$ mean matrix of $\mathbf{X}$ and T is the transpose operator.

Here, the mean matrix is a matrix in which all elements of each column are equal to the mean value of that column. Next, the eigenvector matrix $\mathbf{S}$ and eigenvalue matrix $\mathbf{\Lambda}$ of $\mathbf{A}$ are determined by

$$\mathbf{AS} = \mathbf{\Lambda S}. \qquad (3)$$

The eigenvector matrix $\mathbf{S}$ can be expressed in a block vector form as defined in

$$\mathbf{S} = [s_1 \; s_2 \; s_3], \qquad (4)$$

where $s_1, s_2, s_3 \in \mathbb{R}^{h \times w}$.

At this point, an empirical mean plane can easily be formed from the first two principal components s_1 and s_2 as shown in Fig. 17(b). The third principal component s_3 is orthogonal to s_1 and s_2. Thus, its normal vector is also orthogonal to the empirical mean plane. From this, the perpendicular distance vector d, which represents the distance between s_3 and the
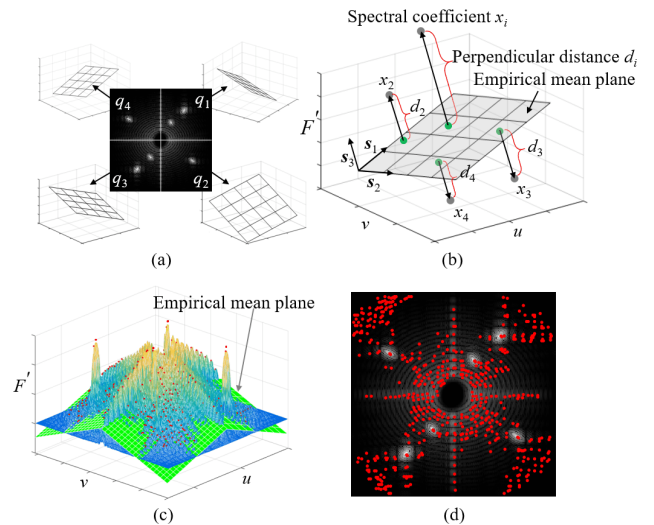

FIGURE 17. Detected candidate central frequencies of monocomponent spectra at a detail region. (a) A detail region divided into four quadrants, (b) perpendicular distances d_i from each magnitude spectrum coefficient to the empirical mean plane, (c) a 3D view of detected candidate local maximum points on the four concatenated empirical mean planes, and (d) a top view of Fig. 17(c).

empirical mean plane, is simply determined by

$$d = (\mathbf{X} - \mathbf{\Gamma})s_3. \qquad (5)$$

In general, elements of the perpendicular distance vector d are positive, negative, or zero numbers. The positive and negative numbers mean that those elements are above and below the empirical mean plane, respectively. In the case of zero numbers, those elements are on the empirical mean plane.

The perpendicular distance vector d of quadrant q_1 is denoted as dq_1 and is converted into a perpendicular distance matrix $\mathbf{D}q_1$ to easily compare to its neighbors in the final step. Each perpendicular distance element d_i of $\mathbf{D}q_1$ is graphically shown in Fig. 17(b). Subsequently, the perpendicular distance matrix of quadrant q_2 is determined in the same way as that of
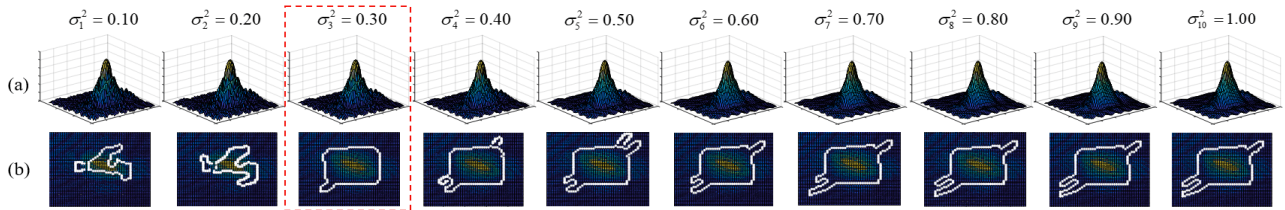
FIGURE 18. Smoothed energy concentration with a Gaussian function at different scales and detected spectrum boundaries of component C-I_4. (a) Results of varying the scale parameter from a small value to a large value, and (b) results of detecting base boundaries.

quadrant q_1. At this point, the symmetry property of Fourier transform is simply applied to the perpendicular distance matrices of quadrants q_1 and q_2 so as to obtain those of quadrants q_3 and q_4. Finally, all perpendicular distance matrices are concatenated into a single matrix $\hat{\mathbf{D}}$ as expressed in

$$\hat{\mathbf{D}} = \begin{bmatrix} \mathbf{D}_{q_4} & \mathbf{D}_{q_1} \\ \mathbf{D}_{q_3} & \mathbf{D}_{q_2} \end{bmatrix}, \qquad (6)$$

where $\mathbf{D}_{q_1}$, $\mathbf{D}_{q_2}$, $\mathbf{D}_{q_3}$, and $\mathbf{D}_{q_4}$ are the perpendicular distance matrices of quadrants q_1, q_2, q_2, and q_4, respectively.

According to *Definitions 1* and *2*, the central frequency of a monocomponent spectrum is a mountain peak of energy concentration. In other words, it is a local maximum point of the spectrum. Therefore, it is easy to detect all candidate local maximum points in a detail region by using the following constraints: if a spectrum coefficient x_i is above the empirical mean surface, which is generated from four empirical mean planes, and is greater than its eight neighbors, that coefficient is defined as a local maximum point $\{p_m\}_{m=1,...,M}$. Figs. 17(c) and 17(d) show the detected candidate central frequencies in the detail region on the empirical mean surface in 3D view and top view, respectively.

2) 2D LOCAL MINIMUM BOUNDARY DETECTION
2D local minimum boundary detection is a process that is designed to adaptively detect candidate monocomponent spectrum boundaries in Fourier support. From *Definitions 1* and *2*, the base of a mountain-shaped energy concentration can perfectly represent the monocomponent spectrum boundary. However, as a result of the 2D local maximum point detection (see Fig. 17(c)), the 2D Fourier spectrum of a typical image always has a rough surface, making it inaccurate in detecting the boundaries of monocomponent spectra. Therefore, we propose *region growing based on a scale-space representation algorithm* to detect the base of a mountain-shaped energy concentration. In this algorithm, Gaussian function G_1 with a scale parameter is applied to 2D Fourier spectrum $F(u, v)$ to automatically smooth its rough surfaces as defined in

$$L\left(u, v; \sigma_l^2\right) = F(u, v) * G_1\left(u, v; \sigma_l^2\right), \qquad (7)$$

$$G_1\left(u, v; \sigma_l^2\right) = \frac{1}{2\pi\sigma^2} e^{-\left(\frac{u^2+v^2}{2\sigma^2}\right)}, \qquad (8)$$

where L is the smoothed 2D Fourier spectrum; $*$ denotes the convolution operator; the semicolon in arguments of L and

G_1 means that the convolution is performed only over the variables u and v; and σ^2 is the variance of the Gaussian function. This variance is defined as a scale parameter whose subscript l is the index.

Fig. 18(a) shows the results of varying the scale parameter from a small value to a large value. Then the proposed region growing based on a scale-space representation algorithm is applied to detect the base boundaries of a mountain-shaped energy concentration. Fig. 18(b) shows the detected base boundaries at different scale values. The best one from these candidate boundaries represents the monocomponent spectrum boundary. It can be seen that the detected base boundary at the scale value 0.30 in Fig. 18(a) is the most suitable. In addition, the criterion for selecting the best detected base boundary by varying the scale parameter is essential. Based on preliminary experiments, it is revealed that the maximum region growth rate as demonstrated in Fig. 19 always provides the best detected base boundary. Therefore, the maximum region growth rate is used as a criterion for selection.
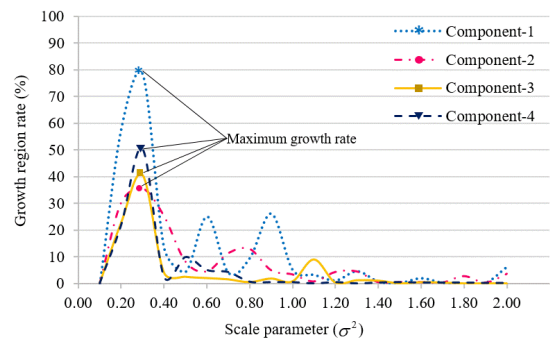
FIGURE 19. Relation between a scale parameter and a growth-region rate.

In order to explain how the region growing based on scale-space representation algorithm works, the key variables and parameters are defined as follows. Let r be a distance from a central frequency to its boundary, L^τ be a boundary element of a local search region, L^ε be a neighboring element of L^τ, $R(u, v; \sigma_l^2)$ be a growth region, and Figs. 20(a)-20(d) be a sample of 2D Fourier spectra. More formally, the algorithm can be presented as having the following steps.

Step 1: Initialize $\sigma_l^2 = 0.10$ and $r = 1$ at p_m.
Step 2: Convolve the 2D Fourier spectrum $F(u, v)$ with Gaussian function $G_1(u, v; \sigma_l^2)$ as defined in Eq. (8).
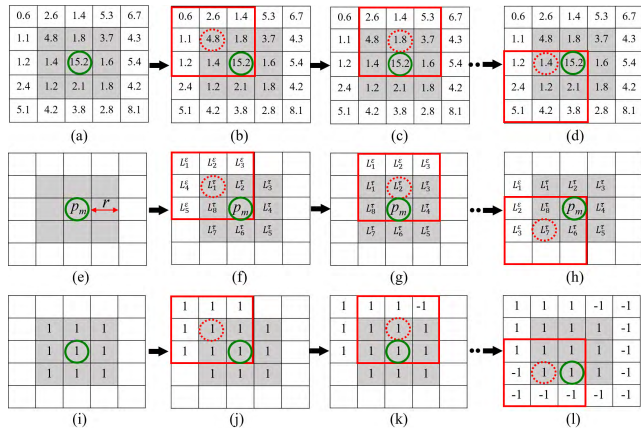
FIGURE 20. 2D local minimum boundary detection on a grey-colored local search region with an initialized distance $r = 1$. (a) - (d) Examples of smoothed 2D Fourier spectrum L for local search region in Steps 1 to 5, (e) - (h) defined spectrum coefficient neighbors of L^τ in Steps 1 to 5, and (j) - (l) growth region in Steps 1 to 5.

Step 2.1: Define the local search region for a local maximum point p_m with a square size of $(2r + 1)^2$ elements, and set a growth region $R(u, v; \sigma_l^2)$ with the constraint that is defined by

$$R\left(u, v; \sigma_l^2\right) = \begin{cases} 1, & \text{if } r = 1 \\ \text{do nothing}, & \text{otherwise,} \end{cases}$$
(9)

where 1 represents an active element. Figs. 20(e) and 20(i) show the initial local search region in gray color and initial growth region $R(u, v; \sigma_l^2)$ at $r = 1$, respectively.

Step 2.2: Set a 3×3 mask on the local search region boundary and define spectrum coefficient neighbors $\{L_y^\varepsilon\}_{y=1,...,Y}$ of L_j^τ but exclude $\{p_m, \{L_j^\tau\}_{j=1,...,J}\}$ where I and J are the maximum number of L_y^ε and L_j^τ, respectively, as demonstrated in Fig. 20 (f).

Step 2.3: Detect the growth region $R(u, v; \sigma_l^2)$ at p_m by using a sign function defined by

$$R\left(u, v; \sigma_l^2\right) = \begin{cases} 1, & \text{if } \{L_i^\varepsilon\}_{i=1,...,I} \le L_j^\tau \\ -1, & \text{otherwise,} \end{cases}$$
(10)

where -1 represents an inactive element; that is, an outbound boundary of a region is detected. Figs. 20(j) to 20(l) demonstrate the detected growth region $R(u, v; \sigma_l^2)$ derived from Figs. 20(f) to 20(h) of Figs. 20 (b) to 20(d), respectively.

Step 2.4: Increase r by 1 and then repeat Steps 2.1-2.4 until all elements in the outbound boundary of the growth region $R(u, v; \sigma_l^2)$ are equal to -1. Fig. 21 shows the results on the growth region of increasing distance r.

Step 3: Increase σ_l^2 with a step size 0.10, and reset r to 1, and then repeat Step 2 until σ_l^2 equals 2.00.

Step 4: Select the argument γ of maximum growth rate of a growth region from all scales σ_l^2 at p_m by using the criterion defined by

$$\gamma = \arg\max_{\sigma_l^2} \left\{ \frac{\sum_{u=1}^{U}\sum_{v=1}^{V}\left|R\left(u, v; \sigma_l^2\right) - R\left(u, v; \sigma_{l-1}^2\right)\right|}{\sum_{u=1}^{U}\sum_{v=1}^{V}R\left(u, v; \sigma_l^2\right)} \times 100 \right\}.$$
(11)

Step 5: Use γ to retrieve the best growth region from all candidates in $R(u, v; \sigma_l^2)$ to represent a monocomponent spectrum boundary of $R(u, v; \sigma_\gamma^2)$. Fig. 22(a) shows the result of a detected monocomponent spectrum boundary of a local maximum point.

The outcome of the local minimum boundary detection algorithm is a set of monocomponent spectrum boundaries surrounding the local maximum points as illustrated in Fig. 22(b). This algorithm can adaptively detect candidate monocomponent spectrum boundaries. Moreover, it automatically provides a suitable scale-space selection based on a Gaussian function. However, there are many overlapping candidate monocomponent spectrum boundaries. Therefore, an actual monocomponent identification algorithm is required to identify actual monocomponent spectrum boundaries as described in the next section.

