2013 Seventh International Conference on Image and Graphics

Variational Earth Mover's Distance for Image Segmentation

Peihua Li
School of Communication and Information Engineering
Dalian University of Technology
Dalian, China
peihuali@dlut.edu.cn

Qilong Wang
School of Computer Science and Technology
Heilongjiang University
Harbin, China
wangqilong.415@163.com

***Abstract*—Image segmentation using similarity or dissimilarity measures between probability distributions has been of great research interest in recent years. It is shown that the cross-bin metrics such as EMD is superior to the bin-wise metrics. However, existing segmentation approaches involving EMD are limited to univariate distributions, or one-dimensional marginal distributions of multidimensional features. This paper presents a novel segmentation method based on the variational EMD (VEMD) model, which can exploit joint distributions of multidimensional features. This method formulates the segmentation problem as the minimization of the EMD-based functional, which measures the distance between the foreground (resp. background) distribution and the reference foreground (resp. background) distribution. Using the simplex method and theory of shape derivative, we minimize the functional and obtain the gradient descent flow. We use a Gaussian filtering level-set method to obtain the numerical solution, in which the level-set re-initialization and smoothness constraint commonly imposed by the contour length are not necessary any more. Experiments show that the proposed method outperforms the state-of-the-art segmentation methods in the presence of illumination changes and noise.**

***Keywords*-image segmentation; variational Earth Mover's Distance; multidimensional distribution;**

I. Introduction

Active contour models for image segmentation have been studied for many years. Region-based segmentation methods that use image region statistics are generally more robust to noise and are less likely trapped in the local minimum than the boundary-based ones[1]. The energy functionals are minimized via variational calculus or shape derivative, and the resulting gradient descent flows drive the evolution of region boundary. The level-set method [2] is widely employed in active contour models. One of its great benefit is that the topological changes of regions, such as splitting and merging, are naturally handled .

The first and second-order statistics, i.e., the mean or covariance, may suffice for nearly homogenous region segmentation [3]. But for noisy, cluttered or textured region segmentation, one may require probability distribution modeling of image regions. In these cases, the energy functionals are often based on similarity or dissimilarity measures between probability distributions. It has been shown that the cross-bin metrics such as the Earth Mover's Distance (EMD) are more suitable for segmentation than the bin-wise metrics, e.g., Bhattacharyya distance, Kullback-Leibler (K-L) divergence or chi-squared distance [4], [5], [6]. The underlying reason is that the cross-bin metrics take into account the possible drift of features to adjacent histogram bins, which may be induced by various effects of, say, noise, illumination variation, viewing angle changes or histogram binning scheme. However, existing segmentation approaches that involve EMD are limited to univariate distributions, or one-dimensional (1-D) marginal distributions of multidimensional features (hereafter abbreviated as 1-D EMD). Being unable to favorably utilize the joint distribution of multidimensional features, 1-D EMD may demonstrate unsatisfactory performance in delineating highly noisy, cluttered or textured regions.

In order to resolve the above problem, this paper presents a novel segmentation method based on the variational EMD (VEMD) model [7]. It can exploit joint probability distributions rather than only 1-D marginal distributions of multidimensional features. In our method, the segmentation problem is expressed as the minimization of an energy functional which measures the foreground (resp. background) distributions with the reference foreground (resp. background) distributions. We minimize the functional through the simplex method and theory of shape derivative to obtain the gradient descent flow. Gaussian filtering level-set method is used to obtain the numerical solution, in which the level-set re-initialization and smoothness constraint commonly imposed by the contour length are not necessary.

The structure of the remaining paper is organized as follows. Section II reviews the related papers. Section III formulates the EMD-based segmentations model. In section IV, a two phase method is presented for minimizing the energy functional. Section V compares the performance of the VEMD with 1-D EMD and Bhattacharyya gradient flows. Conclusion and future work are finally given in section VI.

II. Related Work

In [8], the energy functionals based on the Helling distance or chi-squared distance are proposed which compare the difference between the reference histogram and histogram of the image region. The functionals are minimized

978-0-7695-5050-3/13 $26.00 © 2013 IEEE
DOI 10.1109/ICIG.2013.53

using a shape derivative framework instead of the conventional calculus of variation [9]. Freedman et al. [10] introduced an image segmentation method by matching via K-L divergence the distributions of learned shape and appearance to medical imagery. In [11], image segmentation based on Bhattacharyy distance was investigated which maximizes the discrepancy between the empirical distributions inside and outside the contours.

