

Histogram Based Segmentation Using Wasserstein Distances

Tony Chan¹, Selim Esedoglu², and Kangyu Ni³

¹ Department of Mathematics, UCLA
chan@math.ucla.edu

² Department of Mathematics, University of Michigan
esedoglu@umich.edu

³ Department of Mathematics, UCLA
kni66@math.ucla.edu

Abstract. In this paper, we propose a new nonparametric region-based active contour model for clutter image segmentation. To quantify the similarity between two clutter regions, we propose to compare their respective histograms using the *Wasserstein* distance. Our first segmentation model is based on minimizing the Wasserstein distance between the object (resp. background) histogram and the object (resp. background) reference histogram, together with a geometric regularization term that penalizes complicated region boundaries. The minimization is achieved by computing the gradient of the level set formulation for the energy. Our second model does not require reference histograms and assumes that the image can be partitioned into two regions in each of which the local histograms are similar everywhere.

Keywords: image segmentation, region-based active contour, Wasserstein distance, clutter.

1 Introduction

Parametric region-based active contour models have been widely used in image segmentation. One of their advantages is that they incorporate region information with boundary information. For example, the Chan-Vese model is able to carry out foreground and background segmentation without any explicit reference to edges [5]. However, the standard Chan-Vese model is based on the assumption that the foreground (resp. background) intensity is fairly homogeneous, i.e. the probability density functions of object intensities and background intensities are both Gaussian with the same variance. This can be a significant restriction in applications. Other parametric region-based active contours models, including certain generalizations of the Chan-Vese model, assume that the histogram of image intensities in different regions of the segmentation are Gaussian. For example, in [18], the segmentation models distinguish the object from the background by intensity means and/or variances of image regions.

The purpose of this work is to segment images consisting of clutter features. There are also many other image models that are not within our scope, such as

shape, texture and smooth regions. An effective method of incorporating different image models can be found in [17]. Clutter features are often found in natural scenes, such as trees and grass. They are highly nonhomogeneous in intensity and their corresponding histograms do not necessarily have particular statistical structure – for example, they may not obey a Gaussian distribution. They also usually do not have a particular geometric content. Therefore, parametric methods are not suitable for segmentation of cluttered regions. In this work, we use image intensity histograms to drive the segmentation process, which makes no simplifying assumptions about the statistics of the image intensity values. It also does not rely on any geometric content found in the regions. We thus segment images purely based on histogram information found within its various regions. Image histograms have been used extensively in PDE-based image processing. For instance, [4] and [16] are variational methods in histogram modification for contrast enhancement.

There are a number of nonparametric segmentation models in the literature that are closely related to our work. In [10,7], the authors propose to maximize the mutual information between the region labels and the image intensities. In [3,1], the proposed model is to minimize the chi-2 comparison function between the object (resp. background) histogram and the object (resp. background) reference histogram. Their experimental results show effectiveness in segmenting slightly-textured images, e.g. human faces. However, the chi-2 comparison function is not a metric and is not suitable for comparing histograms in many situations. As a simple demonstration, the chi-2 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. To overcome this issue, we propose to use the Wasserstein distance (Monge-Kantorovich distance) to compare histograms. The Wasserstein distance between two normalized histograms is the least work required to move the region lying under the graph of one histogram to that of the other. It extends as a metric to measures such as the delta function. We believe this to be the more natural and appropriate way to compare histograms, since it does not suffer from the shortcoming mentioned above concerning point-wise metrics such as the standard L^p norms or the chi-2 comparison function. Experimental results show that there is indeed a significant benefit using the Wasserstein distance to compare histograms, and that it is quite effective in segmenting images consisting of cluttered regions. Optimal transport ideas have been used in other context in image processing, such as [8] on image registration and morphing and many others [2], [6] and [15].

The layout of the paper is as follows. Section II presents facts from optimal transportation theory used in this paper; in particular, we describe briefly the Monge-Kantorovich problem and how to solve it. Section III consists of two subsections, each one devoted to one of proposed new models. Also, level set formulations of these new models and their associated optimality conditions and gradient descent equations are given here. Section IV shows the algorithms and

discretization for solving the proposed models. Section V shows experimental results and comparison with other methods in both synthetic and real images.

