Local Histogram based Segmentation using the Wasserstein Distance

Kangyu Ni · Xavier Bresson · Tony Chan · Selim Esedoglu

Abstract We propose and analyze a nonparametric region-based active contour model for segmenting cluttered scenes. The proposed model is unsupervised and assumes that pixel intensity is independently identically distributed. The proposed energy functional consists of a geometric regularization term that penalizes the length of region boundaries, and a region-based image term that uses the probability density function (or histogram) of pixel intensity to distinguish different regions. More specifically, the region data encourages partitioning the image domain so that the local histograms within each region are approximately homogeneous. The solutions of the proposed model do not need to differentiate histograms. The similarity between normalized histograms is measured by the *Wasserstein distance with exponent 1*, which is able to fairly compare two histograms, both continuous and discontinuous. We employ a fast global minimization method based on [11, 6] to solve the proposed model. The advantages of this method include less computational time compared with the minimization method by gradient descent of the associated Euler-Lagrange equation [12] and the ability to find a global minimizer. Moreover, our proposed model has several desired properties due to the use of the Wasserstein distance. We further propose a variant of the proposed model that addresses local illumination changes in an image.

Keywords Image Segmentation · Unsupervised · Wasserstein Distance · Image Processing and Computer Vision · Nonparametric

1 Introduction

Image segmentation plays an important role in computer vision. The process involves partitioning the image domain into several regions either according to edge information or region information so that the image within each region has uniform characteristics. The characterized regions depend on the application and may include one or more of the following: edges, intensities, textures, and shapes. Snake [21], balloon [10], and geodesic active contours based [7, 20] methods use edge detection functions and evolve contours towards sharp gradients of pixel intensity. This classic active contour approach is widely used in medical image processing. However, it is not robust to noise, and typically a noisy image has to be smoothed. Region-based active contours incorporate region and boundary information and are robust to noise. Furthermore, they are able to detect objects with either sharp or smooth edges. One of the first region-based active contours is the Mumford-Shah segmentation model [24], which approximates an image by a piecewise smooth function, with a length penalizing term. However, this model is difficult to solve in practice. The active contours without edges (ACWE) model [13], a variant of the piecewise constant Mumford-Shah model, approximates an image by a two-phase

This research is supported by ONR grant N00014-06-1-0345 and NSF grant DMS-0610079.

Kangyu Ni
University of California Los Angeles
E-mail: kni66 at math.ucla.edu

Xavier Bresson
University of California Los Angeles
E-mail: xbresson at math.ucla.edu

Tony Chan
University of California Los Angeles
E-mail: chan at math.ucla.edu

Selim Esedoglu
University of Michigan Ann Arbor
E-mail: esedoglu at umich.edu

piecewise constant function and is based on a level-set implementation [26]. The minimizing flow is derived by computing the variation of the energy with respect to the level set function. Region competition [37] is a statistical and variational model that is based on minimizing a generalized Bayes and Minimum description length criterion. The model penalizes the boundary length and the Bayes error within each region, in which appropriate probability distributions are chosen. The ACWE, region competition, and other parametric region-based active contour models, such as [34,27], assume the probability density function (pdf) of the pixel intensity in each region up to a few parameters. For example, often a Gaussian distribution is assumed with mean and variance the only unknowns. However, many natural images are not necessarily described by a specific distribution. Nonparametric region-based active contour models, such as [19,16,4,1], use the full pdf, or histogram, of the intensity to drive the segmentation. Therefore, they do not suffer from the above limitations. Our model is related to, yet different from, previous work. In [1,4], the segmentation model is supervised and the data descriptors directly depend on the regions, which consequently involves histogram differentiation in the evolution equations. Unsupervised segmentation models in [19,16] take an information-theoretic approach and their data descriptors also directly depend the regions. In our work, each pixel is initially assigned a local histogram, i.e. a normalized histogram of the pixel intensities in a neighborhood of that pixel. The model finds a partition such that the local histograms in each region are similar to one another. We use an optimal transport distance to measure the similarity between histograms.

Previous models are quite effective in segmenting images when the histograms in each region are distinct. However, the distances used for comparing histograms are pointwise and may not be reliable even under simple circumstances. Furthermore, some distances used are not metrics; for instance, the triangle inequality is not satisfied. As an example of this issue, a pointwise distance between two delta functions with disjoint supports is the same no matter how far apart the supports are; this is a situation that arises often in segmentation applications, since for example images consisting of two objects with approximately constant but different intensities would fall into this category. Previous nonparametric approaches did not address this issue and used the Parzen window method [28] to approximate and smooth histograms. The degree of smoothness has to be controlled by a user-selected parameter. To overcome this issue, we propose to use an optimal transport distance to compare histograms, which extends as a metric to measures such as the delta functions. We believe this to be the more natural and appropriate way to compare histograms.

The optimal transport, or the Monge-Kantorovich, problem is to find the most efficient plan to rearrange one probability measure into another. We will introduce Kantorovich's version [18] here. Let (X, μ) and (Y, ν) be two probability measure spaces. Let π be a probability measure on the product space $X \times Y$ and $\Pi(\mu, \nu) = \{\pi \in P(X \times Y) : \pi[A \times Y] = \mu[A]$, and $\pi[X \times B] = \nu[B]$ hold for all measureable sets $A \in X$ and $B \in Y\}$ be the set of admissible transference plans. For a given cost function $c : X \times Y \to \mathbb{R}$, the total transport cost, associated to plan $\pi \in \Pi(\mu, \nu)$, is

$$I[\pi] = \int_{X \times Y} c(x, y) d\pi(x, y).$$

The optimal transport cost between μ and ν is

$$T_c(\mu, \nu) = \inf_{\pi \in \Pi(\mu,\nu)} I[\pi].$$

More detail can be found in [35] and [30]. In the case when the probability measures μ and ν are on $\mathbb{R}$, with cost function $c(x, y) = |x - y|^p$, the optimal transport cost has a closed-form solution,

$$T_p(\mu, \nu) = \int_0^1 |F^{-1}(t) - G^{-1}(t)|^p dt,$$

where F and G are the cumulative distribution functions of μ and ν, respectively, and F^{-1} and G^{-1} represent their corresponding inverse functions. The optimal transport distance, commonly called the *Wasserstein distance with exponent p*, is $W_p(\mu, \nu) = T_p(\mu, \nu)^{1/p}$. When the cost function is Euclidean distance $c(x, y) = |x - y|$,

$$W_1(\mu, \nu) = \int_0^1 |F^{-1}(t) - G^{-1}(t)| dt = \int_{\mathbb{R}} |F(x) - G(x)| dx.$$

The last equality is obtained by Fubini's Theorem. The Wasserstein distance defines a metric and is insensitive to oscillations [35].

The main contributions of this paper are as follows:

1. the novelty of using the Wasserstein distance to properly compare histograms, both discontinuous and continuous,
2. a segmentation model that does not need to differentiate histograms to find a solution,
3. the use of the fast global minimization method [6] to solve the proposed model, which significantly improves the previous model [12] in two ways, the computational time is less than the standard method and initialization can be arbitrary,
4. mathematical properties of the proposed model are presented.

The proposed model shown in this paper is based on the statistics of image intensity. However, the data term certainly can be replaced by histograms of other features, such as gradient, curvature, orientation and scale.