3) ACTUAL MONOCOMPONENT IDENTIFICATION ALGORITHM

According to two previous algorithms, 2D local maximum point detection and 2D local minimum boundary detection, which provide a large possible number of monocomponent spectrum boundaries as shown in Fig. 22(b), an actual monocomponent identification algorithm is essential for identifying the actual ones from a large number of sets. As mentioned in Sections III-B-1) and III-B-2), a set of local maximum points p_m and monocomponent spectrum boundaries ω are required to identify the actual monocomponents. The algorithm concept is to search for the best of the monocomponent spectrum boundaries. To do this, each local maximum point is compared to its nearest neighbors; if its peak is the highest and its boundary is non-overlapped, that local maximum point is identified as an actual monocomponent. Furthermore, the separated vertical (column) and horizontal (row) searches are very useful for seeking the actual monocomponent spectrum boundaries. This makes the algorithm not only identify actual monocomponents but also consume less computation time. In order to clearly explain how the actual monocomponent identification algorithm works, the key variables are defined as follows. Let p_m be a local maximum point matrix and ωp_m be the spectrum boundary of p_m. The algorithm can be described in the following steps.
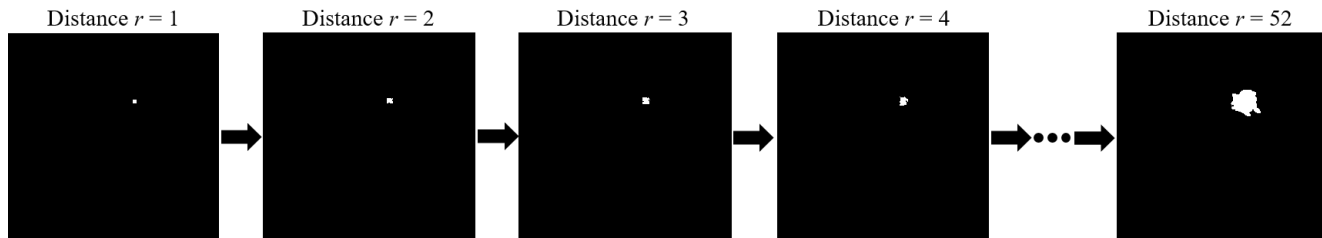
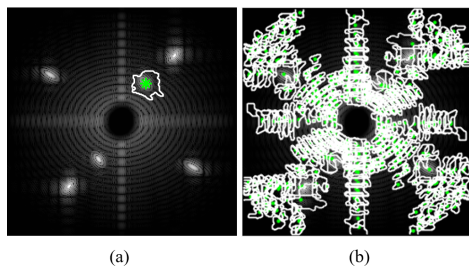**FIGURE 21.** Detected growth regions of increasing distance *r*.

FIGURE 22. Detected monocomponent spectrum boundaries at a detail region. (a) A detected monocomponent spectrum boundary obtained from Fig. 21 and (b) detected monocomponent spectrum boundaries of all local maximum points.

Step 1: Set $t = 1$ where t is the iteration order.

Step 2: Convert a matrix $p_m^{(t)}$ to a 1D array $P^{(t)}$ by concatenating either rows or columns of the matrix.

Step 3: Compare each $p_m^{(t)}$ to the nearest neighbors by using (12) and update array $P^{(t+1)}$.

$$P^{(t+1)} = \{p_m^{(t)} \mid p_{m-1}^{(t)} < p_m^{(t)} \text{ and } p_m^{(t)} > p_{m+1}^{(t)}\}. \quad (12)$$

Step 4: If all monocomponent spectrum boundaries $\omega p_m^{(t+1)}$ indexed by $P^{(t+1)}$ are non-overlapping, then they are actual monocomponent spectrum boundaries and the algorithm is terminated. Otherwise, go to Step 5.

Step 5: If $\omega p_m^{(t+1)}$ indexed by $P^{(t+1)}$ are not empty, then increase t by 1 and go to Step 3. Otherwise, overlapping boundaries are grouped into a single monocomponent spectrum boundary ω_s by means of the following union operator.

$$\omega_s = \bigcup_{m=1}^{V} \omega p_m^{(t)}, \quad (13)$$

where V denotes the maximum number of overlapping elements in each group. If there is more than one group, the overlapping boundaries are independently grouped by isolating them from the other overlapping boundaries. The result of the union operation is identified as the actual monocomponent spectrum boundary and then the algorithm is terminated. A final set of monocomponent spectrum boundaries consists of independent grouped-boundaries and non-overlapping boundaries.

Based on the synthesized image-I, the outcome of each iteration is shown in Fig. 23(a) for the row search and Fig. 23(b) for the column search. The results of identifying actual monocomponent spectrum boundaries in both row and column searches are not different. This algorithm can identify actual monocomponent spectra and solve the inaccurate segmentation problem that causes redundancy or interference components. Therefore, the monocomponent spectrum boundaries detected by our algorithm are more accurate than those obtained by 2D EWT methods. However, in nature scene images, the identification results of row and column searches are probably different. Hence, to support a variety of real-world datasets, this algorithm is able to identify actual monocomponent spectrum boundaries in either row or column search. The central frequency of each monocomponent spectrum is accurately detected as illustrated in Fig. 23, but its boundary shape is still not complete. Therefore, the detected monocomponent spectrum boundaries are used to estimate the complete monocomponent spectrum boundaries prior to creating actual bandpass filter banks called "empirical image filter banks" as presented in the next section.

C. EMPIRICAL IMAGE FILTER BANK CONSTRUCTION

Empirical image filter bank construction is the process of constructing a bandpass filter bank in accordance with the detected monocomponent spectrum boundary which is the base of a mountain-shaped energy concentration. This filter bank is useful for decomposing an input image (2D signal) into a set of monocomponent images with fewer ringing artifacts. Nevertheless, as a result of Section III-B-3), the shape of the actual monocomponent spectrum boundaries is incomplete for decomposing monocomponent images efficiently. Hence, the empirical image filter bank design based on an ellipse function is proposed to provide the complete monocomponent spectrum boundary. Moreover, ringing artifacts which appear in the monocomponent images as illustrated in Fig. 8 are an outcome of the constructed wavelet filter bank by 2D EWT. Typically, these artifacts can be removed by using a simple low-pass filter [37], i.e., a Gaussian filter because it has the property of being non-oscillatory. Thus, we use a Gaussian function to construct an empirical image filter bank in order to reduce the artifacts. The empirical image filter bank construction for decomposing monocomponent images consists of two stages: (i) generation of complete
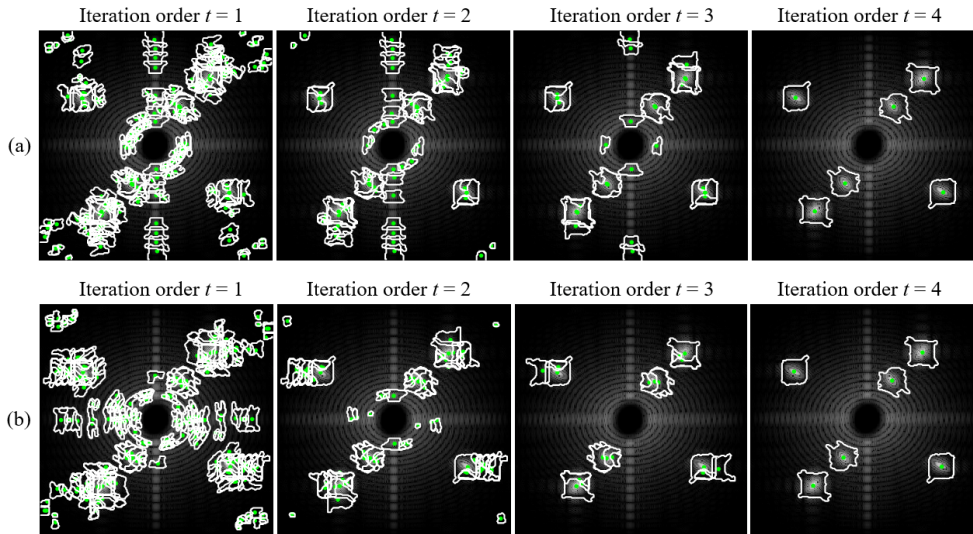
FIGURE 23. Illustration of step-by-step identification of actual monocomponent spectrum boundaries in a detail region. (a) Actual monocomponent identification in row search and (b) actual monocomponent identification in column search.
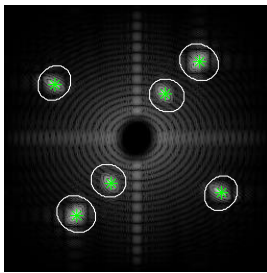
FIGURE 24. Complete monocomponent spectrum boundaries at a detail region of the synthesized image-I.

monocomponent spectrum boundaries and (ii) construction of empirical image filter banks. In the former stage, the u-coordinate and v-coordinate of a monocomponent spectrum boundary derived from the actual monocomponent identification algorithm are used to estimate the corresponding u'-coordinate and v'-coordinate of a complete monocomponent spectrum boundary by means of an ellipse function defined in

$$u' = u'_0 + \alpha \cos \alpha \cos \theta - b \sin \alpha \sin \theta, \qquad (14)$$
$$v' = v'_0 + \alpha \cos \alpha \sin \theta - b \sin \alpha \cos \theta, \qquad (15)$$

where (u'_0, v'_0) is the center of an ellipse; a and b represent the major axis and the minor axis of the ellipse; θ is the orientation of the ellipse; and α is the angular parameter with an interval $[0, 2\pi]$. (u'_0, v'_0), a, b, and θ are estimated by means of least square fitting of the ellipse function as recommended in [38]. Figs. 16(c) and 24 show the complete monocomponent spectrum boundaries of the approximation and detail regions, respectively. Each complete monocomponent spectrum boundary δ is used to generate a complete

monocomponent spectrum region β defined in

$$\beta (u', v') = \begin{cases} 1, & \delta (u', v') \in E \\ 0, & \text{otherwise,} \end{cases} \qquad (16)$$

where E is the ellipse region.

In the latter stage, the complete monocomponent spectrum region β is used to construct an empirical image filter bank ψ defined in

$$\psi (u', v') = G_2 \left(u', v'; \sigma^2 \right) * \beta (u', v'), \qquad (17)$$

where G_2 is the Gaussian function, which is similar to G_1 defined in Eq.(8). The σ^2 value is calculated from the minimum of a and b.
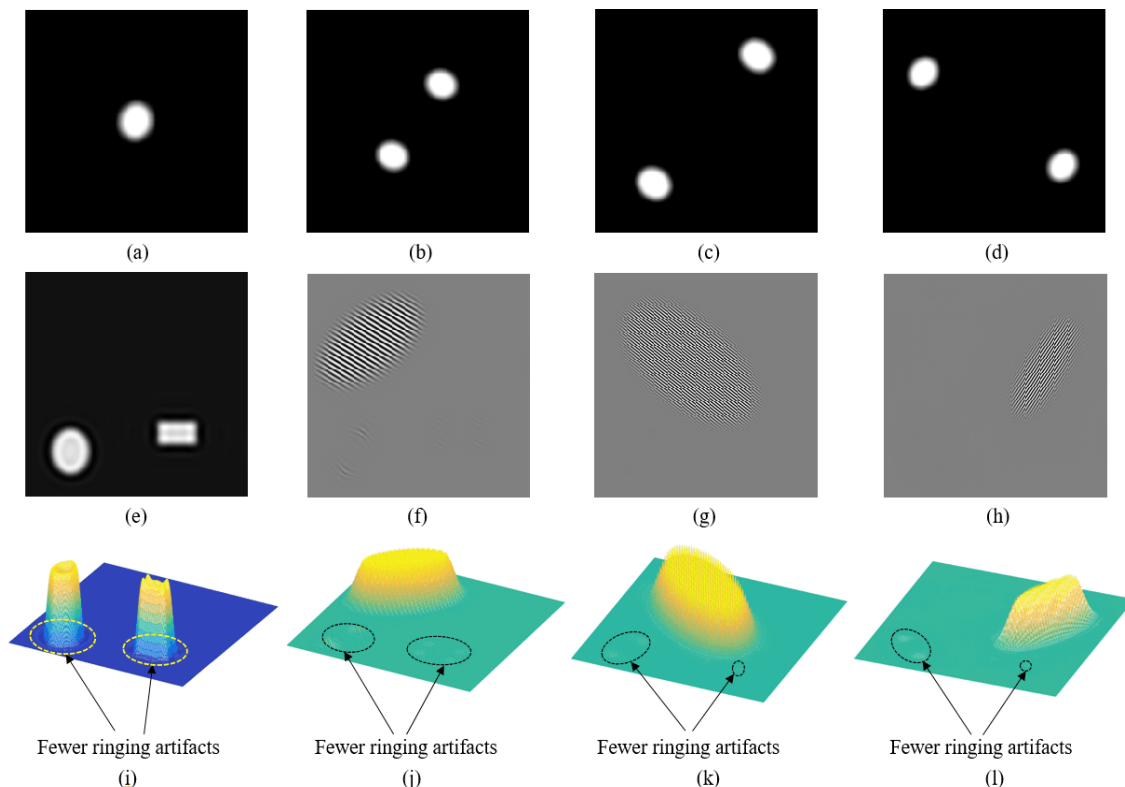
The empirical image filter banks for approximation coefficient ψ_a and detail coefficient ψ_d are constructed from Eq. (15). These filters are used to decompose the monocomponent spectrum from approximation and detail regions. The monocomponent images $\mathcal{M}_a$ and $\mathcal{M}_d$ are decomposed from the approximation and detail regions by applying inverse Fourier transform to the dot product of $F(u, v)$ and $\psi_a(u', v')$, and of $F(u, v)$ and $\psi_b(u', v')$ as expressed in

$$\mathcal{M}_a = F^{-1} \left(F(u, v) \cdot \psi_a (u', v') \right), \qquad (18)$$
$$\mathcal{M}_d = F^{-1} \left(F(u, v) \cdot \psi_d (u', v') \right), \qquad (19)$$

where $(.)$ is the dot product operator.

Based on the synthesized image-I, Fig. 25(a) shows the empirical image filter bank for the approximation coefficient of component C-I$_1$. On the other hand, the empirical image filter banks for detail coefficients of components C-I$_2$, C-I$_3$, and C-I$_4$ are shown in Figs. 25(b), 25(c), and 25(d), respectively. This process yields a set of monocomponent images $\{\mathcal{M}_d, \{\mathcal{M}_d\}_{n=1}^{N-1}\}$, where N is the number of monocomponent images. Based on the synthesized image-I,

FIGURE 25. Illustrations of empirical image filter banks constructed for the proposed EMID method and decomposed components of synthesized image-I. (a) Empirical image filter banks of the approximation region for decomposing component C-I$_1$, (b), (c), and (d) empirical image filter banks of the detail region for decomposing components C-I$_2$, C-I$_3$, and C-I$_4$, (e) $\mathcal{M}_a$ component image of C-I$_1$ in top view, (f), (g), and (h) $\mathcal{M}_d$ component images of components C-I$_2$, C-I$_3$, and C-I$_4$ in top view, (i) reduced ringing artifacts of the $\mathcal{M}_a$ component image of component C-I$_1$ in 3D view, and (j), (k), and (l) reduced ringing artifacts of the $\mathcal{M}_d$ component images in 3D view.

Fig. 25(e) shows the reconstructed monocomponent image of flat objects by taking the inverse Fourier transform of component C-I$_1$ at the approximation region. Although the flat object edge of component C-I$_1$ is blurred, the object shape is similar to that of the reference image-I. In the same way, Figs. 25(f), 25(g) and 25(h) show the reconstructed monocomponent images of textured objects by taking the inverse Fourier transform of components C-I$_2$, C-I$_3$, and C-I$_4$ at the detail region, respectively. It can be seen that the texture objects of those components, C-I$_2$ to -I$_4$, and those of reference image-II to -IV are similar. Moreover, Figs. 25(i) to 25(l) show that ringing artifacts appearing in these four components are greatly reduced due to the efficient empirical image filter bank.

Also, those components are monocomponent images according to *Definitions 1* and *2*. However, partial spectrum loss commonly occurs with real world images in the form of strips. To solve this problem, a set of compensated filter banks Z are constructed as expressed in

$$Z = \{\zeta_1, \zeta_2, \zeta_3, \ldots, \zeta_C\}, \tag{20}$$

where C is the number of compensated filter banks and ζ is a compensated filter bank. More details for this issue are implemented and presented in Section IV-B.