2 Wasserstein Distance

The original Monge-Kantorovich problem was first posed in 1781 by G. Monge in [11]: what is the minimum work required to move a pile of dirt into a hole with the same volume? The original mathematical formulation turned out to be a difficult problem because it requires that no mass be split. Kantorovich proposed a relaxed version in [9], which we summarize in the following. 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 $\Pi(\mu, \nu) = \{\pi \in P(X \times Y) : \pi[A \times Y] = \mu[A], \text{ and } \pi[X \times B] = \nu[B] \text{ hold for all measurable sets } A \in X \text{ and } B \in Y\}$ be the set of admissible transference plans. For a given cost function $c : X \rightarrow Y$, the total transportation cost, associated to $\pi \in \Pi(\mu, \nu)$, is $I[\pi] = \int_{X \times Y} c(x, y) d\pi(x, y)$. The optimal transportation cost between μ and ν is $T_c(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} I[\pi]$. More detail can be found in [19] and [14], which is a good exposition on this subject.

In this paper, we are interested in the case when the probability is on the real line. Let μ and ν be two probability measures on \mathbb{R} , with respective cumulative distribution functions F and G . Then, it is known that for a convex cost function $c(x, y)$, the optimal transportation cost is $T_c(\mu, \nu) = \int_0^1 c(F^{-1}(t), G^{-1}(t)) dt$. In particular, the optimal transportation cost for the linear cost function $c(x, y) = |x - y|$ is $T_1(\mu, \nu) = \int_0^1 |F^{-1}(t) - G^{-1}(t)| dt$ and by Fubini's Theorem, $T_1(\mu, \nu) = \int_0^1 |F(t) - G(t)| dt$.

In the proposed models, we use the Wasserstein distance to determine the similarities between two normalized image histograms. Let $P_a(y)$ and $P_b(y)$ be two normalized histograms and let $F_a(y)$ and $F_b(y)$ be their corresponding cumulative distributions. The linear Wasserstein distance (W_1 distance) between $P_a(y)$ and $P_b(y)$ is defined by

$$W_1(P_a, P_b) = T_1(P_a, P_b) = \int_0^1 |F_a(y) - F_b(y)| dy. \quad (1)$$

An important consequence of this definition is that, unlike chi-2 function, Wasserstein distance is a metric. If two δ -functions are close by, the Wasserstein distance between them is small, because the area between their corresponding cumulative distribution functions is small.

3 Proposed Models

In this paper, we propose two segmentation energy models using the W_1 distance. By minimizing these energies, we hope to find an optimal region such that the region boundaries match the clutter boundaries. The first proposed model requires reference object (resp. background) histograms as inputs; this is the same

setting as in [1]. The first model is to minimize the Wasserstein distance between object (resp. background) histogram and the object (resp. background) reference histogram, together with a geometric regularization term on the interface. The second model do not require any reference histograms and assumes that the local histograms within the object region (resp. background region) are similar everywhere. We assign each pixel a neighborhood histogram, the histogram of a small neighborhood around that pixel. This model is to find an optimal region such that the object (resp. background) histogram is similar to all the neighborhood histograms inside (resp. outside) the region.

Given a grey scale image $I : \Omega \rightarrow [0, 255]$, the normalized image histogram restricted on the region Σ and the associated cumulative distribution function can be written in the following level set representation

$$P_{\Sigma}(y) = \frac{\int_{\Omega} H(\phi(z))\delta(y - I(z))dz}{\int_{\Omega} H(\phi(z))dz} \tag{2}$$

and

$$F_{\Sigma}(y) = \frac{\int_{\Omega} H(\phi(z))H(y - I(z))dz}{\int_{\Omega} H(\phi(z))dz} , \tag{3}$$

where $y \in [0, 255]$ is an intensity value, ϕ is a level set function [13] such that $\Sigma = \{x \in \Omega : \phi(x) > 0\}$, and δ and H are the Dirac and Heaviside function, respectively. Similarly, using the same ϕ for outside the region Σ^c , we have

$$P_{\Sigma^c}(y) = \frac{\int_{\Omega}[1 - H(\phi(z))]\delta(y - I(z))dz}{\int_{\Omega}[1 - H(\phi(z))]dz} \tag{4}$$

and

$$F_{\Sigma^c}(y) = \frac{\int_{\Omega}[1 - H(\phi(z))]H(y - I(z))dz}{\int_{\Omega}[1 - H(\phi(z))]dz} . \tag{5}$$

We use the level set method [13], because it allows changes of topology, such as merging and splitting. We normalize histograms because two identical clutters of different sizes should have zero distance.

