




Research Article

A Robust Image Segmentation Framework Based on Nonlocal Total Variation Spectral Transform

Jianwei Zhang ¹, Yue Shen ¹, Zhaohui Zheng ², and Le Sun ³

¹School of Mathematics and Statistics, Nanjing University of Information Science and Technology, Nanjing 210044, China

²Department of Clinical Immunology, Xijing Hospital, Fourth Military Medical University, No. 127 Changle West Rd., Xi'an 710032, China

³Jiangsu Engineering Center of Network Monitoring, School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing 210044, China

Correspondence should be addressed to Zhaohui Zheng; zhengzh@fmmu.edu.cn

Received 14 September 2021; Revised 27 November 2021; Accepted 22 January 2022; Published 24 February 2022

Academic Editor: Zheng Chu

Copyright © 2022 Jianwei Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Image segmentation plays an important role in various computer vision tasks. Nevertheless, noise always inevitably appears in images and brings a big challenge to image segmentation. To handle the problem, we study the nonlocal total variation (NLTV) spectral theory and build up an image segmentation framework with NLTV spectral transform to segment images with noise. Firstly, we decompose an image into the NLTV flow in the NLTV spectral transform, with which the max response time of each pixel is calculated. Secondly, a separation surface is constructed with the max response time to distinguish the objects and preserve the structure details in segmentation. Thirdly, the image is filtered by the surface in the NLTV spectral domain, and a rough segmentation result is obtained by means of an inverse transform. Finally, we use a binary process and morphological operations to refine the segmentation result. Experiments illustrate that our method can preserve edge structures effectively and has the ability to achieve competitive segmentation performance compared with the state-of-the-art approaches.

1. Introduction

Image segmentation refers to partitioning images into multiple homogeneous parts or objects. It plays a significant role in a broad range of computer vision applications, including scene understanding [1], image compression [2], and image retrieval [3, 4]. To date, two categories of segmentation methods have been widely proposed: data-driven methods [5–7] and model-driven methods [8–28].

Among data-driven methods, the common strategy is to extract the semantic features of images using deep convolutional neural networks, based on which each pixel can obtain a semantic label to realize segmentation. The popular deep neural networks for semantic segmentation consist of FCN [5], U-Net [6], SegNet [7], etc., which can obtain satisfying segmentation results without any postprocess techniques. However, deep neural networks often suffer from high computational resource consumption and need a great mass of labeled data. Moreover, the interpretability of

neural networks is always an Achilles' heel. Therefore, model-driven methods are our research centrality.

According to different segmentation strategies, model-driven methods can be further categorized as boundary-based methods, region-based methods, hybrid methods, and transform-based methods. Boundary-based methods separate objects from the background by edge or shape. The representative methods include edge detection [8–10] and graph-cut methods [11, 12]. The former uses intensity discontinuity to segment an object. Common edge detection operators contain Prewitt [8], Sobel [9], Roberts [9], and Canny [10]. Compared with the edge detection approaches, graph-cut-based methods can achieve better segmentation accuracy. Nonetheless, the extraction of gradients is sensitive to noise, which makes the boundary-based models produce unsatisfying segmentation results for noisy images.

Region-based approaches recognize similar regions and complete segmentation by means of statistical techniques. The Chan-Vese model [13] and FCM [14] are representative

works. The Chan–Vese model makes the contour curve close to the object boundary by minimizing the energy on both sides of the evolution curve [15]. Nevertheless, the Chan–Vese model fails to obtain satisfying results because of the intensity inhomogeneity. FCM improves the tolerance to ambiguity and obtains more reasonable segmentation results by introducing a membership matrix. However, FCM is unrobust to noise because of the fact that it merely considers gray-level information. To solve the problem, many variants of FCM [16–19] have been developed, which bring good segmentation performance. Nonetheless, the improved methods are still sensitive to the complex background and intensity inhomogeneity.

Hybrid methods employ boundary information to detect the region of objects and then use region information to preserve the boundary structures. Recently, transition region (TR)-based image thresholding [20–23] has been proposed as a type of hybrid method. The method, firstly, uses edge detectors or statistical techniques to extract a transition region, which is a structure similar to the image edge, and then, it segments the image by a threshold, which is a gray level mean value of the transition region. TR-based image thresholding additionally exploits the spatial information to acquire more satisfying segmentation results. However, it is a global thresholding method, which is unrobust to intensity inhomogeneity.

The aforementioned model-driven methods segment the image using spatial features, which results in sensitivity to noise. Differently, transform-based approaches, firstly, transform the image to a specific domain according to mathematical theories, where noise and image details have different performances. Then, denoised images are obtained by filtering and inverse transformation, on which post-processing is performed to segment the image. As one of the popular transform approaches, wavelet transform is widely used in diverse computer vision tasks because of its ease of use and multiresolution processing ability. The common operation of the wavelet transform in image processing is to decompose the image to obtain multiscale sub-bands in the wavelet domain with the help of Mallat’s pyramid algorithm [24]. Then, filter the image by low-pass, band-pass, or high-pass filter to obtain the required features. Finally, the processed image can be obtained by inverse transform. To get satisfying segmentation results, wavelet transform is often combined with other segmentation methods, such as watershed segmentation [25], clustering approaches [26], and image thresholding methods [27]. For instance, the method in [25], firstly, decomposes the original image into a multiscale pyramid representation in the wavelet transform domain. Secondly, the watershed algorithm is applied to segment every image of the multiscale pyramid into several regions, including objects and background. Thirdly, the reverse wavelet transform is conducted on the split regions to get the next higher resolution representation. Finally, the size of split regions gradually becomes the same as that of regions in the ground truth to achieve the segmentation result. Nonetheless, wavelet transform-based methods are sensitive to contrast, and the segmentation results are influenced by the selection of wavelet basis functions.

Recently, the NLTV spectral theory has been introduced [28] and has attracted people’s attention. The NLTV spectral transform can transform the image from the spatial domain to the spectral domain, in which objects with different contrast, size, and detailed structures can be distinguished well. Additionally, the NLTV spectral transform can preserve image structures because of its nonlocal operators [28]. To this end, we further discuss the performance of NLTV spectral theory and attempt to further enhance the applicability of the NLTV spectral transform. Inspired by the work [29], we demonstrate the sensitivity of the NLTV spectral transform to size, contrast, and its detailed structures in images with or without noise. We also indicate that the spectral transform is invariance to rotation and translation. Besides, we are motivated to put forward a robust image segmentation framework with NLTV spectral transform. The main process is as follows: firstly, the NLTV flow is imposed on an image to acquire the NLTV spectral transform, by which spectral response and a salient time map of the image are calculated. The elements in the salient time map represent the max response time of each pixel of the image. Secondly, we filter the salient time map by a Gaussian filter to remove the isolated points and perform a least-squares regression using a polynomial on the filtered map to fit a separation surface. Thirdly, the image is filtered by the surface in the NLTV spectral domain, followed by the NLTV inverse transform to obtain a rough segmentation result. Finally, we use morphological operators and a binary process to refine the segmentation result.

It should be noticed that the total variation (TV) spectral transform-based method [30] has a similar idea in segmenting images with noise. However, the TV spectral transform used in [30] calculates the horizontal and vertical gradient of every pixel, which means only local information is selected to describe object features. In reference [30], the TV flow is obtained by iteratively solving the ROF model, and then the TV spectral transform is yielded. Considering that the edge detail of objects is lost for solving the ROF model, the guided filter is adopted to refine the object edge in [30]. In contrast to the spectral transform strategy in [30], our method pays more attention to the difference between one pixel and all other pixels in the image, termed nonlocal gradients, to achieve NLTV spectral transform. With the nonlocal information, the edge details can be effectively preserved when segmenting the object in a variety of noises. In addition, our segmentation framework does not introduce the guided filter, which may bring the noise from the original image to the segmentation result. We perform the experiments on synthetic, natural, and medical cell images, which demonstrate that the proposed method can achieve competitive segmentation performance compared with the state-of-the-art methods.

Overall, the contributions of this work are twofold, which are as follows:

- (i) We illustrate the properties of NLTV spectral transform by theoretical proof and experiments. The analysis demonstrates that objects with varying size, contrast, and detailed structures can be

distinguished in the NLTV spectral domain. Additionally, the transform is invariant to rotation and translation. These properties indicate the feasibility of segmentation based on NLTV spectral transform.

- (ii) We propose an image segmentation framework using NLTV spectral transform, which fits a separation surface to filter sub-bands in the NLTV spectral domain, and it obtains segmentation results by means of postprocessing. Our method can achieve satisfying results for images with diverse noise or complex texture.

The rest of the article is structured as follows: section 2 gives an overview of the NLTV spectral theory. Section 3 discusses the properties of NLTV transform and introduces our segmentation framework based on NLTV spectral transform. Section 4 illustrates the experimental results of the proposed method. At last, the paper is concluded in section 5.

2. Preliminaries

This section introduces the NLTV spectral transform framework [28]. The framework is made of several parts: nonlocal operators, NLTV flow, NLTV spectral transform, and spectral response.

2.1. Nonlocal Operators. According to continuous definitions on the graphs of nonlocal gradient and divergence [31], three nonlocal operators, namely nonlocal derivatives, nonlocal gradients, and nonlocal divergences, are defined as follows:

Let $\Omega \subset R^2$ be a bounded domain and $w(X, Y) \geq 0$ be non-negative weights between any two points, $X, Y \in \Omega$. In the view of graphs, these weights correspond to a certain relationship between these points. For simplicity, we assume that these weights are symmetric, which means $w(X, Y) = w(Y, X)$. Then, Gilboa and Osher [28] extended the local derivative to a nonlocal version by the following definition:

$$\partial_Y u(X) = (u(Y) - u(X))\sqrt{w(X, Y)}, \quad X, Y \in \Omega, \quad (1)$$

where $u(X)$ is a real function, $u: \Omega \rightarrow R, 0 < w(X, Y) < \infty$, and $\partial_Y u(X)$ represents the partial derivatives of $u(X)$ in the direction of point X and Y .

Similar to local gradients derived from local partial derivatives, nonlocal gradient $\nabla_w u(X): \Omega \rightarrow \Omega \times \Omega$ is defined as the vector composed of all partial derivatives.

$$(\nabla_w u)(X, Y) = (u(Y) - u(X))\sqrt{w(X, Y)}, \quad X, Y \in \Omega. \quad (2)$$

Before introducing nonlocal divergence, the definition of inner product for vectors is shown as below. Denoting vectors as $\vec{v}_1 = v_1(X, Y), \vec{v}_2 = v_2(X, Y) \in \Omega \times \Omega$, an inner product is defined as follows:

$$\langle \vec{v}_1, \vec{v}_2 \rangle = \int_{\Omega} v_1(X, Y)v_2(X, Y)dY. \quad (3)$$

Then nonlocal divergence $(\text{div}_w \vec{v})(X): \Omega \times \Omega \rightarrow \Omega$ is defined as the adjoint of nonlocal gradient. $(\text{div}_w \vec{v})(X) = \int_{\Omega} (v(X, Y) - v(Y, X))\sqrt{w(X, Y)}dY$.

2.2. NLTV Flow. The weight matrix \mathbf{W} depends on the patch similarity. For fixed point X and arbitrary point Y in the image, $\mathbf{W}(X, Y)$ represents the weight between the points X and Y , which is defined as follows:

$$\begin{cases} \mathbf{W}(X, Y) = \frac{E(X, Y)}{\sum_{Y \in \Omega} E(X, Y)}, \\ E(X, Y) = \exp\left(-\frac{\|P(X) - P(Y)\|_2^2}{\sigma^2}\right), \end{cases} \quad (4)$$

where $P(X)$ and $P(Y)$ represent the patches centered at points X and Y in the image, respectively. σ is a parameter to control the decay of the exponential function. $E(X, Y)$ describes the similarity between the points X and Y .

NLTV is divided into two types, including isotropic NLTV and anisotropic NLTV. The former is defined as follows:

$$J_{\text{ISO-NLTV}}(u) = \int_{\Omega} \left(\int_{\Omega} (u(X) - u(Y))^2 w(X, Y)dY \right)^{1/2} dX. \quad (5)$$

The latter is defined as follows:

$$J_{\text{ANISO-NLTV}}(u) = \int_{\Omega \times \Omega} |u(X) - u(Y)|\sqrt{w(X, Y)}dYdX. \quad (6)$$

In our work, the anisotropic nonlocal TV is applied to calculate NLTV flow.

$$\begin{cases} \frac{\partial u}{\partial t} \in \partial_u J_{\text{NLTV}}(u), \\ u(0, X) = u(X). \end{cases} \quad (7)$$

2.3. NLTV Transform. The sine and cosine functions are the basic functions in Fourier transform. These basic functions' amplitude forms impulses in the Fourier domain. The work [28] generalized this to NLTV domain. By examining the elementary structures disks for NLTV functional, the second derivative in the time of NLTV flow is considered the representation of the impulse of the elementary structure. Hence, the NLTV transform is defined by the following:

$$\phi(t) = u_{tt}t, \quad (8)$$

where $t \in (0, \infty)$ is a time parameter of the NLTV flow equation (7), and u_{tt} is the second derivative in the time of the NLTV flow.

For NLTV transform, the inverse transform reconstructs a signal or image from all $\phi(t)$ elements.

$$I(X) = \int_0^{\infty} \phi(t, X) dt + \bar{u}, \quad (9)$$

where $\bar{u} = (1/\Omega) \int_{\Omega} u(X) dX$ is the residual part of NLTV transform, and it is also the mean value of the initial condition.

