A Unified Tensor Level Set for Image Segmentation

Bin Wang, Xinbo Gao, Senior Member, IEEE, Dacheng Tao, Member, IEEE, and Xuelong Li, Senior Member, IEEE

Abstract—This paper presents a new region-based unified tensor level set model for image segmentation. This model introduces a three-order tensor to comprehensively depict features of pixels, e.g., gray value and the local geometrical features, such as orientation and gradient, and then, by defining a weighted distance, we generalized the representative region-based level set method from scalar to tensor. The proposed model has four main advantages compared with the traditional representative method as follows. First, involving the Gaussian filter bank, the model is robust against noise, particularly the salt- and pepper-type noise. Second, considering the local geometrical features, e.g., orientation and gradient, the model pays more attention to boundaries and makes the evolving curve stop more easily at the boundary location. Third, due to the unified tensor pixel representation representing the pixels, the model segments images more accurately and naturally. Fourth, based on a weighted distance definition, the model possesses the capacity to cope with data varying from scalar to vector, then to high-order tensor. We apply the proposed method to synthetic, medical, and natural images, and the result suggests that the proposed method is superior to the available representative region-based level set method.

Index Terms—Gabor filter bank, geometric active contour, image segmentation, level set method, partial differential equation (PDE) and tensor field.

I. INTRODUCTION

Image segmentation, a process of subdividing an image into a series of nonintersected subregions with approximately similar properties, is a fundamental step in the process of automatic image analysis since it helps to identify, describe, and understand different interested objects in images.

The field of image segmentation kept continuously developing for almost 40 years, and a number of algorithms came forth steadily in each year [36]. In recent years, active contour models have been increasingly and widely used in image processing. These methods all need to initialize a closed curve, also called the evolving curve [3], [4], [20], [26], in the image to be segmented and then evolve it driven by a partial differential equation (PDE) until the evolving curve converges. According to the representation of the evolving curve, active contour models can be classified as explicit [6], [7], [11], [35] and implicit [1]–[3], [10], [13], [17], [19], [20], [22], [23], [25], [26] categories. Snake [6], [11], [35], a typical explicit active contour model, uses parametric equations to explicitly represent the evolving curve. Implicit active contour models, i.e., level set methods [17], [19], [20], replace the parametric curve with a signed distance function, i.e., the level set function. A level set function [20] is a real-valued function of multiple variables; when the function takes a constant value, e.g., zero, the obtained set is the zero level set which is used to represent the evolving curve. Thus, the evolving curve is implicitly represented by the zero set of the level set function. This representation results implicit active contour models that could handle the topological changing more conveniently than the explicit active contour models [25].

Based on the way defining boundaries, level set methods can further be categorized into either edge-based [1], [2], [10], [13], [17], [19], [22] or region-based [3], [4], [32], [33] algorithms. Edge-based level set methods consider boundaries as a kind of discontinuation in gray values. To detect this discontinuation, these methods all define an edge indicator that is a positive and nondecreasing function. In theory, the evolving curve should stop where the edge indicator is equal to zero. However, in practice, due to the existence of noise, edge-based level set methods often cannot stop at a boundary location. Additionally, it is not easy for this kind of methods to detect weak and interior boundaries in images. To address these problems, Chan and Vese [3] proposed a region-based level set method called Chan-Vese (CV) method which incorporates the Mumford–Shah [15], [16] functional into a level set framework to give a piecewise constant representation to an image. The main idea of this method is to find the similarity in a given image, and it regards the subregion inside the evolving curve as a uniform object and the subregion outside as the background. The evolving curve is driven by an energy functional which is defined on the level set function and incorporates a “fitting” term. The “fitting” term defines the extent that the piecewise constant representation fits the initial image.

The CV method [3] utilizes a scalar, i.e., the gray value, to represent a pixel, but a scalar cannot comprehensively describe other information in an image, such as gradient and orientation information. Additionally, this method does not work well for images with noise, particularly the salt and pepper noise. Chan et al. [4] essentially extended the traditional CV method...
[3] from a scalar field to a vector field, but this method was mainly applied on the segmentation of multichannel images. That is to say, the representation of each pixel is still a scalar in each channel. Due to its vector form, this model ignores the spatial structure information among multichannel images. Wang and Vemuri [32], [33] proposed a level set method, the Wang-Vemuri (WV) method, to segment a symmetrical tensor field, e.g., diffusion tensor magnetic resonance images (DT-MRIs). This method can also be used to segment texture images by using a local structure tensor (LST) [30], [39]. However, an LST just takes into account the horizontal and vertical direction information and ignores a basic important component, i.e., the gray value, which makes the application of this method limited. Additionally, this method is only applicable for the symmetry tensor field.

Aiming at the aforementioned problems, we apply a Gabor filter bank [8], [9], [12], [27], [28] to extract the local geometrical features, e.g., scale and orientation, and construct a unified tensor representation, i.e., a three-order tensor, for each pixel by combining the Gabor features with the gray values of the image and the Gaussian smoothed image. There are three reasons to utilize a Gabor filter bank for feature extraction: 1) Gabor transform can detect multiscale edge information; 2) since a Gabor function provides a Gaussian weighted differential operation, it can be used to obtain the stable gradient information; and 3) for the orientation selectivity of the Gabor basis function, the Gabor filter bank can capture abundant orientation information of object boundaries. This unified tensor preserves more information than a scalar used in the CV method [3]. Based on this tensor representation, we extend the CV method [3] to accept tensors as input for image segmentation.