2 Related Works

In [19], Kim et al. took an information-theoretic approach and proposed a nonparametric region-based active contours model. Given an image $I : \Omega \to [0, L]$ with two regions, in each of which pixel intensities are independently identically distributed, a curve $\overrightarrow{C}$ is evolved towards the boundary. Denote the region inside (resp. outside) the curve $\overrightarrow{C}$ by R_+ (resp. R_-). Define the region labels associated with curve $\overrightarrow{C}$ by

$$L_{\overrightarrow{C}}(x) = \begin{cases} L_+ & \text{if } x \in R_+, \\ L_- & \text{if } x \in R_-. \end{cases}$$

The proposed model maximizes the mutual information between the image pixel intensities and region labels, subject to a constraint on the total length of the region boundaries:

$$\inf_{\overrightarrow{C}} \oint_{\overrightarrow{C}} ds - \lambda|\Omega| M(I(X); L_{\overrightarrow{C}}(X)), \tag{1}$$

where λ is a positive parameter, $|\cdot|$ is the 2-dimensional Lebesgue measure, i.e. area, and M stands for mutual information, defined as:

$$M(I(X); L_{\overrightarrow{C}}(X)) = h(I(X)) - h(I(X)|L_{\overrightarrow{C}}(X)).$$

Since entropy of image $h(I(X))$ is constant, maximizing the mutual information between $I(X)$ and $L_{\overrightarrow{C}}(X)$ minimizes the conditional entropy $h(I(X)|L_{\overrightarrow{C}}(X))$. The curve $\overrightarrow{C}$ is evolved so that knowing which region a pixel belongs to decreases the uncertainty of the pixel intensity. The conditional entropy is

$$\begin{aligned} &h(I(X)|L_{\overrightarrow{C}}(X)) \\ &= -\frac{1}{|\Omega|}\Big(\int_{R_+} \log P_+(I(x))dx + \int_{R_-} \log P_-(I(x))dx \Big), \end{aligned}$$

where the probability density functions $P_+(I(x))$ and $P_-(I(x))$ of each region are approximated using the Parzen window method [28],

$$P_+(I(x)) = \frac{1}{|R_+|} \int_{R_+} K(I(x) - I(\hat{x}))d\hat{x}, \tag{2}$$

$$P_-(I(x)) = \frac{1}{|R_-|} \int_{R_+} K(I(x) - I(\hat{x}))d\hat{x}. \tag{3}$$

The Gaussian function $K(z) = (1/\sqrt{2\pi\sigma^2})e^{-z^2/2\sigma^2}$ is used as a smoothing kernel, where σ is a scalar parameter that controls the smoothness of the approximation. The minimization problem (1) is solved by the following gradient flow:

$$\begin{aligned} \frac{\partial \overrightarrow{C}}{\partial t} = \lambda\Big[&\log \frac{P_+(I(\overrightarrow{C}))}{P_-(I(\overrightarrow{C}))} + \frac{1}{|R_+|}\int_{R_+} \frac{K(I(x) - I(\overrightarrow{C}))}{P_+(I(x))}dx \\ &- \frac{1}{|R_-|}\int_{R_-} \frac{K(I(x) - I(\overrightarrow{C}))}{P_-(I(x))}dx\Big]\overrightarrow{N} - \kappa\overrightarrow{N}, \end{aligned} \tag{4}$$

where $\overrightarrow{N}$ is the outward normal and κ is the curvature of $\overrightarrow{C}$. The implementation for (4) is by the level-set method with narrow band approach.

In [16], Herbulot et al. also took a nonparametric region-based active contours approach and used information entropy as competition between two regions:

$$\inf_{\overrightarrow{C}} \oint_{\overrightarrow{C}} ds + \lambda h(I(X), R_+) + \lambda h(I(X), R_-), \tag{5}$$

where entropy of pixel intensities in each region is

$$h(I(X), R_+) = -\int_{R_+} P_+(I(x)) \log P_+(I(x))dx$$

$$h(I(X), R_-) = -\int_{R_-} P_-(I(x)) \log P_-(I(x))dx.$$

The probability density functions $P_+(I(x))$ and $P_-(I(x))$ are approximated using the Parzen window method as described in (2) and (3). The minimization is solved by the following gradient flow:

$$\begin{aligned} \frac{\partial \overrightarrow{C}}{\partial t} = \lambda\Bigg[&- \Big(P_+(\log P_+ + 1) - P_-(\log P_- + 1)\Big) \\ &- \frac{1}{|\Omega|}\Bigg(h(I(X), R_+) - h(I(X), R_-) \\ &+ \int_{R_+} K(I(x) - I(\overrightarrow{C})) \log P_+(I(x))dx \\ &+ \int_{R_-} K(I(x) - I(\overrightarrow{C})) \log P_-(I(x))dx \Bigg)\Bigg]\overrightarrow{N} - \kappa\overrightarrow{N}, \end{aligned}$$

The curve evolution is implemented by using smoothing B-splines.

3 Proposed Model I

In this section, we discuss an unsupervised segmentation model proposed in our previous work [12] for cluttered images. Suppose the observed gray-scale image $I : \Omega \to [0, L]$ is measurable and has two regions of interests. Let $\mathcal{N}_{x,r}$ be the ball of radius r centered at x. Define the local histogram of a pixel $x \in \Omega$ by

$$P_x(y) := \frac{|\{z \in \mathcal{N}_{x,r} \cap \Omega : I(z) = y\}|}{|\mathcal{N}_{x,r} \cap \Omega|},$$

for $0 \leq y \leq L$. Define the corresponding cumulative distribution function by

$$F_x(y) := \frac{|\{z \in \mathcal{N}_{x,r} \cap \Omega : I(z) \leq y\}|}{|\mathcal{N}_{x,r} \cap \Omega|}, \tag{6}$$

for $0 \leq y \leq L$. These are the image data used in the following proposed segmentation model:

$$\inf_{\Sigma, P_1, P_2} \Big\{ E_1(\cdot,\cdot,\cdot|I) = \mathrm{Per}(\Sigma) \tag{7}$$
$$+\lambda \int_{\Sigma} W_1(P_1, P_x)dx + \lambda \int_{\Sigma^c} W_1(P_2, P_x)dx \Big\},$$

where $\mathrm{Per}(\Sigma)$ is the perimeter of the set Σ. This minimization problem finds an optimal region $\Sigma \subseteq \Omega$ and approximates the local histograms inside Σ (resp. Σ^c) by a constant histogram P_1 (resp. P_2). Recall that W_1 is the Wasserstein distance with exponent 1, described in the introduction:

$$W_1(P_1, P_2) = \int_0^L |F_1(y) - F_2(y)|dy. \tag{8}$$

Energy functional (7) can be formulated in terms of the level set method [26]. The boundary between Σ and Σ^c is represented by the 0-level set of a Lipschitz function $\phi : \Omega \to \mathbb{R}$.

$$\inf_{\phi, F_1, F_2} \Big\{ E_1(\cdot,\cdot,\cdot|I) = \int_{\Omega} |\nabla H(\phi(x))|dx \tag{9}$$
$$+\lambda \int_{\Omega} H(\phi(x)) \int_0^L |F_1(y) - F_x(y)|dy\, dx$$
$$+\lambda \int_{\Omega} [1 - H(\phi(x))] \int_0^L |F_2(y) - F_x(y)|dy\, dx \Big\},$$

where H is the Heaviside function, $\int_{\Omega} |\nabla H(\phi(x))|dx$ represents $\mathrm{Per}(\Sigma)$, and $H(\phi)$ (resp. $1 - H(\phi)$) defines Σ (resp. Σ^c).

The minimization of (9) can be achieved by a two-step scheme, which gives a local minimum. First, we fix ϕ and minimize with respect to F_1 and F_2, respectively. Variations with respect to F_1 and F_2 yield the following optimality conditions that should be held for all $0 \leq y \leq L$,

$$\int H(\phi(x)) \frac{F_1(y) - F_x(y)}{|F_1(y) - F_x(y)|} dx = 0$$

and

$$\int [1 - H(\phi(x))] \frac{F_2(y) - F_x(y)}{|F_2(y) - F_x(y)|} dx = 0,$$

respectively. Therefore,

$$F_1(y) = \text{median of } F_x(y), \text{ over } \{x : \phi(x) \geq 0\} \tag{10}$$

and

$$F_2(y) = \text{median of } F_x(y), \text{ over } \{x : \phi(x) < 0\}. \tag{11}$$

Next, with fixed $F1$ and $F2$, the gradient descent of Euler-Lagrange equation for ϕ gives

$$\phi_t = \delta(\phi) \Big[\nabla \cdot \left(\frac{\nabla \phi}{|\nabla \phi|} \right) - \lambda \int_0^L (|F_1(y) - F_x(y)| \tag{12}$$
$$-|F_2(y) - F_x(y)|)dy \Big],$$

where δ is a regularized Dirac function and $\nabla \cdot \left(\frac{\nabla \phi}{|\nabla \phi|} \right)$ is the curvature of the level sets. Steps (10), (11), and (12) are iterated alternately, until convergence to a steady state solution.