2.4. NLTV Spectral Response. Corresponding to the amplitude of the response in Fourier domain, the NLTV spectral response is defined as follows:

$$S(t) = \int_{\Omega} |\phi(t, X)| dX, \quad t \in (0, \infty). \quad (10)$$

The NLTV spectral response can roughly measure the importance of image information at different time scales in the NLTV spectral domain [28]. The main features of the image emerge at the time scale corresponding to the high response. Otherwise, the NLTV spectral transform could be considered negligible.

3. Proposed Method

This section discusses the properties of the NLTV spectral transform and displays a segmentation method for images with noise using the NLTV spectral transform. Firstly, the seminal works [29, 30], which demonstrate the properties of TV spectral transform in images with or without noise, are extended to the NLTV spectral transform in motivation. Secondly, a segmentation method using NLTV spectral transform for images with noise is introduced.

3.1. Motivation. The section tries to research the properties of NLTV spectrum transform in images with or without noise. Theories and experiments without noise are shown, firstly. Then, the properties are extended to the noise condition by experiments. As known to all, the typical noises in digital images are additive noise, multiplicative noise, and impulse noise. For this reason, we corrupt the images with Gaussian noise, Salt & Pepper noise, and Speckle noise.

3.1.1. Property 1: Sensitivity to Size. A short proof about the property is provided. For the sake of simplicity, we consider scaling with a gray level image $f(X)$, where $X = (x, y) \in \Omega$. Then, the image after scaling can be denoted as $f(aX)$. With the above notations, we explore why NLTV spectral transform values over the time scale of images before and after scaling satisfy the following relationship:

$$\tilde{\phi}(t, X) = a\phi(at, aX), \quad (11)$$

where $\tilde{\phi}(t, X)$ and $\phi(t, X)$ are NLTV spectral transforms corresponding to images before and after scaling, respectively. Notice that for the original image $f(X)$, the NLTV flow can be derived from the following partial differential equation:

$$\begin{cases} -\frac{\partial u}{\partial t} \in \partial_u J_{\text{NL-TV}}(u), \\ u(0, X) = f(X). \end{cases} \quad (12)$$

Inspired by the case of TV, we consider the elementary structures called nonlocal disks for the image $f(X)$. A set \mathbf{A} can be used as a nonlocal disk when two conditions are satisfied [28]: 1) \mathbf{A} is a nonlocal calibrable set. 2) The curvature is constant on the internal boundary of the set \mathbf{A} .

The characteristic function of \mathbf{A} is $\chi_{\mathbf{A}}(X) = \begin{cases} 1, & X \in \mathbf{A} \\ 0, & X \notin \mathbf{A} \end{cases}$. Then, the explicit solution of problem (12) with $u(0, X) = \chi_{\mathbf{A}}(X)$ is expressed as follows:

$$u(t, X) = \begin{cases} (1 - t\lambda_{\mathbf{A}})\chi_{\mathbf{A}}(X), & X \in \mathbf{A}, \\ 0, & X \notin \mathbf{A}, \end{cases} \quad (13)$$

where $\lambda_{\mathbf{A}} = (\text{Per}(\mathbf{A})/|\mathbf{A}|)$ and $\text{Per}(\mathbf{A})$ and $|\mathbf{A}|$ are, respectively, perimeter and normal of \mathbf{A} . In the same way, the NLTV flow of nonlocal disk \mathbf{A}' for the image $f(aX)$ is as follows:

$$\tilde{u}(\tilde{t}, \tilde{X}) = \begin{cases} (1 - \tilde{t}\lambda_{\mathbf{A}'})\chi_{\mathbf{A}'}(\tilde{X}), & \tilde{X} \in \mathbf{A}', \\ 0, & \tilde{X} \notin \mathbf{A}'. \end{cases} \quad (14)$$

The energy of points in the image $f(X)$ and $f(aX)$ decreases with the average speed of $\lambda_{\mathbf{A}}$ and $\lambda_{\mathbf{A}'}$, respectively. It is worth noting that $\lambda_{\mathbf{A}}$ is equal to $\lambda_{\mathbf{A}'}$ because the object patterns before and after scaling are similar. Hence, we have $\tilde{u}(\tilde{t}, \tilde{X}) = u(t, X)$ with $\tilde{t} = t/a$ and $\tilde{X} = X/a$. For time scaling is $\tilde{t} = t/a$, $\tilde{u}_{\tilde{t}\tilde{X}}(\tilde{t}, \tilde{X}) = a^2 u_{tt}(t, X)$ is obvious. Therefore, $\tilde{\phi}(t/a, X/a) = t \tilde{u}_{\tilde{t}\tilde{X}}(\tilde{t}, \tilde{X})/a = t \cdot a^2 u_{tt}(t, X)/a = a\phi(t, X)$.

Figure 1 is an example showing how the NLTV spectral transform separates different size objects. The multiscale NLTV spectral descriptions of the pixels are shown in Figure 1(b), which shows that there is a positive correlation between the size and the time to reach the max spectral response. In addition, we can find that the disappearance order of objects in Figure 1(c) is consistent with the order of reaching max spectral response time in Figure 1(b). Figure 1(d) shows the visualization of subbands in the NLTV spectral domain, and it is a more intuitive interpretation of figure 1(b). Moreover, Figure 2 shows the sensitivity of NLTV spectral transform to size and similar performance in different noises.

3.1.2. Property 2: Sensitivity to Local Contrast. Combing the work [29], we attempt to provide a short proof. The image after gray-scale transformation by factor a is denoted as $af(X)$. Then, we plan to prove that the NLTV spectral signatures of $f(X)$ and $af(X)$ satisfy the following relationship:

$$\tilde{\phi}(at, X) = \phi(t, X). \quad (15)$$

It is noting that $\phi(t, X)$ is still related with characteristic function $\chi_{\mathbf{A}}(X)$ mentioned in property 1. Copying the analysis of property 1, the NLTV flows of $f(X)$ and $f(aX)$ are as follows:

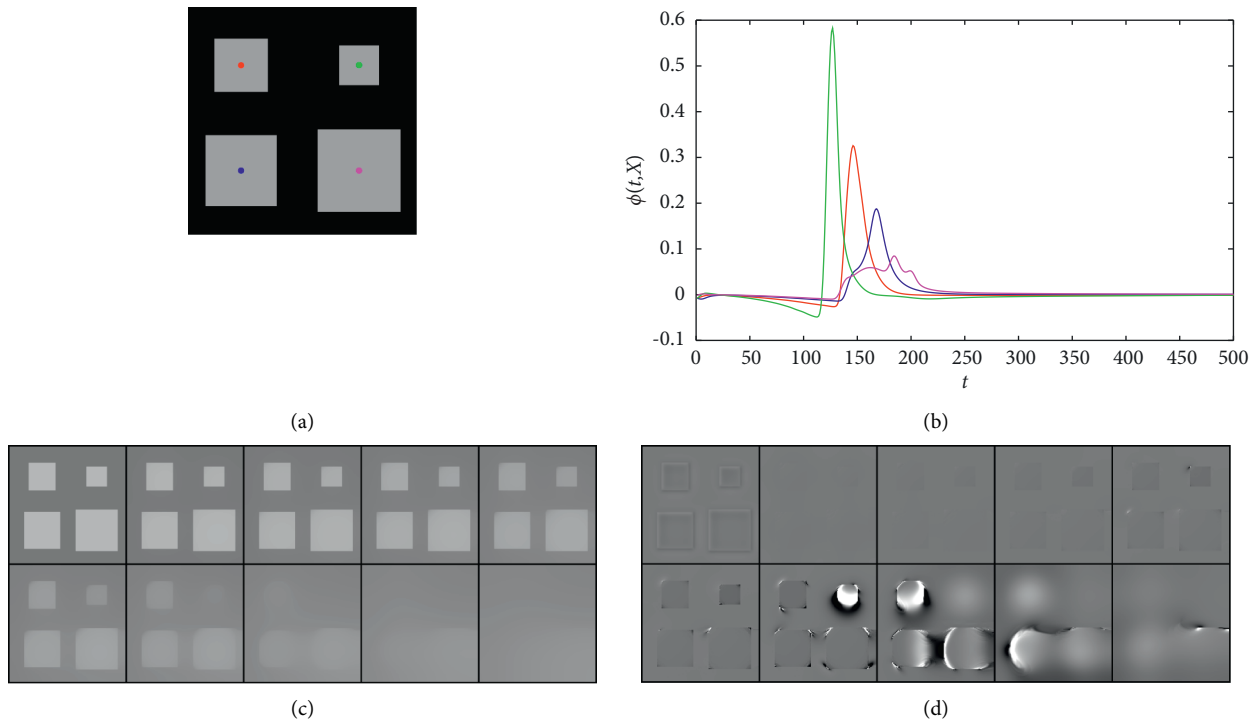


FIGURE 1: Demonstration of property 1. Signatures are distinguished because of their sensitivity to size. (a) Image f . (b) Multiscale NLTV spectral descriptions of different pixels. (c) Results of NLTV flow of f . (d) Multiscale NLTV spectral components.

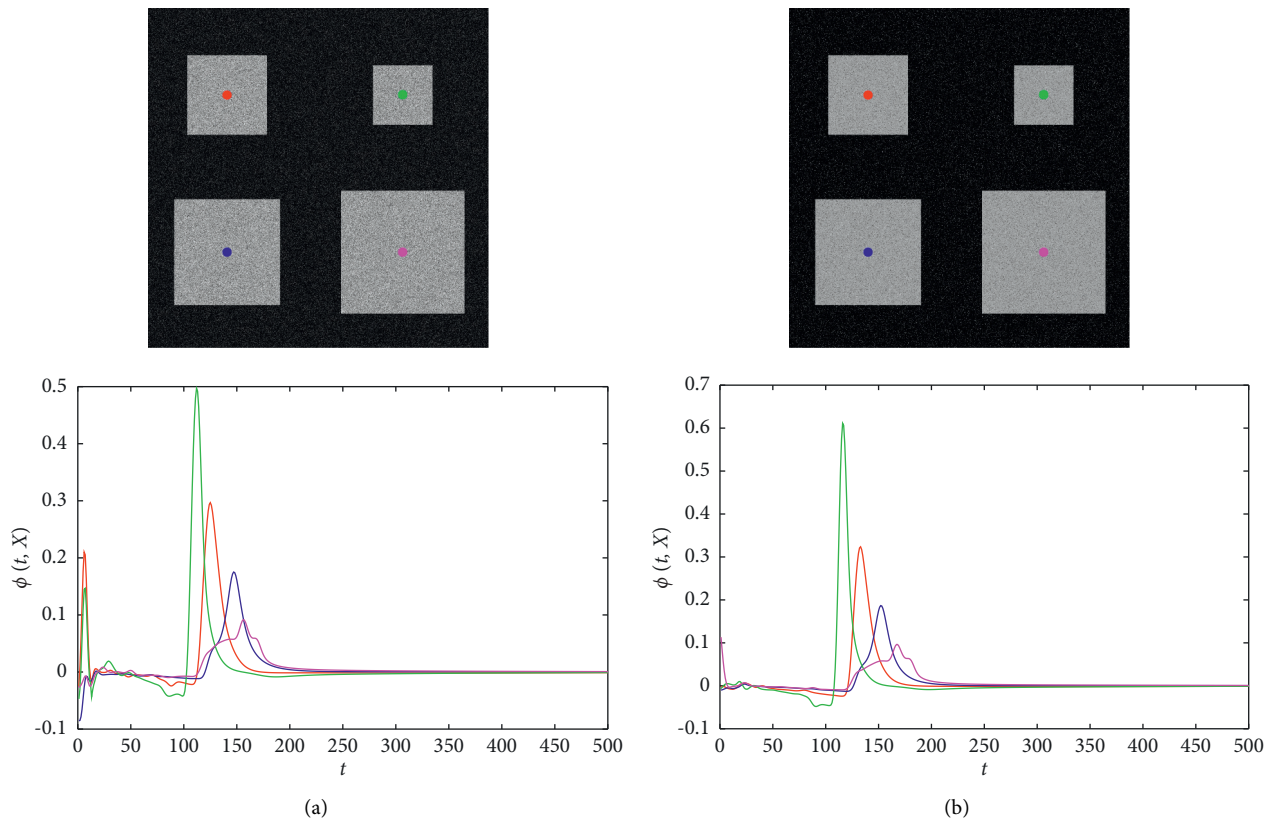


FIGURE 2: Continued.