The proposed model has six advantages as follows. First, this model is robust for noisy images since the unified tensor representation incorporates the gray values of the Gaussian smoothed image. Second, because the three-order tensor representation contains orientation information, the model has better orientation selectivity helping to segment texture images. Third, the model takes into account the gradient information that intrinsically increases the weights of the pixels on the boundaries, which makes the evolving curve stop at the boundaries easier and help to segment objects in the nonhomogenous background. Fourth, since a pixel in an image is represented by a unified tensor, our model can obtain more accurate segmentation result. Fifth, the model possesses the capability to deal with data type varying from scalar to vector and to high-order tensor. Finally, the model is a kind of framework; the Gabor filter bank can be replaced by other analysis to build the tensor field. The proposed method has been applied to a set of synthetic, medical, and natural images. The result proves that, compared with the CV method [3] and the WV method [32], [33], the proposed method is robust against salt- and pepper-type noise, deals with the nonhomogenous image better, has the capacity of orientation selectivity, and detects an object more accurately.

The organization of the remainder of this paper is as follows. In Section II, we first describe the background of the CV model [3], its extensions to multichannel images [4], and the extension to the two-order symmetrical tensor field [32], [33] and then introduce the Gabor filter bank. Section III details the proposed unified tensor level set model, involving the construction of the unified tensor field and the tensor level set method, as well as its special cases. Section IV contains the implementation of the proposed model, including the regularization of the Heaviside and Dirac function, the evolution function, and its numerical scheme. Section V assesses the proposed model in comparing with the CV model [3] on synthetic and medical images. Section VI gives the concluding remarks and future work.

II. BACKGROUNDS

Gabor filter bank and region-based level set methods are both the basis of the proposed method. In this section, we first describe the gradual progress in the region-based level set methods and then introduce the Gabor filter bank utilized to build the unified tensor representation of a pixel.

A. Region-Based Level Set Methods

Compared with edge-based level set methods, region-based level set methods have many advantages [4], e.g., more robustness against noise and insensitivity to the initial position of the evolving curve. Since Chan and Vese proposed the scalar CV method [3], region-based level set methods have been sequentially extended from a scalar field to a vector field and then to a tensor field. In this process, the data structure became more and more complex, but the basic idea is similar. The rest of this section describes these methods.

Let \( u_0 : \Omega \rightarrow R \) be a given image, where \( \Omega \subset R^2 \) and is the image domain. Let \( C \subset \Omega \) be a closed curve implicitly represented by the zero level set of a Lipschitz function \( \phi : \Omega \rightarrow R \). In the scalar CV method [3], the energy functional \( E(c_1, c_2, C) \) is defined by

\[
E(c_1, c_2, C) = \mu \text{Length}(C) + \nu \text{Area}(C)
+ \lambda_1 \int_{\text{inside}(C)} \left| u_0(x, y) - c_1 \right|^2 \, dxdy
+ \lambda_2 \int_{\text{outside}(C)} \left| u_0(x, y) - c_2 \right|^2 \, dxdy
\]

(1)

where \( \mu \geq 0, \nu \geq 0, \lambda_1 > 0 \), and \( \lambda_2 > 0 \) are constant parameters; \( c_1 \) and \( c_2 \) are the gray value averages of inside and outside the evolving curve, respectively; \( \text{Length}(C) \) is the length of the evolving curve; and \( \text{Area}(C) \) is the area of the region closed by the evolving curve.

Chan et al. [4] proposed a region-based level set method for multichannel images, e.g., color image. This method can be regarded as a vector CV method [4]. The energy functional is defined by

\[
E(\vec{c}^+, \vec{c}^-, C) = \mu \text{Length}(C)
+ \int_{\text{inside}(C)} \frac{1}{N} \sum_{i=1}^{N} \left| u_{0,i}(x, y) - c_i^+ \right|^2 \, dxdy
+ \int_{\text{outside}(C)} \frac{1}{N} \sum_{i=1}^{N} \left| u_{0,i}(x, y) - c_i^- \right|^2 \, dxdy
\]

(2)
where $\bar{\sigma}^+$ and $\bar{\sigma}^-$ are the averages in vector form of the regions inside and outside the evolving curve, respectively, and $u_{i,j}(x, y)$ is the $i$th channel image.

Wang and Vemuri [32], [33] proposed the symmetrical tensor level set method for segmenting DT-MRIs; its energy functional is

$$E(T_1, T_2, C) = \mu \text{Length}(C)$$

$$+ \int_{\text{inside}(\Omega)} \text{dist}^2 \left(T(x, y), T_1\right) dx dy$$

$$+ \int_{\text{outside}(\Omega)} \text{dist}^2 \left(T(x, y), T_2\right) dx dy$$

(3)

where $T_1$ and $T_2$ are the averages in two-order symmetrical tensor form of the inside and outside regions, respectively; $\text{dist}(\cdot)$ is the Frobenius norm of matrices.

To sum up, the region-based level set methods extend from a single image [3] to multichannel images [4] and then to tensor images [32], [33]. However, they basically use a scalar to represent a pixel in a given image or a single channel image, and they do not provide a full and comprehensive representation to represent a pixel in a given image or a single channel image, and a single image [3] to multichannel images [4] and then to tensor images [32], [33]. However, they basically use a scalar to represent a pixel in a given image or a single channel image, and they do not provide a full and comprehensive representation to images. In the following section, we introduce the Gabor filter bank and describe the process using the Gabor filter bank to extract the local geometric features of the image.

B. Gabor Filter Bank

A human visual system is similar to a filter bank. Marcelja [18] and Daugman [8], [9] use Gabor functions to model the responses of the visual cortex. Daugman [8], [9] further developed the 2-D Gabor functions which are used by Lee [12] and Tao et al. [27], [28] to give images Gabor-based image representations.