However, equation (12) has a serious time-step restriction, in addition to being a second-order equation. Numerically, the curvature term, the first term of (12), is approximated by

$$\frac{\partial}{\partial x} \left(\frac{\phi_x}{\sqrt{\phi_x^2 + \phi_y^2 + \epsilon^2}} \right) + \frac{\partial}{\partial y} \left(\frac{\phi_y}{\sqrt{\phi_x^2 + \phi_y^2 + \epsilon^2}} \right), \tag{13}$$

where $\epsilon > 0$ so that the denominators are not zero but small enough to stay close to the solution. By the CFL condition, the time-step restriction of the explicit scheme for (12) is [25] $\triangle t \leq c \cdot \epsilon \cdot (\triangle x)^2$, where c is a constant. The factor ϵ comes from (13) when $\phi_x^2 + \phi_y^2 = 0$. This time-step restriction can be improved to $\triangle t \leq c \cdot (\triangle x)^2$ with Chambolle's method [9], where $c = 1/8$. The application of Chambolle's method on the proposed model is presented in Section 4.3.

4 Fast Global Minimization of Model I

4.1 Global Minimization of Model I

Like many variational segmentation models, model (7) suffers from being non-convex (with respect to Σ) and is therefore sensitive to initializations. The requirement of reasonable initializations conflicts the purpose of automatic segmentation. Numerically, a non-compactly supported dirac function is used in [13] to increase the chances of finding global minimizers of the piecewise constant segmentation model. Theoretically, based on the framework of [6,11,23], we propose the following global minimization of Model I:

$$\min_{0 \leq u \leq 1, P_1, P_2} \Big\{ E_2(\cdot,\cdot,\cdot|I) = \int_{\Omega} |\nabla u(x)|dx \tag{14}$$
$$+\lambda \int_{\Omega} W_1(P_1, P_x)u(x)dx$$
$$+\lambda \int_{\Omega} W_1(P_2, P_x)(1 - u(x))dx \Big\}.$$

This problem is closely related to problem (7) but overcomes the non-convexity. Let $\mathbf{1}_S$ denote the characteristic function of set S. Model (14) extends the original

minimization over the non-convex set $\{u \in BV(\Omega) : u = \mathbf{1}_\Sigma$ for some set Σ with finite perimeter$\}$ to the convex set $\{u \in BV(\Omega) : 0 \le u \le 1\}$. Thus, (14) is convex with respect to u and, unlike (7), does not have (non-global) local minima with respect to the geometric unknown.

The major advantage of (14) is that initializations can be arbitrary. The relation between (7) and (14) is that, for fixed F_1 and F_2, a global minimizer of (7) can be found through a global minimizer of (14). This relation is stated in the following theorem, which is based on the geometric properties of TV.

Theorem 1: (Global Minimizers) Suppose $I(x) \in [0,1]$. If P_1, and P_2 are fixed, and $u(x)$ is any minimizer of $E_2(\cdot, P_1, P_2|I)$, then for a.e. $\mu \in [0,1]$, $\mathbf{1}_{\{x:u(x)>\mu\}}(x)$ is a global minimizer of $E_1(\cdot, P_1, P_2|I)$.

Proof: Based on [11], by the coarea formula and setting $\Sigma(\mu) := \{x : u(x) > \mu\}$, we can write E_2 in terms of E_1

$$\begin{aligned} E_2(u, P_1, P_2|I) = \int_0^1 \Big\{ & \mathrm{Per}(\Sigma(\mu)) \\ & +\lambda \int_{\Sigma(\mu)} W_1(P_1, P_x)dx \\ & +\lambda \int_{\Omega-\Sigma(\mu)} W_1(P_2, P_x)dx \Big\} d\mu \\ = \int_0^1 & E_1(\Sigma(\mu), P_1, P_2|I)d\mu, \end{aligned}$$

Therefore, if u is a minimizer of $E_2(\cdot, P_1, P_2|I)$, then for a.e. $\mu \in [0,1]$, $\Sigma(\mu)$ is a minimizer of $E_1(\cdot, P_1, P_2|I)$.□

4.2 Existence of Global Minimization Solutions

In this section, we show the existence of a minimizer for and convexity of model (14).

Theorem 2: (Existence of Solutions for u) For fixed P_1 and P_2,

$$\begin{aligned} \min_{0\le u\le 1} \Big\{ E_2(\cdot, P_1, P_2 \mid I) = & \int_\Omega |\nabla u(x)|dx \\ & +\lambda \int_\Omega W_1(P_1, P_x)u(x)dx \\ & +\lambda \int_\Omega W_1(P_2, P_x)(1-u(x))dx \Big\} \end{aligned} \tag{15}$$

has a solution $u \in BV(\Omega)$ with $0 \le u \le 1$.

Proof: Let $\{u_n\} \in BV(\Omega)$ with $0 \le u \le 1$ be a minimizing sequence. Then, $\int_\Omega |Du_n|$ is uniformly bounded. Since every uniformly bounded sequence in $BV(\Omega)$ is relatively compact in $L^1(\Omega)$, there exists a subsequence $\{u_{n_k}\}$ converging to some $u \in BV(\Omega)$. Since $u_{n_k} \to u$ in $L^1(\Omega)$, we have $u_{n_k} \to u$ in measure, i.e. $|\{x : |u_{n_k}(x) - u(x)| \ge \epsilon\}| \to 0$ as $\epsilon \to 0$. Since we also have $0 \le u_{n_k} \le 1$, u satisfies $0 \le u \le 1$. Finally, one can check easily that u is indeed a minimizer by the lower semicontinuity of $BV(\Omega)$ and Fatou's lemma. □.

For fixed u, the minimizer for F_1 (resp. F_2) has an explicit solution. Variations of E_2 with respect to F_1 and F_2 yield the following optimality conditions that should hold for all $0 \le y \le L$:

$$\int u(x)\frac{F_1(y) - F_x(y)}{|F_1(y) - F_x(y)|}dx = 0$$

and

$$\int [1-u(x)]\frac{F_2(y) - F_x(y)}{|F_2(y) - F_x(y)|}dx = 0,$$

respectively. Therefore,

$$F_1(y) = \text{weighted (by } u(x)\text{) median of } F_x(y), \tag{16}$$

and

$$F_2(y) = \text{weighted (by } 1-u(x)\text{) median of } F_x(y), \tag{17}$$

We will next show that $E_2[u, P_1, P_2|I]$ is convex with respect to each variable. First, E_2 is convex with respect to u because $\int_\Omega |Du(x)|dx$ is convex in u and the set $\{u \in BV(\Omega) : 0 \le u \le 1\}$ is convex. Second,

Theorem 3: The minimization problem

$$\min_{P_1 \in P(\Omega)} E_2[u, \cdot, P_2|I]$$

is convex, where $P(\Omega)$ denotes the set of Borel probability measures on Ω.

Proof: $E_2[u, \cdot, P_2|I]$ is convex in P_1 because the Wasserstein distance is a metric and in particular satisfies the triangle inequality. Since $P(\Omega)$ is a convex set, minimization with fixed u and P_2 is a convex problem.□

Similarly, the minimization $\min_{P_2 \in P(\Omega)} E_2[u, P_1, \cdot|I]$ is convex. Therefore, $E_2[u, P_1, P_2|I]$ is convex with respect to each variable.

4.3 Fast Minimization Scheme

Minimizing the proposed energy E_2 in (14) with respect to u can be efficiently solved by applying methods in [2,6]. The regularization and data terms in (14) can be decoupled by using a new variable v to replace u in the data term and adding a convex term that forces v and u to be significanly close:

$$\begin{aligned} \min_{u, 0\le v\le 1} \int_\Omega |\nabla u(x)|dx + & \frac{1}{2\theta}\int_\Omega (u(x) - v(x))^2 dx \\ & +\lambda \int_\Omega r(x, F_1, F_2)v(x)dx \ , \end{aligned} \tag{18}$$

where

$$r(x, F_1, F_2) = \int_0^L |F_1(y) - F_x(y)| - |F_2(y) - F_x(y)|dy,$$

and $\theta > 0$ is a small parameter. Minimizing the convex variational model (18) can be approached by alternately solving the following coupled problems:

$$\min_u \int_\Omega |\nabla u(x)| + \frac{1}{2\theta}(u(x) - v(x))^2 dx \tag{19}$$

and

$$\min_{0\le v\le 1} \int_\Omega \frac{1}{2\theta}(u(x) - v(x))^2 + \lambda r(x, F_1, F_2)v(x)dx \ . \tag{20}$$

The minimization problem in (19) can be efficiently achieved by Chambolle's method [9], based on the dual formulation of the total variation norm. The derived solution is

$$u(x) = v(x) - \theta \mathrm{div} p(x) \ , \tag{21}$$

where $p = (p^1, p^2)$ solves $\nabla(\theta \mathrm{div} p - v) - |\nabla(\theta \mathrm{div} p - v)|p = 0$ and is solved by a fixed point method,

$$p^{n+1} = \frac{p^n + \delta t \nabla(\mathrm{div} p^n - v/\theta)}{1 + \delta t |(\mathrm{div} p^n - v/\theta)|}. \tag{22}$$

The solution of (20) is found as in [6]:

$$v(x) = \max\{\min\{u(x) - \theta\lambda r(x, F_1, F_2), 1\}, 0\} \ . \tag{23}$$

The proposed fast minimization scheme is to iterate (16), (17), (22), (21), and (23) alternately, until convergence.