3.1 Histogram Segmentation with Reference Histograms

For the first segmentation model, we are given a foreground reference histogram $P_f(y)$ and a background reference histogram $P_b(y)$. The model is

$$\inf_{\Sigma} E_1(\Sigma) = \text{Per}(\Sigma) + \lambda\{W_1(P_{\Sigma}, P_f) + W_1(P_{\Sigma^c}, P_b)\} ,$$

where W_1 is the W_1 distance described in (1). The first term is the length of the boundary of Σ , as a regularization term. The second (resp. third) is a fitting term between the object (resp. background) histogram and object (resp. background) reference histogram. The level set formulation of (6) is

$$\inf_{\phi} E_1(\phi) = \int_{\Omega} |\nabla H(\phi(x))| dx + \lambda \left\{ \int_0^{255} |F_{\Sigma}(y) - F_f(y)| dy + \int_0^{255} |(F_{\Sigma^c}(y) - F_b(y))| dy \right\},$$

where we plug in F_{Σ} and F_{Σ^c} by (3) and (5), respectively.

To minimize the energy, we derive the associated Euler-Lagrange equation. The gradient descent for ϕ is given by the following evolution equations

$$\phi_t = \delta(\phi) \left[\nabla \cdot \left(\frac{\nabla \phi}{|\nabla \phi|} \right) - \lambda(A - B) \right], \tag{6}$$

where

$$A = \frac{1}{Area(\Sigma)} \int_0^{255} \frac{F_{\Sigma}(y) - F_f(y)}{|F_{\Sigma}(y) - F_f(y)|} [H(y - I(x)) - (F_{\Sigma}(y))] dy$$

and

$$B = \frac{1}{Area(\Sigma^c)} \int_0^{255} \frac{F_{\Sigma^c}(y) - F_b(y)}{|F_{\Sigma^c}(y) - F_b(y)|} [H(y - I(x)) - (F_{\Sigma^c}(y))] dy .$$

3.2 Histogram Segmentation with Neighborhood Histograms

We modify the first segmentation model (6) so that input reference histograms are not required. For simplicity, we assume that the image of interest has two regions, object and background region, each of which has the same histograms locally (e.g. clutter features). The histogram restricted on a small region (neighborhood histogram) is similar to either the object histogram or the background histogram. Therefore, we compare the object (resp. background) histogram with all the neighborhood histograms in the object (resp. background) region.

For each point $x \in \Omega$, we compute the neighborhood cumulative distribution function

$$F_{x,r}(y) = \frac{Area(\{x \in B_r(x) : I(x) \leq y\})}{Area(\{B_r(x)\})} .$$

The size r of the neighborhood is chosen according to the clutter features in an image. It needs to be greater than or equal to the size of the clutter feature. For an accurate result, it should not be too large. In this paper, the selection of the size is specified by the user. The proposed model is

$$\inf_{\Sigma} E_2(\Sigma) = Per(\Sigma) + \lambda \left\{ \int_{\Sigma} W_1(P_1, P_{x,r}) dx + \int_{\Sigma^c} W_1(P_2, P_{x,r}) dx \right\} . \tag{7}$$

In a level set formulation, (7) becomes

$$\inf_{\Sigma} E_2(\Sigma) = \int_{\Omega} |\nabla H(\phi(x))| dx$$

$$\begin{aligned}
 & + \lambda \left\{ \int_{\Omega} H(\phi(x)) \int_0^{255} |F_1(y) - F_{x,r}(y)| dy dx \right. \\
 & \left. + \int_{\Omega} [1 - H(\phi(x))] \int_0^{255} |F_2(y) - F_{x,r}(y)| dy dx \right\}. \tag{8}
 \end{aligned}$$

Note that $F_{x,r}(y)$'s need to be computed only once before optimization. $F_1(y)$ and $F_2(y)$ are two constant cumulative distribution to be determined, independent of ϕ .