A 2-D Gabor function is defined as

$$G_{s,d}(\mathbf{x}) = \frac{\|\mathbf{k}\|}{\sigma^2} \cdot \exp \left(-\frac{\|\mathbf{k}\|^2 \cdot \|\mathbf{\pi}\|^2}{2\sigma^2}\right) \cdot \exp(i\mathbf{k} \cdot \mathbf{\pi})$$

(4)

where $\mathbf{k} = (k_{\max} / f)^s e^{i(\pi / d_{\max})d}$ is the frequency vector that determines the scales and directions of the Gabor functions, $\mathbf{x} = (x, y)$ is the variable in a spatial domain, $\sigma$ is a parameter controlling the number of oscillations under the Gaussian envelope, and $s$ and $d$ denote the scale and direction, respectively.

In essence, the Gabor function is the product of a Gaussian function and a complex wave. As shown in Fig. 1, the real part of a 2-D Gabor function with a fixed direction and scale has a Gaussian envelope represented by the red grid line. When we choose a different scale and direction, a series of Gabor functions is obtained. In our model, $k_{\max} = \pi / 2$, $f = \sqrt{2}$, $\sigma = 3\pi / 2$, $d_{\max} = 8$, $d \in \{0, 1, 2, 3, 4, 5, 6, 7\}$, and $s \in \{0, 1, 2, 3\}$; then, a set of Gabor functions with four scales and eight directions is obtained and illustrated in Fig. 2.

From Fig. 2, we can visually find that these Gabor functions have great capacity for spatial localization and orientation selectivity. In our model, the Gabor-based image representa-

III. UNIFIED TENSOR LEVEL SET

This section first details the construction of the unified tensor field and then describes the energy functional and gradient flow (i.e., the evolution function) of the proposed unified tensor level set method. Finally, Section III-C discusses the three special cases of the proposed model.

A. Unified Tensor Representation

To segment an image more accurately, more overall information in the image to be segmented should be considered by the segmentation algorithms, and more suitable representation for the information should be used to depict the image. Early region-based level set methods [3], [4] just use a scalar representing a pixel in an image. Then, Wang and Vemuri’s method [32], [33] uses a symmetrical matrix, i.e., the LST [30], [39],
representing a pixel when this method is applied on texture images, but this method ignores gray values of pixels which is an important feature. None of these methods provide a relative comprehensive representation for images. Moreover, [4], [32], and [33] mainly focus on multichannel images or DT-MRIs and not on an appropriate representation for a single image.

By introducing Gabor features, we build a unified tensor representation for a pixel. This tensor representation contains more information (e.g., average gray value, gradient, and orientation) and is relatively overall. As illustrated in Fig. 3, the construction of the unified tensor representation contains three steps.

Step 1) To make our model more robust against noise, the initial image is smoothed by a Gaussian filter bank, and then, the gray value of each pixel in the smoothed image is included into the unified tensor representation as a matrix written as

\[
[T_{s,d,k=1}]_{S \times D} = \frac{1}{\sqrt{SD}} \begin{bmatrix}
G_{\sigma_1}(u^0_{x,y}) & \cdots & G_{\sigma_1}(u^0_{x,y}) \\
\vdots & \ddots & \vdots \\
G_{\sigma_S}(u^0_{x,y}) & \cdots & G_{\sigma_S}(u^0_{x,y})
\end{bmatrix}_{S \times D}
\]

where \(T_{s,d,k}^{x,y} \) is an element in the three-order tensor representation. \(s \) denotes the scale, and its maximum number is \(S \), and \(S = 4 \). \(d \) is the direction, and its maximum number is \(D \), and \(D = 8 \). \(u^0_{x,y} \) is the image to be segmented, and \(G_{\sigma_1,\ldots,\sigma_S}(\cdot) \) is the output generated by using the Gaussian function having different standard deviations convolving with the image. The standard deviations correspond with the standard deviations used by the Gabor filter bank.

Step 2) The gray value of each pixel in the image to be segmented is embed into the unified tensor representation, and the process is formulated as

\[
[T_{s,d,k=2}]_{S \times D} = \frac{1}{\sqrt{S \times D}} [u^0_{x,y}]_{S \times D}.
\]

Step 3) The Gabor features are used to represent the gradient and orientation of images. Having the Gabor functions defined by (4) convolved with the image to be segmented, the Gabor-based image representation in \(R^{M \times N \times 4 \times 8} \) is obtained. Thus, a rule of correspondence between a pixel of the image and a matrix in \(R^{4 \times 8} \) is built as follows:

\[
Gabor(u^0_{x,y} = G_{s,d}(x,y))
\]

where \(G_{s,d}(\cdot) \) is the Gabor function defined by (4) and Gabor(\cdot) is the outputs generated by convolving the Gabor functions with the image.

Through steps 1)–3), an image is projected on a five-order tensor in \(R^{M \times N \times 4 \times 8 \times 3} \). The first two indexes give the pixel location, and the last three indexes give the three-order tensor representation. That is to say, the third index gives the value of the scale, and the fourth index gives the direction. Since the image varies from coarse to fine along the fifth index, we call this index the fineness. Thus, each pixel in the image is represented by a three-order tensor in \(R^{4 \times 8 \times 3} \) which contains the densities in different fineness, the gradient and orientation information extracted from the neighborhood of the pixel by using the Gabor filter bank.

In a word, we utilize (5)–(7) to build a rule of correspondence between a scalar and a tensor in \(R^{4 \times 8 \times 3} \). This tensor provides pixels a more accurate and flexible tensor representation, and it is a comprehensive unified representation.

B. Unified Tensor Level Set Method

In this section, we propose the unified tensor level set method and detail the energy functional, the associated evolution equation, and a weighted distance function.

