J Math Imaging Vis (2012) 44:330–340
DOI 10.1007/s10851-012-0329-z

Nonparametric Image Segmentation Using Rényi's Statistical Dependence Measure

Haili Zhang · Yunmei Chen · Jiangli Shi

Published online: 9 February 2012

Abstract In this paper, we present a novel nonparametric active region model for image segmentation. This model partitions an image by maximizing the similarity between the image and a label image, which is generated by setting different constants as the intensities of partitioned subregions. The intensities of these two images can not be compared directly as they are of different modalities. In this work we use Rényi's statistical dependence measure, maximum cross correlation, as a criterion to measure their similarity. By using this measure, the proposed model deals directly with independent samples and does not need to estimate the continuous joint probability density function. Moreover, the computation is further simplified by using the theory of reproducing kernel Hilbert spaces. Experimental results based on medical and synthetic images are provided to demonstrate the effectiveness of the proposed method.

Keywords Nonparametric · Image segmentation · Pattern classification · Rényi's measure · Image processing · Experimental results

1 Introduction

Image segmentation or pattern classification is of fundamental importance in the field of medical image processing. During the last few decades, a considerable amount of approaches have emerged to tackle this issue. However, the difficulties caused by intensity inhomogeneity, higher level of noise and unevenly distributed illumination still need to be addressed.

Let Ω be a bounded open subset of $\mathbb{R}^2$, and $I : \Omega \to \mathbb{R}$ be an observed noisy image. For the two-phase case, Mumford and Shah [1] provide the following model for simultaneous smoothing and segmentation:

$$\min_{u,C} \int_{\Omega_i} (u_i - I)^2 \mathrm{d}x + \alpha \int_{\Omega \setminus C} |\nabla u|^2 \mathrm{d}x + \beta |C|, \tag{1}$$

where C separates Ω into two different regions Ω_i $(i = 1, 2)$, $|C|$ represents the length of C, u is a piecewise smooth approximation of I, and u_i is the restriction of u to Ω_i. When u is a constant c_i in each Ω_i, model (1) reduces to the following form,

$$\min_{C, c_i's} \sum_i \int_{\Omega_i} (c_i - I)^2 \mathrm{d}x + \beta |C|. \tag{2}$$

This piecewise constant Mumford-Shah model has been well studied by Chan et al. in [2, 3]. The major advantage of this model is that it can separate two relatively homogeneous regions without using any edge information. However, the homogeneity assumption limits its applications.

A more general approach is parametric region based active contour method. This method is based on the assumption that at each $x \in \Omega_i$, the image intensity $I(x)$ is an independent random variable drawn from the probability density function (p.d.f.) $P(I_i(x)|\lambda_i)$, where I_i is the restriction of I to Ω_i and λ_i is a parameter vector which needs to be estimated. The framework of this method minimizes the negative log-likelihood functional together with the length term,

H. Zhang (✉) · Y. Chen · J. Shi
Department of Mathematics, University of Florida, Gainesville, FL 32611, USA
e-mail: hlzhang@ufl.edu

Y. Chen
e-mail: yun@ufl.edu

J. Shi
e-mail: jlshi@ufl.edu

i.e.,

$$\min_{C,\lambda_i' s} -\sum_i \int_{\Omega_i} \log P\big(I_i(x)|\lambda_i\big)\mathrm{d}x + \beta|C|. \tag{3}$$

In the region competition model by Zhu et al. [4] and geodesic active region models by Rousson et al. [5] and Paragios et al. [6], $P(I_i(x)|\lambda_i)$ is chosen to be a Gaussian distribution:

$$P\big(I_i(x)|c_i,\sigma_i\big) = \frac{1}{\sqrt{2\pi}\sigma_i}\exp\left(\frac{-(I_i(x)-c_i)^2}{2\sigma_i^2}\right). \tag{4}$$

If all the σ_i's are the same and prefixed, model (3) (4) reduces to model (2). This is a global Gaussian model as it assumes all random variables $I(x)$ in the region Ω_i share the same mean c_i and variance σ_i. Model (3) provides desirable segmentation results when the parametric form of the intensity distribution is known. However, a specific assumption of the intensity distribution can be a significant restriction in real applications, especially when the image has heavy noise or is of multi-modal intensity distribution.

To overcome this problem, nonparametric models [7] have been developed to increase the robustness and successfully applied to image segmentation and registration. These methods are featured by using nonparametric density estimation to replace the parametric density estimation. For instance, the nonparametric active contour model in [8] is driven by the disparity of the foreground and background p.d.f.'s, which are approximated by Parzen window density method. In [9] the dynamic segmentation of video image sequences is obtained by minimizing the disparity of the p.d.f. of the current frame with the previous one and the p.d.f.'s are also estimated using Parzen window method. The variational segmentation model in [10] incorporates boundary information with region information, where the boundary information is obtained from the edge map image and the interior region information is represented by the intensity p.d.f. captured using Parzen window density estimation. Mory et al. [11] regard the foreground and background p.d.f.'s to be unknown, which are integrated in the region computation model. The proposed model could simultaneously perform segmentation and nonparametric density estimation, which are updated using the Parzen window density method. [12] regards the foreground and background cumulative distribution function (c.d.f.) as unknown and utilizes the Wasserstein distance to measure the disparity of local c.d.f. with the estimated c.d.f.'s. A work closely related to this paper is [13], in which Kim et al. segment images through maximizing mutual information between the image to be segmented and its corresponding label image defined by setting different constants as image intensities of partitioned subregions (ref. Sect. 2), which turns out to be minimizing the displacement of the Logarithmic of the foreground and background p.d.f.'s and the p.d.f.'s are again estimated using Parzen window method.

Borrowing the idea from [13], in this paper we propose a new approach of nonparametric image segmentation that uses Rényi's statistical dependence measure, maximum correlation coefficient, as a similarity measure of two images in different modalities. By using this measure as an alternative choice of dependence measure to mutual information, we do not need to estimate the continuous joint probability density function of two images, which is sensitive to image quantization and make the optimization process complicated. Moreover, the computation is further simplified by applying the theory of reproducing kernel Hilbert spaces.

Before going further, we want to mention some recent work related to joint segmentation and registration. Yezzi et al. [14] provide a very interesting work on simultaneously image segmentation and registration. In their work, $I:\Omega\to\mathbb{R}$ and $\hat{I}:\hat{\Omega}\to\mathbb{R}$ are the two images to be registered and segmented. The goal is to find a curve $C\subset\Omega$ and $g:\mathbb{R}^2\to\mathbb{R}^2$ such that C and $\hat{C}=g(C)$ give correct segmentation for I and $\hat{I}$ simultaneously. In Le Guyader et al.'s work [15], they assume the template image T has already been segmented by the zero level set of Φ_0 and the reference image R is the image to be segmented. The segmentation is then obtained by looking for a displacement field u such that the zero level set of $\Phi_0(x+u(x))$ fits the boundary of the reference image R. Meanwhile, they use a nonlinear-elasticity based smoother to ensure the smoothness of the displacement field u.