5 Properties of Proposed Models and Comparison with Other Models

The proposed model has several desired mathematical properties as shown in Table 1. In Section 4.2, we show the existence of solution and the convexity of model in each variable. In the discrete setting, if the resolution of an image f is $m \times n$ and L is the number of the gray levels, then $u \in \mathbb{R}^{m\times n}$ and $P \in P(\{0, 1, ...L\}) \subset \mathbb{R}^L$. Therefore, the model in the discretized form is convex in $\mathbb{R}^{m\times n} \times P(\{0, 1, ...L\})$ and thus a global minimizer can be found. Based on Chambolle's dual method regarding the length-penalizing term, the solution converges after a small number of iterations, compared to directly solving the associated Euler-Lagrange equation. Moreover, since the Wasserstein distance is insensitive to oscillations, our model is intrinsically robust to noise. On the other hand, it does not require histograms to be smoothed, which has to be done for many segmentation models even for noiseless images. For instance, the Wasserstein distance is able to distinguish the distance between any pair of delta functions with disjoint supports. Many distances do not tell apart the distance between two disjointly supported histograms unless the histograms are smoothed. The complexity of computing one iteration is O(Lmn). For a 200×150 image, the computational time for a solution to converge is approximately two minutes. Since the partition is implicitly embedded in function u, the model is able to handle topological changes.

Kim et al.'s model [19] also has existence of solution. Their model minimizes over a non-convex set $\{u \in BV(\Omega) : u = 1_\Sigma$ for some set Σ with finite perimeter$\}$, thus does not guarantee to get a global minimizer. The gradient flow (4) has a curvature term and the convergence can be slow, due to the CFL condition discussed in Section 3. The probability density functions are estimated by the Parzen window method. This enables their model to handle noise but introduces a user-selected parameter, i.e. kernel width. They use the fast Gauss transform to compute probability densities, which reduces the complexity of computing one iteration to $O(M)$, where M is the size of the narrow band. The level-set method is used for curve evolution and thus allows topological changes.

Herbulot et al. [16] use smoothing B-splines to implement their derived evolution equation instead of the usual level-set method to avoid extensive computational time. The complexity of each iteration is O(LM), where L is the number of gray levels and M and the size of the narrow band. The parametric method using B-splines does not handle topological changes of the contours. They further use a smoothing B-splines in order to be more robust to noise. The tradeoff between the smoothness and interpolation error is controlled by a parameter that has to be chosen by the user. Their model also minimizes over a non-convex set, thus does not guarantee to get a global minimizer.

6 Description of Model II

We propose a variant of Model I that handles segmentation properly when the captured image has uneven lighting exposure, due to reasons such as the location of the light source and camera. The original model considers the data term globally, i.e. compares all the local histograms within each region. Therefore, when the local lighting changes significantly, local histograms of the same feature may have similar shapes but are far apart by a translation in the intensity axis. As a result, the Wasserstein distance between them is large and thus the original model is not designed to deal with uneven lighting. To model this variation, we introduce a function $a(x)$, representing the translation in the intensity axis, and propose a new model:

Table 1 Properties of the proposed model and Kim et al. [19] and Herbulot et al. [16] models

	Our model	Kim et al. [19]	Herbulot et al. [16]
existence of solution	✓	✓	✓
global minimum/convexity	✓	×	×
fast minimization	✓	×	✓
insensibility to noise	✓	-	-
no need to smooth histograms (noiseless case)	✓	×	×
local change of lighting	✓	×	×
complexity for one iteration	O(Lmn)	O(M)	O(LM)
computational time	2 mins	10 mins	10 mins
handle topological changes	✓	✓	×

$$\inf_{\Sigma,a,F_1,F_2}\Big\{E_3(\Sigma,a,F_1,F_2|I)=\mathrm{Per}(\Sigma) \tag{24}$$
$$+\frac{\alpha}{2}\int|\nabla a(x)|^2dx$$
$$+\lambda\int_{\Sigma}\int_0^L|F_1(y)-F_x(y-a(x))|dy\;dx$$
$$+\lambda\int_{\Sigma^c}\int_0^L|F_2(y)-F_x(y-a(x))|dy\;dx\Big\}.$$

This model allows local histograms to translate on the intensity axis in order to find a best fit among one another within each region. A regularity constraint $\int|\nabla a(x)|^2dx$ is imposed to ensure smoothness of a.

To solve the minimization, we have the following three-step scheme. The evolution equations for F_1, F_2 and ϕ can be derived similarly as in Section 3:

$$F_1(y)=\text{median of } F_x(y-a(x)),\ \text{over } \{\phi\geq 0\} \tag{25}$$

$$F_2(y)=\text{median of } F_x(y-a(x)),\ \text{over } \{\phi< 0\} \tag{26}$$

$$\phi_t=\delta(\phi)\left[\nabla\cdot\left(\frac{\nabla\phi}{|\nabla\phi|}\right)-\lambda\int_0^L\Big(|F_1(y)-F_x(y-a(x))|\right.$$
$$\left.-|F_2(y)-F_x(y-a(x))|\Big)dy\right]. \tag{27}$$

The minimization with respect to $a(x)$ is to solve:

$$\inf_a E_3(\Sigma,\cdot,F_1,F_2|I)=\frac{\alpha}{2}\int|\nabla a(x)|^2dx \tag{28}$$
$$+\lambda\int_{\Sigma}\int_0^L|F_1(y)-F_x(y-a(x))|dy\;dx$$
$$+\lambda\int_{\Sigma^c}\int_0^L|F_2(y)-F_x(y-a(x))|dy\;dx\;.$$

Without the first term, $a(x)$ can be solved explicitly by

$$a_0(x)=\begin{cases}F_1^{-1}(0.5)-F_x^{-1}(0.5) & \text{if } \phi(x)>0\\ F_2^{-1}(0.5)-F_x^{-1}(0.5) & \text{if } \phi(x)\leq 0\end{cases}$$

Therefore, the problem of (28) is transformed into solving the following:

$$\inf_a\frac{1}{2}\int|a(x)-a_0(x)|^2dx+\frac{\alpha}{2}\int|\nabla a(x)|^2dx\;. \tag{29}$$

The solution to (29) is $a(x)-\alpha\triangle a(x)=a_0(x)$, which can be easily solved, for example, by the fast fourier transform. We may also employ the fast global minimization technique for Model II, instead using (27).

7 Experimental Results

7.1 Comparison with other methods

As explained in Section 5, our model does not require histograms to be smoothed for proper segmentation. In contrast, previous methods use Parzen window method [28] to estimate pdfs, which requires a smoothness parameter selection. If the bandwidth of the kernel is too small, point-wise metrics cannot detect similar intensities. Fig.1 (a) is a synthetic image with three regions, in each of which the pixel intensity is independently identically distributed (b). The pixels in the inner region take intensities $3, 110, 140$, and 247, with probability about 0.25 each. The pixels in the middle region take intensities $85, 110, 140$, and 165, with probability about 0.25 each. The pixels in the outer region take intensities $80, 115, 135$, and 170, with probability about 0.25 each. The middle and outer regions are perceptually similar and so are their corresponding intensity histograms, (d) and (e), respectively. A desired partition is to distinguish the inner region from the rest. The initial contour is shown in (f). Our model does not have the smoothing parameter and correctly segments the inner region from the rest because of the use of the Wasserstein distance.