IV. EXPERIMENTS AND DISCUSSION
In this section, the performance of the proposed EMID method is evaluated (i) with a synthesized image-II in terms of the quality of monocomponent images and (ii) with a real image in terms of Thai text localization. Two experiments are set up to test each of these aspects as follows.
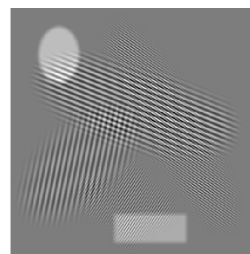
FIGURE 26. Synthesized image-II (a toy image proposed by Gilles *et al.* [6], Gilles [28]).

A. QUALITY OF MONOCOMPONENT IMAGES
The first experiment aims to evaluate the quality of monocomponent images. This experiment is designed with the following steps. In data preparation and measurement, synthesized image-II (a toy image) in Fig. 26, proposed
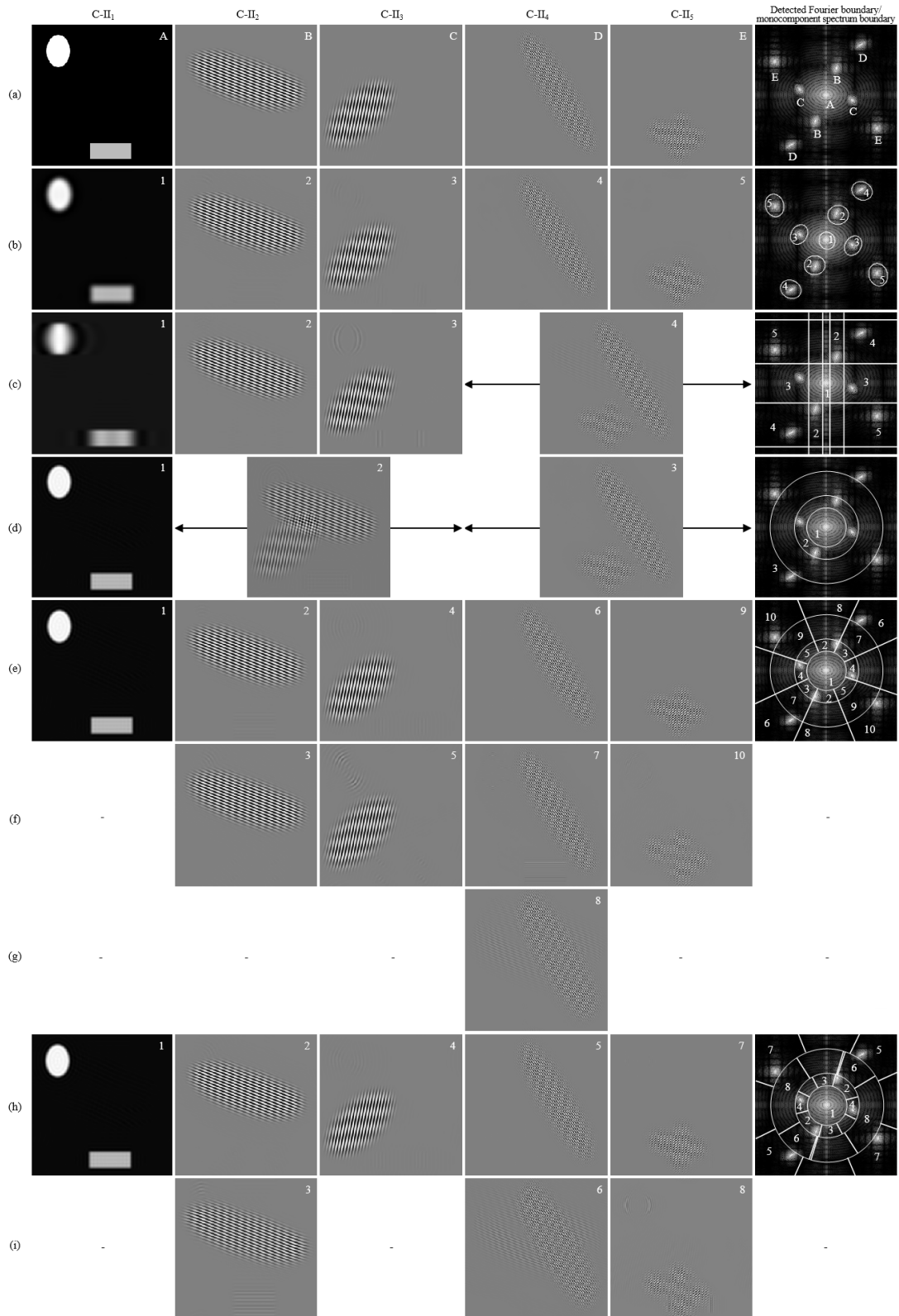
FIGURE 27. Components of synthesized image-II decomposed by the proposed EMID and baseline methods in manual mode. (a) Reference images, (b) images decomposed by the proposed EMID method, (c) 2D Tensor EWT, (d) 2D Littlewood-Paley EWT, (e) 2D Curvelet EWT-I, (f) redundancy components from 2D Curvelet EWT-I, (g) 2D Curvelet EWT-I, (h) 2D Curvelet EWT-II, and (i) redundancy components from 2D Curvelet EWT-II.

by Gilles *et al.* [6] and Gilles [28], is used to evaluate the performance of the proposed and baseline methods. It is generated from a combination of five reference images displayed in columns 1 to 5 of Fig. 27(a). The reference image-I (upper-left corner of Fig. 27) contains a flat object that is composed of low as well as high frequencies. The other reference image-II to -V (the 2nd-5th columns of Fig. 27(a)) contain texture objects that are composed of different high frequencies. When this synthesized image-II is transformed into Fourier domain, its spectrum locations in Fourier support of reference image-I to -V are labeled with letters A to E, respectively, as depicted in column 6 of Fig. 27(a). For ease of observation and explanation, the reference image-I in Fourier domain possessing more low frequencies is defined to be a monocomponent spectrum-1 (see 6th column of Fig. 27(a) labeled with letter A). In the same way, the other reference image-II, -III, -IV and -V possessing high frequencies are defined to be monocomponent spectrum-2, -3, -4 and -5 (see 6th column of Fig. 27(a) labeled with letters B, C, D, and E), respectively. Component-1 to -5 (C-II$_1$ to C-II$_5$) are obtained from decomposing monocomponent spectrum-1 to -5, respectively, of the synthesized image-II. Obtained images C-II$_1$ to C-II$_5$ are monocomponent image-1 to -5, respectively, when they are perfectly decomposed with high image quality. For quality measurement, these reference images are used to assess the quality of monocomponent images decomposed by all test methods in terms of peak signal-to-noise ratio (PSNR) [29] and structural similarity index (SSIM) [30]. The PSNR is a classical metric for measuring image quality based on the intensity of the image. A high PSNR value means a high quality image. On the other hand, the SSIM based on the image structure is more consistent with human-eye perception than the PSNR as reported in [6]. The range of SSIM value is from -1.0 to 1.0 where 1.0 means the best quality and -1.0 means the worst quality. In order to clearly assess the quality of monocomponent images, both PSNR and SSIM metrics are used to test the hypothesis of experiments which seek to determine an efficient method of decomposing monocomponent images from a synthesized image-II with high quality.

In experimental conditions, the baseline methods are tested in both non-adaptive and adaptive decomposition for monocomponent images. The non-adaptive decomposition consists of 2D discrete wavelet transform (2D DWT) and 2D stationary wavelet transform (2D SWT) [39]. The 2D SWT is derived from the 2D DWT by modifying the decimation operation to support the translation invariance property [40]. On the other hand, the adaptive decomposition consists of two main methods: 2D empirical mode decomposition (2D EMD) and 2D empirical wavelet transform (2D EWT). The 2D EMD method [41] is derived from a bidimensional empirical mode decomposition (BEMD) based on radial basis function (RBF). This method is more practical for performance evaluation due to faster decomposition. In addition, all 2D EWT methods are tested in two segmentation modes: manual and automatic. The manual mode requires a suitable number

of components for spectrum segmentation, but the automatic mode does not. In the manual mode, the conventional 2D EWT methods including 2D Tensor EWT, 2D Littlewood-Paley EWT, 2D Curvelet EWT-I, and 2D Curvelet EWT-II introduced by Gilles *et al.* [6] require an appropriate number of components for decomposing an image into monocomponent images. The criteria used for Fourier boundary detection are based on local minimum and/or local maximum detection in conjunction with a preprocessing stage for removing the global trend problem. Additionally, all parameters of the baseline methods are set as recommended by [6]. In the automatic mode, the conventional 2D EWTs automatically select the number of components based on the coarse histogram in conjunction with Otsu's algorithm [31]. Gilles *et al.* introduced a series of 2D Curvelet EWT called 2D Curvelet EWT-III [32]. The type III of this series is also included in the experiment for testing a variety of possible component segmentations of the conventional 2D EWT methods. Our proposed EMID method, which is already designed for fully automatic component segmentation, is tested only in the second mode.

As listed in Table 1, the test results of PSNR clearly demonstrate that the proposed EMID method outperforms the conventional 2D EWT methods in terms of quality of monocomponent images. The PSNR values of components C-II$_2$, C-II$_3$, C-II$_4$, and C-II$_5$ decomposed by the proposed method are higher than those decomposed by other baseline methods in both manual and automatic modes. However, the PSNR value of component C-II$_1$ decomposed by the proposed method is 24.1154 dB which is lower than that of component C-II$_1$ of 2D Littlewood-Paley EWT, 2D Curvelet EWT-I, and 2D Curvelet EWT-II in both manual and automatic modes. When we look inside the decomposed components, the PSNR value of components C-II$_1$ to C-II$_5$ decomposed by 2D DWT, 2D SWT, and 2D EMD is lower than that of the proposed method. Furthermore, the PSNR values of components C-II$_1$, C-II$_4$, and C-II$_5$ decomposed by 2D Tensor EWT in manual mode are much lower than those of the proposed method. Likewise, the PSNR values of components C-II$_2$, C-II$_3$, C-II$_4$, and C-II$_5$ decomposed by 2D Littlewood-Paley EWT in manual and automatic modes are much lower than those of the proposed method. The PSNR value of component C-II$_1$ decomposed by 2D Tensor EWT in automatic mode is significantly lower than that of the proposed method. The average PSNR value obtained by the proposed method is 52.0803 dB which is higher than those of all of the baseline methods. In particular, the average PSNR values decomposed by the proposed method are 20.1123 dB, 30.0208 dB, and 30.0213 dB higher than those decomposed by 2D Tensor EWT in manual mode and 2D Littlewood-Paley EWT in manual and automatic modes, respectively. These are dramatically higher values.

SSIM assessment also demonstrates that the proposed method outperforms all baseline methods. Table 2 shows that the SSIM values obtained by the proposed method are substantially higher than those of all of the baseline methods

TABLE 1. Comparison between the PSNR values of each component of a synthesized image-II provided by the proposed and baseline methods.

Method	Type of decomposition	Mode	PSNR (dB)					
			C-II$_1$	C-II$_2$	C-II$_3$	C-II$_4$	C-II$_5$	Average
2D DWT	Non-adaptive method	-	16.4358	19.4231	20.6738	18.9972	19.5427	19.0145
2D SWT	Non-adaptive method	-	17.3561	21.3468	22.6326	23.1200	20.6532	21.0217
2D EMD	Adaptive method	-	23.4738	21.8131	20.0915	19.7481	18.5427	20.7338
2D Tensor EWT	Adaptive method	Manual	19.7606	50.0937	46.9472	24.3651	18.6737	31.9680
2D Littlewood-Paley EWT			<u>25.8132</u>	22.8584	18.9110	24.1628	18.5520	22.0595
2D Curvelet EWT-I			<u>25.8198</u>	36.7557	49.7259	46.3487	49.2126	41.5725
2D Curvelet EWT-II			<u>25.8198</u>	47.6955	53.5462	46.5355	49.1375	44.5469
2D Tensor EWT	Adaptive method	Automatic	19.0220	44.5943	48.9657	65.5354	50.0321	45.6299
2D Littlewood-Paley EWT			25.7159	22.7739	18.8729	24.2653	18.6669	22.0590
2D Curvelet EWT-I			25.7120	47.9210	53.7183	47.1721	49.7323	44.8512
2D Curvelet EWT-II			25.7120	35.6992	47.6871	46.9266	48.8032	40.9656
2D Curvelet EWT-III			20.7857	41.9059	54.7039	66.8619	49.3472	46.7209
Proposed method	Adaptive method	Automatic	24.1054	**50.7559**	**55.4706**	**71.2563**	**59.8131**	**52.0803**

Bold numbers indicate that the proposed method provides the best results while underlined numbers indicate that the baseline methods provide better results.

TABLE 2. Comparison between the SSIM values of each component of a synthesized image-II provided by the proposed and baseline methods.

Method	Type of decomposition	Mode	SSIM					
			C-II$_1$	C-II$_2$	C-II$_3$	C-II$_4$	C-II$_5$	Average
2D DWT	Non-adaptive method	-	0.0413	0.4231	0.3652	0.3817	0.3031	0.3029
2D SWT	Non-adaptive method	-	0.0909	0.3476	0.2876	0.6617	0.4575	0.3691
2D EMD	Adaptive method	-	0.0336	0.3895	0.2684	0.6241	0.4222	0.3476
2D Tensor EWT	Adaptive method	Manual	0.0149	0.5860	0.7182	0.6252	0.4664	0.4822
2D Littlewood-Paley EWT			0.0256	0.3502	0.2357	0.6475	0.4893	0.3496
2D Curvelet EWT-I			0.0256	0.5365	0.5649	0.7709	0.8618	0.5519
2D Curvelet EWT-II			0.0256	0.5431	0.5752	0.7719	0.8527	0.5537
2D Tensor EWT	Adaptive method	Automatic	0.0167	0.5297	0.5292	0.7060	0.7978	0.5159
2D Littlewood-Paley EWT			0.0255	0.3398	0.2227	0.4636	0.3140	0.2731
2D Curvelet EWT-I			0.0255	0.5307	0.4615	0.6676	0.7975	0.4965
2D Curvelet EWT-II			0.0255	0.4848	0.4521	0.6137	0.6403	0.4433
2D Curvelet EWT-III			0.0135	0.3789	0.3629	0.6442	0.5524	0.3904
Proposed method	Adaptive method	Automatic	**0.2330**	**0.9987**	**0.9964**	**0.9899**	**0.9989**	**0.8434**

Bold numbers indicate that the proposed method provides the best results.

in both manual and automatic modes. Because of the SSIM value based on the structure image, the shape of the object in components decomposed by the proposed method is similar to reference images, thus the image quality of those components decomposed by the proposed method is the highest.

Comparison between five reference images and components decomposed by the proposed and baseline methods are illustrated in Figs. 27, 28, and 29 where different rows are from different decomposition methods. Columns 1 to 5 in these figures depict decomposed components C-II$_1$ to C-II$_5$, respectively. The last column of Figs. 27 and 28 shows the detected Fourier boundaries of the baseline method and monocomponent spectrum boundaries of the proposed methods. The decomposed component labels correspond to the monocomponent spectrum boundary labels. Different image decompositions mostly detect different monocomponent spectrum boundaries.