The EMD [12], known as the Wasserstein distance in mathematics, is a cross-bin metric for measuring the dissimilarity between probability distributions. The EMD based on one-dimensional probability distributions has closed form and its applications to image segmentation are investigated independently by several researchers. In [6], the authors prefer the *cdf distance*, a special case of which has an identical form with 1-D EMD, to bin-wise metrics for distribution matching. They achieved promising segmentation results in difficult medical imagery. Ni et al. [4] formulated the image segmentation problem as the minimization of the EMD between the one-dimensional object histogram (resp. background) and the object reference (resp. background) histogram. Therein the 1-D EMD is shown to outperform the chi-squared distance. In [5], the shape derivative is used to derive the gradient flow associated with the 1-D EMD; combination of multiple EMD's involving 1-D marginal distributions is employed for image and scene segmentation. The above methods are limited to one-dimensional probability distributions and are unable to take advantage of joint distributions of multidimensional features.

The EMD is also employed for spatio-temporal object extraction (tracking) from consecutive frames in video. In [13], the Differential EMD (DEMD) was proposed for robust tracking in the presence of illumination changes. The DEMD is a two-phase optimization method, i.e., the best solution to the transportation problem (TP) followed by derivatives with respect to translation parameters. Based on the signature as a means of probability representation [12], the DEMD naturally takes advantage of multivariate distribution. It however can not follow the complex boundaries and non-rigid transformations of objects. In [14], 1-D EMD was used for object tracking which enables mean shift like iteration; an approach that combines histogram smoothing and bin-wise metrics is studied so that multi-dimensional distributions can be utilized. Li presented a variational EMD (VEMD) model [7] which used multidimensional distributions via signature for contour tracking.

Because of the benefits of the EMD, there is a lot of research interest on, aside from image segmentation and tracking based on EMD mentioned above, the novel EMD variants and approximations for improving its efficiency and robustness [15], [16], [17] or application in feature detector/descriptor [18], among others.

III. Image Segmentation based on VEMD

In this paper we use signature [12] to represent the distribution of an image region. For simplicity, we assume one image is composed of two regions–a foreground region and a background region. Suppose we are given a reference foreground (resp. background) signature, our segmentation method consists in seeking an image region whose boundary match the true object boundary.

Let $\mathbf{I}(\mathbf{z}) : \Omega \to \mathbb{R}^n$ be a color image ($n = 3$) or a feature map, where $\mathbf{z}$ denotes the spatial coordinate. We use the level-set function $\phi(\mathbf{z})$ to represent image region of complex boundary, where $\Gamma = \{\mathbf{z} : \phi(\mathbf{z}) = 0, \mathbf{z} \in \Omega\}$ denotes the boundary between the inner, foreground region $\{\mathbf{z} : \phi(\mathbf{z}) < 0, \mathbf{z} \in \Omega\}$ and outer, background one $\{\mathbf{z} : \phi(\mathbf{z}) > 0, \mathbf{z} \in \Omega\}$.

Suppose that the feature space of the foreground object is divided into N_f subspaces (or histogram bins). By using the Heaviside function H, the signature of the inner region can be represented by $\{\mathbf{h}_u^f, q_u^f\}_{u=1,\ldots,N_f}$, where $\mathbf{h}_u^f$ and q_u^f denote the center and frequency of points that fall into subspace u, respectively, i.e.,

$$\mathbf{h}_u^f(\phi) = \frac{\sum\limits_{\mathbf{z}\in\Omega}(1 - H(\phi(\mathbf{z})))\mathbf{I}(\mathbf{z})\delta(u, \mathbf{I}(\mathbf{z}))}{\sum\limits_{\mathbf{z}\in\Omega}(1 - H(\phi(\mathbf{z})))}, \qquad (1)$$

$$q_u^f(\phi) = \frac{\sum\limits_{\mathbf{z}\in\Omega}(1 - H(\phi(\mathbf{z})))\delta(u, \mathbf{I}(\mathbf{z}))}{\sum\limits_{\mathbf{z}\in\Omega}(1 - H(\phi(\mathbf{z})))},$$