To minimize this energy, we first fix ϕ and minimize with respect to $F_1(y)$ and $F_2(y)$, respectively. Then, we fix $F_1(y)$ and $F_2(y)$ and minimize with respect to ϕ . The evolution equations are

$$\begin{aligned}
 F_1(y) &= \frac{\int_{\Omega} H(\phi(x)) F_{x,r}(y) dx}{\int_{\Omega} H(\phi(x)) dx} \\
 F_2(y) &= \frac{\int_{\Omega} [1 - H(\phi(x))] F_{x,r}(y) dx}{\int_{\Omega} [1 - H(\phi(x))] dx}
 \end{aligned}$$

$$\phi_t = \delta(\phi) \left[\nabla \cdot \left(\frac{\nabla \phi}{|\nabla \phi|} \right) - \lambda \int_0^1 (|F_1(y) - F_{x,r}(y)| - |F_2(y) - F_{x,r}(y)|) dy \right] \tag{9}$$

As the evolution equations suggest, the object (resp. background) cumulative distribution function F_1 (resp. F_2) is the average of all the neighborhood cumulative distribution functions $F_{x,r}$ inside (resp. outside) the curve. The minimization forces the 0-level curve of ϕ to move toward the boundaries of the object, so that the object (resp. background) cumulative distribution function is similar to all the neighborhood cumulative distribution histograms inside (resp. outside) the curve.

4 Numerical Method

For numerical implementation, we use a C^∞ regularized Heaviside function and the corresponding regularized Dirac function as follows

$$H_\varepsilon(z) = \frac{1}{2} \left(1 + \frac{2}{\pi} \arctan \left(\frac{z}{\varepsilon} \right) \right), \text{ and } \delta_\varepsilon(z) = \frac{1}{\pi} \frac{\varepsilon}{\varepsilon^2 + z^2}.$$

The evolution equations (6) and (9) for both proposed models have the following form

$$\phi_t = \delta(\phi) \left[\nabla \cdot \left(\frac{\nabla \phi}{|\nabla \phi|} \right) + \lambda A(\phi) \right].$$

We compute ϕ by the following discretization

$$\frac{\phi^{n+1} - \phi^n}{\Delta t} = \delta_\epsilon(\phi^n) \left[\Delta_x^- \left(\frac{\Delta_x^+ \phi^n}{|\nabla \phi^n|} \right) + \Delta_y^- \left(\frac{\Delta_y^+ \phi^n}{|\nabla \phi^n|} \right) + \lambda A(\phi^n) \right],$$

where

$$\begin{aligned} |\nabla \phi^n| &= \sqrt{(\Delta_x^+ \phi^n)^2 + (\Delta_y^+ \phi^n)^2 + \epsilon}, \\ \Delta_x^- \phi_{i,j} &= \phi_{i,j} - \phi_{i-1,j}, \Delta_x^+ \phi_{i,j} = \phi_{i+1,j} - \phi_{i,j}, \\ \Delta_y^- \phi_{i,j} &= \phi_{i,j} - \phi_{i,j-1}, \Delta_y^+ \phi_{i,j} = \phi_{i,j+1} - \phi_{i,j}. \end{aligned}$$

In the evolution equation (6), the corresponding $A(\phi)$ term can be written as

$$\int B(y) [H(y - f(x)) - C(y)] dy = \int B(y) C(y) dy + \int B(y) H(y - f(x)) dy, \quad (10)$$

for some functions $B(y)$ and $C(y)$.

Note that the first term is independent of x , while the second term can be simplified as

$$\int B(y) H(y - f(x)) dy = \int_{f(x)}^{255} B(y) dy.$$

Now, we only need to compute once

$$G(i) = \int_i^{255} A(y) dy$$

for $i \in \{0, 1, \dots, 255\}$. Then, the second term of the right hand side of (10) can be obtained fast by looking up $G(f(x))$ and by linear interpolation.

5 Experimental Results

We show and compare the proposed segmentation methods with some of existing methods. Figure 1 shows a 144×144 synthetic image, which has three regions with different distributions, as shown in Fig. 2. The inner region and the middle region look distinct, as well as their corresponding histograms, even though the histograms overlap 50 percents. On the other hand, the middle region and the outer region look similar, as well as their corresponding histograms, even though the histograms do not overlap at all. In both cases, the degree of similarity in image regions agree with the degree of similarity in their corresponding histograms. Figure 3 shows results of proposed and existing segmentation methods. The first row is the final contour, corresponding histograms, and cumulative distributions (from left to right) of proposed segmentation with reference histograms. The foreground and background reference histograms are obtained by

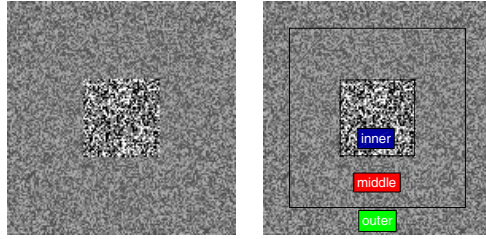


Fig. 1. Left: synthetic image. Right: boundaries between inner, middle, and outer regions.