The proposed method is able to detect the boundaries of monocomponent spectrum-1 to -5 accurately and completely as shown in column 6 of Fig. 27(b). For this reason, our decomposed components shown in columns 1 to 5

of Fig. 27(b) bear a very close resemblance to the reference images in columns 1 to 5 of Fig. 27(a), respectively. The components are monocomponent images according to *Definitions 1* and *2* in which objects in each component contain a single frequency. Also, there are no interferences or redundancy components, and fewer ringing artifacts. This leads to higher image quality as shown in Tables 1 and 2.

On the other hand, the components decomposed by the baseline methods still exhibit interference and redundancy components. The interference occurs in some components decomposed by 2D Tensor EWT in manual mode and 2D Littlewood-Paley EWT in manual and automatic modes as shown in Figs. 27(c), 27(d), and 28(c). In the same way, the interference occurs in all components decomposed by 2D DWT, 2D SWT, and 2D EMD as illustrated in Fig. 29.

Column 6 of Fig. 27(c) shows detected Fourier boundaries from 2D Tensor EWT in manual mode. The components C-II$_1$, C-II$_2$, and C-II$_3$ in columns 1, 2, and 3 of Fig. 27(c) are from detected Fourier boundaries labeled with numbers 1, 2, and 3 as shown in column 6 of Fig. 27(c), respectively. Obtained components C-II$_1$ and C-II$_3$ are different from the
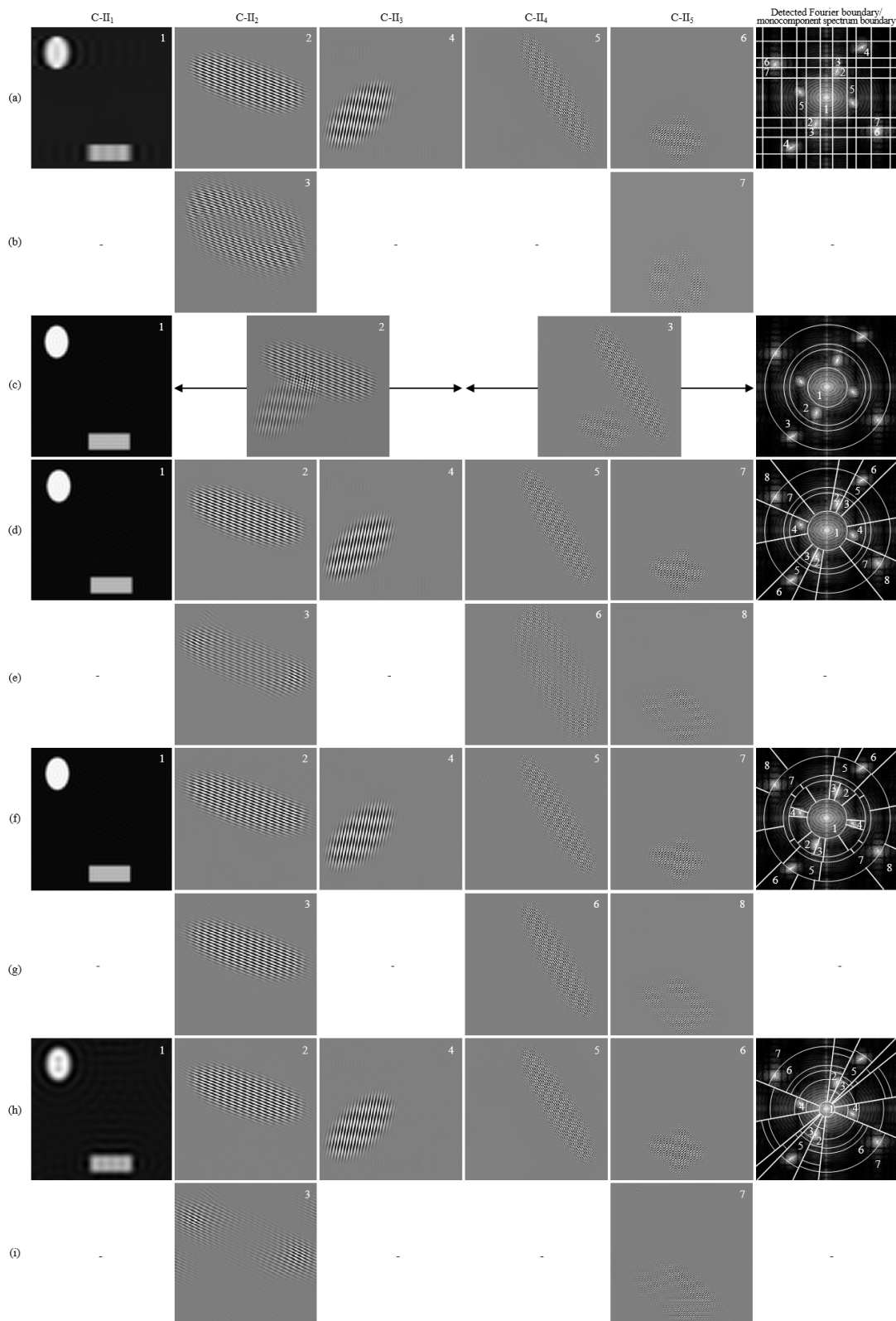
FIGURE 28. Components of the synthesized image-II decomposed by the baseline methods in automatic mode. (a) 2D Tensor EWT, (b) redundancy components created by 2D Tensor EWT, (c) 2D Littlewood-Paley EWT, (d) 2D Curvelet EWT-I, (e) redundancy components created by 2D Curvelet EWT-I, (f) 2D Curvelet EWT-II, (g) redundancy components created by 2D Curvelet EWT-II, (h) 2D Curvelet EWT-III, and (i) redundancy component created by 2D Curvelet EWT-III.
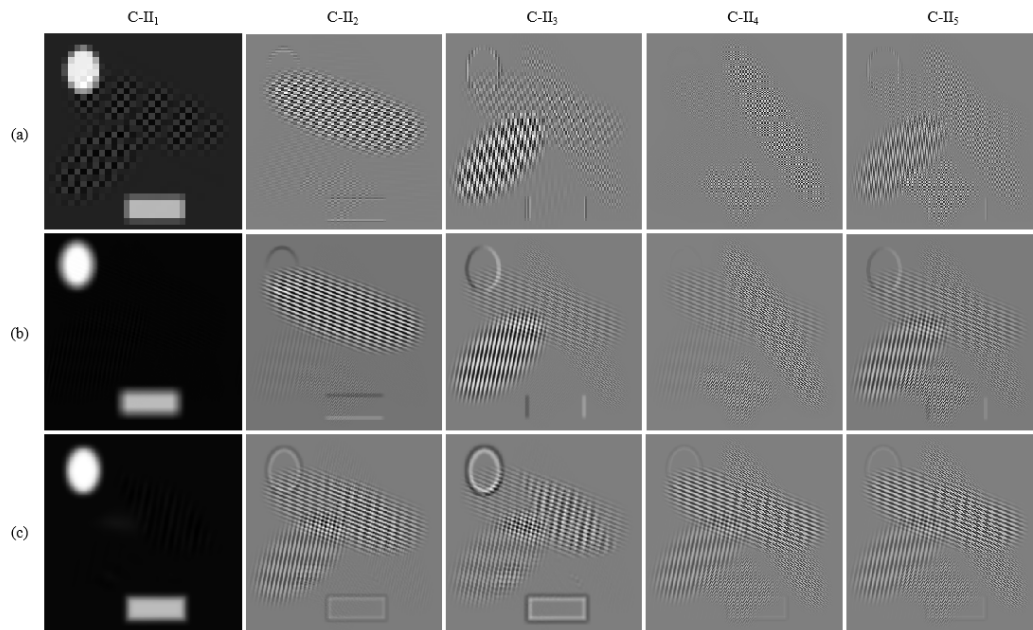
FIGURE 29. Components of the synthesized image-II decomposed by (a) 2D DWT, (b) 2D SWT, and (c) 2D EMD.

reference image-I and -III, and ringing artifacts appear in component C-II$_3$; this leads to their low quality. On the contrary, component C-II$_2$ is very similar to reference image-II and so is high in quality. Other detected Fourier boundaries labeled with numbers 4 and 5 result in an interference component between C-II$_4$ and C-II$_5$. This component (labeled with number 4) is shown in the image between columns 4 and 5 in Fig. 27(c). It is used to assess image quality of components C-II$_4$ and C-II$_5$ by comparing it with reference image-IV and -V. The components C-II$_4$ and C-II$_5$ are low quality as shown in Tables 1 and 2 due to interference components. Moreover, the interference component labeled with number 4 is not a monocomponent image.

The Fourier boundaries detected by 2D Littlewood-Paley EWT in manual mode are illustrated in column 6 of Fig. 27(d). Component C-II$_1$, shown in column 1 of Fig. 27(d), is from detected Fourier boundaries labeled with number 1. It is similar to the reference image-I. On the contrary, components C-II$_2$ to C-II$_5$ differ from reference image-II to -V. As can be seen in the image between columns 2 and 3 in Fig. 27(d), the component labeled with number 2, composed of two components C-II$_2$ and C-II$_3$, results from detected Fourier boundary number 2. Also, a component labeled with number 3 composed of C-II$_4$ and C-II$_5$, shown in between columns 4 and 5, is obtained from detected Fourier boundary labeled with number 3. Both are interference components, causing very low image quality.

The 2D Littlewood-Paley EWT in automatic mode can detect Fourier boundaries shown in column 6 of Fig. 28(c). The detected Fourier boundary labeled with number 1 results in component C-II$_1$ depicted in column 1 of Fig. 28(c). Component C-II$_1$ is similar to the reference image-I and

is a high quality image. Nevertheless, the detected Fourier boundary labeled with number 2 results in interference between components C-II$_2$ and C-II$_3$ as illustrated in the image between columns 2 and 3 of Fig. 28(c). In the same way, the component in the image between columns 4 and 5 in Fig. 28(c) results from interference between components C-II$_4$ and C-II$_5$. This component is from the detected Fourier boundary labeled with number 3. Those interference components are clearly different from the reference images and so are low in quality.

These instances of interference mostly come from inaccurate spectrum segmentation due to the use of the average magnitude spectrum technique of the baseline methods. Consequently, the different monocomponent spectra are grouped into the same Fourier boundary. Interference components obtained from this kind of Fourier boundary have low image quality and also are not monocomponent images.

For the non-adaptive method, Fig. 29(a) shows the results of image decomposition of the 2D DWT method. Columns 1 to 5 of Fig. 29(a) depict components C-II$_1$ to C-II$_5$ which are obtained from approximation coefficients of level 3, horizontal detail coefficients of level 2, vertical detail coefficients of level 2, diagonal detail coefficients of level 1, and vertical detail coefficients of level 1, respectively. Component C-II$_1$ is not similar to reference image-I. Components C-II$_2$ to C-II$_5$ occur interference components. This leads to low image quality. In the same way, Fig. 29(b) shows components which are decomposed by the 2D SWT method. Columns 1 to 5 of Fig. 29(b) show components C-II$_1$ to C-II$_5$ which are obtained from approximation coefficients of level 4, horizontal detail coefficients of level 2, vertical detail coefficients of level 2, diagonal detail coefficients of

level 1, and vertical detail coefficients of level 1, respectively. Component C-II$_1$ is similar to reference image-I but the object edge is blurred. Interference components occur in components C-II$_2$ to C-II$_5$. This leads to low image quality. Both non-adaptive methods, 2D DWT and 2D SWT, are unsuccessful in decomposing monocomponent images with high quality. In addition, the results of component decomposition by 2D EMD method are illustrated in Fig. 29(c). Component C-II$_1$ is similar to reference image-I but the object edge is blurred. Also, interference components occur in components C-II$_2$ to C-II$_5$. This leads to low image quality.

Regarding redundancy, some redundancy components decomposed by 2D Curvelet EWT-I, 2D Curvelet EWT-II in Curvelet EWT-II, and 2D Curvelet EWT-III in automatic mode can be seen clearly in Figs. 27(e), 27(f), 27(h), 27(i), 28(a) 28(b), and 28(d) to 28(i).

Column 6 of Fig. 27(e) depicts Fourier boundaries detected by 2D Curvelet EWT-I in manual mode. The boundaries result in components C-II$_1$ to C-II$_5$ in columns 1 to 5 of Fig. 27(e). The component C-II$_1$ from 2D Curvelet EWT-I in manual mode is similar to that from 2D Littlewood-Paley EWT in manual mode because they are both decomposed from equal radius width of Fourier boundaries. The quality of component C-II$_1$ in terms of PSNR and SSIM from 2D Curvelet EWT-I is equal to that from 2D Littlewood-Paley EWT as listed in Tables 1 and 2. Monocomponent spectrum-2 to -5 are segmented into different Fourier boundaries, thus these boundaries result in redundant components C-II$_2$ to C-II$_5$. The monocomponent spectrum-2 is segmented into two different Fourier boundaries labeled with numbers 2 and 3 in column 6 of Fig. 27(e). These boundaries result in two components C-II$_2$ in spatial domain as illustrated in column 2 of Figs. 27(e) and 27(f). The monocomponent spectrum-3 is detected in two Fourier boundaries labeled with numbers 4 and 5 which result in two components C-II$_3$ in column 3 of Figs. 27(e) and 27(f). Redundant components C-II$_4$ shown in column 3 of Figs. 27(e), 27(f), and 27(g) are the result of different Fourier boundaries labeled with numbers 6, 7, and 8. Similarly, two components C-II$_5$ illustrated in column 5 of Figs. 27(e) and 27(f) are the results of different Fourier boundaries labeled with numbers 9 and 10.

Among redundant components, the highest-quality one is used to assess image quality by comparing it with the reference image. For example, component C-II$_2$ in column 2 of Fig. 27(e) has a higher image quality than component C-II$_2$ in column 2 of Fig. 27(f), so it is used to assess image quality by comparing it with the reference image-II. The components in columns 1 to 5 of Fig. 27(e) are used to assess image quality by comparing them with reference images shown in Tables 1 and 2. The image quality depends on accurate spectrum segmentation. The spectrum of a monocomponent is segmented into more than one Fourier boundary which leads to a loss of monocomponent spectrum. Thus, the image quality of the monocomponents is low. The Fourier boundary of component C-II$_3$ is more accurately and completely detected than those of components C-II$_5$, C-II$_4$, and

C-II$_2$. Consequently, the quality of component C-II$_3$ is higher than that of components C-II$_5$, C-II$_4$, and C-II$_2$. Moreover, the redundant components C-II$_2$ to C-II$_5$ produced by 2D Curvelet EWT-I in manual mode are lower quality than those produced by the proposed method.