Let us define a tensor \(T \in R^{M \times N \times S \times D \times K} \) and then unfold \(T \) along mode 1 and mode 2 simultaneously. Thus, \(T \) becomes a tensor field \([5], [14], [21], [27], [28], [31], [34], [37], [38]\) with elements in the form of a three-order tensor in \(R^{S \times D \times K} \), and each element corresponds to a pixel in the image to be segmented. There is an evolving curve \(C \subseteq \Omega \in R^{M \times N} \) that divides the field \(T \) into two regions, i.e., \(\omega \in R^{M \times N}, \overline{\omega} \subseteq \Omega \), and \(C = \partial \omega \). Then, we assume that the field \(T \) is composed of these two homogenous regions and further assume that the object to be detected in the field is with the similar value. The fitting error between this piecewise constant representation and the field \(T \) is \(E_g \). Adding a regularizing term \(E_g \), the energy functional is defined as

\[
E(C) = E_g + E_e.
\]

\(E_g \) denotes the geometrical feature of the evolving curve, i.e., the length of the curve. Accompanied with the decreasing of this energy, the fitting term is minimized, and the segmentation result is obtained. This process is formulated by

\[
\inf_{C} \{E(C)\}.
\]

By manually initializing a closed curve \(\partial \omega \) in the image domain \(\Omega \), the level set function \(\phi : \Omega \rightarrow R \) in higher dimension is given by

\[
\phi(x,y) = \begin{cases} 
  d, & (x,y) \in \omega \\
  0, & (x,y) \in \partial \omega \\
  -d, & (x,y) \in \overline{\omega} 
\end{cases}
\]
where \( d \) is the shortest distance from point \((x, y)\) to the initial curve \( \partial \omega \) and \( \omega \) is the region enclosed by the initial curve \( \partial \omega \). Then, evolving the curve \( \omega \) in the tensor field, illustrated in Fig. 4, can be implicitly represented by the zero curve of the function \( \phi \) and written as [3]

\[
\begin{align*}
C &= \partial \omega = \{(x, y) \in \Omega : \phi(x, y) = 0\} \\
\text{inside}(C) &= \omega = \{(x, y) \in \Omega : \phi(x, y) > 0\} \\
\text{outside}(C) &= \Omega \setminus \omega = \{(x, y) \in \Omega : \phi(x, y) < 0\}.
\end{align*}
\]

Meanwhile, the regions inside and outside of the evolving curve \( C \) are homogenous, and their gray values are represented, respectively, by

\[
\begin{align*}
\bar{s}^{+}_{d, k} &= \frac{\int_{\Omega} T^{s, d, k}_{x,y} H(\phi) dx dy}{\int_{\Omega} H(\phi) dx dy} \\
\bar{s}^{-}_{d, k} &= \frac{\int_{\Omega} T^{s, d, k}_{x,y} (1 - H(\phi)) dx dy}{\int_{\Omega} (1 - H(\phi)) dx dy}
\end{align*}
\]

where \( \bar{s}^{+}_{d, k} \) and \( \bar{s}^{-}_{d, k} \) are the averages in tensor form of the regions inside and outside the evolving curve \( C \), respectively. Substituting (11) and (12) into (8), the energy functional is rewritten as

\[
E\left(\phi, \bar{s}^{+}_{d, k}, \bar{s}^{-}_{d, k}\right) \]

\[
= \mu \text{Length}(\phi) + \lambda_+ E_{\text{in}}(\phi, \bar{s}^{+}_{d, k}) + \lambda_- E_{\text{out}}(\phi, \bar{s}^{-}_{d, k})
\]

\[
= \mu \int_{\Omega} \delta(\phi(x, y)) |\nabla \phi(x, y)| dx dy
\]

\[
+ \lambda_+ \int_{\Omega} \text{dist}^2_{x,y} \left(T^{s, d, k}_{x,y}, \bar{s}^{+}_{d, k}\right) H(\phi(x, y)) dx dy
\]

\[
+ \lambda_- \int_{\Omega} \text{dist}^2_{x,y} \left(T^{s, d, k}_{x,y}, \bar{s}^{-}_{d, k}\right) (1 - H(\phi(x, y))) dx dy
\]

\[
(13)
\]

where \( \mu > 0 \) and is the weight of the regularizing term, \( \lambda_+ > 0 \) and \( \lambda_- > 0 \) are the weights of the fitting term. Length(\( \phi \)) is the regularizing term denoting the length of the evolving curve \( C \), i.e.,

\[
\text{Length}(\phi) = \int_{\Omega} |\nabla H(\phi(x, y))| dx dy
\]

\[
= \int_{\Omega} \delta(\phi(x, y)) |\nabla \phi(x, y)| dx dy.
\]

\[
(14)
\]

\( H(\cdot) \) is the Heaviside step function, i.e.,

\[
H(x) = \begin{cases} 
1, & \text{if } x \geq 0 \\
0, & \text{if } x < 0.
\end{cases}
\]

\[
(15)
\]

\( \delta(\cdot) \) is the Dirac delta function, i.e.,

\[
\delta(x) = \frac{d}{dx} H(x).
\]

\[
(16)
\]

The distance function is defined as

\[
\text{dist}_{x,y} \left(T^{s, d, k}_{x,y}, C^{s, d, k}_{+/-}\right) = \sqrt{\sum_{s=1}^{S} \alpha_s \sum_{d=1}^{D} \beta_d \sum_{k=1}^{K} \gamma_k \left(T^{s, d, k}_{x,y} - C^{s, d, k}_{+/-}\right)^2}
\]

\[
(17)
\]

where

\[
\alpha_s \geq 0 \sum_{s=1}^{S} \alpha_s = 1
\]

\[
\beta_d \geq 0 \sum_{d=1}^{D} \beta_d = 1
\]

\[
\gamma_k \geq 0 \sum_{k=1}^{K} \gamma_k = 1
\]

\[
(18)
\]

The definition of distance can be considered as a kind of weighted Frobenius norm or Hilbert–Schmidt norm.