On the other hand, Kim et al.'s model [19] needs a careful selection of the smoothness parameter σ (variance of the Gaussian kernel) in order to segment correctly. Fig.2(a) is the final contour with $\sigma=5$, which in-

correctly groups the inner and middle regions together. This is because the histograms of the inner and middle regions overlap 50% but the histograms of the middle and outer region do not overlap. In (b), the segmentation with $\sigma = 10$ is correct because the intensity pdf is greatly smoothed and thus mutual information is able distinguish the inner region from the rest. When $\sigma = 50$, the final contours (c) incorrectly separate pixels with intensity $= 3,247$ from pixels with intensity $= 110, 115, 135, 140$.

We emphasize here that nonparametric models are able to deal with a greater variety of images than parametric models. In this experiment, the object and background have the same intensity mean and variance. In Fig.3(a), we show the boundaries of the objects in red curves and the corresponding histograms in each region. Fig.3(c) and (b) are the final contours of our proposed model and ACWE, respectively. The proposed model is able to distinguish the objects from the background. On the other hand the ACWE model cannot handle this case due to its parametric nature.

7.2 Comparison between original model and fast global minimization

The proposed fast global minimization in Section 4 improves the original minimization in [12] described in Section 3 of model I. Fig.4 is a downsized 175×135 image of cheetah. In Section 4.1, we explain that the global minimization model does not have local minima and thus is guaranteed to find a global minimizer. We experiment with several images with different and arbitrary initializations and all arrive at similar results for each image. This is a nice consequence of the global minimization model being convex with respect to each variable. On the other hand, the original minimization is non-convex and thus requires initializations to be reasonably close to the final contours. Moreover, the fast global minimization improves the speed from two hours to two minutes.

7.3 Robustness to noise/More results of Model I and Model II

Fig.5(a) is a clean image of cheetah and (b) is with noise. The final contour shown in (d) by the global minimization of Model I is able to segment the cheetah patterns and is nearly as good as the result in (c) of the clean image. In this experiment, the neighborhood radius is 11.

Fig.6 shows other experiments of Model I. The first experiment is a 285×281 image consisting of two Brodatz textures. The final contours are shown in (a) and the corresponding histograms on each region are plotted in (c). Model I is able distinguish these two Brodatz textures, even though their intensity distributions are highly discontinuous. The second is a 481×321 image of tiger; (b) shows the final contours by Model I and (d) shows the histograms in each region. The final contour successfully selects the tiger patterns.

Fig.7 shows that Model II improves Model I when there are local lighting changes in the image. The first experiment is a 384×223 image of cheetah. In (a), Model I is able to capture some of the cheetah patterns but not near the back legs, due to the local lighting difference. Final contours of Model II, in (b), are more accurate. Another experiment is a 282×218 image of fish. The final contours by Model I, in (d), do not select the fish patterns accurately, because the local illumination is significantly uneven. Model II, on the other hand, is able to overcome this difficulty, as shown in (e) the final contours separates the fish patterns from the background.

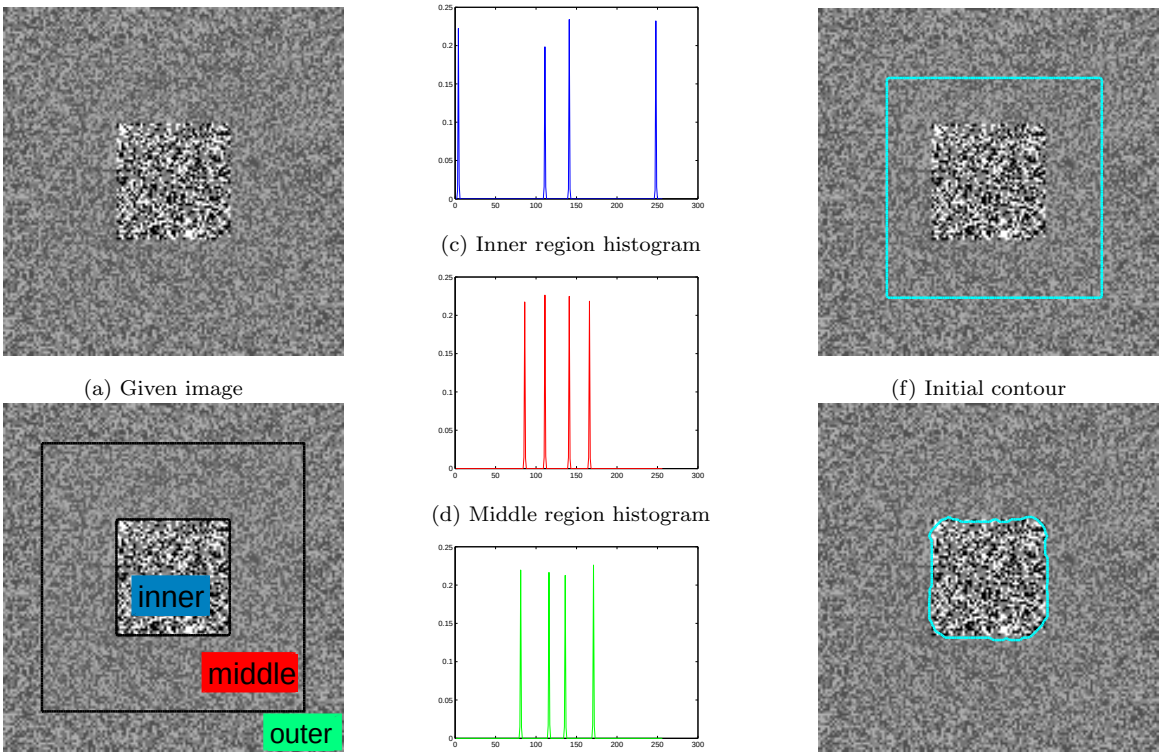
7.4 Implementation issues

We show a method to solve the weighted median for $F_1(y)$ in equation (16) in the discrete case.
For each $y = 0, 1, ..., L$,

1. Compute the weighted histogram, H_y, of value $F_x(y)$ with weight $u(x)$. More precisely, for all pixels $x \in \Omega$, each value $F_x(y)$ is counted $u(x)$ times. Then, normalize the weighted histogram, H_y, by dividing by the total count, $\sum_{x\in\Omega} u(x)$.
2. For each weighted histogram H_y, compute the cumulative distribution C_y.
3. The weighted median is then $F_1(y) = C_y^{-1}(0.5)$.

The calculation of $F_2(y)$ is similar and with weight $1 - u(x)$.

We empirically demonstrate that segmentation results are not sensitive to the size of local neighborhood histograms, within a reasonable range. The experiment is on a 384×223 image of cheetah, shown in Fig.5(a). Fig.8 shows final contours by global minimization of Model I with different neighborhood sizes, radius ranging from 1 to 25. If the neighborhood size is smaller than the clutter features, the final contour partitions clutter features into smaller regions, an undesired result. If the neighborhood size is large enough, our results show the cheetah patterns are segmented for a large range of neighborhood sizes.



(a) Given image

(b) Intensity is i.i.d in each region

(c) Inner region histogram

(d) Middle region histogram

(e) Outer region histogram

(f) Initial contour

(g) Model I

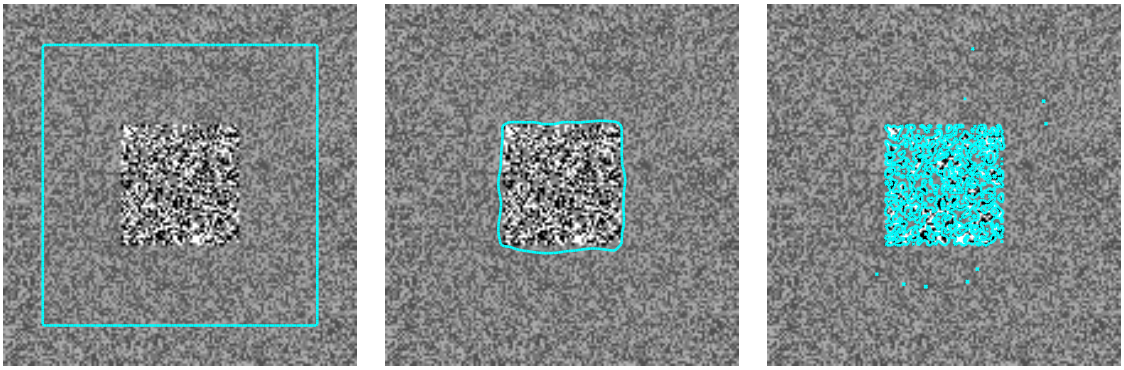
Fig. 1 The given image (a) has three regions (b), in each of which pixel intensity is independently identically distributed. (c), (d), and (e) are the intensity histograms of the pixels in the inner, middle, and outer regions, respectively. The pixels in the inner region take intensities $3, 110, 140$, and 247, with probability about 0.25 each; the pixels in the middle region take intensities $85, 110, 140$, and 165; and the pixels in the outer region take intensities $80, 115, 135$, and 170. The middle and outer regions look similar, as well as their corresponding histograms. Wasserstein distance does not require histograms to be smoothed in order to compare histograms in a reasonable manner. The final contour of proposed model I, in (g), correctly distinguish the inner region from the rest.