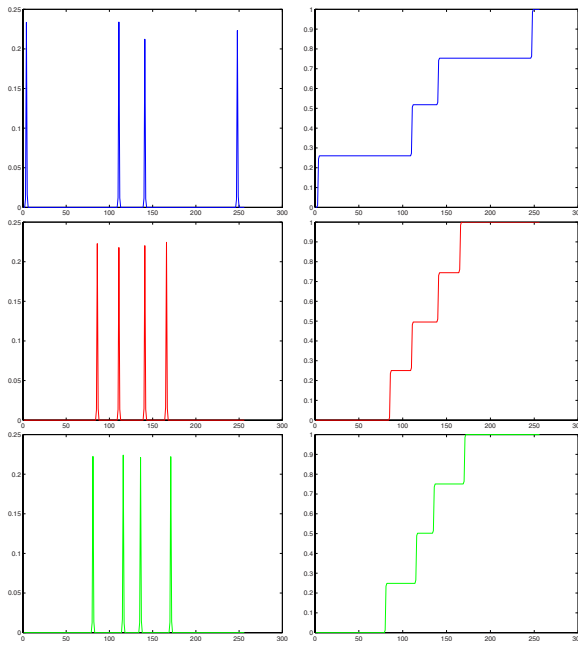


Fig. 2. From top to bottom. Left column: inner, middle, and outer region histograms. Right column: inner, middle, and outer region cumulative distributions.

calculating on the inner and outer region, respectively. The final contour shows that the proposed model is able to segment the middle and the outer region as background. The second row is the final contours, corresponding histograms, and cumulative distributions of proposed segmentation model with neighborhood histograms. This proposed method is also able to distinguish the foreground (inner region) from the background (middle and outer region). This shows that the W_1 distance is effective in histogram segmentation. The third row is the final contour, corresponding histograms, and cumulative distributions of segmentation with reference histograms using chi-2 function. Since this model strongly favors