Fig. 27(h) shows components and detected Fourier boundaries by 2D Curvelet EWT-II in manual mode. Components C-II$_1$ and C-II$_3$ in columns 1 and 3 of Fig. 27(h) are the results of decomposing Fourier boundaries labeled with numbers 1 and 3, respectively. Component C-II$_1$ from 2D Curvelet EWT-II in manual mode is similar to components C-II$_1$ from 2D Curvelet EWT-I and 2D Littlewood-Paley EWT in manual mode, so their components C-II$_1$ have equal image quality.

In 2D Curvelet EWT-II in manual mode, the boundary of monocomponent spectrum-3 is better detected than those of monocomponent spectrum-2, -4, and -5; thus, the image quality of component C-II$_3$ labeled with number 4 is higher than those of components C-II$_2$, C-II$_4$, and C-II$_5$ labeled with numbers 2, 5, and 7, respectively. However, it is still of lower quality than component C-II$_3$ from the proposed method. Detected Fourier boundaries labeled with numbers 2 and 3 result in two redundant components C-II$_2$ in column 2 of Figs. 27(h) and 27(i). Two redundant components C-II$_4$ in column 4 of Figs. 27(h) and 27(i) are the results of decomposing Fourier boundaries labeled with numbers 5 and 6. Two redundant components C-II$_5$ in column 5 of Figs. 27(h) and 27(i) are the results of detected Fourier boundaries labeled with number with numbers 7 and 8. The obtained monocomponent images lose some component information and are of low quality.

The components and Fourier boundaries detected by 2D Tensor EWT in automatic mode are shown in Figs. 28(a) and 28(b). Components C-II$_1$, C-II$_3$, and C-II$_4$ in spatial domain in columns 1, 3, and 4 of Fig. 28(a) are the results of Fourier boundaries labeled with numbers 1, 4, and 5, respectively. Two redundant components C-II$_2$ in spatial domain in column 2 of Figs. 28(a) and 28(b) are the results of Fourier boundaries labeled with numbers 2 and 3. In column 5 of Figs. 28(a) and 28(b) are two components C-II$_5$ that are the results of Fourier boundaries labeled with numbers 6 and 7. The image quality of C-II$_4$ (component labeled with number 5) is higher than those of Components C-II$_5$, C-II$_3$, C-II$_2$, and C-II$_1$ because components C-II$_5$ and C-II$_2$ are redundant components that lose some monocomponent spectrum information from inaccurate spectrum segmentation before image decomposition, and component C-II$_3$ has ringing artifacts.

Column 6 of Fig. 28(d) shows Fourier boundaries detected by 2D Curvelet EWT-I in automatic mode. The detected Fourier boundaries labeled with numbers 1 and 4 result in components C-II$_1$ and C-II$_3$ in columns 1 and 3 of Fig. 28(d), respectively. The obtained component C-II$_3$ has ringing artifacts. The detected Fourier boundaries labeled with numbers 2 and 3 result in two redundant components C-II$_2$ in column 2 of Figs. 28(d) and 28(e). Two components C-II$_4$ in column 4 of Figs. 28(d) and 28(e) are the results of

Fourier boundaries labeled with numbers 5 and 6, respectively. Fourier boundaries labeled with numbers 7 and 8 result in two components C-II$_5$ in column 5 of Figs. 28(d) and 28(e), respectively. The obtained components C-II$_2$, C-II$_4$, and C-II$_5$ in columns 2, 4, and 5 of Fig. 28(d) are of low image quality due to loss of partial monocomponent spectrum information.

Detected Fourier boundaries from 2D Curvelet EWT-II in automatic mode are illustrated in column 6 of Fig. 28(f). Component C-II$_1$ in column 1 of Fig. 28(f) is the result of detected Fourier boundaries labeled with number 1. Component C-II$_1$ from 2D Curvelet EWT-II is similar to that from 2D Curvelet EWT-I in automatic mode. This leads to their equal image quality. The detected Fourier boundary labeled with number 4 results in component C-II$_3$ in column 3 of Fig. 28(f) which has ringing artifacts. Redundant components C-II$_2$ in column 2 of Figs. 28(f) and 28(g) are the result of detected Fourier boundaries labeled with numbers 2 and 3, respectively. Redundant components C-II$_4$ in column 4 of Fig. 28(f) and 28(g) are the result of detected Fourier boundaries labeled with numbers 5 and 6. Redundant components C-II$_5$ in column 5 of Figs. 28(f) and 28(g) are the result of detected Fourier boundaries labeled with numbers 7 and 8. The redundant components have low image quality.

Column 6 of Fig. 28(h) shows detected Fourier boundaries by 2D Curvelet EWT-III in automatic mode. The detected Fourier boundaries labeled with numbers 1, 4, and 5 result in components C-II$_1$, C-II$_3$, and C-II$_4$ which are illustrated in columns 1, 3, and 4 of Fig. 28(h), respectively. Components C-II$_1$ and C-II$_3$ are different from reference image-I and -III and are of low image quality. Detected Fourier boundaries labeled with numbers 2 and 3 result in two components C-II$_2$ in column 2 of Figs. 28(h) and 28(i). Two redundant components C-II$_5$ in column 5 of Figs. 28(h) and 28(i) are the results of detected Fourier boundaries labeled with numbers 6 and 7. Thus, components C-II$_2$ and C-II$_5$ in columns 2 and 5 of Fig. 28(h) lose some monocomponent information and so have low image quality.

This kind of redundant component occurs when the spectrum of a single component is segmented into different Fourier boundaries. This results in more than one decomposed monocomponent image, but they are of low image quality. However, the proposed method is specifically designed to detect monocomponent spectra in two directions—u- and v-axes—to support the spectra of an image for obtaining monocomponent images, thus it achieves more accurate spectrum segmentation than the conventional methods do in one dimension for each u- and v-axis.

To evaluate the effectiveness of our empirical image filter banks for ringing artifact reduction, the Fourier boundaries labeled with number 1 (see Figs. 28 and 29) from all 2D EWT segmentations are used to construct the empirical image filter banks for component decomposition. Components C-II$_1$ are selected for demonstration because of frequent occurrence of ringing artifacts.

Fig. 30 shows that the PSNR values of components C-II$_1$ provided by the proposed empirical image filter banks
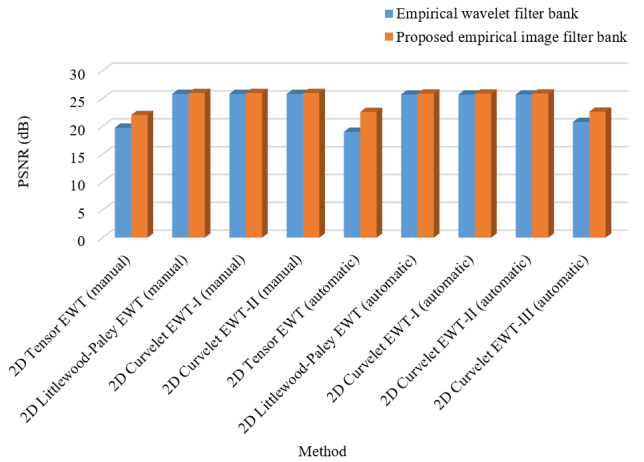

FIGURE 30. Comparison of PSNR values of empirical wavelet filter banks and proposed empirical image filter banks constructed from a variety of 2D EWT methods.

are higher than those provided by the empirical wavelet filter banks. In particular, when 2D Tensor EWT in manual and automatic modes and 2D Curvelet EWT-III in automatic mode are used, the PSNR values of components C-II$_1$ provided by the empirical image filter banks are significantly higher than those of the empirical wavelet filter banks. On the other hand, the PSNR values of components C-II$_1$ provided by the empirical image filter banks are slightly higher than those of the empirical wavelet filter banks when 2D Littlewood-Paley EWT, 2D Curvelet EWT-I, and 2D Curvelet EWT-II in manual and automatic modes are used.
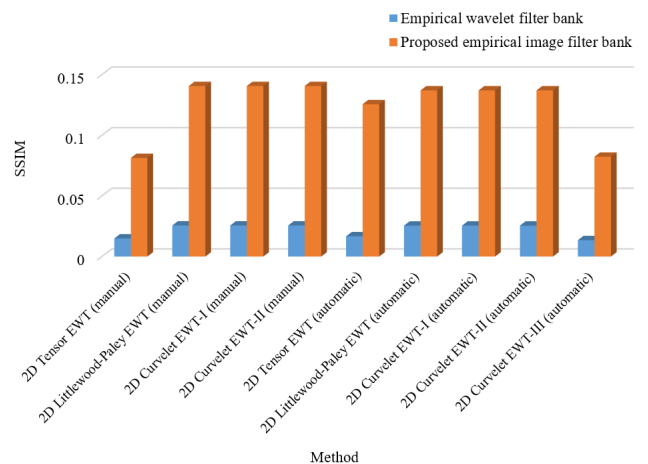
FIGURE 31. Comparison of SSIM values of empirical wavelet filter banks and proposed empirical image filter banks constructed from a variety of 2D EWT methods.

As shown in Fig. 31, SSIM values of components C-II$_1$ provided by the empirical image filter banks are much higher than those of the empirical wavelet filter banks because the structure of components C-II$_1$ decomposed by the empirical image filter banks is similar to the reference image-I.
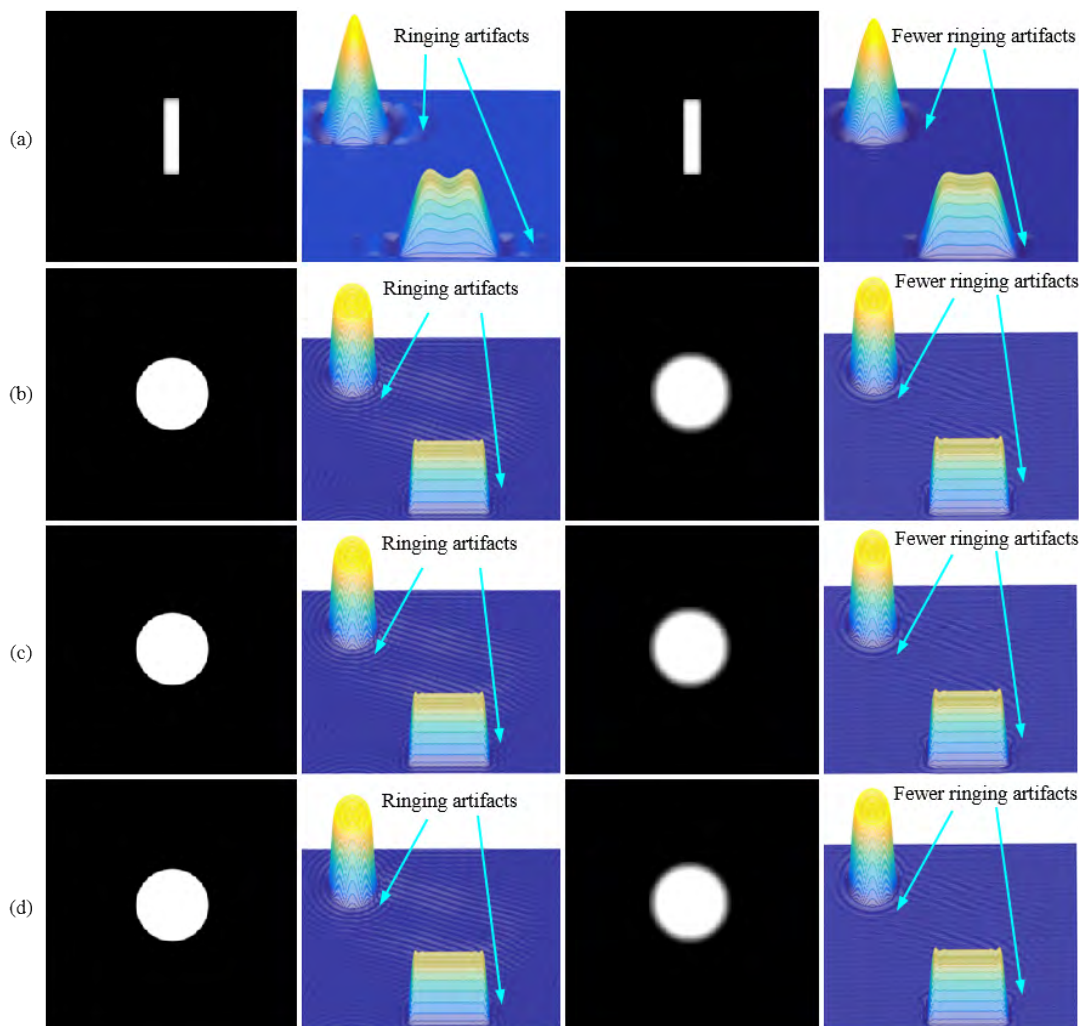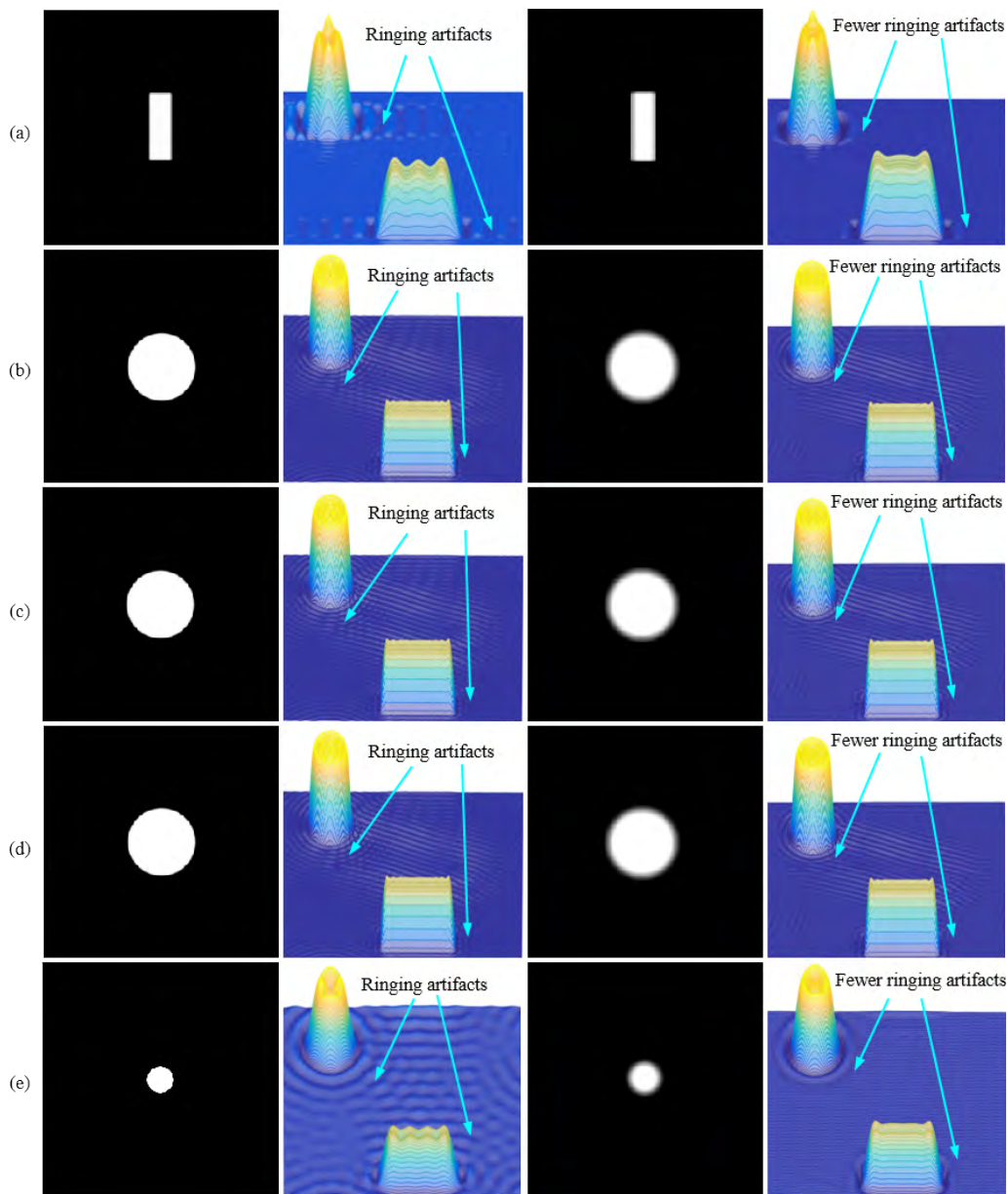
FIGURE 32. Ringing artifact reduction by proposed empirical image filter bank compared with empirical wavelet filter bank segmented by 2D EWT in manual mode. (a) 2D Tensor EWT, (b) 2D Littlewood-Paley EWT, (c) 2D Curvelet EWT-I, and (d) 2D Curvelet EWT-II. Columns from the leftmost to the rightmost display empirical wavelet filter banks, 3D view components C-I$_1$ provided by empirical wavelet filter banks, empirical image filter banks, and 3D view components C-I$_1$ provided by empirical image filter banks, respectively.