Another interesting work we would like to mention is the diffeomorphic active contours by Arrate et al. [16]. They use a geometric flow approach to segment three-dimensional medical images and use the reproducing kernels to represent the shape gradient.

The remainder of this paper is organized as follows: In Sect. 2, we introduce our problem and give a brief review of the mutual information based nonparametric image segmentation approach; Sect. 3 contains some background information about Rényi's statistical dependence measure and reproducing kernel Hilbert space associated with Gaussian kernels; We propose our approach and numerical schemes in Sect. 4; the numerical experimental results are presented in Sect. 5 and we conclude this paper in Sect. 6.

2 Problem Statement and Related Works

Let Ω be a bounded Lipschitz domain, $I:\Omega\to\mathbb{R}$ be a given image and C be an arbitrary curve in the domain Ω. The segmentation problem is to move C such that it separates the foreground from the background. To do so we generate a binary image L corresponding to the position of C in

the following way

$$L(x) = \begin{cases} F & \text{if } x \in R; \\ B & \text{if } x \in R^c. \end{cases} \tag{5}$$

R and R^c denote the region inside and outside C respectively.

We want to mention that the label image L in this setting changes as the curve C evolves. When C reaches the position of the right segmentation, the intensities inside C and outside C have different statistics, for instance, different means or/and variances. Obviously, the intensities inside C and outside C for the label image L are always two different constants. Therefore, when C provides a good segmentation, the image I and label image L should be better matched statistically. This is the basic idea of using matching I and L to assist segmentation.

However, it is usually not easy to match these two images as they are of different modalities and it does not make sense to directly compare their intensities. To cope with this difficulty, a number of similarity measures based on statistical dependence have been proposed. For instance, [13] chooses to maximize the mutual information between original image I and its label image L, together with a constraint of a length term, i.e.,

$$E(C) = -MI(I, L) + \lambda \oint_C \mathrm{d}s. \tag{6}$$

In (6), I and L are viewed as random variables. At each point $x \in \Omega$, the image intensity $I(x)$ (or $L(x)$) is a sample drawn from the random variable I (or L) and all the samples $\{I(x)|x \in \Omega\}$ (or $\{L(x)|x \in \Omega\}$) are assumed to be independent. The mutual information $MI(I, L)$ is defined as follows:

$$\begin{aligned} MI(I, L) &= h(I) - h(I|L) \\ &= h(I) - Pr(L = F)h(I|L = F) \\ &\quad - Pr(L = B)h(I|L = B), \end{aligned} \tag{7}$$

where the entropy of a continuous random variable Z is

$$h(Z) = -\int_{R^N} p_Z(z) \log p_Z(z) \mathrm{d}z. \tag{8}$$

Since $h(I)$ is independent of the curve C, we only need to estimate $h(I|L = F)$ and $h(I|L = B)$, which are estimated by using the nonparametric Parzen window density strategy, i.e.,

$$h(I|L = F) \approx -\frac{1}{|R|} \int_R \log\left(\frac{1}{|R|} \int_R K\big(I(x) - I(\hat{x})\big) \mathrm{d}\hat{x} \right) \mathrm{d}x, \tag{9}$$

$$h(I|L = B) \approx -\frac{1}{|R^c|} \int_{R^c} \log\left(\frac{1}{|R^c|} \int_{R^c} K\big(I(x) - I(\hat{x})\big) \mathrm{d}\hat{x} \right) \mathrm{d}x. \tag{10}$$

In the above equations, $K(\cdot)$ is the so-called window function or kernel, which is symmetric, vanishing at infinity and satisfyies

$$\int_{\mathbb{R}^2} K(s) \mathrm{d}s = 1. \tag{11}$$

For instance, we can choose K to be the Gaussian p.d.f., i.e.,

$$K(s) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{s^2}{2\sigma^2} \right), \tag{12}$$

which is the Parzen-window density estimation kernel. Note that this kernel is infinitely differentiable and thus lending the same property to the estimated p.d.f.

The mutual information could be effectively used as a similarity measure to match the image to be segmented and its label image. However, it requires to estimate the joint p.d.f. of I and L, which is sensitive to image quantization and increases the complexity of computation. In this work, we choose to use Rényi's statistical measure, maximum cross correlation, as a similarity measure. This measure deals directly with samples and does not need to estimate the continuous joint p.d.f. A brief review of this measure is provided in the next section.

3 Rényi's Statistical Measure

In [17] Rényi proposed a set of postulates for a suitable dependence measure Q of two random variables/vectors X and Y, which has drawn much attention ever since. These postulates include

1. $Q(P_{X,Y})$ is well-defined;
2. $0 \le Q(P_{X,Y}) \le 1$;
3. $Q(P_{X,Y}) = 0$ if and only if X, Y are independent;
4. $Q(P_{X,Y}) = 1$ if $Y = f(X)$ or $X = g(Y)$, where f and g are Borel measurable functions.

Then Rényi showed that one measure satisfying these conditions is

$$Q(P_{X,Y}) = \sup_{f,g \in V} CC\big(f(X), g(Y)\big), \tag{13}$$

where V is the space of all Borel measurable functions with finite positive variance, and $CC(f(X), g(Y))$ is the correlation coefficient of $f(X)$ and $g(Y)$, i.e.,

$$CC\big(f(X), g(Y)\big) = \frac{\mathrm{cov}(f(X), g(Y))}{\sqrt{\mathrm{var}(f(X))}\sqrt{\mathrm{var}(g(Y))}}. \tag{14}$$

In the above formula, the covariance between $f(X)$ and $g(Y)$ is defined as

$$\operatorname{cov}\big(f(X), g(Y)\big) = \mathrm{E}\big[\big(f(X) - E\big[f(X)\big]\big)\big(g(Y) - E\big[g(Y)\big]\big)\big], \tag{15}$$

where $E[f(X)]$ is the expectation value of $f(X)$.

The difficulty of using Rényi's measure lies in the fact that we need to find the optimal f and g in the space V, which is the set of all Borel measurable functions with finite positive variance. It is extremely difficult to search f and g in such a huge space. Fortunately, we have shown in [18] that the supremum in V could be attained in a much smaller space, which is a reproducing kernel Hilbert space (RKHS) associated with a reproducing kernel that is continuous, symmetric, positive definite and vanishing at infinity. For completeness, we include the RKHS theory in the Appendix. In this work, we choose the Gaussian function to be the reproducing kernel, i.e.,

$$K(x, y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-y)^2}{2\sigma^2}\right). \tag{16}$$

According to the theory of RKHS (see [18] or Appendix), any two functions f and g in the RKHS associated with the Gaussian kernel can be approximated by functions p and q of the form,

$$p(x) = \sum_{i=1}^{n} \frac{\alpha_i}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-y_i)^2}{2\sigma^2}\right), \tag{17}$$

and

$$q(x) = \sum_{j=1}^{m} \frac{\beta_j}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-z_j)^2}{2\sigma^2}\right), \tag{18}$$

for some parameters σ, y_i, α_i, z_j, β_j, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$. In practice, we can choose f and g to be of the above form and fix σ, y_i, z_j. Therefore we only need to estimate the coefficients α_i and β_j, which could significantly simplify the computation.