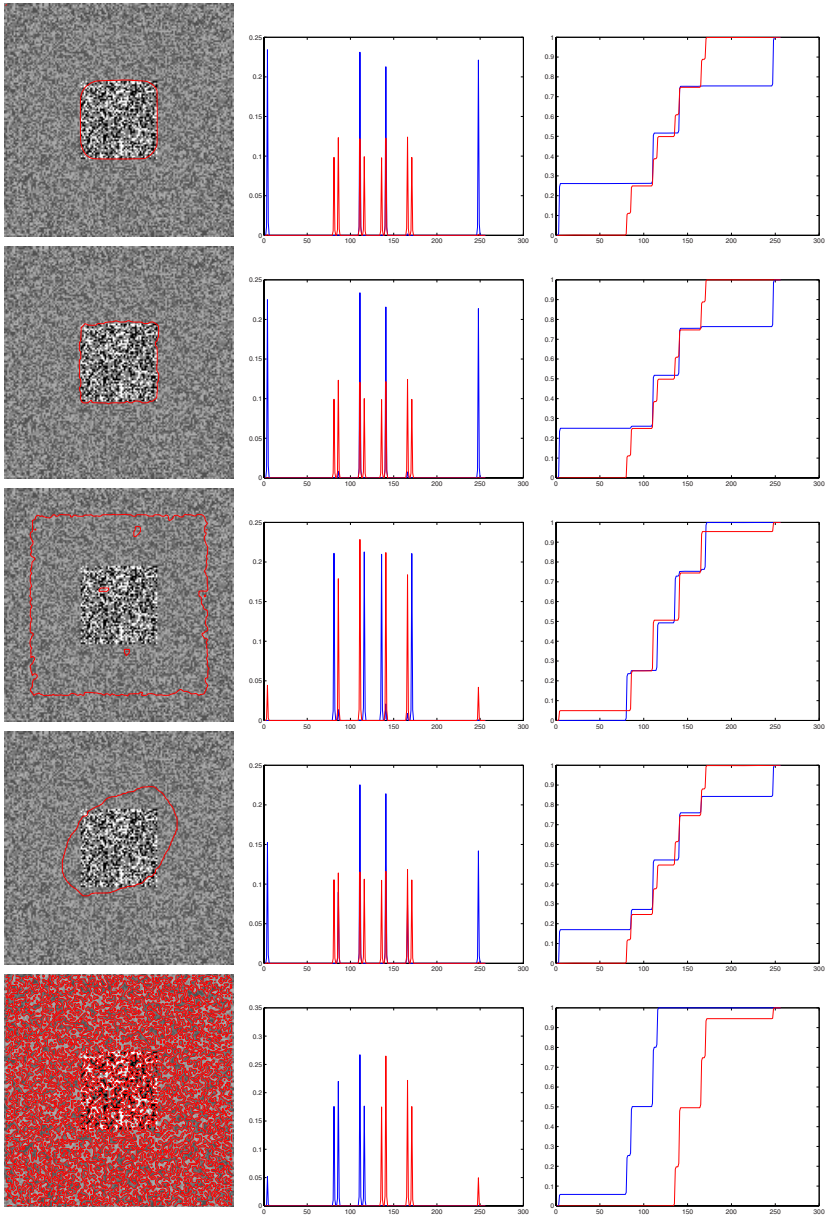


Fig. 3. Comparison of proposed histogram segmentation and existing methods. The first column is the final contour of different segmentation methods. The second (resp. third) columns are corresponding foreground and background histograms (resp. cumulative distributions). First row: proposed histogram segmentation with reference histograms. Second row: proposed histogram segmentation with neighborhood histograms. Third row: histogram segmentation using chi-2 function with reference histograms. Fourth row: Chan-Vese segmentation. Fifth row: Chan-Vese segmentation.

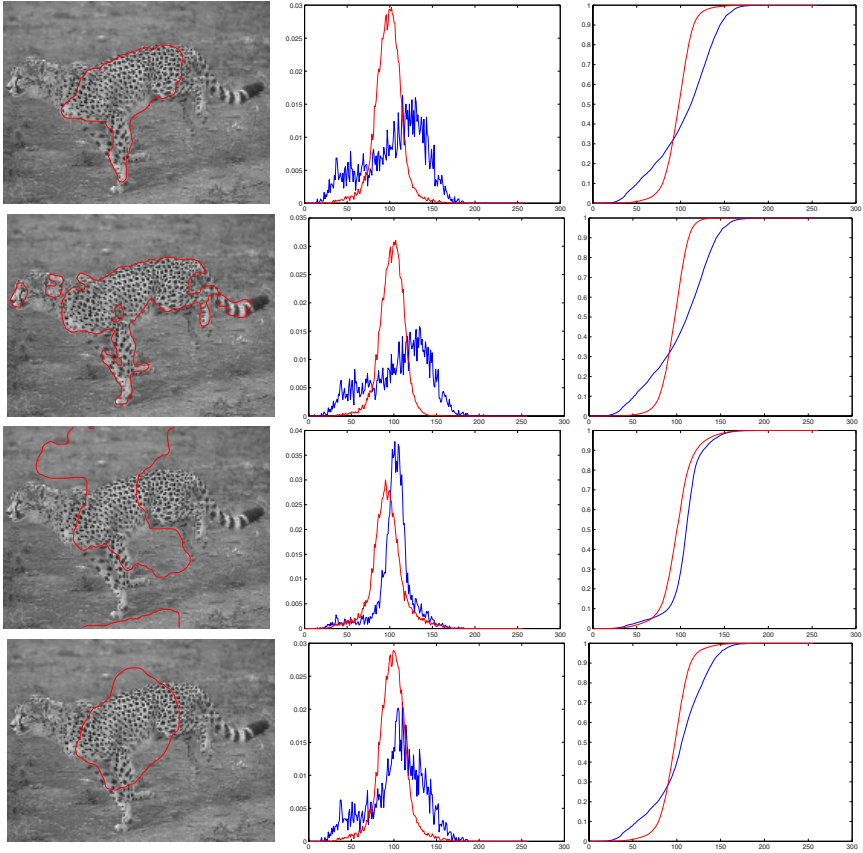


Fig. 4. Comparison of proposed histogram segmentation and existing methods. The first column is the final contour of different segmentation methods. The second (resp. third) columns are corresponding foreground and background histograms (resp. cumulative distributions). First row: proposed histogram segmentation with reference histograms. Second row: proposed histogram segmentation with neighborhood histograms. Third row: histogram segmentation using chi-2 function with reference histograms. Fourth row: Chan-Vese segmentation.

overlapping histograms, the middle region is segmented falsely as foreground. The fourth and fifth columns are the final contours, corresponding histograms, and cumulative distributions of Chan-Vese segmentation, with different fidelity parameters. The fourth row shows that the Chan-Vese segmentation is not able to come close to a correct segmentation. The fifth row shows that the Chan-Vese segmentation, with larger fidelity parameters, segments at a pixel level, in order to distinguish foreground and background intensity values. In any case, the standard Chan-Vese segmentation fails the task because the average intensity of any region in this image is the same.