where $\delta(u, \mathbf{I}(\mathbf{z}))$ is the Kronkecker delta function which equals unitary if $\mathbf{I}(\mathbf{z})$ belongs to subspace u and equals zero otherwise. Likewise, the signature of the background region can be represented by $\{\mathbf{h}_u^b, q_u^b\}_{u=1,\ldots,N_b}$, where N_b denotes the subspace number of the background feature space and $\mathbf{h}_u^b$, q_u^b have the following forms:

$$\mathbf{h}_u^b(\phi) = \frac{\sum\limits_{\mathbf{z}\in\Omega}H(\phi(\mathbf{z}))\mathbf{I}(\mathbf{z})\delta(u, \mathbf{I}(\mathbf{z}))}{\sum\limits_{\mathbf{z}\in\Omega}H(\phi(\mathbf{z}))}, \qquad (2)$$

$$q_u^b(\phi) = \frac{\sum\limits_{\mathbf{z}\in\Omega}H(\phi(\mathbf{z}))\delta(u, \mathbf{I}(\mathbf{z}))}{\sum\limits_{\mathbf{z}\in\Omega}H(\phi(\mathbf{z}))}.$$

Let $\{\mathbf{s}_u^f, p_u^f\}_{u=1,\ldots,N_f}$ and $\{\mathbf{s}_u^b, p_u^b\}_{u=1,\ldots,N_b}$ be the reference signatures of the foreground and background, respectively. The segmentation model is to seek a region boundary $\Gamma = \{\mathbf{z} : \phi(\mathbf{z}) = 0, \mathbf{z} \in \Omega\}$ such that the following

functional is minimized:

$$\underset{\phi}{\operatorname{argmin}} \Big(\alpha \underbrace{\min_{r^f_{uv}} \sum_{u=1}^{N_f} \sum_{v=1}^{N_f} d^f_{uv} r^f_{uv}(\phi)}_{g_f(r^f_{uv};\phi)} + (1-\alpha) \underbrace{\min_{r^b_{uv}} \sum_{u=1}^{N_b} \sum_{v=1}^{N_b} d^b_{uv} r^b_{uv}(\phi)}_{g_b(r^b_{uv};\phi)} \Big), \tag{3}$$

$s.t.$

$$\begin{cases} \sum_{u=1}^{N_f} r^f_{uv}(\phi) = q^f_v(\phi) & 1 \le v \le N_f \\ \sum_{v=1}^{N_f} r^f_{uv}(\phi) = p^f_u & 1 \le u \le N_f \\ r^f_{uv} \ge 0 & 1 \le u, v \le N_f \end{cases} \quad \text{and} \tag{4}$$

$$\begin{cases} \sum_{u=1}^{N_b} r^b_{uv}(\phi) = q^b_v(\phi) & 1 \le v \le N_b \\ \sum_{v=1}^{N_b} r^b_{uv}(\phi) = p^b_u & 1 \le u \le N_b \\ r^b_{uv} \ge 0 & 1 \le u, v \le N_b \end{cases}, \tag{5}$$

where $0 \le \alpha \le 1$ is a constant regulating the relative weights between the two terms (in this paper we set $\alpha = 0.65$ to emphasize the role of foreground), d^f_{uv} (resp. d^b_{uv}) denotes the cost to transport unit mass from the subspace u to the subspace v for the inner (resp. outer) region. It may be computed as the Euclidean distance between $\mathbf{h}^f_u$ (resp. $\mathbf{h}^b_u$) and $\mathbf{s}^f_v$ (resp. $\mathbf{s}^b_v$).

IV. Functional Optimization

Following [7], we use a two phase method to optimize the functional (3). We first fix $\phi(\mathbf{z})$ and minimize $g(r^f_{uv};\phi)$ (resp. $g(r^b_{uv};\phi)$) with respect to r^f_{uv} (resp. r^b_{uv}), subject to (*s.t.*) the corresponding constraint equations using the simplex algorithm. Next, we fix r^f_{uv} and r^b_{uv} and minimize with respect to $\phi(\mathbf{z})$ to derive the gradient descent flow.

A. Simplex Algorithm to Solve TP

Minimization of $g(r^f_{uv};\phi)$ *s.t.* constraints (4) and that of $g(r^f_{uv};\phi)$ *s.t.* constraints (5) are independent and identical. In the following, without ambiguity, we omit the subscript or superscript f or b that denote foreground or background.