Replacing the energy in (9) by (13), (9) is rewritten as

\[
\inf_{\phi, \bar{s}^{+}_{d, k}, \bar{s}^{-}_{d, k}} \left\{ E\left(\phi, \bar{s}^{+}_{d, k}, \bar{s}^{-}_{d, k}\right) \right\}
\]

\[
(19)
\]

which means that the energy functional, i.e., (13), should be minimized accompanied with the evolution of the level set function \( \phi \), which is a problem of calculus of variations. In order to solve this problem, we first fix \( \bar{s}^{+}_{d, k} \) and then compute the associated Euler–Lagrange equation for the unknown level set function \( \phi \). By adding an artificial time variable \( t \geq 0 \), the evolution equation is obtained as

\[
\frac{\partial \phi}{\partial t} = \delta(\phi) \left[ \mu \text{div} \left( \frac{\nabla \phi}{|\nabla \phi|} \right) \\
- \lambda_+ \sum_{s=1}^{S} \alpha_s \sum_{d=1}^{D} \beta_d \sum_{k=1}^{K} \gamma_k \left(T^{s, d, k}_{x,y} - \bar{s}^{+}_{s, d, k}\right)^2 \\
+ \lambda_- \sum_{s=1}^{S} \alpha_s \sum_{d=1}^{D} \beta_d \sum_{k=1}^{K} \gamma_k \left(T^{s, d, k}_{x,y} - \bar{s}^{-}_{s, d, k}\right)^2 \right]
\]

\[
(20)
\]
where \( \text{div}(\cdot) \) is the divergence. We use (20) as the gradient flow to minimize the energy functional (13).

### C. Special Cases of the Proposed Model

The proposed unified tensor level set model is a generalized version of the region-based level set method [3]. This section presents three special cases of the proposed model.

By taking the integral sign into the summation, \( E_{\text{in/out}}(\phi, c_+, c_-) \) in (13) is rewritten as

\[
E_{\text{in/out}}(\phi, c_+, c_-) = \int_{\Omega} \left( \sum_{s=1}^{S} \alpha_s \sum_{d=1}^{D} \beta_d \sum_{k=1}^{K} \gamma_k \left( T_{s,d,k}^x - c^x_{s,d,k} \right) \right)^2 \phi(x,y)dx dy
\]

\[
= \sum_{s=1}^{S} \alpha_s \sum_{d=1}^{D} \beta_d \sum_{k=1}^{K} \gamma_k \left( T_{s,d,k}^x - c^x_{s,d,k} \right)^2 \phi(x,y)dx dy
\]

\[
= \sum_{s=1}^{S} \alpha_s \sum_{d=1}^{D} \beta_d \sum_{k=1}^{K} \gamma_k E_{\text{in/out}}^{s,d,k}.
\]

Substituting (21) into (13), the energy functional is rewritten as

\[
E(\phi, c_+, c_-)
= \mu \text{Length}(\phi)
+ \sum_{s=1}^{S} \alpha_s \sum_{d=1}^{D} \beta_d \sum_{k=1}^{K} \gamma_k \left( \lambda_+ E_{\text{in}}^{s,d,k} + \lambda_- E_{\text{out}}^{s,d,k} \right)
\]

In our model, \( s \) denotes the scale, \( d \) the direction, and \( k \) the fineness. The maximum numbers of these parameters are \( S = 4, D = 8, \) and \( K = 3 \), respectively. Applying this method to the unified tensor field, we can reduce the unified tensor level set model into the following three cases.

**Case 1)** When \( S = D = K = 1 \), \( T_{s,d,k} \) is simplified as a scalar. Additionally, if \( \alpha_1 = \beta_1 = \gamma_1 = 1 \) and \( T_{s,d,k} = u_{s,d,k} \), the proposed model reduces into the scalar CV model [3].

**Case 2)** When \( S = D = 1 \), \( T_{s,d,k} \) is simplified as a vector. Additionally, if \( \alpha_1 = \beta_1 = 1, \gamma = [1/K, 1/K, \ldots, 1/K] \), and \( T_{s,d,k} = u_{s,d,k} \), our model reduces into the vector CV model [4].

**Case 3)** When \( S = D = 2, K = 1 \), \( T_{s,d,k} \) is simplified as a two-order symmetry tensor. Additionally, if \( \alpha = [1, 1, \ldots, 1], \beta = [1, 1, \ldots, 1] \), and \( T_{x,y} = \begin{bmatrix} \partial_x u_{x,y} \cdot \partial_x u_{x,y} & \partial_x u_{x,y} \cdot \partial_y u_{x,y} & \partial_y u_{x,y} \cdot \partial_y u_{x,y} \end{bmatrix} \) (i.e., the LST [30], [39]), the proposed model reduces into the model proposed by Wang and Vemuri [32], [33].

In brief, the proposed tensor level set method has the capacity to deal with the data varying from scalar to three-order tensor that is a full-information-contained representation for the pixels in images. Many methods [3], [4], [32], [33] can be obtained by reducing the proposed model.

### IV. IMPLEMENTATION

The evolution of the level set function is driven by the evolution equation, i.e., (20). To solve this equation, we first regularize the Heaviside step function and the Dirac delta function and then detail the numerical scheme of the proposed method.