4 Proposed Model and Numerical Method

In this section, we propose our model and corresponding numerical schemes. Our aim is to find a curve C such that the resulted label image L defined in (5) matches the best with the original image I. In this work we use the Rényi's statistical dependence measure, maximum correlation coefficient, as a similarity measure to align I and L. By using this measure we don't need to estimate the continuous joint p.d.f. of the two images as in the models based on mutual information.

As shown in the previous postulates, when the image I and its label image L are functions of each other, i.e., $L = f(I)$ or $I = g(L)$ for some function f or g, the maximum cross correlation attains its maximum value 1. Note that L is piecewise constant, so is the resulting image $g(L)$. It does not make a big difference by maximizing the cross correlation between $f(I)$ and L or $g(L)$, so in the following we choose to maximize the cross correlation between $f(I)$ and the label image L. The objective energy functional is obtained by combining the cross correlation of $f(I)$ and L and the length of C, i.e.,

$$E(C, a_1, \ldots, a_n) = \oint_C \mathrm{d}s + \frac{\lambda}{2}\big(1 - CC\big(f(I), L\big)\big)^2, \tag{19}$$

where

$$f\big(I(x)\big) = \sum_{i}^{n} a_i K(I, y_i), \tag{20}$$

and the corresponding label image $L(x)$ is defined by

$$L(x) = \begin{cases} c_1 & \text{if } x \in R; \\ c_2 & \text{if } x \in R^c. \end{cases} \tag{21}$$

4.1 Level Set Formulation and Numerical Method

Energy functional (19) can be minimized using the level set approach [2, 19–21]. The curve C is represented by the zero level of a Lipschitz function $\phi : \Omega \to \mathbb{R}$ and the resulting energy functional becomes

$$E(\phi, a_1, \ldots, a_n) = \int_{\Omega} \big|\nabla H\big(\phi(x)\big)\big|\mathrm{d}x + \frac{\lambda}{2}\big(1 - CC\big(f(I), L\big)\big)^2, \tag{22}$$

where H is the Heaviside function and

$$L(x) = c_1 H\big(\phi(x)\big) + c_2\big(1 - H\big(\phi(x)\big)\big). \tag{23}$$

The alternate minimization (AM) approach [22] is employed to solve this problem. First, we keep $a_1, \ldots, a_n$ fixed and solve for ϕ using the gradient descent approach, i.e.,

$$\frac{\partial \phi}{\partial t} = \delta(\phi)\left[\operatorname{div}\left(\frac{\nabla \phi}{|\nabla \phi|}\right) + \lambda\big(1 - CC\big(f(I), L\big)\big)F\right], \tag{24}$$

where δ is the regularized Dirac function and

$$F = \frac{(f(I) - \overline{f(I)})\operatorname{var}(L) - \operatorname{cov}(f(I), L)(L - \overline{L})}{\operatorname{var}(f(I))^{\frac{1}{2}}\operatorname{var}(L)^{\frac{3}{2}}} \times (c_1 - c_2). \tag{25}$$

In (24), $\delta(\phi)\operatorname{div}(\frac{\nabla \phi}{|\nabla \phi|})$ is the first variation of the smoothing term (TV-term) in (22) with respect to ϕ. In the same way, $\delta(\phi)\lambda(1 - CC(f(I), L))F$ is the first variation of the fidelity term in (22) with respect to ϕ.

Note that $\delta(\phi)F$ is the first variation of $CC(f(I),L)$ with respect to ϕ and this derivation can be obtained by applying the chain rule, that is, the first variation of $CC(f(I),L)$ with respect to L times the first variation of L with respect to ϕ, i.e.,

$$\delta(\phi)F=\frac{\mathrm{d}CC(f(I),L)}{\mathrm{d}\phi}=\frac{\mathrm{d}CC(f(I),L)}{\mathrm{d}L}\cdot\frac{\mathrm{d}L}{\mathrm{d}\phi}. \tag{26}$$

Note that

$$\begin{aligned}CC\big(f(I),L\big)&=\frac{\operatorname{cov}(f(I),L)}{\sqrt{\operatorname{var}(f(I))}\sqrt{\operatorname{var}(L)}}\\&=\frac{\mathrm{E}[(f(I)-\overline{f(I)})(L-\overline{L})]}{\sqrt{\mathrm{E}[(f(I)-\overline{f(I)})^2]}\sqrt{\mathrm{E}[(L-\overline{L})^2]}},\end{aligned} \tag{27}$$

thus

$$\begin{aligned}&\frac{\mathrm{d}CC(f(I),L)}{\mathrm{d}L}\\&=\frac{(f(I)-\overline{f(I)})\operatorname{var}(L)-\operatorname{cov}(f(I),L)(L-\overline{L})}{\operatorname{var}(f(I))^{\frac{1}{2}}\operatorname{var}(L)^{\frac{3}{2}}}.\end{aligned} \tag{28}$$

On the other hand, $\frac{\mathrm{d}L}{\mathrm{d}\phi}=\delta(\phi)(c_1-c_2)$. Therefore, we can get (25) by applying (26).

Then we keep ϕ fixed and minimize (22) with respect to a_i's. In this paper, we present two algorithms to update a_i's. The first one is to use the gradient descent method

$$\frac{\partial a_i}{\partial t}=\big(1-CC\big(f(I),L\big)\big)E. \tag{29}$$

In the above equation, E is the first variation of $CC(f(I),L)$ with respect to a_i, i.e.,

$$E=\frac{\operatorname{cov}(p_i,L)\operatorname{var}(f(I))-\operatorname{cov}(f(I),L)\operatorname{cov}(f(I),p_i)}{\operatorname{var}(f(I))^{\frac{3}{2}}\operatorname{var}(L)^{\frac{1}{2}}}, \tag{30}$$

where

$$p_i=K(I,y_i). \tag{31}$$

The second method is based on the formulation of the energy functional. To minimize (22) with respect to a_i's is equivalent to maximizing $CC(f(I),L)$ with respect to a_i's as the length term is independent of a_i's.

For simplicity, we first introduce some notations. The image I and L are viewed as vectors of length N, where N is the number of pixels. For each $i=1,2,\ldots,n$, set $p_i=K(I,y_i)$ be a vector of length N, and define $P=[p_1,p_2,\ldots,p_n]$, $a=[a_1,a_2,\ldots,a_n]$. For any vector x, we denote the mean of x to be $\overline{x}$. Thus

$$f(I)=\sum_i^n a_iK(I,y_i)=\sum_{i=1}^n a_ip_i=Pa, \tag{32}$$

and

$$CC\big(f(I),L\big)=\frac{\langle Pa-\overline{Pa},L-\overline{L}\rangle}{|Pa-\overline{Pa}||L-\overline{L}|}=\frac{\langle P_0a,L_0\rangle}{|P_0a||L_0|}, \tag{33}$$

where $P_{0ij}=P_{ij}-\frac{1}{N}\sum_{k=1}^N P_{k,j}$ and $L_0=L-\overline{L}$.