Let ϕ be fixed, the minimization of $g(r_{uv};\phi)$ *s.t.* the corresponding constraints reduces to a classical balanced TP. It is known that the optimal solutions to such a class of problems always exist, and any one of the constraint equations can be dropped while the remaining system of $M = 2N - 1$ equations is linearly independent [19, Sec. 6.1].

We write $g(r_{uv};\phi)$ and the corresponding constraint equations in matrix form as (neglect the last equation in the constraints):

$$\begin{aligned} &\min_{\mathbf{x}} \ f = \mathbf{c}^T \mathbf{x}, \\ &s.t. \ \ \mathbf{A}\mathbf{x} = \mathbf{b}, \ \mathbf{x} \ge \mathbf{0}, \end{aligned} \tag{6}$$

where $\mathbf{c}$ and $\mathbf{x}$ are N^2-dimensional column vectors, $\mathbf{b}$ is an M-dimensional column vector and $\mathbf{A}$ is an M-by-N^2 matrix. $\mathbf{c}$, $\mathbf{x}$, $\mathbf{b}$ and $\mathbf{A}$ have the following forms, respectively:

Let $\mathbf{x}^T = [\mathbf{x}^T_{\mathbf{B}} \ \mathbf{x}^T_{\mathbf{N}}]$ be an initial feasible solution, where $\mathbf{x}_{\mathbf{B}}$ and $\mathbf{x}_{\mathbf{N}}$ denote the vectors of basic and non-basic variables, respectively. The matrix $\mathbf{A}$ can be accordingly written as $\mathbf{A} = [\mathbf{B} \ \mathbf{N}]$, where $\mathbf{B}$, comprising columns in $\mathbf{A}$ that corresponds to the basic variables, is a basis of Euclidean space R^M, and $\mathbf{N}$ consists of columns in $\mathbf{A}$ corresponding to the non-basic variables.

In the simplex algorithm, first, we compute the relative cost vector $\mathbf{r}^T_N = \mathbf{c}^T_{\mathbf{B}}\mathbf{B}^{-1}\mathbf{N} - \mathbf{c}^T_{\mathbf{N}}$. If $\mathbf{r}^T_N \not\le \mathbf{0}$, determine its largest, positive component and the corresponding column in $\mathbf{N}$, denoted by $\mathbf{n}_j$, is selected as the vector to enter the basis $\mathbf{B}$; otherwise, the current solution is optimal and stop the iteration. Next, we compute $\mathbf{y}_j = \mathbf{B}^{-1}\mathbf{n}_j$ and $\tilde{\mathbf{b}} = \mathbf{B}^{-1}\mathbf{b}$ and decide the row index i for which $i = \max_k\{\tilde{b}_k / y_{kj}, y_{kj} > 0\}$, where y_{kj} and $\tilde{b}_k$ are k^{th} entries of $\mathbf{y}_j$ and $\tilde{\mathbf{b}}$, respectively. The column vector in $\mathbf{B}$ that corresponds to the ith basic variable of $\mathbf{x}_{\mathbf{B}}$ is selected to leave the basis. Finally, Select y_{ij} as the pivot element, use the Gaussian elimination to transform the column $\mathbf{y}_j$ to the unit vector. In this way, the transformation of the basis variables are completed and in the meantime the coefficient matrix of basis variables remains the identity matrix. The above process iterates until we obtain the optimum solution.

The initial feasible solution can be obtained by the method introduced in [19, Sec. 6.2]. Through the simplex algorithm, we obtain the optimal solution to the TP and the corresponding cost function can be written as

$$g(r_{uv};\phi) = \mathbf{c}^T_{\mathbf{B}}\mathbf{B}^{-1}\mathbf{b} = \sum_{v=1}^{N} l_v q_v(\phi) + \sum_{u=1}^{N-1} h_u p_u, \tag{7}$$

where $l_v = \sum_{i=1}^{M} \mathbf{c}_{\mathbf{B}}(i)\mathbf{B}^{-1}(i,v)$, $h_u = \sum_{i=1}^{M} \mathbf{c}_{\mathbf{B}}(i)\mathbf{B}^{-1}(i,u+N)$. Here $\mathbf{c}_{\mathbf{B}}(i)$ denotes the i^{th} entry of the vector $\mathbf{c}_{\mathbf{B}}$, and $\mathbf{B}^{-1}(i,v)$ the $(i,v)^{\text{th}}$ entry of the matrix $\mathbf{B}^{-1}$.