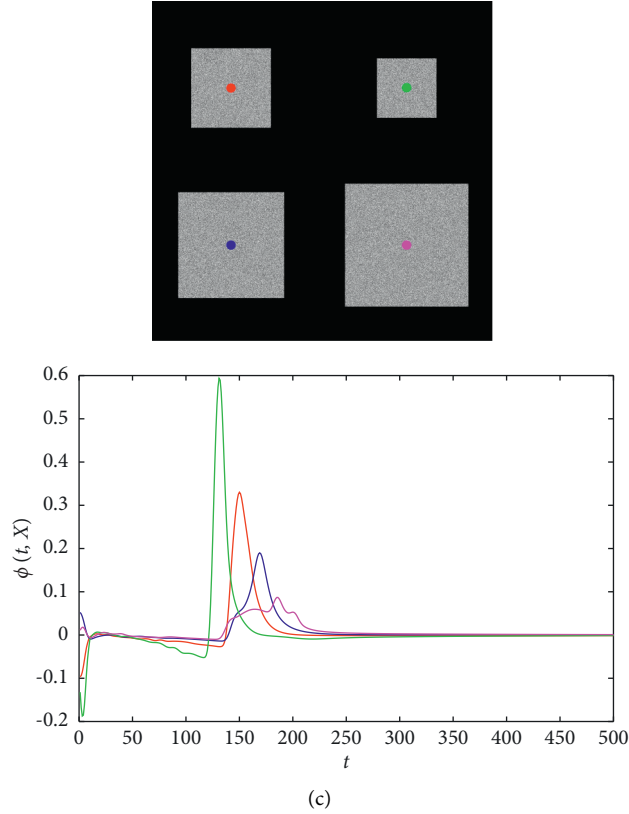


FIGURE 2: NLTV spectral transform on different size objects corrupted with different noises. (a) is the image corrupted with Gaussian (10% variance), Salt & Pepper (10% density), and Speckle noise (10% variance), respectively. (b) is the multiscale NLTV spectral descriptions of different pixels corrupted with noises. (a) Gaussian noise. (b) Salt & Pepper noise. (c) Speckle noise.

$$\begin{aligned}
 u(t, X) &= \begin{cases} (1 - t\lambda_A)\chi_A(X), & X \in \mathbf{A}, \\ 0, & X \notin \mathbf{A}, \end{cases} \\
 \tilde{u}(\tilde{t}, \tilde{X}) &= \begin{cases} a(1 - \tilde{t}\lambda_{A'})\chi_{A'}(\tilde{X}), & \tilde{X} \in \mathbf{A}', \\ 0, & \tilde{X} \notin \mathbf{A}', \end{cases}
 \end{aligned} \quad (16)$$

where $\tilde{t} = at$, $\mathbf{A} = \mathbf{A}'$, and $X = \tilde{X}$. $\tilde{u}(\tilde{t}, \tilde{X}) = au(t, X)$ and $\tilde{u}_{\tilde{t}\tilde{t}}(a\tilde{t}, \tilde{X}) = u_{tt}(t, X)/a$ are given. Therefore, $\tilde{\phi}(at, X) = at\tilde{u}_{\tilde{t}\tilde{t}}(\tilde{t}, \tilde{X}) = at \cdot u_{tt}(t, X)/a = \phi(t, X)$.

An example is demonstrated on a synthetic image without noise, as shown in Figure 3. The image exhibited in figure 3(a) contains four different contrast squares with a black background. The NLTV spectral transform is calculated, and multiscale NLTV spectral descriptions of different pixels are shown in Figure 3(b). Figures 3(c) and 3(d) show more intuitive performance, which indicates that the low contrast squares disappear first. In addition, the NLTV spectral transform is implemented on different noises to verify its performance. As shown in Figure 4, except for small time scales, the NLTV spectral description has a similar performance, which demonstrates the sensitivity of the NLTV spectral transform to contrast images with noise.

3.1.3. Property 3: Sensitivity to Detailed Structures. Figure 5 shows objects with diverse structures. Figure 5(b) shows that different objects have different time scales when reaching the max spectral response. Figures 5(c) and 5(d) show

an intuitive description. The center square with high contrast is decomposed, firstly. Then, the square ring to which the blue point belongs starts to be decomposed. The black square ring is decomposed finally. The experiment indicates the sensitivity of the NLTV spectral transform to detailed structures. The phenomena are caused by the nonlinear property of the NLTV spectral transform. Assuming that images f and g make up the image h , the response of these images satisfies the following:

$$\phi_h \neq \phi_f + \phi_g. \quad (17)$$

To observe the decomposition process of NLTV spectral transform within noise, examples are carried out on different noises. Figure 6 shows the decomposition results of different pixels in diverse noises. It can be seen that, except for small time scales, the NLTV spectral description is similar to the case shown in figure 5(b). The experiments demonstrate that the NLTV spectral transform has a sensitivity to detailed structures.

3.1.4. Property 4: Invariance to Rotation and Translation. Suppose the original image is denoted as $f(X)$, $X \in \Omega$. Then, the image after rotation by angle θ about the origin is $f(\mathbf{R}X)$, where \mathbf{R} is the rotation matrix.

$$\begin{aligned}
 \mathbf{R} &= \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}, \\
 f(X) &= f(\mathbf{R}X).
 \end{aligned} \quad (18)$$

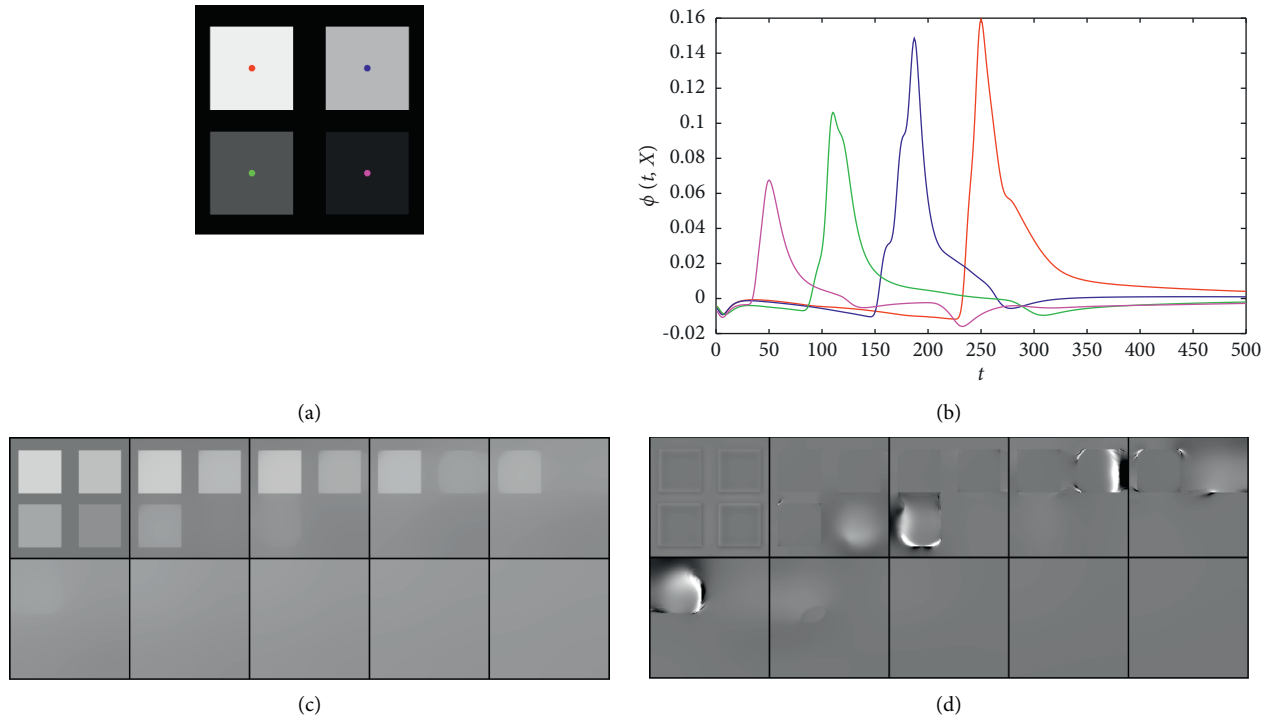


FIGURE 3: Demonstration of property 2. Signatures are distinguished because of their sensitivity to contrast. (a) Image f . (b) Multiscale NLTV spectral descriptions of different pixels. (c) Results of NLTV flow of f . (d) Multiscale NLTV spectral components.

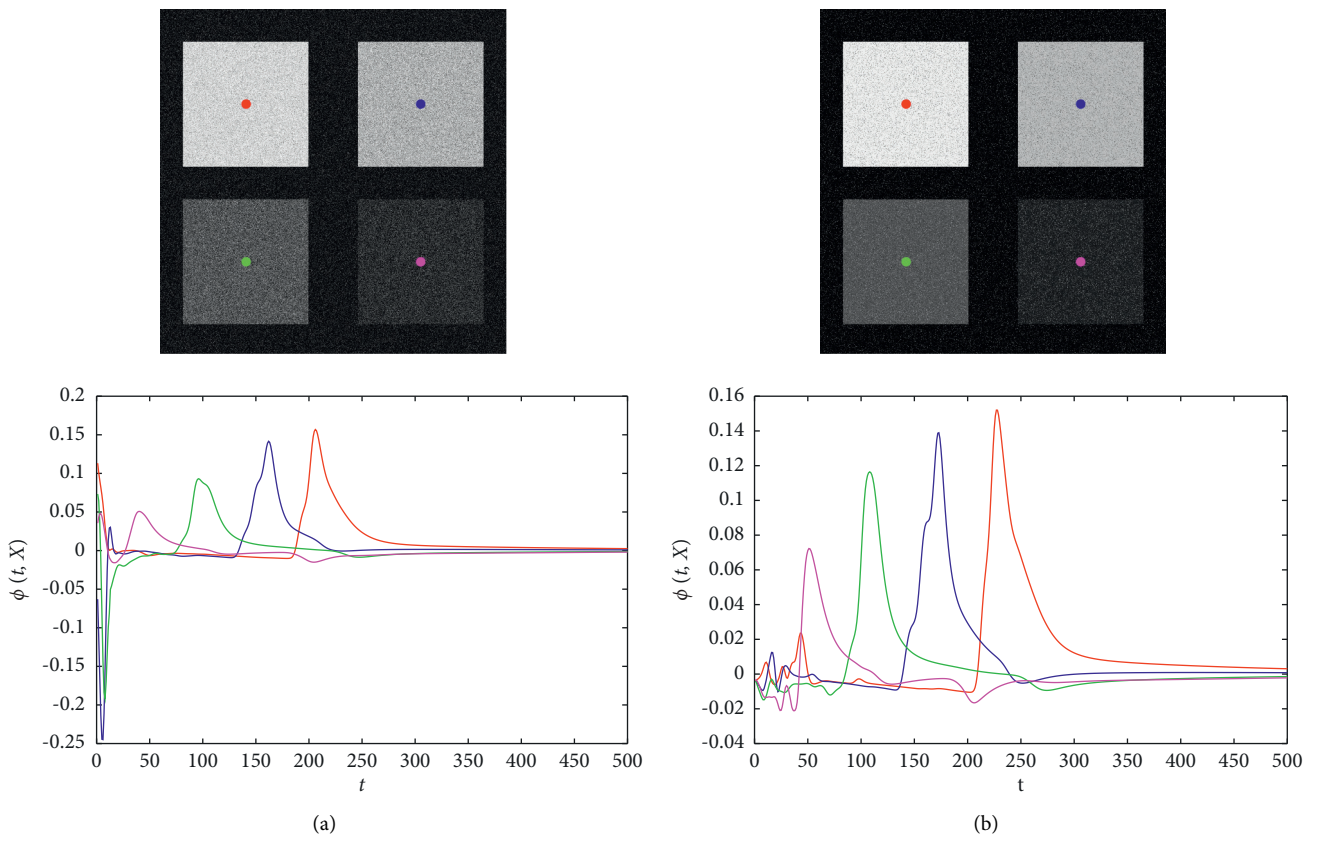


FIGURE 4: Continued.

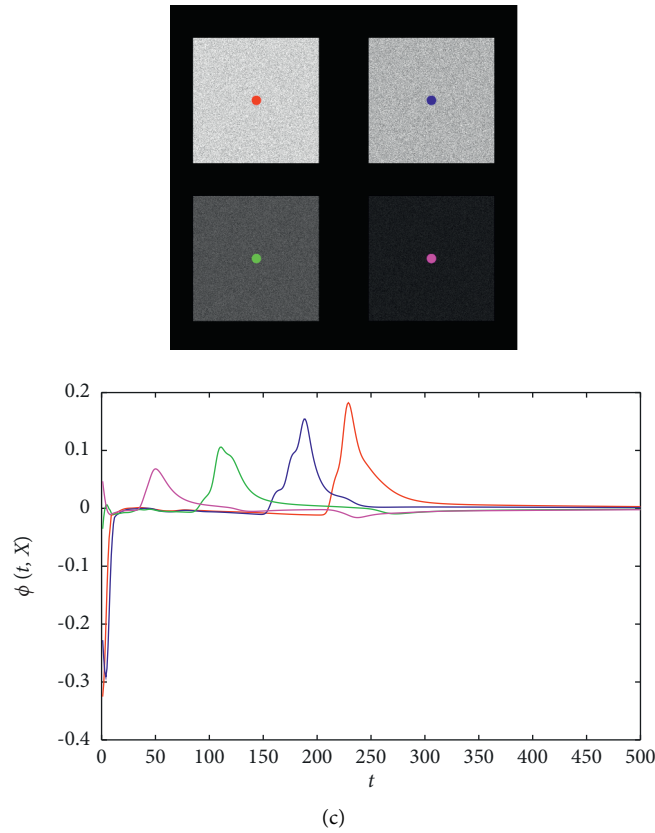


FIGURE 4: NLTV spectral transform on different local contrasts corrupted with different noises. (a) is the image corrupted with Gaussian (10% variance), Salt & Pepper (10% density), and Speckle noise (10% variance) respectively. (b) is the multiscale NLTV spectral descriptions of different pixels corrupted with noises. (a) Gaussian noise. (b) Salt & Pepper noise. (c) Speckle noise.