Let $\hat{P}$ be a matrix with orthonormal column vectors which spans the column space of P_0, then we have $P_0a=\hat{P}\alpha$ for some vector α. Under these formulations, we get

$$\begin{aligned}CC\big(f(I),L\big)&=\frac{\langle P_0a,L_0\rangle}{|P_0a||L_0|}=\frac{\langle\hat{P}\alpha,L_0\rangle}{|\hat{P}\alpha||L_0|}=\frac{\alpha^T\hat{P}^TL_0}{|\alpha||L_0|}\\&=\frac{\alpha^TU\Lambda V^TL_0}{|\alpha||L_0|}=\frac{(U^T\alpha)^T\Lambda(V^TL_0)}{|U^T\alpha||V^TL_0|}.\end{aligned} \tag{34}$$

In the above equation, $U\Lambda V^T$ is the singular value decomposition of the matrix $\hat{P}^T$, where U is an $n\times n$ unitary matrix, V^T is an $N\times N$ unitary matrix and Λ is an $n\times N$ diagonal matrix with nonnegative diagonal entries $\lambda_1,\lambda_2,\ldots,\lambda_n$, which are listed in decreasing order.

Now let $x=\frac{U^T\alpha}{|U^T\alpha|}$, $y=\frac{V^TL_0}{|V^TL_0|}$, then $x\in\mathbb{R}^n$, $y\in\mathbb{R}^N$, $|x|=|y|=1$, thus

$$CC(f(I),L)=x^T\Lambda y=x^T\Theta\Theta z=\langle\Theta x,\Theta z\rangle,$$

where $z=[y_1,y_2,\ldots,y_n]$ and Θ is an $n\times n$ diagonal matrix with diagonal entries $\sqrt{\lambda_1},\sqrt{\lambda_2},\ldots,\sqrt{\lambda_n}$.

Therefore, $CC(f(I),L)$ is maximized when Θx and Θz have the same direction. Note that $|x|=1$, we get $x=\frac{\Theta^+(\Theta z)}{|\Theta^+(\Theta z)|}$, where Θ^+ refers to the pseudoinverse of Θ. From the definition of x, we can pick a particular $\alpha=Ux$ and then we can update a from the relation $P_0a=\hat{P}\alpha$.

In this work, ϕ and a_i's are alternatively updated until we reach a satisfactory result. To increase the rate of convergence, the semi-implicit difference scheme is applied in (24), i.e.,

$$\begin{aligned}&\frac{\phi^{n+1}-\phi^n}{\Delta t}\\&=\delta\big(\phi^n\big)\left[\operatorname{div}\left(\frac{\nabla\phi^{n+1}}{|\nabla\phi^n|}\right)+\lambda\big(1-CC\big(f(I)^n,L^n\big)\big)F^n\right].\end{aligned} \tag{35}$$

This equation could be effectively solved by using the additive operator splitting (AOS) method. Note that the curvature term in (24) is approximated by

$$\frac{\partial}{\partial x}\left(\frac{\phi_x}{\sqrt{\phi_x^2+\phi_y^2+\varepsilon^2}}\right)+\frac{\partial}{\partial y}\left(\frac{\phi_y}{\sqrt{\phi_x^2+\phi_y^2+\varepsilon^2}}\right), \tag{36}$$

where ε is a small positive number in case that the denominators become zero. However, it may still cause stability issues and limit the convergence rate.

4.2 A Soft Formulation and Numerical Method

To avoid local minimum problem we propose a soft formulation of the energy functional (19) by using the same strategy in [23], and use Chambolle's dual method [24, 25] to solve it.

Let $u : \Omega \to [0, 1]$ be a fuzzy membership function and rewrite the energy functional as

$$E(u, a_1, \ldots, a_n) = \int_\Omega \big|\nabla u(x)\big|\mathrm{d}x + \frac{\lambda}{2}\big(1 - CC\big(f(I), L\big)\big)^2, \tag{37}$$

where

$$L(x) = c_1 u(x) + c_2\big(1 - u(x)\big). \tag{38}$$

In real applications, we want L be a binary image so that it could give a reasonable segmentation. So during each iteration, we reset

$$L(x) = c_1 \chi(u > 0.5) + c_2\big(1 - \chi(u > 0.5)\big), \tag{39}$$

and χ refers to the characteristic function.

Following the strategy in [24, 25], we introduce an auxiliary variable $v : \Omega \to [0, 1]$ and consider the following approximated energy functional

$$\begin{aligned} &E(u, v, a_1, a_2, \ldots, a_n) \\ &\quad= \int_\Omega \big|\nabla u(x)\big|\mathrm{d}x \\ &\qquad+ \frac{1}{2\theta} \parallel u - v \parallel^2 + \lambda\big(1 - CC\big(f(I), L\big)\big)^2, \end{aligned} \tag{40}$$

where

$$L(x) = c_1 v(x) + c_2\big(1 - v(x)\big), \tag{41}$$

and θ is chosen to be small enough such that the minimizers $u^\star$ and $v^\star$ are close to each other.

We still employ the alternate minimization (AM) approach to solve this minimization problem, i.e., we go on to alternatively solve the following two problems:

$$\min_u \int_\Omega \big|\nabla u(x)\big|\mathrm{d}x + \frac{1}{2\theta} \parallel u - v \parallel^2 \tag{42}$$

and

$$\min_{0 \le v \le 1, a_i\text{'s}} \frac{1}{2\theta} \parallel u - v \parallel^2 + \lambda\big(1 - CC\big(f(I), L\big)\big)^2. \tag{43}$$

The minimization problem (42) could be effectively solved by applying Chambolle's method [24] and the solution is

$$u(x) = v(x) - \theta \mathrm{div} p(x), \tag{44}$$

where $p = (p^1, p^2)$ is given by

$$-\nabla(\theta \mathrm{div} p - v) + \big|\nabla(\theta \mathrm{div} p - v)\big| p = 0. \tag{45}$$

Equation (45) could be solved by a fixed point method, i.e.,

$$p^{n+1} = \frac{p^n + \tau \nabla(\mathrm{div} p^n - v/\theta)}{1 + \tau \nabla|\mathrm{div} p^n - v/\theta|}. \tag{46}$$

Following the same strategy in [25], the solution v of (43) is given by

$$v = \min\big(\max(u + \lambda\theta G, 0), 1\big), \tag{47}$$

where

$$\begin{aligned} G = &\big(1 - CC\big(f(I), L\big)\big) \cdot (c_1 - c_2) \\ &\times \frac{(f(I) - \overline{f(I)})\mathrm{var}(L) - \mathrm{cov}(f(I), L)(L - L)}{\mathrm{var}(f(I))^{\frac{1}{2}} \mathrm{var}(L)^{\frac{3}{2}}}. \end{aligned} \tag{48}$$

a_i's are only included in the cross correlation term, so the optimization scheme is exactly the same as the level set approach.