B. Minimization with respect to Level-set Function

Remember that, by using the simplex algorithm, we have the optimal solutions to $g(r^f_{uv};\phi)$ *s.t.* constraints (4) regarding the foreground, and that to $g(r^b_{uv};\phi)$ *s.t.* constraints (5) regarding the background. Hence, leaving out the terms that are irrelevant to ϕ from Eq. (7), the functional (3) can be represented as

$$\underset{\phi}{\operatorname{argmin}} \Big(\alpha \underbrace{\sum_{u=1}^{N_f} l^f_u q^f_u(\phi)}_{g_f(r^f_{uv};\phi)} + (1-\alpha) \underbrace{\sum_{u=1}^{N_b} l^b_u q^b_u(\phi)}_{g_b(r^b_{uv};\phi)} \Big). \tag{8}$$

1) Gradient descent flow: For computing the Euler-Lagrange equation, we adopt the following C^∞–regularized

Heaviside and the corresponding Delta functions:

$$H_\varepsilon(x) = \frac{1}{2}\left(1 + \frac{1}{\pi}\arctan\left(\frac{x}{\epsilon}\right)\right), \quad (9)$$
$$\delta_\varepsilon(x) = \frac{dH_\varepsilon(x)}{dx} = \frac{1}{\pi}\frac{\varepsilon}{\varepsilon^2 + x^2}.$$

Using H_ε and δ_ε above, the algorithm has the tendency to obtain the global minimizer [3]. We set $\varepsilon = 3$ in all the experiments of this paper. We write the $q_u^f(\phi)$ and $q_u^b(\phi)$ in continuous forms as

$$q_u^f(\phi) = \frac{G_u^f(\phi)}{A_f(\phi)} \triangleq \frac{\int_\Omega (1 - H_\varepsilon(\phi))\delta(u, \mathbf{I})d\mathbf{z}}{\int_\Omega (1 - H_\varepsilon(\phi))d\mathbf{z}}, \quad (10)$$
$$q_u^b(\phi) = \frac{G_u^b(\phi)}{A_b(\phi)} \triangleq \frac{\int_\Omega H_\varepsilon(\phi)\delta(u, \mathbf{I})d\mathbf{z}}{\int_\Omega H_\varepsilon(\phi)d\mathbf{z}}.$$

The functional (8) can be minimized by the theory of shape derivative. After some derivation, we obtain the gradient descent flow

$$\frac{\partial \phi}{\partial \tau} = \big(-\alpha F_f + (1-\alpha)F_b\big)\delta_\varepsilon(\phi) \quad (11)$$

where

$$F_f = 1/A_f \sum_{u=1}^{N_f} l_u^f\big(\delta(u, \mathbf{I}) - q_u^f\big),$$
$$F_b = 1/A_b \sum_{u=1}^{N_b} l_u^b\big(\delta(u, \mathbf{I}) - q_u^b\big),$$

and τ denotes the artificial time.

2) Numerical implementation: Discretizing Eq. (11) in space and time, we have

$$\phi_{\tau+1}(i,j) = \phi_\tau(i,j) + \Delta\tau \quad (12)$$
$$\times \big(-\alpha F_f(i,j) + (1-\alpha)F_b(i,j)\big)\delta_\epsilon(\phi(i,j))$$

where $\Delta\tau$ denotes the discrete time step, $\phi_\tau(i,j)$ denotes the value of function ϕ at pixel point (i,j) at time τ. To guarantee numerical stability of the discrete equation, $\Delta\tau$ should satisfy the CFL condition[20]:

$$\Delta\tau \max_{(i,j)\in\Omega} \Big|\big(-\alpha F_f(i,j) + (1-\alpha)F_b(i,j)\big)\delta_\epsilon(\phi(i,j))\Big| < 1$$

The level-set function ϕ can be arbitrary but in practice the signed distance function (SDF) is usually preferred because of its good properties. During contour evolution, the level-set function may deviate from the SDF. So periodical re-initialization is required which makes the level-set function a true SDF while keeping the zero level-set unchanged. However, there lacks explicit criterion to determine how often one needs to perform such re-initialization. In addition, it is time consuming despite the fast algorithms like fast-marching method.