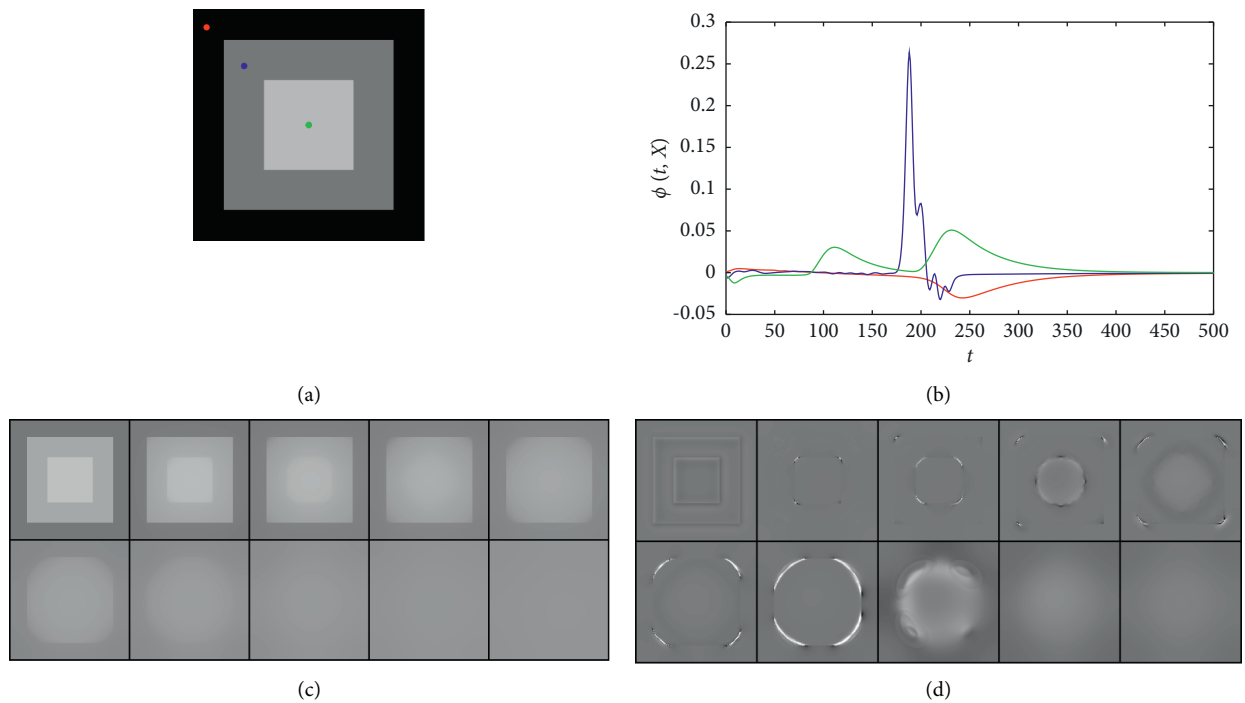


FIGURE 5: Demonstration of property 3. Signatures are distinguished because of their sensitivity to detailed structures. (a) Image f . (b) Multiscale NLTV spectral descriptions of different pixels. (c) Results of NLTV flow of f . (d) Multiscale NLTV spectral components.

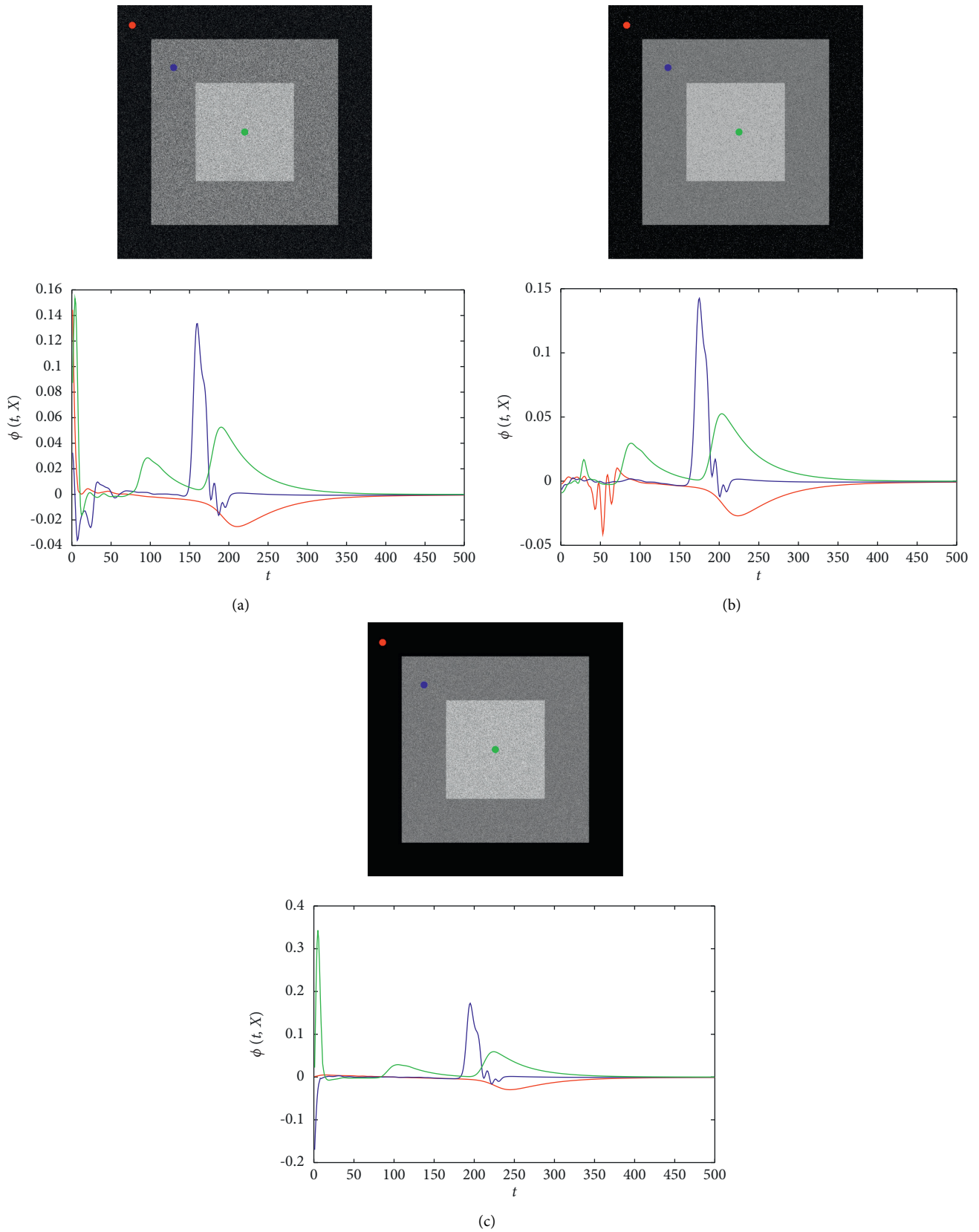


FIGURE 6: NLTV spectral transform on different structures corrupted with different noises. (a) is the image corrupted with Gaussian (10% variance), Salt & Pepper (10% density), and Speckle noise (10% variance), respectively. (b) is the multiscale NLTV spectral descriptions of different pixels corrupted with noises. (a) Gaussian noise. (b) Salt & Pepper noise. (c) Speckle noise.

Moreover, the image after translation by spatial shift on the original image is $f(X-d)$ and $f(X) = f(X-d)$. In essence, the rotation or translation of the image is equal to rotating or translating the coordinate system in the original image. On the other hand, the NLTV spectral transform is invariant to the coordinate system and sensitive to derivatives. Therefore, the NLTV spectral transform is invariant to rotation and translation, i.e.,

$$\begin{cases} \tilde{\phi}_{f(\mathbf{R}X)}(t, X) = \phi_f(t, \mathbf{R}X), \\ \tilde{\phi}_{f(X-a)}(t, X) = \phi_f(t, X-a). \end{cases} \quad (19)$$

There are three groups of objects with different shapes in figure 7(a). The objects in the same group have the same shape and contrast. Different objects have been translated in different positions and rotated at different angles. As figure 7(b) shows, the objects in the same group have a similar NLTV spectral description. More intuitive illustrations are displayed in figures 7(c) and 7(d), which present that the objects within the same group disappear simultaneously. Figure 8 shows the NLTV spectral descriptions of different pixels corrupted with noises. The bottom row of Figure 8 shows that the objects with the same shape have similar descriptions in large time scales, even though they have distinct rotations and translations.

3.2. NLTV Spectral Transform for Robust Image Segmentation

3.2.1. Overview of the Proposed Segmentation Flowchart.

Figure 9 shows the flowchart of the proposed method. The method starts with the decomposition of an original image in the NLTV spectral domain. Then, the available information dimension of every pixel in the image increases from one to the number of time scales. To better get appropriate components, a soft threshold band-pass filter is selected to replace the traditional hard threshold band-pass filter. After obtaining the separation surface result, an inverse transform is used to get an abstract structure. The segmentation result is obtained with the help of the binary process and morphological operations.

3.2.2. NLTV Spectral Decomposition. In the subsection, the process of image decomposition using the NLTV spectral transform is illustrated in detail. Assuming that the number of decomposition components is N , the NLTV flow xxx can be calculated with the help of formulae (6) and (7). According to the definition of the NLTV spectral transform described in formula (8), the second derivative of the element $u(i)$ with respect to time scale needs to be computed. To speed up the calculation, the first and second derivatives are combined, expressed by formula (20).

$$u_{tt}(i, X) = \frac{(u(i+1, X) + u(i-1, X) - 2 \cdot u(i, X))}{\Delta t^2}, \quad (20)$$

where Δt is the time interval. NLTV transform is obtained based on u_{tt} by equation (21).

$$\phi(i, X) = u_{tt}(i, X) \cdot i \cdot \Delta t. \quad (21)$$

The NLTV spectral response can also be calculated using equation (10). The residual can be computed by equation (9). If the forward time difference $u_t(i) = (u(i+1) - u(i))/\Delta t$ is used to calculate the first derivatives, the residual part \bar{f} can be transformed into formula (22).

$$\bar{f} = (N+1) \cdot u(N) - N \cdot u(N+1). \quad (22)$$

3.2.3. Object and Background Separation. After the decomposition of the original image in the NLTV spectral domain, the available information dimension of every pixel in the image increases from one to the number of time scales, i.e., the information used before decomposition is just pixel value. Inspired by the work [29], a separation surface is selected to effectively reduce the interference of noise on segmentation.

To better characterize the feature of objects in the image, time parameters t_1 and t_2 are chosen to construct a time range $[t_1, t_2]$. By the above analysis of the four properties of NLTV spectral transform, the max response time is computed to describe the image. The max response time here is different from the spectral response of equation (10). As equation (10) shows, the spectral response calculates the element $\phi(t)$ of the image in the NLTV spectral domain and can reflect the significant part of the image. The NLTV element $\phi(t)$ on the time scale t corresponding to the low response contains unimportant features, which can be discarded. However, formula (10) demonstrates that it fails to reflect the spatial information of the objects. To better analyze the performance of pixels in the NLTV spectral domain, the max response time is calculated. Specifically, the NLTV spectral transform, firstly, decomposes the image into several spectral components on a time scale, as shown in Figure 9. Then, every pixel in the image corresponds to a set of spectral responses. The time scale of the maximum spectral response is selected to indicate the performance of the local spatial information in the NLTV spectral domain. The maximum response time of pixels inside the same target tends to be close. Therefore, different objects of the image can be extracted by analyzing the max response time corresponding to each pixel. In other words, a salient time map $T(X)$ for each point X is calculated by equation (23).

$$T(X) = \arg \max_i \phi(i, X), \quad i \in [t_1, t_2]. \quad (23)$$

To extract more meaningful information about the segmentation target, we fit a separation surface whose role is a band-pass filter to separate the target from undesired information. Firstly, the filtered max response map $T_{\text{filter}}(X)$ is obtained by performing the Gaussian filtering on $T(X)$ to ensure the smoothness of separation surface. Then, the time scale corresponding to the maximum spectral response is stored as scatters, on which the least square regression is performed to finish fitting the surface $T_{\text{sur}}(X)$. The fitted surface can be regarded as a soft threshold in the range of

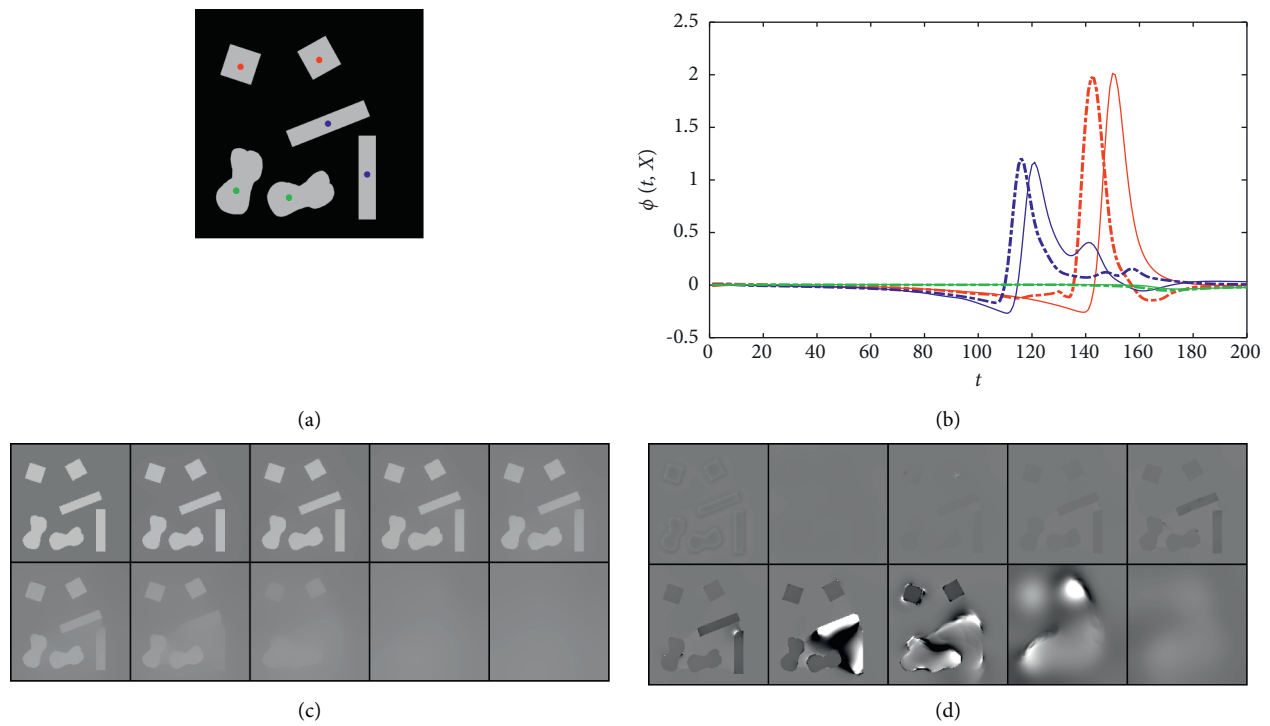


FIGURE 7: Demonstration of property 4. Signatures are similar to different rotations and translations. (a) Image f . (b) Multiscale NLTV spectral descriptions of different pixels. (c) Results of NLTV flow of f . (d) Multiscale NLTV spectral components.