At the end of this section, we want to mention that both of these two methods (level set method and the soft segmentation by using Chambolle's method) are quite effective. However, it is still possible that we may result in a local minimum. This is due to the non-convexity of the cross correlation term in the energy functional (19). Therefore, (19) may have more than one minimizer and we cannot guarantee this approach converges to a global minimizer.

5 Experimental Results

In this section we show our experimental results on various images to demonstrate the performance of the proposed model for segmentation. All the simulations are preformed in Matlab 7.9 (R2009b) on a PC with an Intel Core 2 Duo CPU at 2.4 GHz and 3 GB RAM.

We compare the proposed nonparametric model with two parametric models, namely, the piecewise constant Mumford-Shah model (2) and also the parametric Gaussian model (3), (4). For completeness, we rewrite these models as the following

$$\min_{C, c_1, c_2} \int_{\Omega_1} (c_1 - I)^2 \mathrm{d}x + \int_{\Omega_2} (c_2 - I)^2 \mathrm{d}x + \beta|C|, \tag{49}$$

and

$$\min_{C, c_1, c_2, \sigma_1, \sigma_2} \sum_{i=1}^{2} \int_{\Omega_i} \left(\frac{(I - c_i)^2}{2\sigma_i^2} + \log \sigma_i \right) \mathrm{d}x + \beta|C|. \tag{50}$$

Table 1 Number of iterations and CPU time (s) for the experiments (Figs. 1, 2, and 3)

Images	Size	Iterations			CPU time (s)		
		Model (49)	Model (50)	Proposed	Model (49)	Model (50)	Proposed
Fig. 1(a)	100×100	50	63	35	2.93	2.61	2.31
Fig. 1(e)	100×100	176	182	83	4.48	7.21	3.81
Fig. 1(i)	256×256	86	90	58	7.07	9.16	5.90
Fig. 2(a)	210×180	–	98	20	–	4.63	1.40
Fig. 2(f)	210×180	–	100	50	–	7.14	3.24
Fig. 3(a)	336×406	–	–	20	–	–	5.89
Fig. 3(e)	336×406	–	–	33	–	–	7.34

We apply the same soft formulation to the above two models and compare the results with those got from the proposed model. The numerical algorithm would be terminated once $\|u^{k+1} - u^k\|/\|u^k\| < 10^{-5}$, where u^k refers to the level set function or the membership function during the kth iteration. The corresponding number of iterations and CPU time for all the following experiments are summarized in Table 1. From this table, we can see that compared to the parametric models, the proposed nonparametric model needs less time and iterations to obtain good results.

For the level set approach, the initial ϕ is set to be the signed distance function of the initial circles in the image. However, Chambolle's approach is more robust to the initialization. The initial u and v can be generated as random fields in the range $[0, 1]$. All of the test images are rescaled to the interval $[0, 1]$ and the parameters c_1, c_2 in the label image are fixed to be 1 and 2. Unless otherwise stated, in each figure, we include the test image I, the final transformed image $f(I)$, the final label image L and the segmentation result, i.e., the contours ($\phi = 0$ or $u = 0.5$) superimposed on the original image.

5.1 Comparison with Two Parametric Models

The purpose of Experiment I (Fig. 1) is to test the ability of the proposed model on three synthetic images, where the foreground and foreground are generated by two different distributions. Meanwhile, we also compare the segmentation results with two parametric models (49), (50).

In each row, from left to right, we present the original synthetic test image; corresponding boundary overlaid on the image obtained from the piecewise constant Mumford-Shah model (49), the parametric Gaussian model (50), the proposed model; the distributions of the foreground and background.

In the first test image (a), the intensities of foreground and background are drawn from two Gaussian distributions with different mean and the same variance (e). While the intensities of the foreground and background in the second test image (f) share the same mean and different variance (j). The last test image (k) shows a unimodal Gaussian foreground over a bi-modal Gaussian background (o).

We can easily see that all the above-mentioned three models work effectively for the first test image. However, the piecewise constant Mumford-Shah model fails for the second and third test images as the piecewise constant assumption does not hold for these two cases. As for the parametric Gaussian model, it could automatically estimate the mean and variance between and outside the curve and then use these information to aid the segmentation. Therefore, the result for the second test image (f) is satisfactory even if the image intensities overlaps. However, it fails for the third test image (k) as the background is bi-modal Gaussian distributed, which makes the assumption for the parametric Gaussian model do not hold.

Our proposed nonparametric model successfully separates the foreground from the background in all these three cases. The image intensities of object and background in the first test image vary a lot in the first test image (a) and it would be easily segmented. Regarding the last two images, the background or the foreground is comparatively homogeneous. Hence, we could select y_i's from the intervals where the intensities of the background or the background lie. Thus after applying the function f, linear combination of a series Gaussian functions centered at y_i's, to the image I, the values of $f(I)$ in the inhomogeneous region become extremely small and the transformed images $f(I)$ are more homogenous. Therefore, by carefully choosing y_i's in the function f and maximizing the correlation coefficient of the transformed images $f(I)$ with the label image L, we get the desirable results.

The comparison with these two models indicates that the parametric models fails when the image intensity distribution is more complicated. That is why we need to explore for more general nonparametric models.