#### A. Regularization of the Heaviside and Dirac Functions

The Heaviside step function and the Dirac delta function are discontinuous functions. To minimize the energy functional defined by (13), the Heaviside step function and the Dirac delta function should be regularized to be continuous functions

\[
H_{1,\varepsilon}(x) = \begin{cases} 
1, & x > \varepsilon \\
0, & \frac{\varepsilon}{2} < x < -\varepsilon \\
\frac{1}{2} \left[ 1 + \frac{x}{\varepsilon} + \frac{1}{2} \sin \left( \frac{\pi x}{\varepsilon} \right) \right], & |x| < \varepsilon
\end{cases} \quad (23)
\]

\[
\delta_{1,\varepsilon}(x) = H_{1,\varepsilon}(x) - \begin{cases} 
1, & x > \varepsilon \\
0, & \frac{\varepsilon}{2} < x < -\varepsilon \\
\frac{1}{2} \left( 1 + \cos \left( \frac{\pi x}{\varepsilon} \right) \right), & |x| < \varepsilon
\end{cases}
\]

\[
H_{2,\varepsilon}(x) = \frac{1}{2} \left( 1 + \frac{\varepsilon}{2} \arctan \left( \frac{x}{\varepsilon} \right) \right) \quad (24)
\]

\[
\delta_{2,\varepsilon}(x) = H_{2,\varepsilon}(x) - \begin{cases} 
1, & x > \varepsilon \\
0, & \frac{\varepsilon}{2} < x < -\varepsilon \\
\frac{1}{2} \left( 1 + \cos \left( \frac{\pi x}{\varepsilon} \right) \right), & |x| < \varepsilon
\end{cases}
\]

In practice, (23) and (24), illustrated in Fig. 5, are both applicable to the approximations of the Heaviside and Dirac functions. Although (23) looks more similar to the Heaviside
and Dirac function than (24) does, our model employs (24) to replace the corresponding function since (23) has small support interval which sometimes causes the whole algorithm to more easily compute a local minimum value of the energy. For more details, please refer to [3].

B. Numerical Scheme

There are many types of numerical scheme for the solution of PDEs, e.g., explicit scheme [1], [19], [20] and implicit scheme [3]. Here, since the semi-implicit scheme [4], [24] has the longer time step than the explicit scheme does and it is more easily implemented than the implicit scheme, this kind of scheme is used in the proposed model. The semi-implicit numerical scheme for the evolution equation, i.e., (20), is

\[
\phi_{x,y}^{n+1} = C \left[ \phi_{x,y}^n + m \left( C_1 \phi_{x+1,y}^n + C_2 \phi_{x-1,y}^n + C_3 \phi_{x,y+1}^n + C_4 \phi_{x,y-1}^n \right) \right] \\
- \Delta t \delta_{x,y} \left( \sum_{s=1}^{S} \alpha_s \sum_{d=1}^{D} \beta_d \sum_{k=1}^{K} \gamma_k \left( T_{s,d,k}^{x,y} - C_s^{s,d,k} \right)^2 \right) \\
- \Delta t \delta_{x,y} \left( \sum_{s=1}^{S} \alpha_s \sum_{d=1}^{D} \beta_d \sum_{k=1}^{K} \gamma_k \left( T_{s,d,k}^{x,y} - C_s^{s,d,k} \right)^2 \right) \right] \\
\]  

(25)

where \( \Delta t \) is the time step

\[
m = \frac{\Delta t}{h^2} \delta_{x,y} \left( \phi_{x,y}^n \right) \mu \]  

(26)

and \( C = 1 + m(C_1 + C_2 + C_3 + C_4) \), where

\[
C_1 = \frac{1}{\left( \frac{\phi_{x+1,y}^n - \phi_{x-1,y}^n}{2h} \right)^2 + \left( \frac{\phi_{x,y+1}^n - \phi_{x,y-1}^n}{2h} \right)^2} \\
C_2 = \frac{1}{\left( \frac{\phi_{x+1,y}^n - \phi_{x-1,y}^n}{2h} \right)^2 + \left( \frac{\phi_{x,y+1}^n - \phi_{x,y-1}^n}{2h} \right)^2} \\
C_3 = \frac{1}{\left( \frac{\phi_{x+1,y}^n - \phi_{x-1,y}^n}{2h} \right)^2 + \left( \frac{\phi_{x,y+1}^n - \phi_{x,y-1}^n}{2h} \right)^2} \\
C_4 = \frac{1}{\left( \frac{\phi_{x+1,y}^n - \phi_{x-1,y}^n}{2h} \right)^2 + \left( \frac{\phi_{x,y+1}^n - \phi_{x,y-1}^n}{2h} \right)^2}. 
\]  

(27)

The inside and outside averages of the evolving curve are updated, respectively, with the following regularized equations:

\[
c_{s,d,k}^{s,d,k} = \int_{\Omega} T_{s,d,k}^{x,y} H_s(\phi) dx dy \int_{\Omega} H_s(\phi) dx dy \\
c_{s,d,k}^{s,d,k} = \int_{\Omega} T_{s,d,k}^{x,y} \left( 1 - H_s(\phi) \right) dx dy \int_{\Omega} \left( 1 - H_s(\phi) \right) dx dy. 
\]  

(28)

V. EXPERIMENTAL RESULTS

We conducted several experiments on synthetic, medical, and natural images to illustrate the effectiveness of the proposed unified tensor level set method compared with the CV method [3] and the WV method [32], [33].