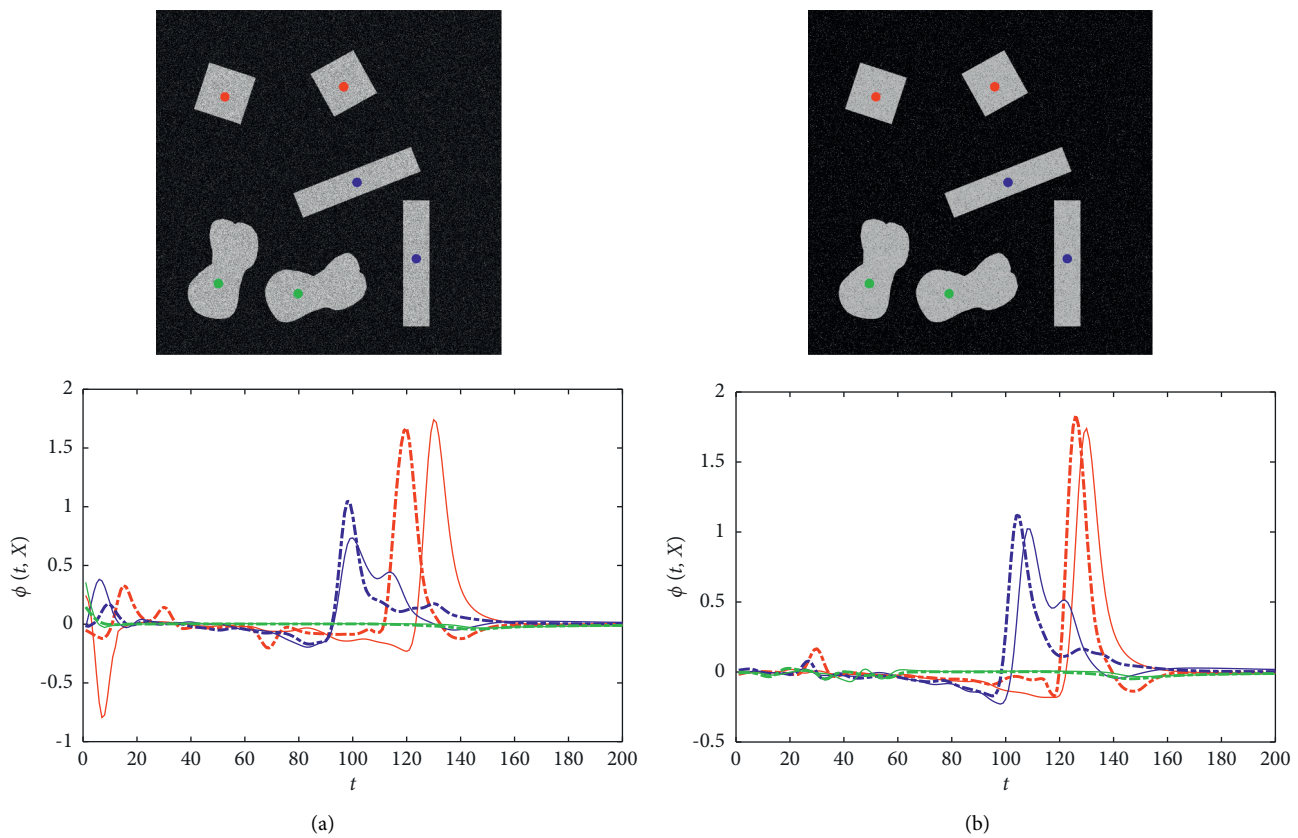


FIGURE 8: Continued.

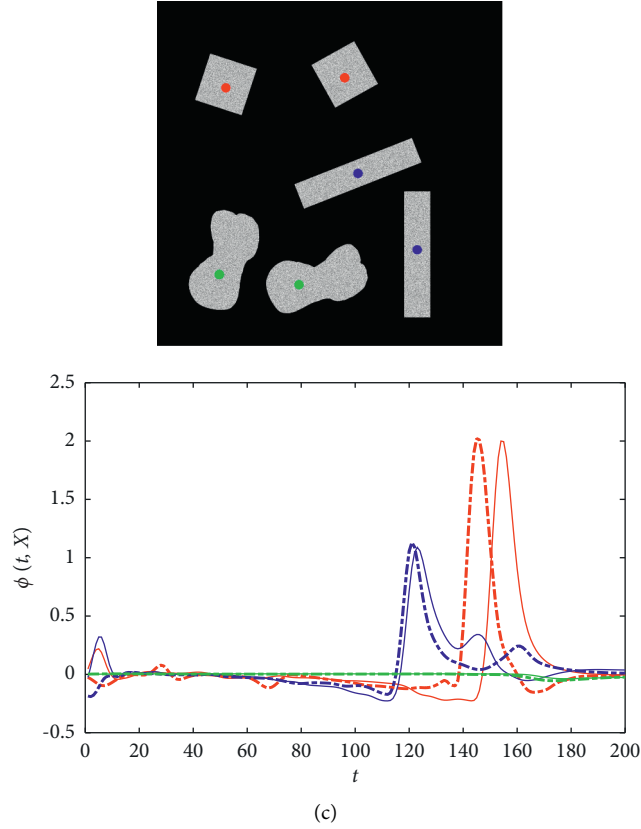


FIGURE 8: NLTV spectral transform on different groups corrupted with different noises. (a) is the image corrupted with Gaussian (10% variance), Salt & Pepper (10% density), and Speckle noise (10% variance), respectively. (b) is the multiscale NLTV spectral descriptions of different pixels corrupted with noises. (a) Gaussian noise. (b) Salt & Pepper noise. (c) Speckle noise.

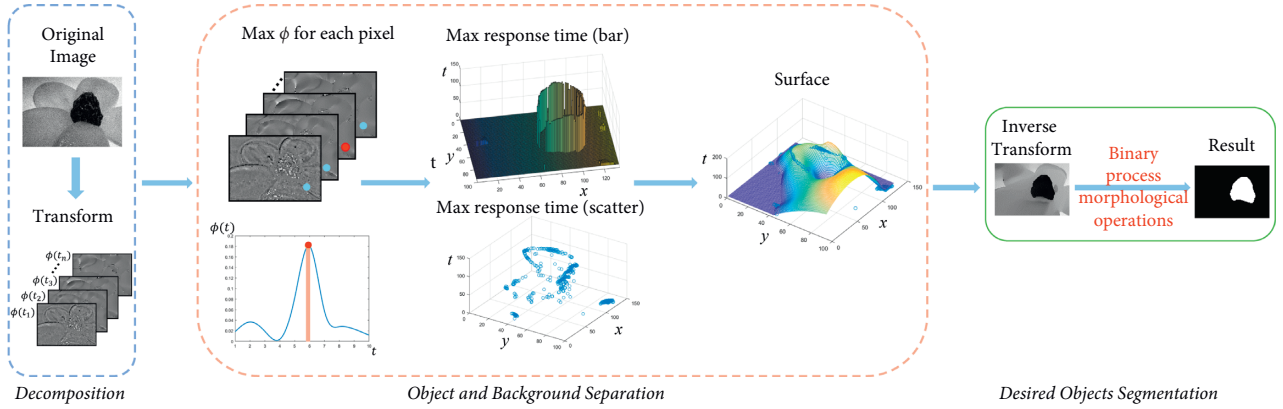


FIGURE 9: Flowchart of the proposed segmentation method using NLTV spectral transform.

$[t_1, t_2]$. For a certain point X , the surface divides it into two parts, $[t_1, T_{\text{sur}}(X)]$ and $[T_{\text{sur}}(X), t_2]$. The latter time range is usually chosen for image description to reduce the effect of noise. Therefore, the band-pass filter for each point X with time range $[t_1, t_2]$ can be denoted as follows:

$$HH_{\text{BPF}, t_1, t_2}(i) = \begin{cases} 0, & 1 \leq i < T_{\text{sur}}(X), \\ 1, & T_{\text{sur}}(X) \leq i \leq t_2, \\ 0, & t_2 < i \leq N. \end{cases} \quad (24)$$

3.2.4. Desired Objects Segmentation. Image reconstruction, which is also called inverse transform, is implemented after surface fitting. The time scale band represents the integration times of each pixel for the object. The target in the original image is easily obtained by integrating over a specific time scale using reconstruction formula (25).

$$I(x) = \sum_{t=T_{\text{sur}}(X)}^T \phi(t, X) + \bar{f}. \quad (25)$$

Binary processing is performed after inverse transform to obtain the segmentation mask. Finally, morphological operations are used to refine the final mask. By the above operations, the desired segmentation mask f_{output} is obtained. To exhibit more details of the proposed method, Algorithm 1 shows the specific process of the NLTV spectral transform-based method for robust image segmentation.

4. Experiment Results

4.1. Data and Settings. To evaluate the performance of the proposed method, synthetic, natural, and medical images are used for experiments. 1) The first experiment contains 3 groups of synthetic images whose textures are taken from the Brodatz Textures dataset [32]. Speckle, Salt & Pepper, and Gaussian noises are added to each group of synthetic images separately. 2) The second experiment contains 3 groups of natural images taken from the MSRA-1000 dataset [33]. 3) The third experiment contains 1 group of cell images, which is taken from the Fluo-N2DH-SIM+ dataset [34]. Three different types of noises are also added to natural and medical images.

We compare our segmentation method with four classical methods, i.e., the C-V model [13], FCM [14], FRFCM [19], and wavelet segmentation method (WSM) [27], which are used in the experiments. The experiments are implemented using the MATLAB R2020b platform and a PC with 16 GB RAM.

The parameter settings for the proposed method are as follows: experiments show that when the image is transformed into the NLTV domain, detailed information is located in a low time scale. Large scale, which is close to T , contains less important information. Objects are mostly distributed in the middle scale. Hence, a middle-scale time range $[t_1, t_2]$ is selected. In the following experiments, t_1 is set to $T/5$ and t_2 is set to $3T/5$. The parameters T and Δt are set to 9 and 0.03, respectively.

4.2. Quantitative Metrics. To quantitatively evaluate the performance of segmentation effect, four different metrics are chosen: FPR [21], FNR [21], dice similarity coefficient (DICE) [35], and segmentation accuracy (SA) [36].

To measure the difference between segmentation results and ground truths, FPR and FNR are chosen in the subsequent experiments. The former calculates the number of background pixels classified as object pixels relative to the total background pixels. FNR measures the number of object pixels classified as background pixels relative to the total object pixels. FPR and FNR are defined as follows:

$$\begin{aligned} \text{FPR} &= \frac{|B_R \cap O_G|}{|B_G|}, \\ \text{FNR} &= \frac{|O_R \cap B_G|}{|O_G|}, \end{aligned} \quad (26)$$

where B_R and B_G represent the number of background pixels in the segmentation results and ground truths, respectively.

Additionally, O_R and O_G are the number of object pixels in the segmentation results and ground truths, respectively.

DICE measures segmentation accuracy by calculating the degree of spatial overlap. Specifically, for the result region \mathbf{A} and target region \mathbf{B} ,

$$\text{DICE}(\mathbf{A}, \mathbf{B}) = \frac{2(\mathbf{A} \cap \mathbf{B})}{\mathbf{A} + \mathbf{B}}, \quad (27)$$

where \cap means the intersection of two sets. The value range of DICE is $[0, 1]$. The higher DICE indicates that the segmentation result is more precise. $\text{DICE}(\mathbf{A}, \mathbf{B}) = 1$ demonstrates that the segmentation result is the most complete, while $\text{DICE}(\mathbf{A}, \mathbf{B}) = 0$ shows that the segmentation result is the worst.