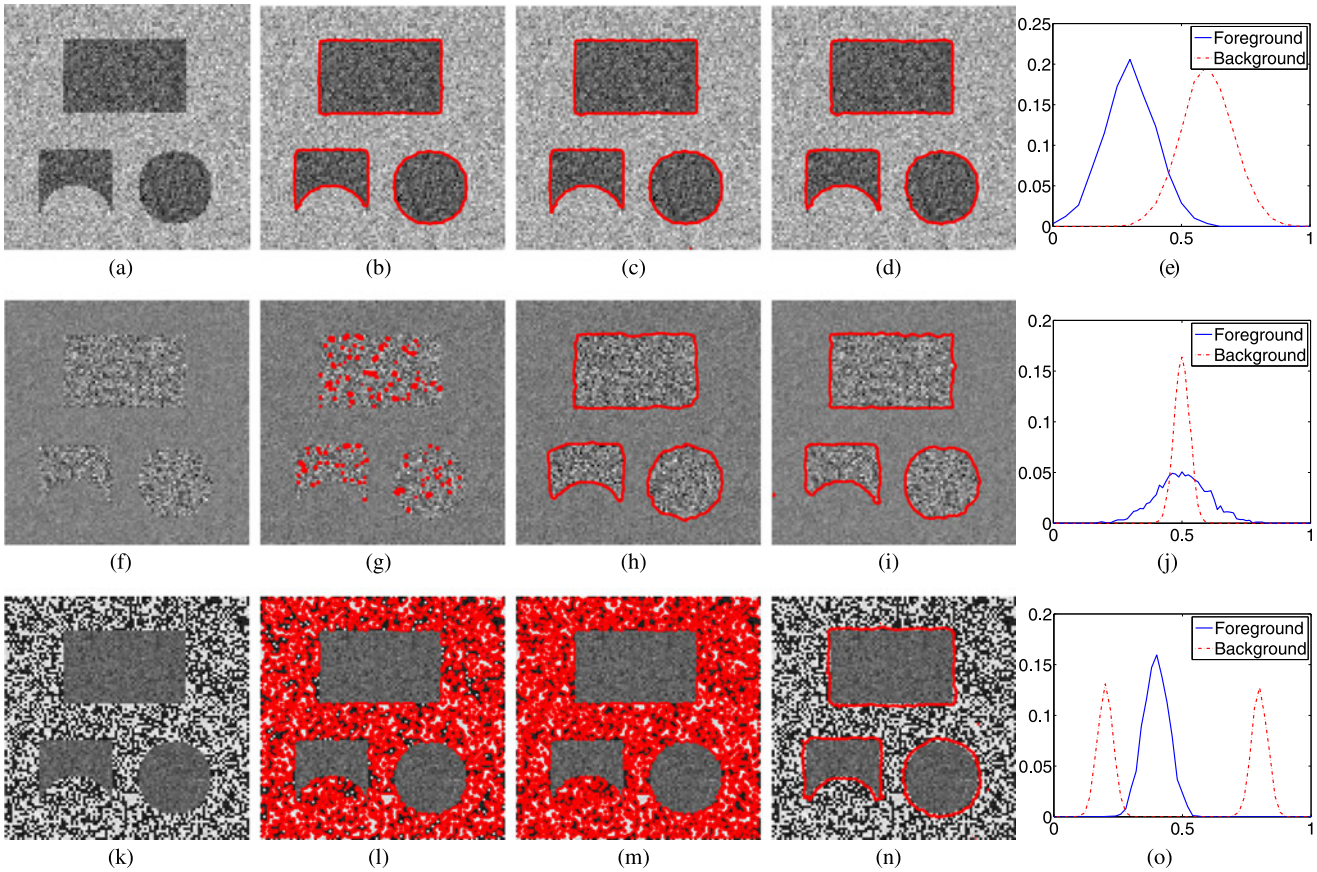

Fig. 1 Segmentation results of three synthetic images. (**a**), (**f**), (**k**) Test images I; (**b**), (**g**), (**l**) Segmentation results of the piecewise constant Mumford-Shah model; (**c**), (**h**), (**m**) Results of the parametric Gaussian model; (**d**), (**i**), (**n**) Results of the proposed model; (**e**), (**j**), (**o**) Distributions of the foreground and background

5.2 Test on Real Medical Images

In this subsection, we test the proposed model on two real medical images, an MRI brain image and a lung image.

In experiment 2 (Fig. 2), the first test image (a) is a clean brain image and the second test image (f) is generated by adding Gaussian noise with zero mean and variance 0.1 to the clean image (a). Meanwhile, we also compare the results with those obtained from the parametric Gaussian model (50).

The segmentation results of the parametric Gaussian model (50) are placed at the end of each row (e), (j). We can see that it successfully separates background, cerebrospinal fluid (c.s.f.) from white matter and gray matter. However, this is quite different from the results (d), (i) obtained from the proposed model. From (d), (i), we can see that gray matter is separated from the rest and the whole image is actually segmented into three parts even if we only do the two phase segmentation.

This is reasonable because we utilize the histogram information (k) of the test image. As indicated in the graph (k), there are three peaks, which from left to right stand for background and c.s.f., gray matter, white matter. Note that background and c.s.f. are considered as a whole and the intensities of the gray matter lie in the interval $[0.4, 0.6]$. We uniformly select y_i's from the interval $[0.4, 0.6]$, then after applying the function f to the test images I, intensities which are not in this interval (white matter, background and c.s.f.), would become almost zero while intensities of the gray matter are enlarged. That is why white matter, c.s.f. and background look dark while the gray matter looks bright in the transformed images $f(I)$ (b), (g). In other words, by doing this transformation, we can view white matter, c.s.f. and background as a whole and separate them from the gray matter.

The last image (l) shows that the cross correlation between the image $f(I)$ and the label image L, i.e., $CC(f(I), L)$, keeps increasing as the iteration process goes. And this trend coincides with the mutual information between the test image I and the label image L. So we can conclude that the proposed method is consistent with the mutual information based nonparametric image segmentation method [13].