Figs. 32 and 33 show ringing artifact reduction by using the proposed empirical image filter banks. Column 1 of Figs. 32 and 33 depict the empirical wavelet filter banks provided by a variety of 2D EWT segmentations. The empirical wavelet filter banks are used to decompose components C-II$_1$ as can be seen in column 2 of Fig. 32 in a 3D view. Ringing artifacts appear in those components C-II$_1$. In particular, many ringing artifacts appear in components C-II$_1$ as shown in column 2 of Figs. 32(a), 33(a), and 33(e) which are of 2D Tensor EWT in manual and automatic modes and 2D Curvelet EWT-III in automatic mode. On the other hand, when the proposed empirical image filter banks as shown in column 3 of Figs. 32 and 33 are applied to components C-II$_1$, the ringing artifacts are greatly reduced as shown in column 4 of Figs. 32 and 33. This illustrates that the proposed empirical image filter banks are able to reduce ringing artifacts and increase image quality.

B. THAI TEXT LOCALIZATION

The second experiment aims to evaluate the performance of the proposed method when applied to text-texture segmentation for text region detection that plays an important role in text localization of natural scene images. The use of the texture-based approach for text region detection is an effective way as reported in [42]–[45]. In prior work, Ali and Hashim [42] presented a wavelet transform-based technique for detecting text edge in text regions. Kim *et al.* [43] introduced a texture-based method for text region detection using a continuously adaptive mean shift algorithm and support vector machine. Angadi and Kodabagi [44] used texture features such as homogeneity and contrast for text region detection. Sin *et al.* [45] detected text regions in natural scene images using frequency features; the edge of text strokes is extracted in the frequency domain. Here, we present a different point of view; that is, all such

FIGURE 33. Ringing artifact reduction by proposed empirical image filter bank compared with empirical wavelet transform segmented by 2D EWT in automatic mode. (a) 2D Tensor EWT, (b) 2D Littlewood-Paley EWT, (c) 2D Curvelet EWT-I, (d) 2D Curvelet EWT-II, and (e) 2D Curvelet EWT-III. Columns from the leftmost to the rightmost display empirical wavelet filter banks, 3D view components C-I$_1$ provided by empirical wavelet filter banks, empirical image filter banks, and 3D view components C-I$_1$ provided by empirical image filter banks, respectively.

text features can be represented as text-texture components when they are transformed into Fourier support. Therefore, the text-texture component can be defined by the integration of the following key features: text edge [42]–[45], text alignment [46], text stoke width [45], [46], text space [51], and text intensity [51]. These features of text usually create a unique text texture that is different from the image background. In this way, the energy concentration of the text-texture component is evident in the Fourier support, making it easy to segment the text region from the complex background.

In order to evaluate the performance of the proposed method, an experimental framework of Thai text localization is designed as schematically shown in Fig. 34. This framework has two main purposes: (i) to evaluate the effectiveness of text region detection that plays a role as a preprocessing procedure of Thai text character localization and (ii) to compare the efficiency of the conventional 2D EWT methods in automatic mode to that of our proposed EMID method. The dataset [47] used in this experiment contains 300 images with a resolution of 480×360 pixels. This dataset is categorized
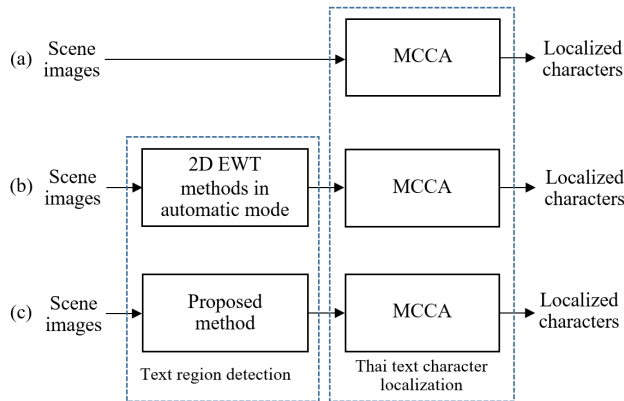
FIGURE 34. Experimental framework for Thai text localization. (a) Typical scheme of Thai text character localization, (b) use of 2D EWT methods in automatic mode for text region detection, and (c) use of the proposed method for text region detection.

into three groups based on background complexity: Group#1, Group#2, and Group#3 as explained in [47]. Group#1 consists of general natural scene images while Group#2 and Group#3 consist of various challenging scenery images and illumination effects, respectively. Furthermore, this dataset contains a collection of Thai characters for which it is always difficult to detect a complete text region, thus leading to the low precision in text localization. This difficulty mainly derives from the Thai alphabet structure that contains three levels: upper, middle, and lower as reported in [48]–[50]. Therefore, the challenges of this dataset are used to test the performance of the conventional 2D EWT and proposed methods.

For fairness of comparison between the conventional 2D EWT methods and the proposed EMID method, only the automatic mode is tested. The reason is that in this dataset it is difficult to identify an appropriate number of components with high efficiency of decomposition for the conventional methods in manual mode. The 2D EWT methods in automatic mode can automatically identify the number of components according to each natural scene image. Thus, their effectiveness can be justly compared with that of the EMID method for Thai text localization application.

1) TEXT REGION DETECTION

Text region detection is an important procedure for identifying the location of the text region prior to localizing text characters in natural scene images. The proposed EMID method is applied to segment textures of the text region from the image background. The strength of the proposed method is its ability to adaptively identify text-texture components, which are a kind of monocomponent in Fourier support at a detail region. The steps of text region detection based on the EMID method can be described as follows.

Step 1: Convert an input color image into a grayscale image (see Figs. 35(a) and 35(b)).

Step 2: Apply the EMID to such a grayscale image to identify text-texture components in Fourier support at a detail region. The result is shown in Fig. 35(c).

Step 3: Decompose text-texture components $\mathcal{M}_d$ and compensated components $\mathcal{M}_z$ by using the proposed empirical image filter banks for detail coefficient ψ_d and compensated filter banks Z, respectively, then combine them into a text-texture image $\mathcal{M}'$ as

$$\mathcal{M}' = \sum_{n=1}^{N-1} |\mathcal{M}_{d_n}| + \sum_{c=1}^{c} |\mathcal{M}_{z_c}|, \quad (21)$$

where C is the number of compensated components such that C is equal to $(N-1) \times 2$. Fig. 35(d) shows complete text-texture components in Fourier support at a detail region. Figs. 35(e) to 35(h) show text-texture component-1 to -4 in spatial domain. They are decomposed according to monocomponent spectrum boundaries labeled with numbers 1 to 4. Magnitudes (absolute values) of Figs. 35(e) to 35(h) are shown in Figs. 35(i) to 35(l), respectively. Fig. 35(m) shows a combination of Figs. 35(i) to 35(l) and Fig. 35(o) shows the text-texture image $\mathcal{M}'$.

Step 4: Apply k-means clustering, such that k is set to 2, to a text-texture image $\mathcal{M}'$ in order to separate text regions from the background. The output of this step is the text region illustrated in Fig. 35(q).

As already mentioned in Section III-C, a compensated filter bank designed to compensate for strips in text regions is essential for real world datasets. Based on preliminary experiments, a pair of compensated filter banks in a vertical orientation as shown in Fig. 35(n) is a successful case. In addition, the appropriate centroid difference between the empirical image filter bank and the compensated filter bank is 10 points. Fig. 35(q) shows detected text regions enhanced by a morphological closing operator from Fig. 35(p).

2) THAI TEXT CHARACTER LOCALIZATION

Thai text character localization is a procedure for detecting isolated Thai characters from the detected text region. In this experiment, a modified connected component analysis (MCCA) method [51] designed for text character localization based on Thai language structure is adopted. First, object edges in the detected text region are detected by using Canny edge detection [52]. Fig. 35(r) shows object edges marked in green bounding boxes. Then, Thai text characters from those object edges are localized by using the MCCA method as shown in Fig. 35(s).
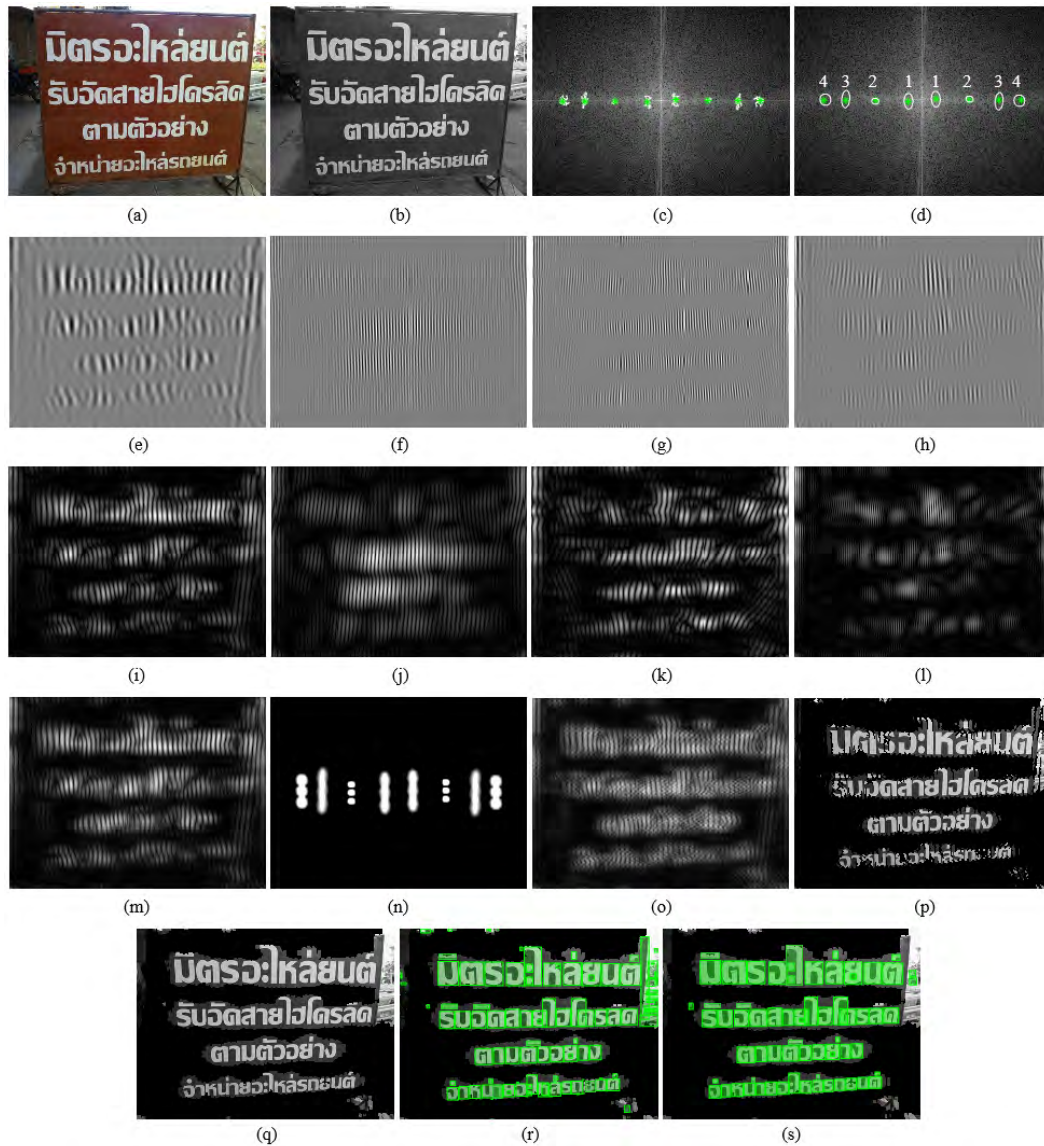
The performance validation of all test methods is typically measured in terms of precision, recall, and F-measure as defined in

$$\text{Precision} = \frac{TP}{TP + FP} \times 100, \quad (22)$$

$$\text{Recall} = \frac{TP}{TP + FN} \times 100, \quad (23)$$

$$\text{F-measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \quad (24)$$

FIGURE 35. Step-by-step demonstration of text region detection and Thai text character localization. (a) Original scene image, (b) grayscale image of Fig. 35(a), (c) identified text-texture components in Fourier support, (d) complete text-texture components derived from Fig. 35(c), (e) text-texture component-1 obtained from Fig. 35(d) labeled with number 1, (f) text-texture component-2 obtained from Fig. 35(d) labeled with number 2, (g) text-texture component-3 obtained from Fig. 35(d) labeled with number 3, (h) text-texture component-4 obtained from Fig. 35(d) labeled with number 4, (i) the magnitude of text-texture component of Fig. 35(e), (j) the magnitude of text-texture component of Fig. 35(f), (k) the magnitude of text-texture component of Fig. 35(g), (l) the magnitude of text-texture component of Fig. 35(h), (m) the text-component image, (n) an empirical image filter bank with a pair of compensated filters, (o) text-texture image, (p) text regions separated from the background by using *k*-means clustering algorithm, (q) text region detected with a morphological operation, (r) object edge detected in text region, and (s) localized Thai text characters.

where *TP*, *FP*, and *FN* are the number of correctly detected text characters, the number of detected non-text objects, and the number of incorrectly detected text characters, respectively.

A high precision value means that the algorithm can detect a large number of correct text characters and a smaller number of non-text objects. A high recall value implies a larger number of detected text characters. F-measure is also used to

measure the overall performance of text character localization in all test methods.