Figure 4 shows segmentation results of various methods for a 135×175 real image. This image has complicated features, i.e. cheetah patterns, that are impossible to segment by existing methods. The first row is the final contour, corresponding histograms, and cumulative distributions of proposed histogram segmentation with reference histograms. The given foreground reference histogram is obtained by calculating the histogram on a small patch of the cheetah. The second row is the final contours, corresponding histograms, and cumulative distributions by segmentation with neighborhood histograms. As the results show, our proposed models proficiently segment the cheetah patterns. The third row is the final contour, corresponding histograms, and cumulative distributions of segmentation with reference histograms using chi-2 function. The fourth row is the final contour, corresponding histograms, and cumulative distributions of Chan-Vese segmentation. Both the existing segmentation methods fail to segment the cheetah pattern from the background. In this example, our proposed models outperform in segmenting clutters.

6 Conclusion

In this work, we propose a novel nonparametric region-based active contour model for segmenting clutter images. It is based on the use of Wasserstein mass transfer metrics in comparing histograms of different regions in the image. Our numerical results corroborate that these metrics are more suitable for histogram comparisons than what has been utilized previously in the existing literature, and lead to substantially better segmentations. Wasserstein metrics can be incorporated into a variety of histogram and curve evolution based segmentation models; we give two examples of such in this paper in order to substantiate our claims.

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, No. 2, pp.2128-2145, 2005.
2. R.E.Broadhurst, Statistical estimation of histogram variation for texture classification, In Texture 2005: Proceedings of the 4th International Workshop on Texture Analysis and Synthesis, pp. 25-30, 2005.
3. S.Jehan-Besson, M.Barlaud, G.Aubert, O.Faugeras, Shape Gradients for Histogram Segmentation Using Active Contours, in Proc. Int. Conf. Computer Vision, Nice, France, 2003, pp.408-415.
4. 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.
5. T.F.Chan, L.A.Vese, Active contours without edges, IEEE Transactions on Image Processing. 10:2(2001), pp. 266-277.
6. R.Chartrand, K.Vixie, B.Wohlber, E.Bollt, A Gradient Descent Solution to the Monge-Kantorovich Problem.

7. 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.
8. S.Haker, L.Zhu, and A.Tannenbaum, Optimal Mass Transport for Registration and Warping, International Journal of Computer Vision 60(3), 225-240,2004.
9. L.V.Kantorovich, On the translocation of masses, C.R.(Doklady) Acad. Sci. URSS(N.S.), 37(1942),pp.199-201.
10. J.Kim, J.W.Fisher, A.Yezzi, M.Cetin, and A.S.Willsky, Nonparametric methods for image segmentation using information theory and curve evolution, in ICIP, 2002, pp. III:797-800.
11. G.Monge, Mémoire sur la théorie des déblais at des remblais, Histoire de l'Académie Royale des Sciences de Paris, pp.666-704, 1781.
12. 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.
13. S.Osher and J.A.Sethian, Fronts propagation with curvature-dependent speed: Algorithms based on Hamilton-Jacobi Formulation, J. Comput. Phys., vol. 79, pp. 12-49, 1988.
14. S.Rachev and L.Rüschendorf, Mass transportation problems. Vol.I: Theory, Vol.II: Applications.Probability and its applications. Springer-Verlag, New York, 1998.
15. Y.Rubner, C.Tomasi, and L.J.Guibas, A metric for distributions with applications to image databases, In IEEE International Conference on Computer Vision, pp. 59-66, Bombay, India, January 1998.
16. G.Sapiro and V.Caselles, Histogram modification via differential equations, Journal of Differential Equations 135:2, pp. 238-268, 1997.
17. Z.Tu and S.Zhu, Image segmentation by data-driven Markov Chain Monte Carlo, in IEEE transactions on pattern analysis and machine intelligenc, Vol.24, No.5, May 2002.
18. 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.
19. C.Villani, Topics in Optimal Transportation. Graduate Studies in Mathematics, Vol. 58, American Mathematical Society, Providence Rhode Island, 2003.