7.5 Limitations and Extensions

Our segmentation model is formulated for gray-scale images but can be extended to color images. The data term can be generalized because the Wasserstein distance is defined on any space of probability measures. However, the implementation would be much more complicated because there is no closed form for the Wasserstein distance between probability measures on Euclidean spaces with dimensions larger than one. The Earth Movers Distance between signatures is equivalent to the Wasserstein distance when signatures have the same total mass (or normalized discrete pdfs) and the optimization has been investigated in [31]. This can be a possible direction to extend our segmentation model. Works in [15,17] numerically solve the the optimal maps of the optimal transport problem on $\mathbb{R}^2$ and may also be applied to our extension. Another limitation is that our model assumes the given image has two regions of clutters. Many natural images have more than two regions and requires a multi-phase segmentation model. This limitation can be easily overcome, since our model has a natural extension to multi-phase segmentation as in [36]. Moreover, since our model only uses the intensity probability density, it does not take into account higher-order characteristics, such as gradient, scale, and orientation. For example, if two textures have the same intensity probability density, our model is not able to distinguish them. However, histograms of suitable descriptors can be used instead of intensity or combined. On the other hand, our segmentation model can be contributed to segmentation algorithms [32,33] that incorporate many image characteristics, one of which is clutter.

8 Conclusions

In this paper, we proposed a fast global minimization of a local histogram based model using the Wasserstein distance with exponent 1 to segment cluttered images. Our model is different from previous nonparametric region-based active contours in three ways. The first is the use of the Wasserstein distance, which is able to



(a) Kim et al. with $\sigma = 1$ (b) Kim et al. with $\sigma = 10$ (c) Kim et al. with $\sigma = 50$

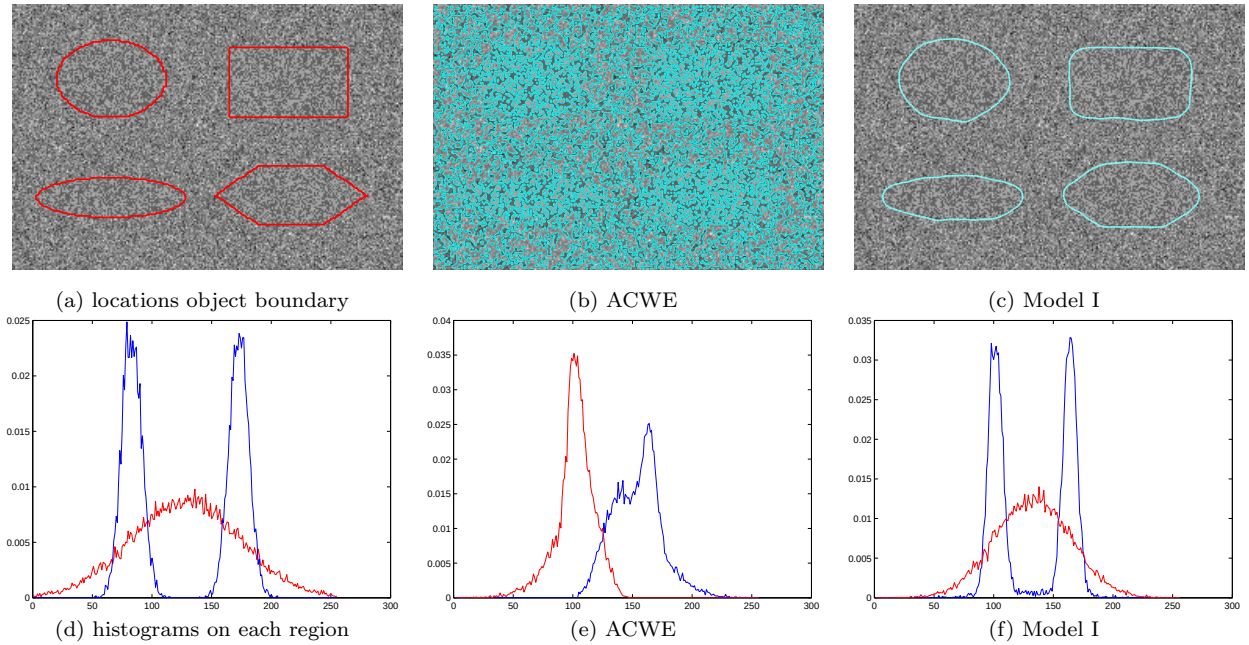
Fig. 2 Kim et al.'s model [19] needs a careful selection of the smoothness parameter σ in order to segment correctly. (a) is the final contour with $\sigma = 5$, which incorrectly groups the inner and middle regions (see Fig. 1(b)). The segmentation with $\sigma = 10$ (b) is correct because the intensity pdf is greatly smoothed and thus mutual information is able distinguish the inner region from the rest. When $\sigma = 50$ (c), the final contours separate pixels with intensity 3 and 247 from pixels with intensity 110, 115, 135, and 140.

compare both continuous and discontinuous histograms properly. The second is that the proposed model does not need to differentiate histograms to find the solutions. The third is the application of the global minimization method and consequently the segmentation results are not sensitive to initializations. We have proved a number of desired mathematical properties of the model and provided experimental verifications. In the future, we will generalize our model to color images and multi-phase segmentation. The former can be achieved by using the fast minimization of vectorial total variation in [5] and adapting the numerical scheme for computing the optimal transport distance in [31, 15, 17]. The later can be approached by applying methods such as the multi-phase level set framework [36].

References

1. G.Aubert, M.Barlaud, O.Faugeras, S.Jehan-Besson, Image Segmentation Using Active Contours: Calculus of Variations or Shape Gradients?, SIAM APPL. Math. vol.1(2), pp.2128-2145, 2005.
2. J.F.Aujol, G.Gilboa, T.Chan, and S.Osher, Structure-texture image decomposition - modeling, algorithms, and parameter selection, Internationl Journal of Computer Vision, vol.67(1), 2006.
3. R.E.Broadhurst, Statistical estimation of histogram variation for texture classification, In Texture 2005: Proc. of the 4th Int. Workshop on Texture Analysis and Synthesis, pp. 25-30, 2005.
4. A. Herbulot, S.Jehan-Besson, S. Duffner, M.Barlaud, and G.Aubert, Segmentation of Vectorial Image Features Using Shape Gradients and Information Measures, J. Math. Imaging and Vision, Vol. 25(3), 2006, pp.365-386.
5. X.Bresson and T.Chan, Fast Minimization of the Vectorial Total Variation Norm and Applications to Color Image Processing, UCLA CAM report 07-25, August 2007.
6. X.Bresson, S.Esedoglu, P.Vandergheynst, J.P.Thiran, S.Osher, Fast Global Minimization of the Active Contour/Snake Model, Journal of Mathematical Imaging and Vision (JMIV), Vol. 28(2), pp. 151-167, 2007.
7. V.Caselles, R.Kimmel and G.Sapiro, Geodesic active contours. Int. J. Comp. Vis., 22(1):61-79, 1997.
8. V.Caselles,J-L.Lisani,J-M.Morel, and G.Sapiro, Shape preserving local histogram modification, IEEE Trans. Image Proc. 8, pp. 220-230, 1999.
9. A.Chambolle, An algorithm for total variation minimization and applications, Jornal of Mathematica Imaging and Vision, vol. 20(1-2), 00, 89-97, 2004.
10. L.Cohen, On active contour models and balloons, Compt. Vision Graphics Image Process., 53(1991), pp.211-218.
11. T.Chan, S.Esedoglu and M.Nikolova, Algorithms for finding global minimizers of image segmentation and denoising models. SIAM Journal on Applied Mathematics, 66(5): 1632-1648, 2006.
12. T.Chan, S.Esedoglu and K.Ni, Histogram Based Segmentation Using Wasserstein Distances, in Proc. of SSVM, pp.697-708, 2007.
13. T.F.Chan, L.A.Vese, Active contours without edges, IEEE Transactions on Image Processing. 10:2(2001), pp. 266-277.
14. G.Chung and L.A.Vese, Energy Minimization Based Segmentation and Denoising Using a Multilayer Level Set Approach, EMMCVPR 2005, LNCS Vol. 3757/2005, pp. 439 - 455.
15. R.Chartrand, K.Vixie, B.Wohlber, E.Bollt, A Gradient Descent Solution to the Monge-Kantorovich Problem, submitted, 2007.
16. A.Herbulot, S.Jehan-Besson, M.Barlaud, G.Aubert, Shape Gradient for Image Segmentation using Information Theory, in ICASSP, may 2004, Vol. 3, pp.21-24.
17. S.Haker, L.Zhu, and A.Tannnenbaum, Optimal Mass Transport for Registration and Warping, International Journal of Computer Vision 60(3), 225-240,2004.
18. L.V.Kantorovich, On the translocation of masses, C.R.(Doklady) Acad. Sci. URSS(N.S.), 37(1942),pp.199-201.
19. J.Kim, J.W.Fisher, A.Yezzi, M.Cetin, and A.S.Willsky, A nonparametric statistical method for image segmentation using information theory and curve evolution , IEEE Trans. on Image Processing, vol. 14, pp.1486-1502, 2005. III:797-800.
20. S.Kichenesamy, A.Kumar, P.Olver, A.Tannenbaum, and A.Yezzi. Conformal curvature flows: from phase transitions to active vision, in Archive for Rational Mechanics and Analysis, vol. 134, pp. 275-301, 1996.
21. M.Kass, A.Witkin and D.Terzopoulos, Snakes: active contours model, Int. J. Comput. Vis., 1:1167-1186, 1991.