Another evaluation metric is SA, which can assess the number of well-classified pixels in the image. The definition of SA is given as follows:

$$\text{SA} = \frac{\sum_{i=1}^N f_i^{\text{truth}}}{N}, \quad (28)$$

where f_i^{truth} means the correctly segmented pixel and N denotes the total number of pixels in an image.

4.3. Parameter Analysis. This section analyzes the effects of Δt and T on the segmentation results of the proposed method through an experiment. The experiment was carried out on MSRA-1000, and the average SA was used as an indicator to show the influence of two parameters on the segmentation accuracy. The average SA was calculated by averaging the SA of all images on the dataset. The parameter Δt ranges from 0.01 to 0.1, and the step is 0.01. Additionally, the maximal time scale T ranges from 1 to 10, and the interval is 1. Figure 10 demonstrates the results for different Δt and T . The proposed method achieves the best performance when $\Delta t = 0.03$ and $T = 9$.

4.4. Synthetic Images. The first experiment was implemented on three synthetic images, which are shown in Figure 11. The first row shows a synthetic image containing multiple repeating structures and a dark grid-like background. A simple synthetic image, which has an irregular object, is arranged in the middle row. The object in the bottom row is complex and has a texture with inhomogeneous contrast. Moreover, three images are separately contaminated with Speckle (10% variance), Salt & Pepper (10% density), and Gaussian (10% variance) noise.

Table 1 lists the quantitative evaluations of different segmentation methods on various images. Combining with Figure 11 and Table 1, FCM got wrong segmentation results because of its sensitivity to noise. FRFCM achieved a good result on the first image and got a high DICE and SA value as shown in Table 1. However, it failed to distinguish the second and the third image because of the inhomogeneous contrast. WSM, which is based on spectral analysis, can remove the influence of noise. However, as Figure 11 shows, WSM oversmoothed the edge and damaged the edge details. Meanwhile, WSM was

Input: gray image f .
Output: segmentation mask f_{output} .

- (1) Initialize: maximal time scale T , time step Δt .
- (2) Calculate the number of decomposition components $N = T/\Delta t$.
- (3) Compute NLTV flow $\{u(i)\}_{i=0}^{N+1}$ using equations (6) and (7).
- (4) Calculate NLTV residual part \bar{f} using equation (22).
- (5) **for** $i = 1, 2, \dots, N$ **do**
- (6) Compute the second derivatives in time of flow for each pixel X by equation (20).
- (7) Achieve NLTV transform by equation (21).
- (8) Calculate NLTV spectral response using equation (10).
- (9) **end for**
- (10) Select time parameters t_1 and t_2 according to the NLTV spectral response.
- (11) Compute the salient time map $T(X)$ by equation (23).
- (12) Obtain $T_{\text{filter}}(X)$ by performing Gaussian filtering on $T(X)$.
- (13) Get the fitted surface $T_{\text{sur}}(X)$ by performing least square regression on $T_{\text{filter}}(X)$.
- (14) Reconstruct the result $I(X)$ using equation (25).
- (15) Get the segmentation mask $f_{bw}(X)$ by thresholding segmentation on $I(X)$.
- (16) Get the final mask $f_{\text{output}}(X)$ by performing morphological operations on $f_{bw}(X)$.

ALGORITHM 1: NLTV spectral transform-based method for robust image segmentation.

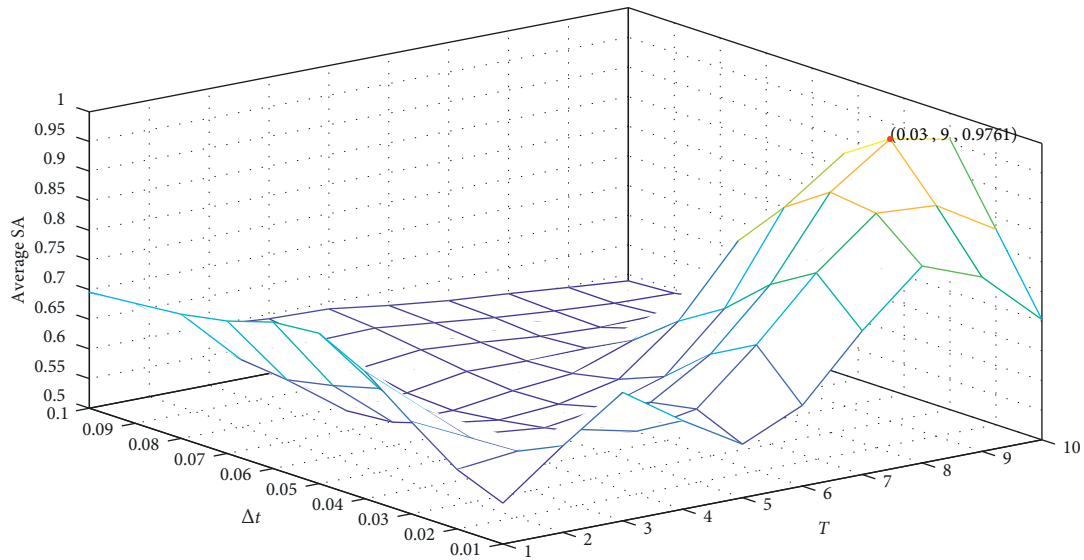


FIGURE 10: Effects of different Δt and T on the average SA.

unable to segment objects accurately on the second and third images. The reason is that WSM is sensitive to inhomogeneous contrast. The C-V model obtained the segmentation results of all images more correctly. One of the reasons was that the C-V model relies on an initial contour, which provides prior information about the approximate position of the object. Nevertheless, the C-V model was sensitive to noise. On the second and third images, the C-V model was unable to accurately segment the targets. The noises slowed down the convergence speed of the algorithm and made the method fall into the local minimum problem. However, the proposed method achieved the best results in all methods. The NLTV spectral transform-based method can segment the objects exactly and can reduce the influence of inhomogeneous contrast at the same time. The reason is that our method

can segment objects, combining object size, contrast, and structures. As shown in Table 1, the proposed method got a high FNR on the second synthetic image, which intended an under-segmentation. The problem was caused by the morphological operators in the output of the proposed method, which may cause edge corrodes.

4.5. Nature Images. To further discuss the proposed method's segmentation ability for images with various noises, the second experiment was performed on three natural images, which are shown in Figures 12, 13, and 14. The object that has a similar contrast to the surroundings is shown in Figure 12. Figure 13 displays a complex scene that has lots of tiny structures in the background. The object in Figure 14 is a piece of paper containing words, and the

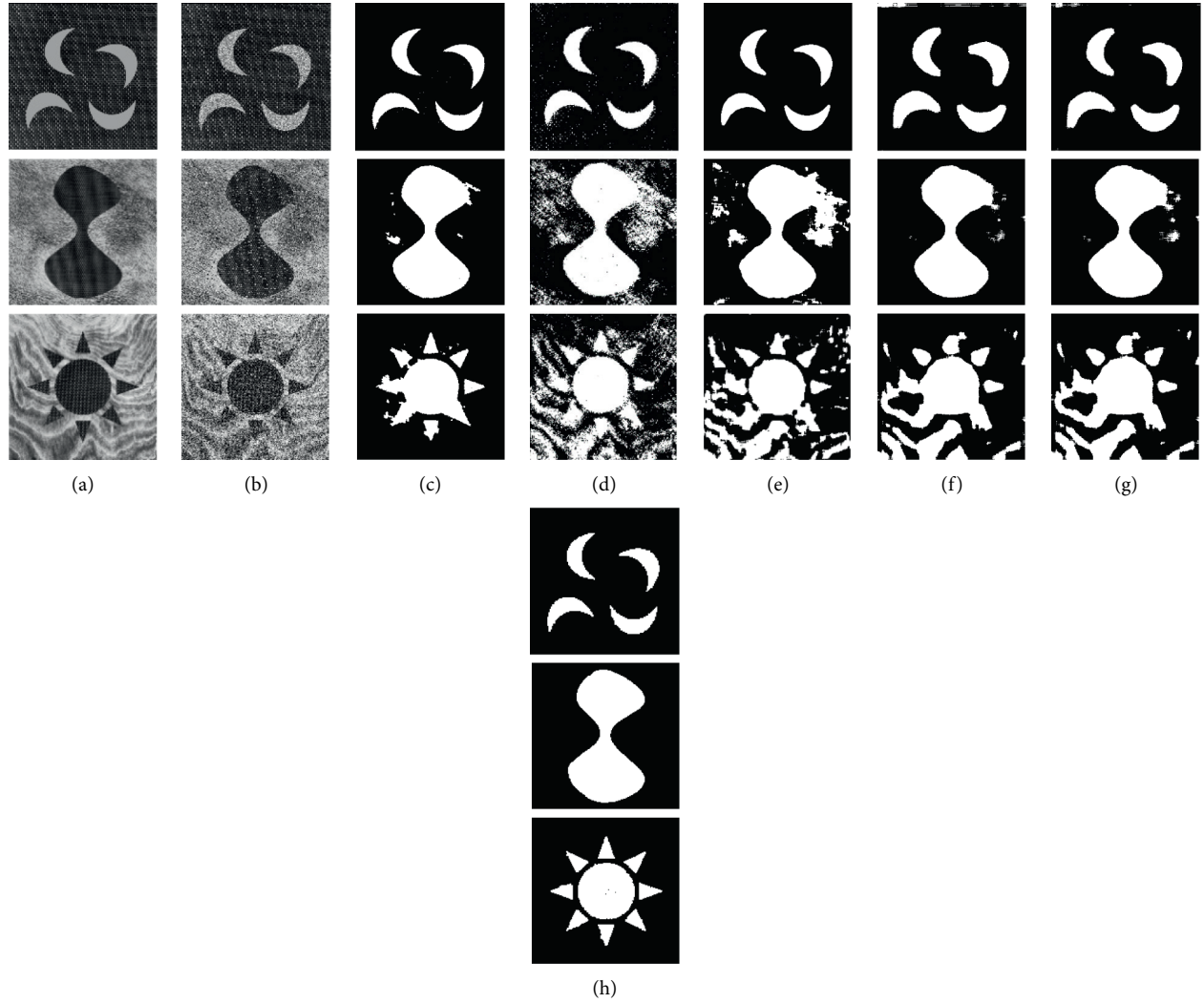


FIGURE 11: Segmentation results on synthetic images corrupted by different noises. (a) Original images. (b) Images (from top to bottom) that are corrupted with Speckle (10% variance), Salt & Pepper (10% density), and Gaussian noises (10% variance), respectively. (c) C-V Model. (d) FCM. (e) FRFCM. (f) WSM (db). (g) WSM (haar). (h) Proposed method.

TABLE 1: Evaluation metrics of compared methods for synthetic images, which are corrupted with Speckle (10% variance), Salt & Pepper (10% density), and Gaussian noise (10% variance), respectively.

Image	Metric	C-V model	FCM	FRFCM	WSM (db)	WSM (haar)	Proposed method
1	FPR	0.0025	0.0412	<i>0.0010</i>	0.0580	0.0402	0.0009
	FNR	0.0501	0.1977	0.0655	<i>0.0428</i>	0.0492	0.0290
	DICE	0.9596	0.7564	<i>0.9604</i>	0.7889	0.8328	0.9821
	SA	0.9907	0.9397	<i>0.9910</i>	0.9408	0.9558	0.9956
2	FPR	<i>0.0167</i>	0.2111	0.1355	0.0241	0.0326	0.0038
	FNR	<i>0.0231</i>	0.0743	0.0187	0.0389	0.0368	0.0273
	DICE	<i>0.9666</i>	0.7531	0.8427	0.9510	0.9415	0.9798
	SA	<i>0.9808</i>	0.8273	0.8959	0.9718	0.9659	0.9884
3	FPR	<i>0.0358</i>	0.3151	0.2352	0.1963	0.1984	0.0235
	FNR	0.0289	0.0901	0.0217	0.0253	0.0255	<i>0.0229</i>
	DICE	<i>0.9075</i>	0.5446	0.6402	0.6775	0.6751	0.9333
	SA	<i>0.9641</i>	0.7249	0.8015	0.8326	0.8308	0.9748

The best two results are highlighted in bold and italics fonts.

TABLE 2: Evaluation metrics of compared methods for “star,” which are corrupted with Speckle (10%, 20%, and 30% variance) noise.

Noise level (%)	Metric	C-V model	FCM	FRFCM	WSM (db)	WSM (haar)	Proposed method
10	FPR	<i>0.0014</i>	0.3299	0.2172	0.2953	0.2934	0.0010
	FNR	0.0636	0.1170	0.1456	<i>0.0228</i>	0.0044	0.0544
	DICE	<i>0.9612</i>	0.2633	0.3553	0.3067	0.3117	0.9667
	SA	<i>0.9953</i>	0.6823	0.7910	0.7189	0.7213	0.9957
20	FPR	<i>0.0014</i>	0.3431	0.2677	0.2979	0.2935	0.0011
	FNR	0.0573	0.1296	0.0354	<i>0.0121</i>	0.0107	0.0578
	DICE	<i>0.9616</i>	0.2534	0.3294	0.3071	0.3109	0.9649
	SA	<i>0.9950</i>	0.6696	0.7461	0.7169	0.7212	0.9955
30	FPR	<i>0.0010</i>	0.3454	0.2668	0.3001	0.2913	0.0010
	FNR	0.0718	0.1155	0.0947	0.0112	<i>0.0170</i>	0.0584
	DICE	<i>0.9599</i>	0.2558	0.3207	0.3059	0.3113	0.9652
	SA	<i>0.9949</i>	0.6683	0.7460	0.7150	0.7231	0.9950

The best two results are highlighted in bold and italics fonts.