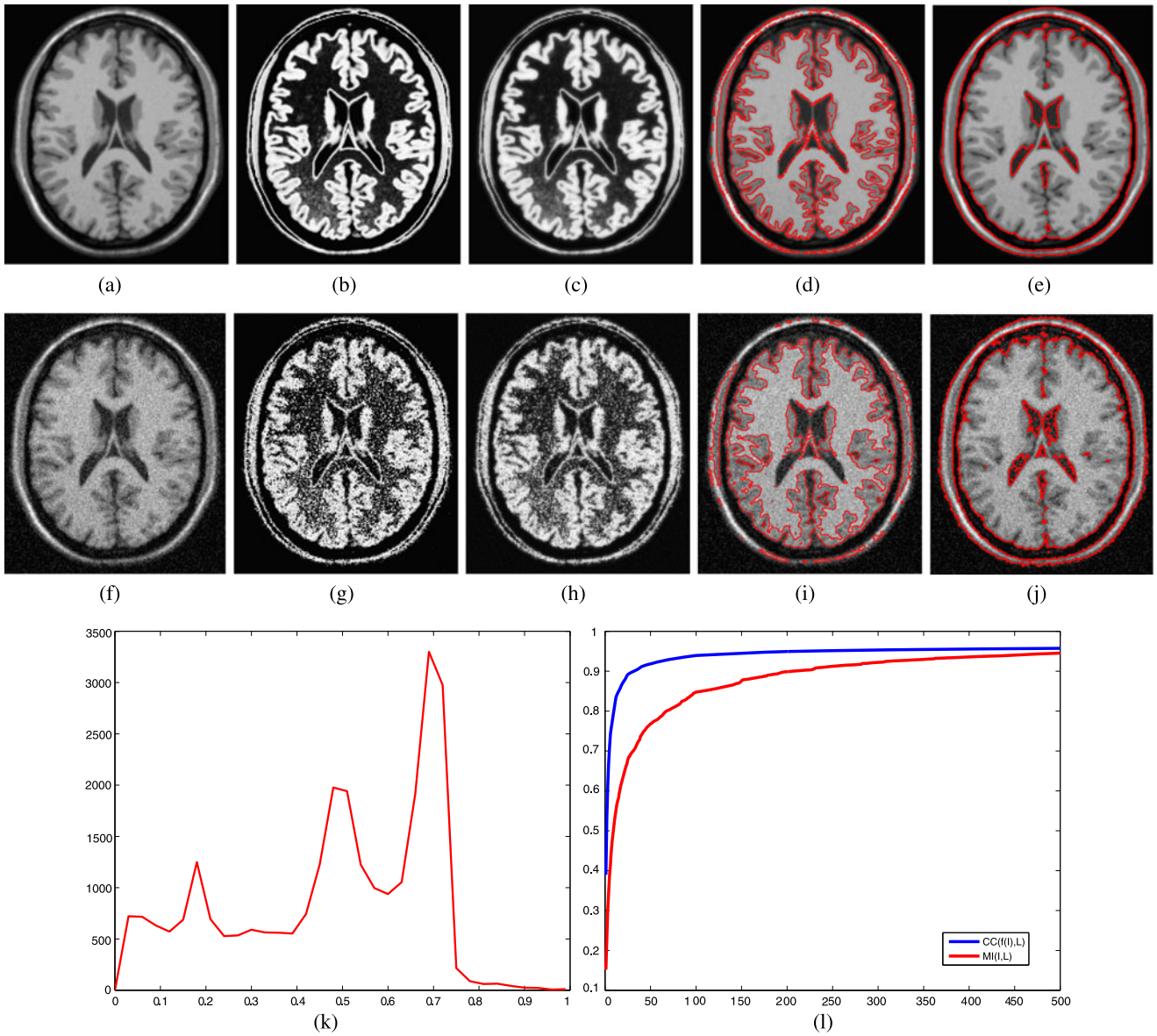

Fig. 2 Segmentation results of a clean brain image (*first row*) and its noisy version (*second row*). *From left to right*: Test image I (**a**), (**f**); Final transformed images $f(I)$ (**b**), (**g**); Final label images L (**c**), (**h**); Segmentation results of the proposed model (**d**), (**i**); Segmentation results obtained from the parametric Gaussian model (**e**), (**j**). *Third row*: (**k**) Histogram of the test image (**a**); (**l**) the information of $CC(f(I), L)$ and $MI(I, L)$ during each iteration

Experiments 1 and 2 indicates that by choosing specific y_i's, the proposed model works well for images with inhomogeneity, unevenly distributed illumination and it can get multiphase segmentation results while only using two phases. In the following experiment, we do not pay too much attention on the selection of y_i's and let them to be equally spaced in the interval $[0, 0.5]$.

Experiment 3 (Fig. 3) aims to test whether the proposed model works for images with fine structures. We choose the test image to be a lung image with lots of fine details. This first test image (a) is a clean image and the second one (e) is more inhomogeneous, which is generated by adding Gaussian noise with zero mean and variance 0.1 to the clean image. After applying the function f to the original test images I (a), (e), the resulted images $f(I)$ (b), (f) have more strong contrast between different features while still preserving the detailed structures. The same parameters are applied for these two tests. The final results (d), (h) show that most of the fine structures are captured and the noise inhomogeneity does not exert a big difference.

6 Conclusion

In this paper, we propose a novel image segmentation framework based on Rényi's statistical dependence measure. The

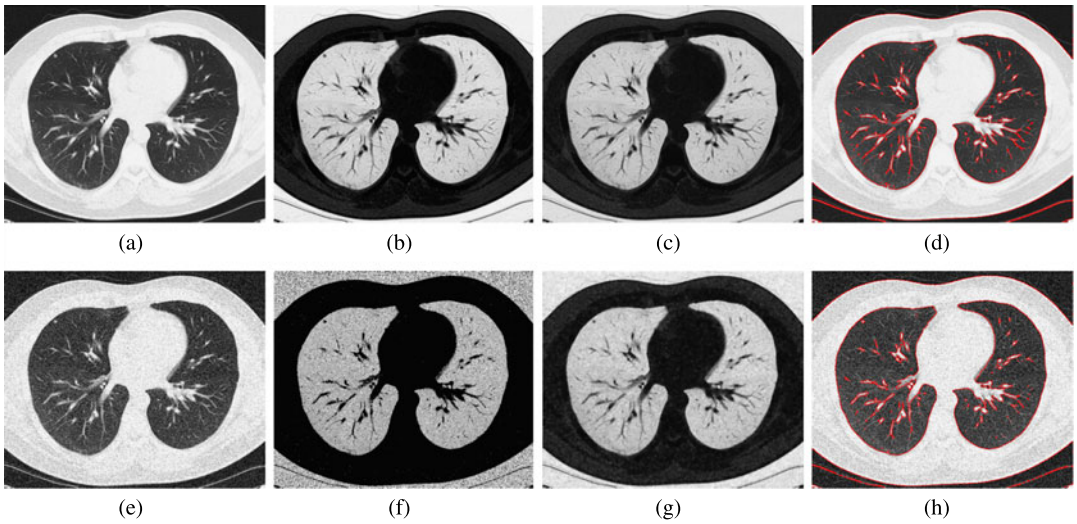

Fig. 3 Segmentation results of a clean lung image (*first row*) and a noisy one (*second row*). *From left to right*: Test images I (**a**), (**e**), transformed images $f(I)$ (**b**), (**f**), label images L (**c**), (**g**) and the final contour ($u = 0.5$) superimposed on the test images (**d**), (**h**)

computation is greatly simplified by applying the theory of reproducing kernel Hilbert space. Two numerical approaches, the level set method and Chambolle's dual approach, are employed during the implementation. Finally, the proposed model is applied to different kinds of images and gets satisfactory results.

Appendix

In this section, we give the basic theories of RKHS we have used in this paper.

Let E be an arbitrary set and H be a Hilbert space of real functions on E. We say that H is a RKHS if the linear map $F_x : f \to f(x)$ is a bounded functional for any $x \in E$.

By this definition, $F_x \in H^\star$, which is the dual space of H. Therefore, Reisz representation theorem shows that there exists a unique $K_x \in H$, such that

$$f(x) = \langle F_x, f\rangle = \langle K_x, f\rangle, \quad \forall f \in H.$$

Define $K : E \times E \to \mathbb{R}$ as $K(x, y) = \langle K_x(\cdot), K_y(\cdot)\rangle$. It is easy to see that K has the following properties:

1. K is symmetric: $K(x, y) = K(y, x)$.
2. Reproducing property: $f(x) = \langle K(x, \cdot), f(\cdot)\rangle$.
3. K is positive definite: $\sum_{i,j}^{n} a_i a_j K(x_i, x_j) \geq 0$ holds for all $x_1, x_2, \ldots, x_n \in E$, $a_1, a_2, \ldots, a_n \in \mathbb{R}$ and the equality holds if and only if $a_i = 0$, $i = 1, 2, \ldots, n$.

We call such an K the reproducing kernel for the Hilbert space H.

On the other hand, suppose $K : E \times E \to \mathbb{R}$ is symmetric and positive definite, then according to the Moore-Aronszajn theorem [26], there is a unique Hilbert space of functions on E for which K is a reproducing kernel. In fact, let $H_0(E)$ be the linear span of the functions $\{K(x, \cdot)|x \in E\}$ and define the inner product in $H_0(E)$ to be

$$\left\langle \sum_{i=1}^{n} a_i K(x_i, \cdot), \sum_{j=1}^{m} b_j K(y_i, \cdot) \right\rangle = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j K(x_i, x_j).$$

Let $H(E)$ be the completion of $H_0(E)$ with respect to this inner product. It is not difficult to check that $H(E)$ is the unique RKHS with reproducing kernel K.