Experiment 1 applied the CV method [3] and the proposed method on the simulated MRIs with different nonhomogenous background. The medical images used in this experiment are built by adding a bias field to the initial images with homogeneous backgrounds. This bias field is created by

\[
BF(x,y) = \left[ \tilde{p}_y, \tilde{p}_y, ..., \tilde{p}_y \right]^T, \quad x \in \{1, 2, ..., M\} \\
\tilde{p}_y = \left[ 0, \frac{\text{ratio} \times G_{\max}}{N-1}, ..., \frac{\text{ratio} \times G_{\max}}{N-1} \right], \\
y \in \{1, 2, ..., N\} 
\]  

(29)

where \( \text{ratio} \) is the bias ratio, \( BF(\cdot) \) is the bias field, and \( G_{\max} \) is the maximum gray value of the initial image. Fig. 6 shows that the CV method [3] has a limited capability to segment objects from the nonhomogenous background. This method can still make a correct segmentation when the bias ratio in (29) is equal to 0.3, but it fails when the bias ratio increases to 0.4. The proposed method incorporates the Gabor features which reduce the influence from the nonhomogenous background. Therefore, the proposed method segments the image correctly until the bias ratio increases to 0.5.

Experiment 1 suggests that the proposed method is more insensitive to the nonhomogeneity than the CV method [3]. Additionally, the proposed method detects more tiny boundaries.

Fig. 6. First to the third row represent the evolution applying the CV method [3] on the images with different nonhomogenous background whose bias ratio varies from 0 to 0.3 then to 0.4. The fourth to the sixth row present the evolution applying the proposed method on the images with a bias ratio varying from 0.4 to 0.45 then to 0.5. The CV method [3] cannot make a correct segmentation when the bias ratio is equal to 0.4, but the proposed method still can segment the nonhomogenous images until the bias ratio increases to 0.5.
than the CV method [3] for images with the nonhomogenous background.

Experiment 2 applied the CV method [3] and the proposed method on images containing a paper-cut snowflake with salt and pepper noise. Fig. 7 shows that, even that the noise density is very small, i.e., equals 0.0005, the CV method [3] still wrongly classifies the positive impulse points as the object. The proposed method correctly segments the paper-cut snowflake from the background when the noise density increases to 0.1. Even when the noise density increases to 0.5, the proposed method still can demarcate the snowflake, although few of the positive impulse points are separated as the object by mistake.

Experiment 2 shows that the proposed method is more robust against salt and pepper noise than the CV method [3] since the proposed method reduces the influence from the salt and pepper noise by involving the gray values of the associated smoothed image.

In experiment 3, the CV method [3], the WV method [32], and the proposed method were applied to a synthetic texture image. In this image, the only difference between the object and the background is the orientation, as shown in Fig. 8. According to Chan and Vese [3], we replace the gray values in the CV method [3] by \( O(x, y) = \tan^{-1}(I_y'/I_x') \) to represent the orientation, but the method does not work well for the segmentation of both two images. Using the LST [30], [39] to represent the local texture, the WV method [32] correctly segments the object. Due to the incorporation of the Gabor features, the proposed method is sensitive to specialized orientations and also obtains the desired results.

Experiment 3 illustrates that the WV method [32] and the proposed method both have the feature of the orientation selectivity. Although eight orientations are used in the unified tensor representation, this is not the maximum number that the Gabor filter bank supports. If necessary, we can use more orientations to accurately extract the orientation information in the images.

In experiment 4, the CV method [3] and the proposed method were applied on a real image with the texture background to verify the capacity of the proposed method to segment a texture image accurately, as shown in Fig. 9. The image contains a paper-cut snowflake on a wood desktop. Because the weighted coefficient of the regularizing term \( \mu \) in (1) relatively influences the result of the segmentation, this coefficient is involved in this experiment. When the CV method [3] is applied on the image, the coefficient varies from \( 0.005 \times 255^2 \) to \( 0.01 \times 255^2 \) then to \( 0.05 \times 255^2 \). Fig. 9 shows that, when the coefficient is small (i.e., \( 0.005 \times 255^2 \)), parts of the texture background are wrongly segmented by the CV method [3]. When the
coefficient increases to $0.05 \times 255^2$, the CV method [3] misses some boundaries in the paper-cut snowflake. The WV method [32] does not segment the snowflake quite correctly. As to the proposed method, the paper-cut snowflake is segmented from the texture background correctly.

Experiment 4 shows that the proposed method segments the object from a texture background more accurately than the WV method [32] does and that the CV method [3] has no ability to deal with the texture segmentation. This is because the proposed method corresponds a pixel in the image to a three-order tensor which contains more information of the images, e.g., orientation, gradient, and gray value.

In experiment 5, the CV method [3], the WV method [32], and the proposed method were applied to an image containing a zebra on the grassland, as shown in Fig. 10. Because the proposed unified tensor contains the Gabor features, the image intensity, and the Gaussian smoothed images, it can duly separate the zebra from the grassland. However, because the CV method [3] only considers the intensity information, it cannot work as well as the proposed method. The LST [30], [39] used by the WV method [32] lacks scale information, which results in the failure of the segmentation of the zebra.

In experiment 6, the CV method [3], the WV method [32], and the proposed method were applied to an image containing a leopard in the underbrush, as shown in Fig. 11. Because the proposed unified tensor contains the Gabor features, the image intensity, and the Gaussian smoothed images, it can duly separate the leopard from the underbrush. However, because the CV method [3] only considers the intensity information, it cannot work as well as the proposed method. The WV method [32] fails to segment the leopard because of the same reason as the aforementioned experiment.

Experiments 5 and 6 suggest that the unified tensor representation is effective for the real texture segmentation, e.g., the zebra and leopard images. This is because the Gabor features incorporating into the unified tensor describe the local texture better than the LST [30], [39] used by the WV method [32].