As can be seen in the experimental results shown in Table 3, for Group#1 dataset, the precision values of all baseline methods and the proposed EMID method are higher than that of MCCA, while the recall values of all baseline methods are much lower than that of MCCA, but that of EMID is just a little lower (88.47% compared to 89.79% of MCCA).

TABLE 3. Performance comparison of the text character localization of all methods tested with three groups of datasets.

Method (Automatic mode)	Group#1			Group#2			Group#3			Average		
	Precision (%)	Recall (%)	F-measure (%)	Precision (%)	Recall (%)	F-measure (%)	Precision (%)	Recall (%)	F-measure (%)	Precision (%)	Recall (%)	F-measure (%)
MCCA	81.49	<u>89.79</u>	85.44	63.55	78.83	70.37	58.55	<u>66.68</u>	62.35	67.86	<u>78.43</u>	72.77
2D Tensor EWT	84.25	60.72	70.57	68.93	55.16	61.28	62.39	39.67	48.50	71.85	51.85	60.23
2D Littlewood-Paley EWT	84.73	85.03	84.88	69.69	72.02	70.84	64.56	54.45	59.08	72.99	70.50	71.73
2D Curvelet EWT-I	84.76	78.04	81.27	<u>70.77</u>	67.58	69.13	64.21	48.81	55.46	73.25	64.81	68.77
2D Curvelet EWT-II	84.62	75.04	79.54	<u>70.78</u>	65.47	68.02	63.83	46.79	53.99	73.08	62.43	67.34
2D Curvelet EWT-III	84.96	77.32	80.96	69.88	66.41	68.10	66.20	47.35	55.21	73.68	63.69	68.32
Proposed method	**85.12**	88.47	**86.77**	70.12	**79.83**	**74.66**	**67.88**	64.18	**65.98**	**74.37**	77.50	**75.90**

Bold numbers indicate that the proposed method provides the best results while underlined numbers indicate that the baseline methods provide better results.
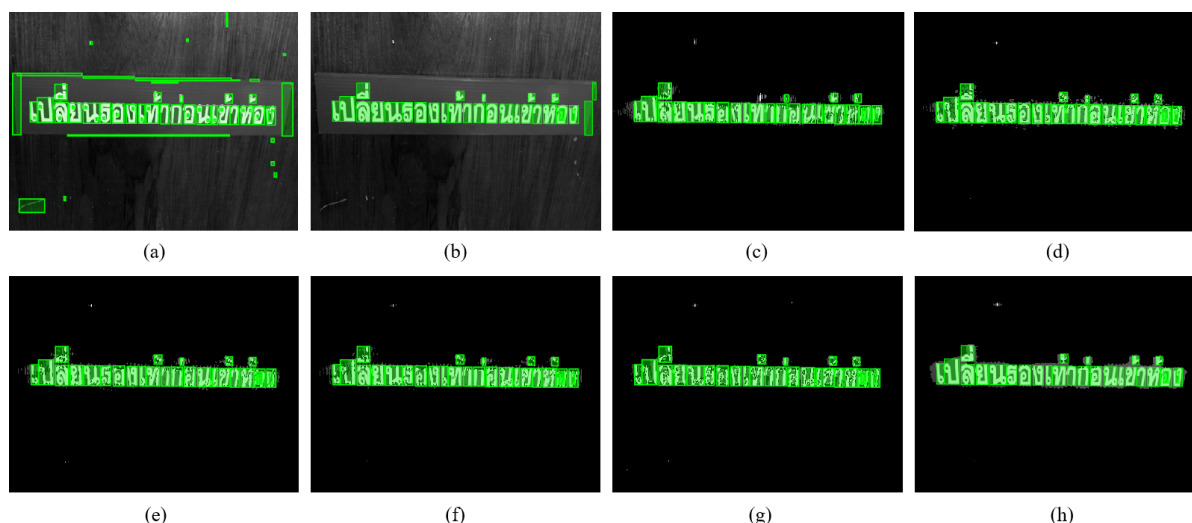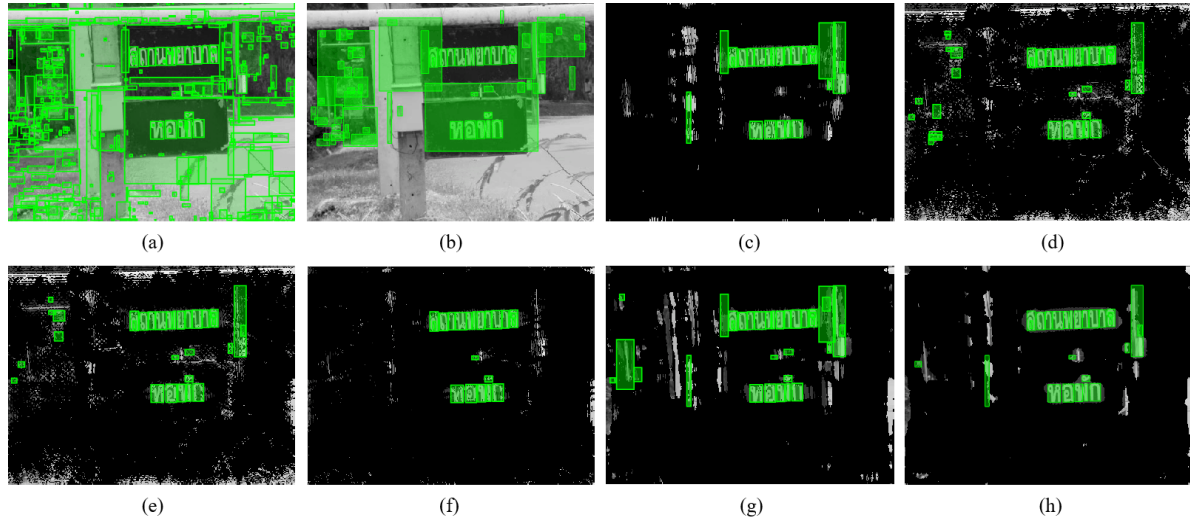

FIGURE 36. Results of Thai text localization of Group#1 dataset provided by (a) Canny edge detection, (b) MCCA, (c) 2D Tensor EWT, (d) 2D Littlewood-Paley EWT, (e) 2D Curvelet EWT-I, (f) 2D Curvelet EWT-II, (g) 2D Curvelet EWT-III, and (h) proposed EMID method.

However, regarding the F-measure, the overall performance of the proposed method is dramatically greater than those of the baseline methods and is significantly greater than that of the MCCA. For Group#2 and Group#3 datasets which consist of challenging images, the precision values of the proposed method for both datasets are dramatically greater than those of the MCCA method. When we look into the Group#2 dataset, the precision values of the proposed method are significantly better than those of the three baseline methods, 2D Tensor EWT, 2D Littlewood-Paley EWT, and 2D Curvelet EWT-III. In the same way, the precision values of the proposed method are better than those of all baseline methods for the Group#3 dataset. As for recall testing, the recall value of the proposed method is much greater than those of all of the baseline methods, but comparable to the MCCA method. For F-measure, the proposed method outperforms every test method. Overall, the average precision value of the proposed method is 74.37% which is a great improvement over that of MCCA at 67.86%, but it is only slightly better than those of all of the baseline methods. However, the average recall value of the proposed method is a bit lower than MCCA. A trade-off between precision and recall was unavoidable. As expected,

the average F-measure of the proposed EMID method is much greater than those of all of the baseline methods.

Fig. 36 demonstrates the results of locating Thai text characters in a sample of Group#1. Fig. 36(a) shows that all objects composed of text and many non-text objects are detected in the scene image. Even though MCCA can localize the text characters, some non-text objects are also detected as shown in Fig. 36(b). This affects the precision performance. On the other hand, the proposed method and the 2D EWT methods are able to efficiently detect text regions from the complex background as shown in Figs. 36(c) to 36(h). In other words, these methods can efficiently remove non-text objects that always occur from the complex background. As the non-text objects are eliminated, the precision performance of these methods is improved. Nevertheless, the 2D Tensor EWT does not detect tone marks as shown in Fig. 36(c). The loss of some text character localization is the cause of the low recall value.

The localized Thai text character results of a sample of Group #2 are shown in Fig. 37. With various scenery objects in this dataset, many non-text objects as well as text characters are detected as shown in Fig. 37(a). Although MCCA can localize the text characters shown in Fig. 37(b), many

FIGURE 37. Results of Thai text localization of Group#2 dataset provided by (a) Canny edge detection, (b) MCCA, (c) 2D Tensor EWT, (d) 2D Littlewood-Paley EWT, (e) 2D Curvelet EWT-I, (f) 2D Curvelet EWT-II, (g) 2D Curvelet EWT-III, and (h) proposed EMID method.
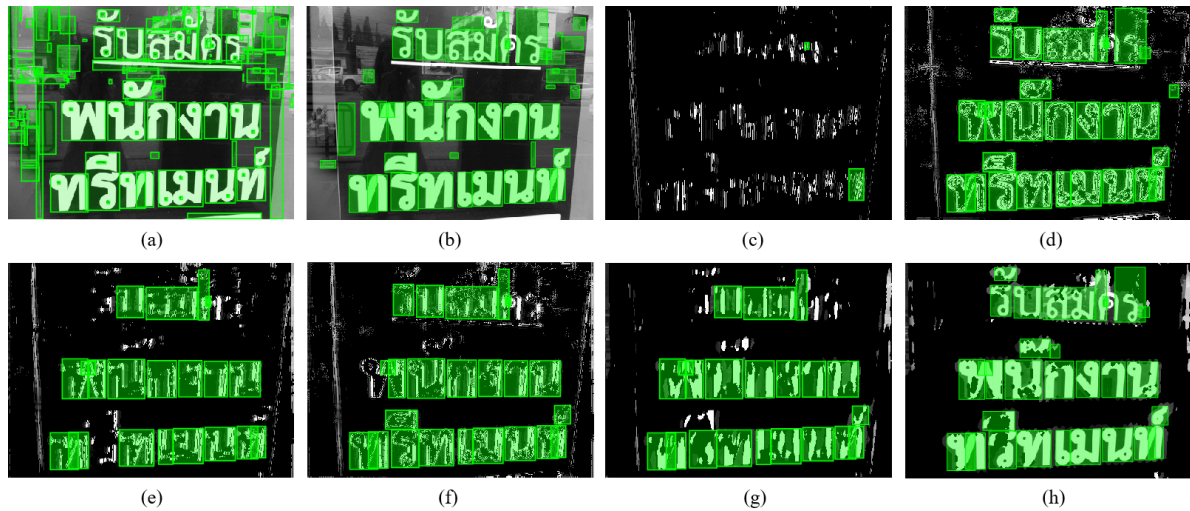
FIGURE 38. Results of Thai text localization of Group#3 dataset provided by (a) Canny edge detection, (b) MCCA, (c) 2D Tensor EWT, (d) 2D Littlewood-Paley EWT, (e) 2D Curvelet EWT-I, (f) 2D Curvelet EWT-II, (g) 2D Curvelet EWT-III, and (h) proposed EMID method.

non-text objects still appear in this scene image. This leads to low precision. On the other hand, the proposed method and 2D EWT methods are able to efficiently detect text regions and eliminate the complex background as shown in Figs. 37(c) to 37(h). In particular, the 2D Curvelet EWT-II and the proposed method are able to eliminate the complex background better than the other methods because these methods are able to detect accurate text-texture components and to identify the text region. Hence, these methods deliver high precision. On the contrary, the identified text regions decomposed by 2D Littlewood-Paley EWT included a lot of unwanted regions shown in Fig. 37(d), because there is interference between text and non-text components with this method. Thus, many non-text components are localized.

Fig. 38 demonstrates the localized Thai text character results of a sample of Group#3. The illumination effects of this dataset affect the detection of all objects as shown in Fig. 38(a). This includes Thai text characters and many non-text objects. Although MCCA can localize text characters and filter out some non-text characters, many non-text objects are still detected as shown in Fig. 38(b). However, the proposed method and 2D EWT methods seen in Figs. 38(c) to 38(h) are able to decompose text-texture components and distinguish text from non-text regions. This leads to increased precision. However, the region identified from the detected Fourier boundary of 2D Tensor EWT is incomplete, so some text characters are not detected. This is the cause of low recall of 2D Tensor EWT and thus its low overall performance. As shown in Figs. 38(e) to 38(g),

the text characters localized by 2D Curvelet EWT-I, 2D Curvelet EWT-II, 2D Curvelet EWT-III are incomplete but the text characters localized by the proposed method are more complete. Moreover, more text regions are detected by the proposed method than by 2D EWT. Therefore, the proposed method is more efficient in localizing text characters than the 2D EWT and MCCA methods.

The proposed method's precision comes from its exceptional ability to identify and decompose text-texture components from non-text regions. When we look inside the EMID method, the energy concentration-based segmentation plays a key role in detecting and identifying accurate text-texture components in Fourier support. Then, the proposed empirical image filter bank is able to decompose text-texture components from a scene image. A text-component image is obtained from those text-texture components while non-text components are eliminated. A text region is easily obtained when the k-means clustering algorithm is used to separate a text component image from the background. Moreover, the text preservation ability of the proposed method derives from efficient compensated components decomposed from a pair of compensated filter banks, thus text characters are localized better than the other test methods.

V. CONCLUSION

In this paper, an EMID method is proposed to decompose monocomponent images with less interference, no redundancy, and fewer ringing artifacts. The novel contribution of the proposed EMID method is that the use of energy concentration-based segmentation in Fourier support detects monocomponent spectrum boundaries. The proposed algorithm is designed based on an empirical concept with the following two processes: (i) energy concentration-based segmentation and (ii) empirical image filter bank construction. In the former, an empirical mean plane generated from PCA is used to detect the central frequency of candidate monocomponent spectra; the detected central frequency is then used to define the candidate monocomponent spectrum boundaries; and finally, the actual monocomponent spectrum boundaries are identified. In the latter, the actual monocomponent spectrum boundaries are used to construct an empirical image filter bank for decomposing the complete monocomponent spectra. Based on the experimental results, the proposed EMID method achieves high quality monocomponent images. This achievement mainly comes from the following reasons: (i) the energy concentration-based segmentation can provide more accurate spectrum boundary detection for extracting monocomponents, and (ii) the empirical image filter bank can reduce ringing artifacts in monocomponent images. Furthermore, when the EMID is applied to text region detection for Thai text localization in natural scene images, the performance of text character localization is significantly improved. This improvement comes from the clearer text region decomposition, thus leading to

TABLE 4. Mathematical notations and their descriptions.