For the particular case $E = \mathbb{R}$. Let $C_0(\mathbb{R})$ be the space of real valued continuous functions vanishing at infinity with the supremum norm. Then we have the following result

$$\sup_{f,g\in V(\mathbb{R})} CC\big(f(X), g(Y)\big) = \sup_{f,g\in H_0(\mathbb{R})} CC\big(f(X), g(Y)\big), \tag{51}$$

where $V(\mathbb{R})$ is the space of all real Borel measurable functions with finite positive variance. This is the main result we have used in this paper and (17), (18) follows directly from this result.

The proof of this result can be obtained through the following three steps (We omit the details here).

1. $H_0(\mathbb{R})$ is dense in $C_0(\mathbb{R})$.
2. Let $V(B)$ be the space of all real bounded Boreal measurable functions, then

$$\sup_{f,g\in V(B)} CC(f(X), g(Y)) = \sup_{f,g\in C_0(\mathbb{R})} CC(f(X), g(Y)).$$

3.

$$\sup_{f,g\in V(\mathbb{R})} CC(f(X), g(Y)) = \sup_{f,g\in V(B)} CC(f(X), g(Y)).$$

References

1. Mumford, D., Shah, J.: Optimal approximation by piecewise smooth functions and associated variational problems. Commun. Pure Appl. Math. **42**, 577–685 (1989)
2. Chan, T., Vese, L.: Active contours without edges. IEEE Trans. Image Process. **10**(2), 266–277 (2001)
3. Vese, L., Chan, T.: A multiphase level set framework for image segmentation using the Mumford and shah model. Int. J. Comput. Vis. **50**(3), 271–293 (2002)
4. Zhu, S.C., Yuille, A.: Region competition: unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **18**(9), 884–900 (1996)
5. Rousson, M., Deriche, R.: A variational framework for active and adaptive segmentation of vector values images. In: Proceedings of Workshop on Motion and Video Computing, pp. 56–61 (2002)
6. Paragios, N., Deriche, R.: Geodesic active regions and level set methods for supervised texture segmentation. Int. J. Comput. Vis. **46**, 223–247 (2002)
7. Duda, R., Hart, P.: Pattern Classification and Scene Analysis. Wiley, New York (1973)
8. Abd-Almageed, W., Smith, C.E., Ramadan, S.: Kernel snakes: non-parametric active contour models. IEEE Int. Conf. Syst. Man Cybern. **1**, 240–244 (2003)
9. Aubert, A., Barlaud, M., Faugeras, O., Jehan-Besson, S.: Image segmentation using active contours: calculus of variations or shape gradients. SIAM Appl. Math. **63**(6), 2128–2154 (2003)
10. Huang, X., Metaxas, D., Chen, T.: Metamorphs: deformable shape and texture models. In Proceedings of IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, vol. 1, pp. 496–503 (2004)
11. Mory, B., Ardon, R., Thiran, J.-P.: Variational segmentation using fuzzy region competition and local non-parametric probability density functions. In: IEEE International Conference on Computer Vision (ICCV), Rio, Brazil (2007)
12. Ni, K., Bresson, X., Chan, T., Esedoglu, S.: Local histogram based segmentation using the Wasserstein distance. Int. J. Comput. Vis. **84**, 97–111 (2009)
13. Kim, J., Fisher, J., Yezzi, A., Cetin, M., Willsky, A.: A nonparametric statistical method for image segmentation using information theory and curve evolution. IEEE Trans. Image Process. **14**(10), 1486–1502 (2005)
14. Yezzi, A., Zollei, L., Kapur, T.: A variational framework for joint segmentation and registration. In Proc. IEEE MMBIA, pp. 44–51 (2001)
15. Le Guyader, C., Vese, L.: A Combined Segmentation and Registration Framework with a Nonlinear Elasticity Smoother. Lecture Notes in Computer Science, vol. 5567, pp. 600–611 (2009)
16. Arrate, F., Ratnanather, J.T., Younes, L.: Diffeomorphic active contours. SIAM J. Imaging Sci. **3**(2), 176–198 (2010)
17. Rényi, A.: On measure of dependence. Acta Math. Acad. Sci. Hung. **10**, 441–451 (1959)
18. Shi, J., Chen, Y., Rao, M., Lee, J.S.: A statistical similarity measure for non-rigid multi-modal image registration. Proc. SPIE **7623**, 762307 (2010)
19. Osher, S., Fedkiw, R.: Level Set Methods and Dynamic Implicit Surfaces. Applied Mathematica Sciences, vol. 153. Springer, New York (2002)
20. Chan, T., Vese, L.: A level set algorithm for minimizing the Mumford-Shah functional in image processing. In: IEEE/Computer Society Proceedings of the 1st IEEE Workshop on "Variational and Level Set Methods in Computer Vision", pp. 161–168 (2001)
21. Osher, S., Sethian, J.A.: Fronts propagating with curvature dependent speed: algorithms based on Hamilton-Jacobi formulations. J. Comput. Phys. **79**(1), 12–49 (1988)
22. Chan, T., Wong, C.K.: Total variation blind deconvolution. IEEE Trans. Image Process. **7**, 370–375 (1998)
23. Chan, T.F., Esedoglu, S., Nikolova, M.: Algorithms for finding global minimizers of denoising and segmentation models. SIAM J. Appl. Math. **66**, 1632–1648 (2006)
24. Chambolle, A.: An algorithm for total variation minimization and applications. J. Math. Imaging Vis. **20**, 89–97 (2004)
25. Bresson, X., Esedoglu, S., Vandergheynst, P., Thiran, J.-P., Osher, S.: Fast global minimization of the active contour/snake model. J. Math. Imaging Vis. **28**, 151–167 (2007)
26. Aronszajn, N.: Theory of reproducing kernels. Trans. Am. Math. Soc. **68**, 337–404 (1950)

Haili Zhang is a Ph.D. student in the Department of Mathematics at University of Florida, USA. She received her B.E. (2005) in Mathematics from Henan Normal University, China, and M.S. (2008) in Mathematics from Beijing Normal University, China. Her research interests include digital image processing, numerical analysis and partial differential equations.

Yunmei Chen is a professor in the Department of Mathematics at University of Florida, USA. She received her B.E. (1967) in Mathematics from Fudan University, China, M.S. (1981) in Mathematics from Tongji University, China, and Ph.D. (1985) in Mathematics from Fudan University, China. Her research interests include partial differential equations, nonlinear analysis and mathematical methods in image analysis.

Jiangli Shi is a Ph.D. student in the Department of Mathematics at University of Florida, USA. He received his B.E. (2004) and M.S. (2007) in Mathematics from Shanghai Jiao Tong University, China. His research interests include partial differential equations, PDE-based image analysis and medical image processing.