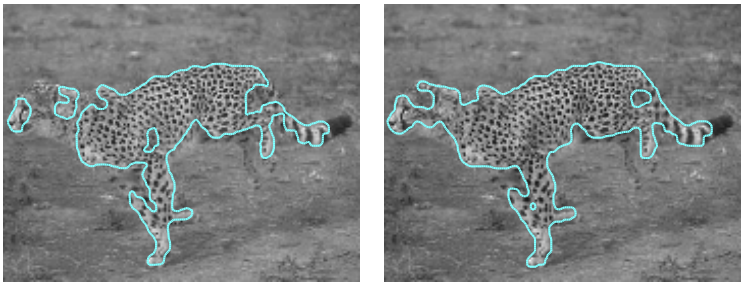
22. G.Monge, Mémoire sur la théorie des déblais at des remblais, Histoire de l'Académie Royale des Sciences de Paris, pp.666-704, 1781.
23. B.Mory and R.Ardon, Fuzzy Region Competition: A Convex Two-Phase Segmentation Framework, in Proc. of SSVM, pp.214-226, 2007.
24. D.Mumford and J.Shah, Optimal approximation by piecewise smooth functions and associated variational problems, Commun. Pure Appl. Math, vol. 42, pp. 577-685,1989.
25. S.Osher and R.Fedkiw, Level Set Methods and Dynamic Implicit Surfaces, Springer-Verlag New York, Applied Mathematica Sciences, 153, 2002.
26. S.Osher and J.A.Sethian, Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi Formulation, J. Comput. Phys., vol. 79, pp. 12-49, 1988.
27. N.Paragios and R.Deriche, Geodesic active regions: a new paradigm to deal with frame partition problems in computer vision, Journal of Visual Communication and Image Prepresentation, 13(1-2):249-268, 2002.
28. Parzen E, On estimation of a probability density function and mode, Ann. Math. Stat. 33, pp. 1065-1076, 1962.
29. Y.Rubner,J.Puzicha,C.Tomasi,J.M.Buhmann, Empirical evaluation of dissimilarity measures for color and texture, Computer Vision and Image Understanding 84, pp. 25-43, 2001.
30. S.Rachev and L.Rüschendorf, Mass transportation problems. Vol.I: Theory, Vol.II: Applications.Probability and its applications. Springer-Verlag, New York, 1998.
31. Y.Rubner, C.Tomasi, and L.J.Guibas, A metric for distributions with applications to image databases, In IEEE Int'l Conf. on Computer Vision, pp. 59-66, Bombay, India, January 1998.
32. G.Sapiro and V.Caselles, Histogram modification via differential equations, Journal of Differential Equations 135:2, pp. 238-268, 1997.
33. Z.Tu and S.Zhu, Image segmentation by data-driven Markov Chain Monte Carlo, in IEEE trans. on pattern analysis and machine intelligenc, Vol.24, No.5, May 2002.
34. A.Yezzi,Jr.,A.Tsai, and A.Willsky, A statistical approach to snakes for bimodal and trimodal imagery, In Int. Conf. on Computer Vision, pp.898-903, 1999.
35. C.Villani, Topics in Optimal Transportation. Graduate Studies in Mathematics, Vol. 58, American Mathematical Society, Providence Rhode Island, 2003.
36. L.A.Vese and T.F.Chan, A Multiphase Level Set Framework for Image Segmentation Using the Mumford and Shah Model, Int'l Journal of Computer Vision, Vol. 50, No. 3, pp. 271-293, 2002.
37. S.C.Zhu and A.Yuille, Region Competition: Unifying Snakes, Region Growing, and Bayes/MDL for Multiband Image Segmentation, IEEE trans. on Pattern Analysis and Machine Intelligence, Vol. 18, No. 9, pp. 884-900, 1996.



(a) locations object boundary (b) ACWE (c) Model I

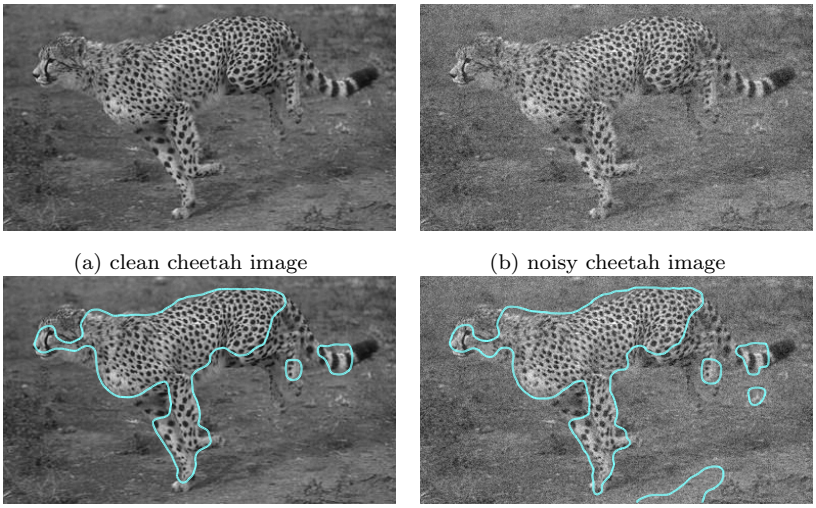
(d) histograms on each region (e) ACWE (f) Model I

Fig. 3 Objects and background regions have the same intensity mean and variance. (b) Final contours of ACWE model. (c) Final contours of proposed model I



(a) result of [12] (b) result of (14)