Experiment 7 applied the CV method [3] and the proposed method on a real MRI of a human brain, as shown in Fig. 12. The boundaries detected by the CV method [3] are not smooth, and some tiny and obvious boundaries are not detected. The proposed method segments the image accurately.

Essentially increased so that the evolving curve can stop at the boundaries easier. Fig. 10 shows that the proposed method detects the obvious and tiny boundaries, and the segmentation result seems to be more accurate than the result of the CV method [3].

Experiment 7 illustrates that the proposed method is effective to segment real MRIs, and the segmentation result is more accurate than that of the CV method [3]. Additionally, the result obtained by the proposed method looks rational because the Gabor-based image representation coincides with a human vision system.

In experiment 8, the CV method [3] and the proposed method were applied on a real computerized tomography (CT) image of a human abdomen with nonhomogenous background, as shown in Fig. 13. The CV method [3] does not segment the image correctly because this method just considers the gray value of the pixels, and that classifies some light parts of the background as objects by mistake. Involving the gradient information by the Gabor features, the proposed method makes the evolving curve stop at the boundaries easier. Meanwhile, the Gabor filter bank makes the proposed method insensitive to the illumination. Therefore, the proposed method segments the image correctly.
Experiment 8 suggests that the proposed method is practical for the segmentation of CT images, moreover, achieves better effectiveness than the CV method [3] for nonhomogenous CT images.

Experiment 9 applied the CV method [3] and the proposed method on a natural image with a horse running on the beach, as shown in Fig. 14. The background of this image is more complex than the one of the images used by the aforementioned experiments. The CV method [3] cannot completely separate the horse from the background, whereas the proposed one basically does. This is because the proposed method uses a unified tensor representing each pixel in the image. This tensor contains information from the smoothed image which smooths weak boundaries in the background; meanwhile, the gradient and orientation extracted from the neighbor of the pixels give a more accurate depiction to pixels. These all result in a natural result.

Experiment 9 shows that the proposed method is effective for the segmentation of natural images, particularly for the natural image with the complex background.

VI. CONCLUSION

In this paper, we built a unified tensor representation for each pixel in an image to comprehensively depict the information of the image and then generalized the scalar CV method [3] to a unified tensor level set method by proposing a weighted tensor distance definition. By involving the Gabor features into the unified tensor representation, our model has the capacity of orientation selectivity and better sensitivity to gradient. Meanwhile, by incorporating the gray values in different fineness into the tensor pixel representation, the proposed method is more robust against noise, particularly against the salt- and pepper-type noise. At last, we have to say that the unified tensor pixel representation results in the computational inefficiency, and this is exactly our future work.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for the helpful comments and suggestions.

REFERENCES


**Bin Wang** received the B.Sc. and M.Sc. degrees in electronics and information system from Northwest University, Xi’an, China, in 1999 and 2002, respectively. He is currently working toward the Ph.D. degree in pattern recognition and intelligent system at Xidian University, Xi’an, China. His research interest is image segmentation.

Xinbo Gao (M’02–SM’07) received the B.Sc., M.Sc., and Ph.D. degrees in signal and information processing from Xidian University, Xi’an, China, in 1994, 1997, and 1999, respectively. From 1997 to 1998, he was a Research Fellow with the Department of Computer Science, Shizuoka University, Shizuoka, Japan. From 2000 to 2001, he was a Postdoctoral Research Fellow with the Department of Information Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong. Since 2001, he has been with the School of Electronic Engineering, Xidian University, where he is currently a Professor of pattern recognition and intelligent system and the Director of the VIPS Lab. His research interests are computational intelligence, machine learning, computer vision, pattern recognition, and artificial intelligence. In these areas, he has published four books and around 100 technical articles in refereed journals and proceedings, including IEEE TIP, TCSVT, TNN, TSMC, etc.

Dr. Gao is on the Editorial Boards of many journals, including EURASIP Signal Processing (Elsevier). He served as General Chair/Cochair or Program Committee Chair/Cochair or PC member for around 30 major international conferences.

**Daecheng Tao** (M’07) received the B.Eng. degree from the University of Science and Technology of China, Hefei, China, the M.Phil. degree from The Chinese University of Hong Kong, Sha Tin, Hong Kong, and the Ph.D. degree from the University of London, London, U.K.

He is currently a Nanyang Assistant Professor with the School of Computer Engineering, Nanyang Technological University, Singapore, and holds a visiting post in London. He is a Visiting Professor with Xidian University, Xi’an, China, and a Guest Professor with Wuhan University, Wuhan, China. He is also a Visiting Research Fellow with Birkbeck College, University of London. He has published more than 100 scientific papers, including IEEE TPAMI, TIP, TKDE, NIPS, CVPR, ECCV, and ACM TKDD and SIGKDD with best paper runner up awards and finalists. One of his TPAMI papers received an interview with ScienceWatch.com (Thomson Scientific). He has authored/edited six books and eight journal special issues. His research is mainly on applying statistics and mathematics for data analysis problems in computer vision, data mining, machine learning, multimedia, and visual surveillance.

Dr. Tao is a member of the IEEE Computer Society, the IEEE Signal Processing Society, and the IEEE System, Man, and Cybernetics Society. He is an Associate Editor of the IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING and the Computational Statistics and Data Analysis (Elsevier). He has chaired/cochaired for special sessions, invited sessions, workshops, panels, and conferences. He has served with more than 80 major international conferences and more than 30 prestigious international journals. His Erdős number is three. He was the recipient of several Meritorious Awards from the International Interdisciplinary Contest in Modeling, which is the highest level mathematical modeling contest in the world, organized by COMAP.

Xuelong Li (M’02–SM’07) is a researcher (Full Professor) with the State Key Laboratory of Transient Optics and Photonics, Xi’an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi’an, China.