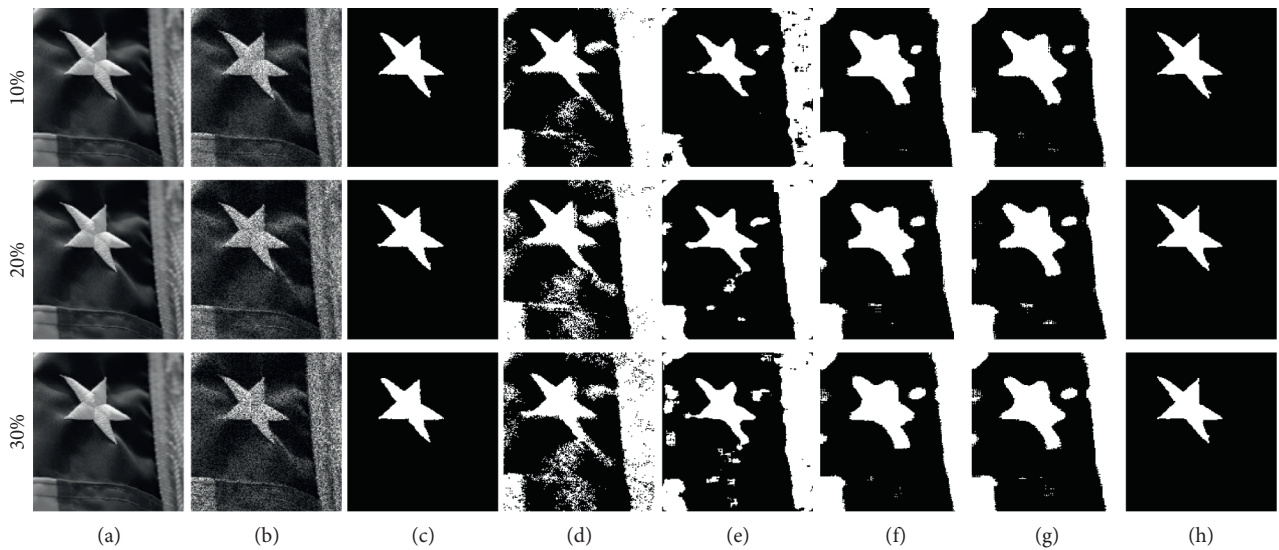


FIGURE 12: Segmentation results on image “star” corrupted by Speckle noise. (a) Original images. (b) Images (from top to bottom) that are corrupted with 10%, 20%, and 30% variance of Speckle noise, respectively (c) C-V model. (d) FCM. (e) FRFCM; (f) WSM (db); (g) WSM (haar); (h) Proposed method.

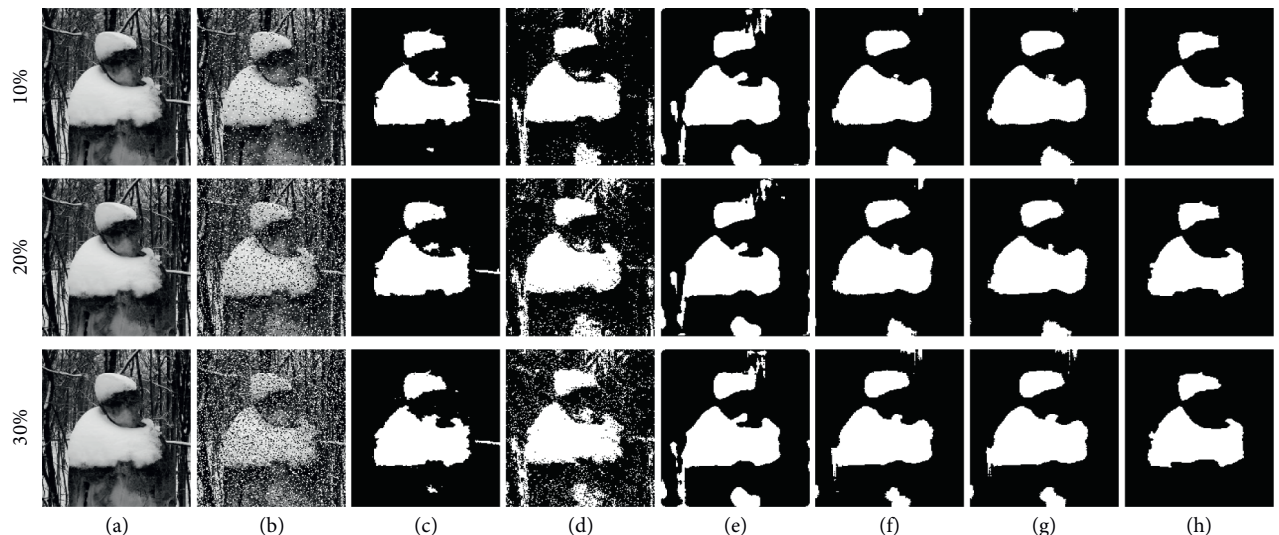


FIGURE 13: Segmentation results on image “snowman” corrupted by Salt & pepper noise. (a) Original images. (b) Images (from top to bottom) that are corrupted with 10%, 20%, and 30% variance of Speckle noise, respectively. (c) C-V Model. (d) FCM; (e) FRFCM. (f) WSM (db). (g) WSM (haar). (h) Proposed method.

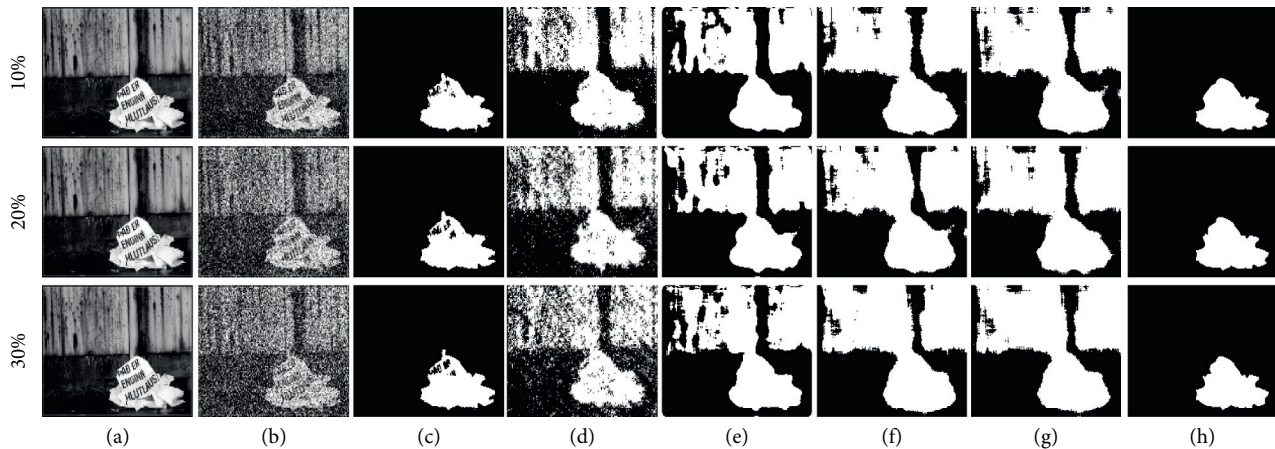


FIGURE 14: Segmentation results on image “paper scrap” corrupted by Gaussian noise. (a) Original images. (b) Images (from top to bottom) that are corrupted with 10%, 20%, and 30% variance of Speckle noise, respectively. (c) C-V Model. (d) FCM. (e) FRFCM. (f) WSM (db). (g) WSM (haar). (h) Proposed method.

words will interfere with segmentation methods. Moreover, three images are separately contaminated with Speckle (10%, 20%, and 30% variance), Salt & Pepper (10%, 20%, and 30% density), and Gaussian (10%, 20%, and 30% variance) noise.

As Figure 12 shows, the object has a similar contrast to the surrounding border. FCM separated the noise while segmenting the object because of its sensitivity to noise. FRFCM had better results than FCM, however, it still had wrong segmentation for noise. WSM can remove the influence of noise. However, WSM failed to remove the impact of inhomogeneous contrast. The C-V model achieved accurate segmentation results because of its initial contour. From Table 2, it can be seen that the C-V model had similar DICE and SA values with the proposed method. However, the C-V model was difficult to segment corner structure because of noises. The proposed method can better preserve structural information while segmenting.

Figure 13 shows the algorithms’ performances on the natural images, which are corrupted by Salt & Pepper noise. Table 3 shows the corresponding quantitative metrics. In Figure 13, there are lots of small objects in the background, which have a similar contrast to the object. When these small targets are contaminated with Salt & Pepper noise, they cause serious interference with segmentation methods, which mainly rely on contrast. Figure 13 shows that WSM has a good result; however, it is unable to segment the areas surrounding the object correctly. Table 4 shows that the C-V model has better results than WSM; however, it still has an incorrect segmentation of the background. Because of the sensitivity of the NLTV spectral transform to contrast, size, and structures, the proposed method can still separate objects when the background has small size structures.

Figure 14 shows the segmentation results on the natural image when it is corrupted with different levels of Gaussian noise. Table 4 shows the corresponding quantitative metrics of algorithms. The natural image is difficult for segmentation methods because it has complex texture like

words inside, which will affect the integrity of the segmentation results. The C-V model was capable of dealing with the background, however, it was unable to handle the interference of the internal texture of the object. FRFCM and WSM dealt with the effect of noise and internal texture but failed to remove the interference caused by contrast. Moreover, WSM cannot obtain accurate edge information of targets. As shown in Figure 14, WSM expanded the object and the edge details disappeared. However, our method can deal with the interference made by noise. The NLTV spectral transform was sensitive to local contrast and size. Hence, it can separate the low-contrast words on the paper scrap. Because of the contrast and structure difference between the paper scrap and the background, the proposed method can separate the object from the background and extract the object’s edge details correctly. Table 4 shows that the proposed method has high FNR values. From Figure 14, the bottom edge in the results of the proposed method is a little expanded, and the left edge is obviously corroded. The main reason is that the morphological operator makes the segmentation result corroded.

4.6. Medical Image. The proposed method was evaluated on a medical image in this part. Because the medical image has a black background and the inference of speckle noise on the image is not obvious, the experiment was implemented on an image with Gaussian noise and Salt & Pepper noise. As Figure 15 shows, the top row is a cell image, which is contaminated with Gaussian noise, and the bottom row is the cell image corrupted with Salt & Pepper noise. On account of the noise, the initial contour of the C-V model generated a local minimum problem and was unable to be iteratively converged. As a result, the segmentation results of the C-V model can only be around the initial contour. FCM had wrong results because of its sensitivity to noise. FRFCM obtained the best result on the cell image corrupted with Salt & Pepper noise. However, Gaussian noise can cause FRFCM to generate an over-segmentation. WSM can

TABLE 3: Evaluation metrics of compared methods for “snowman,” which are corrupted with Speckle (10%, 20%, and 30% variance) noise.

Noise level (%)	Metric	C-V model	FCM	FRFCM	WSM (db)	WSM (haar)	Proposed method
10	FPR	<i>0.0137</i>	0.1406	0.0737	0.0532	0.0467	0.0049
	FNR	<i>0.0577</i>	0.1756	0.0408	0.0848	0.0983	0.0890
	DICE	<i>0.9472</i>	0.6988	0.8556	0.8642	0.8675	0.9524
	SA	<i>0.9783</i>	0.8531	0.9332	0.9406	0.9431	0.9810
20	FPR	<i>0.0140</i>	0.1809	0.0763	0.0570	0.0497	0.0053
	FNR	<i>0.0553</i>	0.1768	0.0402	0.0764	0.0883	0.0979
	DICE	<i>0.9483</i>	0.6546	0.8520	0.8622	0.8682	0.9505
	SA	<i>0.9787</i>	0.8207	0.9312	0.9391	0.9428	0.9804
30	FPR	<i>0.0214</i>	0.2263	0.0785	0.0542	0.0548	0.0175
	FNR	<i>0.0529</i>	0.1651	0.0380	0.0880	0.0764	0.0997
	DICE	<i>0.9345</i>	0.6178	0.8505	0.8603	0.8656	0.9479
	SA	<i>0.9725</i>	0.7870	0.9301	0.9389	0.9408	0.9732

The best two results are highlighted in bold and italics fonts.

TABLE 4: Evaluation metrics of compared methods for “paper scrap,” which are corrupted with Speckle (10%, 20%, and 30% variance) noise.

Noise level (%)	Metric	C-V model	FCM	FRFCM	WSM (db)	WSM (haar)	Proposed method
10	FPR	<i>0.0013</i>	0.4130	0.4138	0.5043	0.5063	0.0002
	FNR	<i>0.1302</i>	0.0962	0.0265	0.0154	0.0230	0.0688
	DICE	<i>0.9265</i>	0.3813	0.4023	0.3585	0.3557	0.9754
	SA	<i>0.9809</i>	0.6271	0.6340	0.5557	0.5533	0.9939
20	FPR	<i>0.0016</i>	0.4236	0.4158	0.5059	0.5077	0.0005
	FNR	<i>0.1310</i>	0.1144	<i>0.0202</i>	0.0225	0.0162	0.0743
	DICE	<i>0.9250</i>	0.3686	0.4037	0.3562	0.3568	0.9725
	SA	<i>0.9805</i>	0.6152	0.6333	0.5538	0.5527	0.9932
30	FPR	<i>0.0016</i>	0.4239	0.4164	0.5170	0.5137	0.0011
	FNR	<i>0.1331</i>	0.1129	<i>0.0230</i>	0.0168	0.0188	0.0893
	DICE	<i>0.9239</i>	0.3693	0.4034	0.3525	0.3534	0.9661
	SA	<i>0.9803</i>	0.6154	0.633	0.5446	0.5472	0.9917

The best two results are highlighted in bold and italics fonts.