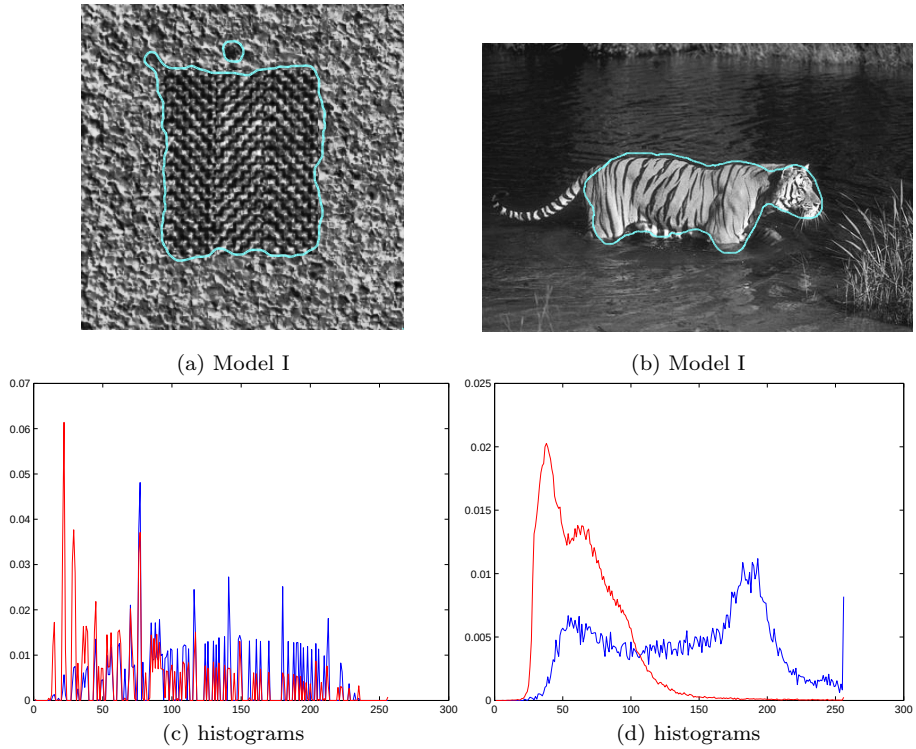
Fig. 4 Down-sized cheetah image. Global minimization improves segmentation result



(a) clean cheetah image (b) noisy cheetah image

(c) final contour of clean image (d) final contour of noisy image

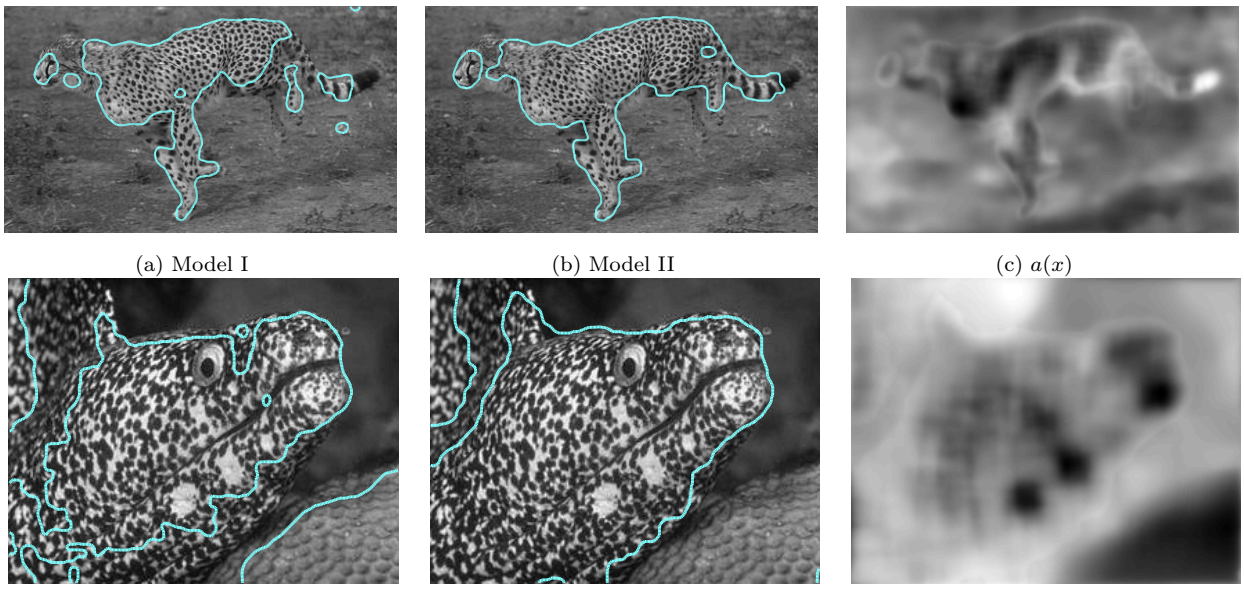
Fig. 5 The performance of Model I is nearly as good even with added noise to the clean image



(a) Model I (b) Model I

(c) histograms (d) histograms

Fig. 6 Final contours of Model I and their corresponding histograms on each region.



(a) Model I (b) Model II (c) $a(x)$

(d) Model I (e) Model II (f) $a(x)$

Fig. 7 The smoothness component allows local illumination changes and captures more of the cheetah pattern

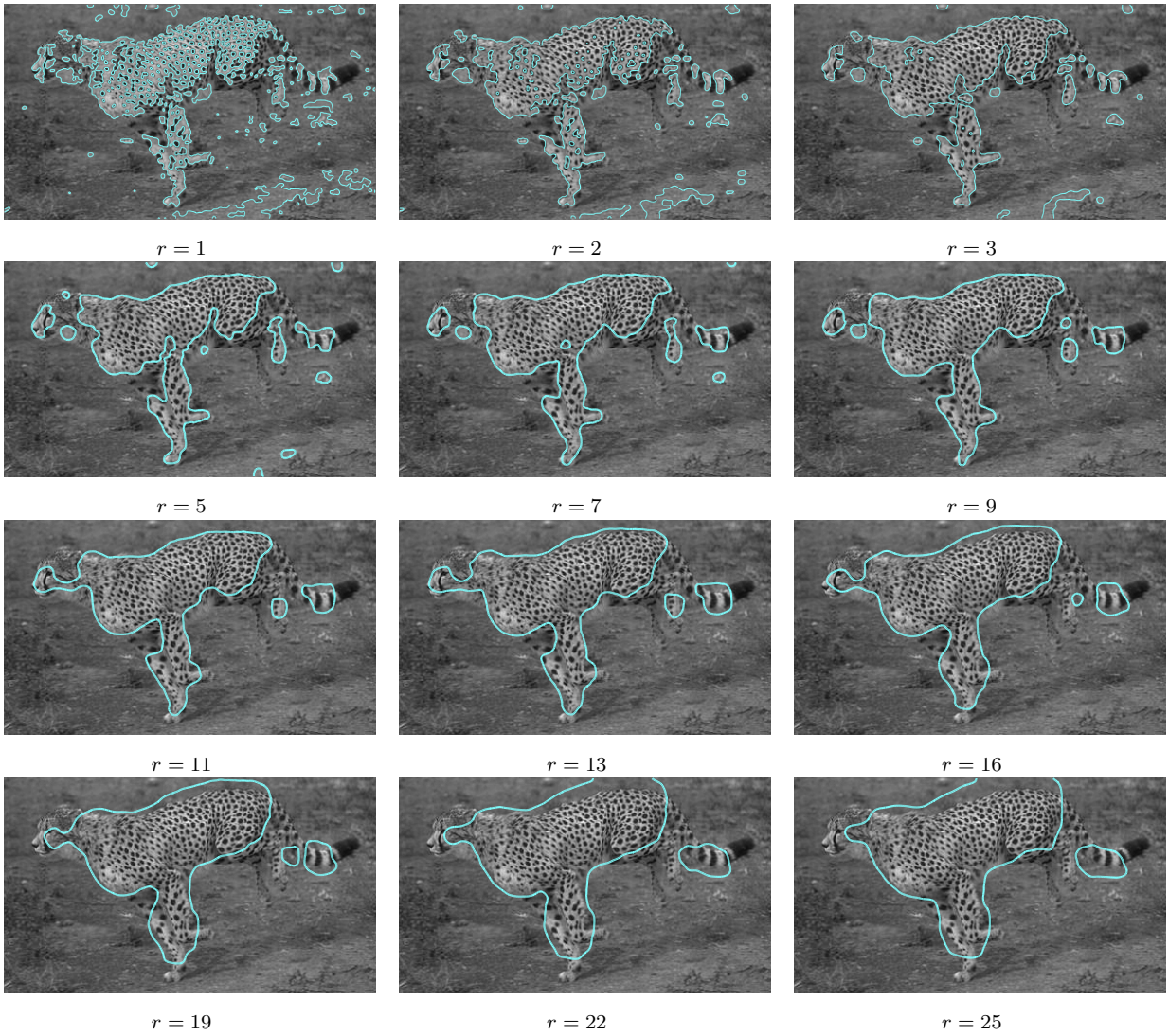


Fig. 8 The neighborhood size in model (14) needs to be equal or bigger than the smallest features of interest in the given image. The segmentation results are not too sensitive to the size r of the neighborhood, but are more accurate when the size is closer to that of the smallest image features of interest