Notation	Description
N	Number of monocomponent images/spectra.
O_s	Single flat object image in Fourier support.
F_s	Single frequency image in Fourier support.
O_m	Multiple flat object image.
F	2D Fourier spectrum.
F'	Detail region in Fourier support.
q_1	Upper-right quadrant of Fourier support.
q_2	Lower-right quadrant of Fourier support.
q_3	Lower-left quadrant of Fourier support.
q_4	Upper-left quadrant of Fourier support.
(u,v)	Coordinate of 2D Fourier spectrum.
$\mathbf{X}$	Spectral coefficient matrix.
h	Number of rows of matrix $\mathbf{X}$ in each quadrant.
w	Number of columns of matrix $\mathbf{X}$ in each quadrant.
$\mathbf{A}$	Square symmetric covariance matrix.
$\mathbf{\Gamma}$	Mean matrix of $\mathbf{X}$.
$\mathbf{\Lambda}$	Eigenvalue matrix.
$\mathbf{S}$	Eigenvector matrix.
s	Eigenvector.
$\mathbb{R}^{h \times w}$	Set of real $h \times w$ matrices.
d	Perpendicular distance vector.
$\mathbf{D}$	Perpendicular distance matrix of a quadrant.
d	Perpendicular distance element.
$\hat{\mathbf{D}}$	Concatenated perpendicular distance matrix from four quadrants derived from $\mathbf{D}$.
x_i	Spectral coefficient.
p	Local maximum point.
m	Index of local maximum points.
M	Number of local maximum points.
G_1, G_2	Gaussian function.
L	Smoothed 2D Fourier spectrum.
σ^2	Variance of Gaussian function or scale parameter.
l	Index of scale parameter.
r	Distance from a central frequency to its boundary.
L^τ	Boundary element of local search region.
L^ε	Neighboring element of L^τ.
R	Growth region.
y	Index of L^ε.
Y	Maximum number of L^ε.
j	Index of L^τ.
J	Maximum number of L^τ.
γ	Maximum growth rate of a growth region.
ω	Monocomponent spectrum boundary.
t	Iteration order of actual monocomponent identification algorithm.
ω_s	Single monocomponent spectrum boundary.
V	Maximum number of overlapping elements.
(u',v')	Element of the complete monocomponent spectrum boundary estimated by the ellipse function.
(u_0',v_0')	Center of the ellipse.
a	Major axis of the ellipse.
b	Minor axis of the ellipse.
θ	Orientation of the ellipse.
α	Angular parameter.
δ	Complete monocomponent spectrum boundary.
β	Complete monocomponent spectrum region.
E	Ellipse region.
ψ	Empirical image filter bank.
ψ_a	Empirical image filter banks for approximation coefficient.
ψ_d	Empirical image filter banks for detail coefficient.
$\mathcal{M}_a$	Monocomponent images decomposed from an approximation region.
$\mathcal{M}_d$	Monocomponent images decomposed from a detail region.

TABLE 4. *(Continued.)* Mathematical notations and their descriptions.

Notation	Description
Z	Set of compensated filter bank.
ζ	Compensated filter bank.
C	Number of compensated filter banks or number of compensated components.
$\mathcal{M}_z$	Compensated components.
$\mathcal{M}'$	Text-texture image.
c	Index of compensated components.
k	Number of clusters in k-means clustering algorithm.

effectively localize text characters and eliminate non-text objects. In future work, the proposed EMID method will be applied in various applications such as image denoising and object detection.

APPENDIX
See Table 4.

REFERENCES
[1] S. G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Trans. Image Process.*, vol. 9, no. 9, pp. 1532–1546, Sep. 2000.

[2] M. Nasri and H. Nezamabadi-Pour, "Image denoising in the wavelet domain using a new adaptive thresholding function," *Neurocomputing*, vol. 72, nos. 4–6, pp. 1012–1025, Jun. 2009.

[3] M. Nikpour and H. Hassanpour, "Using diffusion equations for improving performance of wavelet-based image denoising techniques," *IET Image Proces.*, vol. 4, no. 6, pp. 452–462, Dec. 2010.

[4] G. Chen, W.-P. Zhu, and W. Xie, "Wavelet-based image denoising using three scales of dependency," *IET Image Process.*, vol. 6, no. 6, pp. 756–760, Aug. 2012.

[5] A. Fathi and A. R. Naghsh-Nilchi, "Efficient image denoising method based on a new adaptive wavelet packet thresholding function," *IEEE Trans. Image Process.*, vol. 21, no. 9, pp. 3981–3990, Sep. 2012.

[6] J. Gilles, G. Tran, and S. Osher, "2D empirical transforms. wavelets, ridgelets, and curvelets revisited," *SIAM J. Imag. Sci.*, vol. 7, no. 1, pp. 157–186, Jan. 2014.

[7] S. Lahmiri, "Image denoising in bidimensional empirical mode decomposition domain: The role of student's probability distribution function," *Healthcare Technol. Lett.*, vol. 3, no. 1, pp. 67–71, Dec. 2015.

[8] M. Srivastava, C. L. Anderson, and J. H. Freed, "A new wavelet denoising method for selecting decomposition levels and noise thresholds," *IEEE Access*, vol. 4, pp. 3862–3877, Jul. 2016.

[9] R. Strickland and H. I. Hahn, "Wavelet transform methods for object detection and recovery," *IEEE Trans. Image Process.*, vol. 6, no. 5, pp. 724–735, May 1997.

[10] M. Khare, R. K. Srivastava, and A. Khare, "Single change detection-based moving object segmentation by using Daubechies complex wavelet transform," *IET Image Process.*, vol. 8, no. 6, pp. 334–344, Jun. 2014.

[11] M. Jian, K.-M. Lam, J. Dong, and L. Shen, "Visual-patch-attention-aware saliency detection," *IEEE Trans. Cybern.*, vol. 45, no. 8, pp. 1575–1586, Aug. 2015.

[12] Y. Qin, L. H. Qiao, X. Z. Ren, and Q. F. Wang, "Using bidimensional empirical mode decomposition method to identification buried objects from GPR B-scan image," in *Proc. GPR*, Hong Kong, Jun. 2016, pp. 1–5.

[13] L. Zhang and Y. Zhang, "Airport detection and aircraft recognition based on two-layer saliency model in high spatial resolution remote-sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 4, pp. 1511–1524, Apr. 2017.

[14] M. P. Shleymovich, M. V. Medvedev, and S. A. Lyasheva, "Object detection in the images in industrial process control systems based on salient points of wavelet transform analysis," in *Proc. ICIEAM*, Chelyabinsk, Russia, May 2016, pp. 1–6.

[15] S. Arivazhagan and L. Ganesan, "Texture segmentation using wavelet transform," *Pattern Recognit. Lett.*, vol. 24, no. 16, pp. 3197–3203, Dec. 2003.

[16] Y. Han and P. Shi, "An adaptive level-selecting wavelet transform for texture defect detection," *Image Vis. Comput.*, vol. 25, no. 8, pp. 1239–1248, Aug. 2007.

[17] Z.-Z. Wang and J.-H. Yong, "Texture analysis and classification with linear regression model based on wavelet transform," *IEEE Trans. Image Process.*, vol. 17, no. 8, pp. 1421–1430, Aug. 2008.

[18] L. Liang, J. Pu, and Z. Ping, "Texture segmentation using window empirical mode decomposition," in *Proc. ICAL*, Zhengzhou, China, Aug. 2012, pp. 373–377.

[19] R. Garnavi, M. Aldeen, and J. Bailey, "Computer-aided diagnosis of melanoma using border- and wavelet-based texture analysis," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 6, pp. 1239–1252, Nov. 2012.

[20] A. P. Nanthagopal and R. Sukanesh, "Wavelet statistical texture features-based segmentation and classification of brain computed tomography images," *IET Image Process.*, vol. 7, no. 1, pp. 25–32, Feb. 2013.

[21] B. Boashash, "Time-frequency and instantaneous frequency concepts," in *Time-Frequency Signal Analysis and Processing: A Comprehensive Reference*, 2nd ed. London, U.K.: Academic, 2015.

[22] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-11, no. 7, pp. 674–693, Jul. 1989.

[23] I. Daubechies, "Discrete wavelet transforms: Frames," in *Ten Lectures on Wavelets*, Philadelphia, PA, USA: SIAM, 1992, pp. 53–106.

[24] J. Pan, J. Chen, Y. Zi, Y. Li, and Z. He, "Mono-component feature extraction for mechanical fault diagnosis using modified empirical wavelet transform via data-driven adaptive Fourier spectrum segment," *Mech. Syst. Signal Process.*, vols. 72–73, pp. 160–183, May 2016.

[25] R. R. Coifman, Y. Meyer, S. Quake, and M. V. Wickerhauser, "Signal processing and compression with wavelet packets," in *Wavelets and Their Applications*, vol. 442. Dordrecht, The Netherlands: Springer, 1994, pp. 363–379.

[26] A. Linderhed, "2D empirical mode decompositions in the spirit of image compression," *Proc. SPIE*, vol. 4738, pp. 1–8, Mar. 2002.

[27] J. C. Nunes, Y. Bouaoune, E. Delechelle, O. Niang, and P. Bunel, "Image analysis by bidimensional empirical mode decomposition," *Image Vis. Comput.*, vol. 21, no. 12, pp. 1019–1026, 2003.

[28] J. Gilles, "Empirical wavelet transform," *IEEE Trans. Signal Process.*, vol. 61, no. 16, pp. 3999–4010, Aug. 2013.

[29] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," *Electron. Lett.*, vol. 44, no. 13, pp. 800–801, Jun. 2008.

[30] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[31] J. Gilles and K. Heal, "A parameterless scale-space approach to find meaningful modes in histograms—Application to image and spectrum segmentation," *Int. J. Wavelets Multiresolution Inf. Process.*, vol. 12, no. 6, p. 1450044, Nov. 2014.

[32] J. Gilles. (2017). *Empirical Wavelet Transforms*. [Online]. Available: https://www.mathworks.com/matlabcentral/fileexchange/42141-empirical-wavelet-transforms

[33] H. Lim and H. Park, "A ringing-artifact reduction method for block-DCT-based image resizing," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 7, pp. 879–889, Jul. 2011.

[34] G. S. Shin and M. G. Kang, "Ringing artifact reduction in the wavelet-based denoising," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, vol. 3. Montreal, QC, Canada, May 2004, pp. 233–236.

[35] I. T. Jolliffe, "Properties of sample principal components," in *Principal Component Analysis*, 2nd ed. New York, NY, USA: Springer-Verlag, 2002, pp. 29–42.

[36] M. Hamd and R. Rasool, "Dynamic restoration of periodic noisy image using: Upper-half spectrum," in *Proc. CIICS*, Sharjah, United Arab Emirates, Mar. 2016, pp. 1–4.

[37] J. Kim, M. Choi, and J. Jeong, "Adaptive deblocking algorithm based on image characteristics for low bit-rate video," in *Proc. ICIAR*, Montreal, QC, Canada, 2007, pp. 740–751.

[38] A. W. Fitzgibbon, M. Pilu, and R. B. Fisher, "Direct least squares fitting of ellipses," in *Proc. ICPR*, Vienna, Austria, Aug. 1996, pp. 253–257.

[39] G. P. Nason and B. W. Silverman, "The stationary wavelet transform and some statistical applications," in *Wavelets and Statistics* (Lecture Notes in Statistics), vol. 103. New York, NY, USA: Springer-Verlag, 1995, pp. 281–299.

[40] R. R. Coifman and D. L. Donoho, "Translation-invariant de-noising," in *Wavelets and Statistics* (Lecture Notes in Statistics), vol. 103. New York, NY, USA: Springer-Verlag, 1995, pp. 125–150.

[41] S. M. A. Bhuiyan, R. R. Adhami, and J. F. Khan, "A novel approach of fast and adaptive bidimensional empirical mode decomposition," in *Proc. ICASSP*, Las Vegas, CA, USA, Mar./Apr. 2008, pp. 1313–1316.

[42] S. A. Ali and A. T. Hashim, "Wavelet transform based technique for text image localization," *Karbala Int. J. Mod. Sci.*, vol. 2, no. 2, pp. 138–144, Jun. 2016.

[43] K. I. Kim, K. Jung, and J. H. Kim, "Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 1631–1639, Dec. 2003.

[44] S. A. Angadi and M. M. Kodabagi, "A texture based methodology for text region extraction from low resolution natural scene images," *Int. J. Image Process.*, vol. 3, no. 5, pp. 229–245, Nov. 2009.

[45] B.-K. Sin, S.-K. Kim, and B.-J. Cho, "Locating characters in scene images using frequency features," in *Proc. ICPR*, Quebec City, QC, Canada, Aug. 2002, pp. 489–492.

[46] C. Yi and Y. Tian, "Text string detection from natural scenes by structure-based partition and grouping," *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 2594–2605, Sep. 2011.

[47] K. Woraratpanya *et al.*, "Text-background decomposition for Thai text localization and recognition in natural scenes," in *Proc. ICITEE*, Yogyakarta, Indonesia, Oct. 2014, pp. 138–143.

[48] U. Suttapakti, K. Woraratpanya, and K. Pasupa, "Font descriptor construction for printed Thai character recognition," in *Proc. MVA*, Kyoto, Japan, 2013, pp. 45–48.

[49] T. Kobchaisawat and T. H. Chalidabhongse, "Thai text localization in natural scene images using Convolutional Neural Network," in *Proc. APSIPA*, Siem Reap, Cambodia, Dec. 2014, pp. 1–7.

[50] C. Yingsaeree and A. Kawtrakul, "Rule-based middle-level character detection for simplifying Thai document layout analysis," in *Proc. ICDAR*, Seoul, South Korea, Aug./Sep. 2005, pp. 885–892.

[51] K. Woraratpanya, P. Boonchukusol, Y. Kuroki, and Y. Kato, "Improved Thai text detection from natural scenes," in *Proc. ICITEE*, Yogyakarta, Indonesia, Oct. 2013, pp. 137–142.

[52] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, no. 6, pp. 679–698, Nov. 1986.

UNGSUMALEE SUTTAPAKTI received the B.S. degree from the Suranaree University of Technology, Nakhon Ratchasima, Thailand, in 2006, and the M.S. degree in information technology from the King Mongkut's University of Technology North Bangkok, Bangkok, Thailand, in 2010. She is currently pursuing the Ph.D. degree in information technology with the King Mongkut's Institute of Technology Ladkrabang, Thailand. Her research interests include signal analysis, image processing, computer vision, pattern recognition, and machine learning.

KITSUCHART PASUPA (M'12–SM'16) received the B.Eng. degree in electrical engineering from the Sirindhorn International Institute of Technology, Thammasat University, Thailand in 2003, and the M.Sc. (Eng.) and Ph.D. degrees in automatic control and systems engineering from the Department of Automatic Control and Systems Engineering, The University of Sheffield, in 2004 and 2008, respectively. He was a Research Fellow with the University of Southampton and The University of Sheffield. He is currently an Assistant Professor with the Faculty of Information Technology, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand. His main research interests lie in the application of machine learning techniques in the real world application.

KUNTPONG WORARATPANYA received the B.Ind.Tech. degree in computer technology, the M.Eng. degree in computer engineering, and the D.Eng. degree in electrical engineering from the King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand, in 1992, 1996, and 2005, respectively. He is currently an Assistant Professor with the Faculty of Information Technology, King Mongkut's Institute of Technology Ladkrabang. His research interests include stereoscopic acquisition and compression, multi-media coding and processing, signal processing, speech recognition and processing, pattern recognition and image processing, computer vision, and machine learning.

● ● ●