TABLE 5: Evaluation metrics of compared methods for “cell” which are corrupted with gaussian (5% variance) and salt & pepper (5% density) noise.

Noise	Metric	C-V model	FCM	FRFCM	WSM (db)	WSM (haar)	Proposed method
Gaussian	FPR	0.4835	0.2775	0.1360	0.1170	<i>0.1020</i>	0.0019
	FNR	0.7958	0.1577	0.0213	<i>0.0277</i>	0.0325	0.3540
	DICE	0.0926	0.4688	0.6638	0.6892	<i>0.7112</i>	0.7409
	SA	0.4779	0.7332	0.8640	0.8796	<i>0.8920</i>	0.9436
Salt & Pepper	FPR	0.1256	0.1488	0.0009	0.2666	0.2491	<i>0.0180</i>
	FNR	0.2415	0.1140	0.3895	0.0376	<i>0.0502</i>	0.2780
	DICE	0.6000	0.6216	0.8119	0.5163	0.5299	<i>0.8026</i>
	SA	0.8640	0.8484	0.9562	0.7523	0.7672	<i>0.9459</i>

The best two results are highlighted in bold and italics fonts.

better remove the influence of Gaussian noise. However, WSM achieved high FPR values in the cell image corrupted with Salt & Pepper noise, which intends the over-segmentation. Nevertheless, our method obtained good segmentation performances in both noises. Our method achieved the best results on the image corrupted with

Gaussian noise and obtained the second-best performance in Salt & Pepper noise. Table 5 shows that the proposed method achieves a high FNR value, which implies under-segmentation. As shown in the bottom row in Figure 15, the proposed method is difficult to segment the cells that have both small size and low contrast.

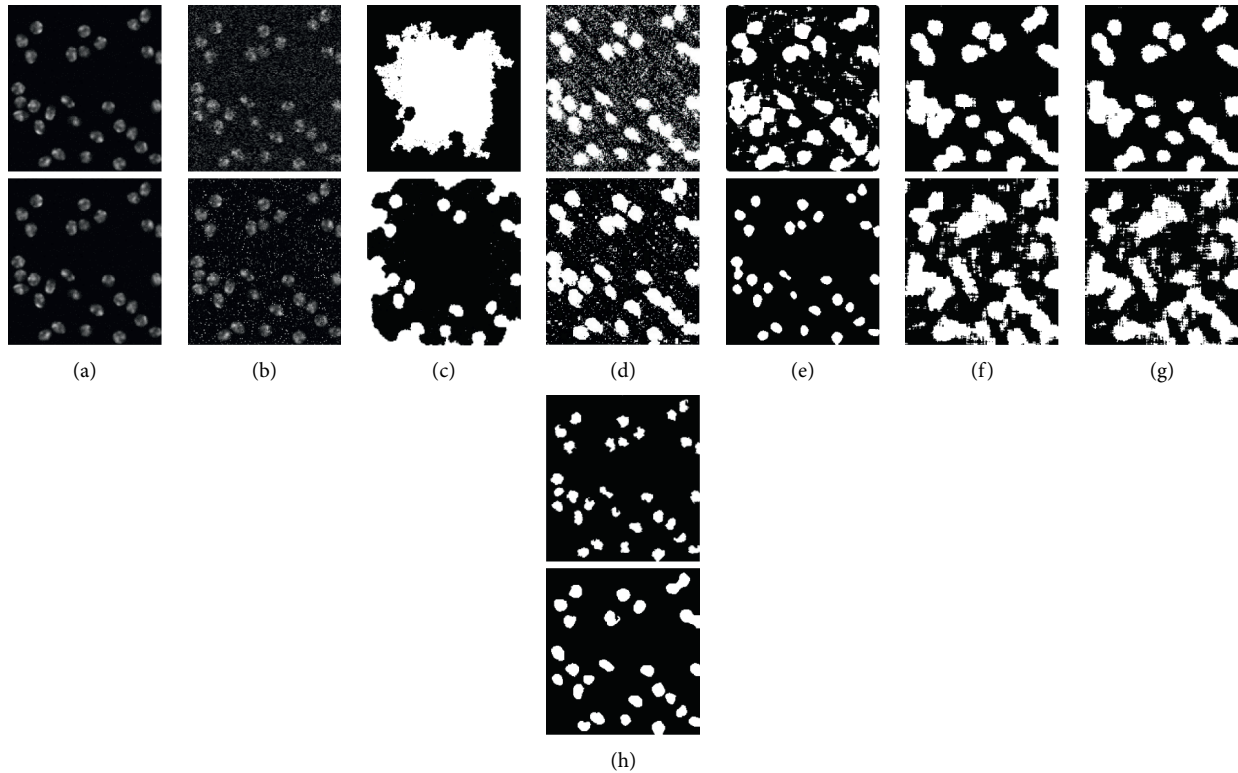


FIGURE 15: Segmentation results on image “cell” corrupted by Gaussian and Salt & Pepper noise. (a) Original images. (b) Images (from top to bottom) that are corrupted with Gaussian (5% variance) and Salt & Pepper (5% density) noise, respectively. (c) C-V Model. (d) FCM. (e) FRFCM. (f) WSM (db). (g) WSM (haar). (h) Proposed method.

5. Conclusion

We have analyzed the properties of NLTV spectral transform with the help of theoretical proof and experiments. Our analyses demonstrate that the object in an image corrupted with various noises can be separated its size, contrast, and detailed structure. The analyses also illustrate that the objects with same structures have similar descriptions in the NLTV spectral domain.

Furthermore, we have developed a novel transform-based method that segments images based on the NLTV spectral transform. The approach, firstly, decomposes an image into many sub-bands in the NLTV spectral domain and utilizes the max response time to represent the image features. Then, to better divide the object and background, the sub-bands in the NLTV spectral domain are filtered by fitting the separation surface, which is calculated based on maximum response time. Next, the filtered image is reconstructed by an inverse transform to obtain the rough segmentation result. Finally, the segmentation mask is calculated using postprocess methods. Subjective and objective evaluations show that the proposed method effectively protects the edge details while segmenting the object in a variety of noises.

However, one limitation of the proposed method is the high computational cost since the computation of nonlocal operators needs a long time and large memory storage. The other limitation of the method is the difficulty in fitting multiple separation surfaces accurately. We attempt to solve the aforementioned problems and develop a fast multiobject segmentation method in future work.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest.

Acknowledgments

This study was partly supported by the National Natural Science Foundation of China (62076137, 61972206, 61971233, and 62011540407) and was also partly supported under the framework of international cooperation program managed by the National Research Foundation of Korea (NRF-2020K2A9A2A06036255 and FY2020).

References

- [1] L. Sless, G. Cohen, S. B. El, and S. Oron, “Road scene understanding by occupancy grid learning from sparse radar clusters using semantic segmentation,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, Seoul, Korea, October 2019.
- [2] M. Akbari, J. Liang, and J. Han, “DSSLIC: deep semantic segmentation-based layered image compression,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 2042–2046, Brighton, UK, May 2019.

- [3] C. Bai, H. Li, J. Zhang, L. Huang, and L. Zhang, "Unsupervised adversarial instance-level image retrieval," *IEEE Transactions on Multimedia*, vol. 23, pp. 2199–2207, 2021.
- [4] C. Bai, L. Huang, X. Pan, J. Zheng, and S. Chen, "Optimization of deep convolutional neural network for large scale image retrieval," *Neurocomputing*, vol. 303, pp. 60–67, 2018.
- [5] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440, Massachusetts, MA, USA, June 2015.
- [6] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241, Munich, Germany, October 2015.
- [7] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: a deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.
- [8] J. M. S. Prewitt, "Object enhancement and extraction," *Picture Processing and Psychopictorics*, vol. 10, pp. 15–19, 1970.
- [9] R. Boyle and R. Thomas, *Computer Vision: A First Course*, pp. 48–51, Blackwell Scientific Publications, Hoboken, NJ, USA, 1988.
- [10] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 6, pp. 679–698, 1986.
- [11] M. N. Reza, I. S. Na, S. W. Baek, and K.-H. Lee, "Rice yield estimation based on K-means clustering with graph-cut segmentation using low-altitude UAV images," *Biosystems Engineering*, vol. 177, pp. 109–121, 2019.
- [12] D. Xiang, U. Bagci, C. Jin et al., "CorteXpert: a model-based method for automatic renal cortex segmentation," *Medical Image Analysis*, vol. 42, pp. 257–273, 2017.
- [13] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Transactions on Image Processing*, vol. 10, no. 2, pp. 266–277, 2001.
- [14] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: the fuzzy c-means clustering algorithm," *Computers & Geosciences*, vol. 10, no. 2-3, pp. 191–203, 1984.
- [15] C. Liu, W. Liu, and W. Xing, "An improved edge-based level set method combining local regional fitting information for noisy image segmentation," *Signal Processing*, vol. 130, pp. 12–21, 2017.
- [16] M. S. Yang and Y. Nataliani, "A feature-reduction fuzzy clustering algorithm based on feature-weighted entropy," *IEEE Transaction Fuzzy System*, vol. 26, pp. 817–835, 2017.
- [17] L. Guo, L. Chen, C. L. P. Chen, and J. Zhou, "Integrating guided filter into fuzzy clustering for noisy image segmentation," *Digital Signal Processing*, vol. 83, pp. 235–248, 2018.
- [18] Y. Jiang, K. Zhao, K. Xia et al., "A novel distributed multitask fuzzy clustering algorithm for automatic MR brain image segmentation," *Journal of Medical Systems*, vol. 43, no. 5, pp. 118–119, 2019.
- [19] N. Mahata, S. Kahali, S. K. Adhikari, and J. K. Sing, "Local contextual information and gaussian function induced fuzzy clustering algorithm for brain MR image segmentation and intensity inhomogeneity estimation," *Applied Soft Computing*, vol. 68, pp. 586–596, 2018.
- [20] P. Parida and N. Bhoi, "2-D Gabor filter based transition region extraction and morphological operation for image segmentation," *Computers & Electrical Engineering*, vol. 62, pp. 119–134, 2017.
- [21] P. Parida and N. Bhoi, "Wavelet based transition region extraction for image segmentation," *Future Computing and Informatics Journal*, vol. 2, no. 2, pp. 65–78, 2017.
- [22] D. Palani and K. Venkatalakshmi, "An IoT based predictive modelling for predicting lung cancer using fuzzy cluster based segmentation and classification," *Journal of Medical Systems*, vol. 43, no. 2, pp. 1–12, 2019.
- [23] Y. Wang, Q. Yuan, and C. He, "Indirect diffusion based level set evolution for image segmentation," *Applied Mathematical Modelling*, vol. 69, pp. 714–722, 2019.
- [24] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, CA, USA, 1998.
- [25] H. Liu, Z. Chen, X. Chen, and Y. Chen, "Multiresolution medical image segmentation based on wavelet transform," in *Proceedings of the IEEE Engineering in Medicine and Biology 27th Annual Conference*, pp. 3418–3421, Shanghai, China, January 2006.
- [26] H. Castillejos, V. Ponomaryov, L. Nino-de-Rivera, and V. Golikov, "Wavelet transform fuzzy algorithms for dermoscopic image segmentation," *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 578721, 2012.
- [27] J. Gao, B. Wang, Z. Wang, Y. Wang, and F. Kong, "A wavelet transform-based image segmentation method," *Optik*, vol. 208, Article ID 164123, 2020.
- [28] J. F. Aujol, G. Gilboa, and N. Papadakis, "Fundamentals of non-local total variation spectral theory," in *Proceedings of the 5th International Conference on Scale Space and Variational Methods in Computer Vision*, pp. 66–77, Lège-Cap Ferret, France, June, 2015.
- [29] G. Gilboa, "A total variation spectral framework for scale and texture analysis," *SIAM Journal on Imaging Sciences*, vol. 7, no. 4, pp. 1937–1961, 2014.
- [30] J. Zhang, J. Qi, Z. Zheng, and L. Sun, "A robust image segmentation framework based on total variation spectral transform," *Pattern Recognition Letters*, vol. 153, pp. 159–167, 2022.
- [31] M. D'Elia, Q. Du, C. Glusa, M. Gunzburger, X. Tian, and Z. Zhou, "Numerical methods for nonlocal and fractional models," *Acta Numerica*, vol. 29, pp. 1–124, 2020.
- [32] S. Abdelmounaime and H. Dong-Chen, *New Brodatz-Based Image Databases for Grayscale Color and Multiband Texture Analysis*, ISRN, 2013.
- [33] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1597–1604, Miami Beach, FL, USA, June 2009.
- [34] M. Maška, V. Ulman, D. Svoboda et al., "A benchmark for comparison of cell tracking algorithms," *Bioinformatics*, vol. 30, no. 11, pp. 1609–1617, 2014.
- [35] S. Minaee, Y. Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: a survey," *IEEE Transaction Pattern Analysis Machine Intelligence*, 2021.
- [36] S. Hao, Y. Zhou, and Y. Guo, "A brief survey on semantic segmentation with deep learning," *Neurocomputing*, vol. 406, pp. 302–321, 2020.