We adopt the Gaussian filtering level-set method [21]. One of the main steps in this method is to filter the level-set function, after each iteration, with a Gaussian function whose standard variance should satisfy $\sigma < \sqrt{\Delta\tau}$. Based on this method, we do not require the re-initialization process in the iteration; the smoothness of the level-set function is also ensured so that the traditional smoothness constraint, commonly imposed by the contour length term, is not necessary. In practice, $\sigma = 0.5$ and the windows size $n = 5$ of the Gaussian kernel are well suited for our problem.

C. Computational Complexity

In the following, we analyze the computational complexity of one iteration involved in the VEMD. The main computations are induced by the simplex algorithm, called two times–one for the foreground and the other for the background, the update of level-set function and the Gaussian filtering. It is known that, in the worst case, the computational complexity of the simplex algorithm is exponential in the problem size M, i.e., $M = 2N_f - 1$ (resp. $M = 2N_b - 1$) for the foreground (resp. background). Fortunately, if M is not large, the iteration number N_S for the simplex algorithm to converge from a feasible solution to the optimal one will be $N_S \approx 2M \sim 3M$ [19] ($N_S \approx 3M$ in our experiments). Hence, the computational complexity of simplex algorithm is $O(6(N_f + N_b - 2))$. It is clear that the computational complexity of the level-set function update is $O(N_I)$, where N_I is the image size. With regard to the Gaussian filtering, its complexity is $O(n^2 N_I)$, where n denotes the width of Gaussian kernel.

The running time of one iteration of VEMD is in average 265ms in the cheetah image with $N_f = N_b = 28$, $N_I = 77648$. The program is written in matlab code on a IBM T42 laptop equipped with Intel© Pentium© M Processor 1.80GHz and 1.5GB memory.

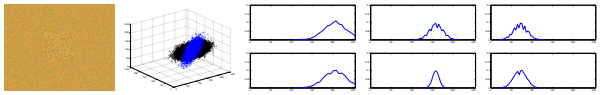
V. Experiments

All the programs are written in matlab code. Throughout the experiments, we use the RGB color space. We compare three gradient descent flows: the VEMD flow, 1-D EMD flow[4] and Bhattacharyya flow (BHAT)[11]. In 1-D EMD flow, every gray-level of three channels is quantized into 16 bins, and the sum of three 1-D EMD's is used based on the marginal 1D histograms of three channels. In BHAT flow, the $16 \times 16 \times 16$ joint histograms are used. In VEMD, the color spaces of the foreground (resp. background) is partitioned into 28 subspaces by K-Means algorithm.

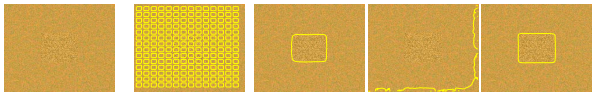
A. Comparisons in Synthetic Images

1) Synthetic image #1: The reference foreground and background distributions are obtained in synthetic image #1 as shown in Fig. 1(a). By zooming in the reference image (a) one can observe the foreground–the small rectangular region located at the center. The probability density function of the foreground is a 3-D Gaussian $N(\mu, \mathbf{\Sigma})$ with mean vector $\mu = [205.69\ 158.95\ 70.96]^T$ and covariance matrix $\mathbf{\Sigma} = [747.40\ 542.43\ 216.53; 542.43\ 65.18\ 328.50; 216.53\ 329.50\ 241.27]$

(written in the matlab notation), while that of the background is $N(\mu, \mathrm{diag}\{\mathbf{\Sigma}\})$, where $\mathrm{diag}\{\mathbf{\Sigma}\}$ denotes a diagonal matrix that has identical diagonal entries to $\mathbf{\Sigma}$. Hence, the foreground and background have identical marginal distributions. Fig. 1(b) shows the scatter plots of foreground (blue) and background (black) pixels, respectively; their respective marginal histograms are shown in Fig. 1(c). Fig. 1(e) shows the initialization curve, and Fig. 1(f) presents the convergence results of BHAT, 1-D EMD and VEMD (from left to right). The 1-D EMD fails because it can not distinguish the foreground from background since only marginal distributions are used. In contrast, the BHAT and VEMD which utilize the joint distributions converge to the true boundary.



(a) Reference image (b) Scatter plot (c) Marginal distributions of foreground (top) and background (bottom)



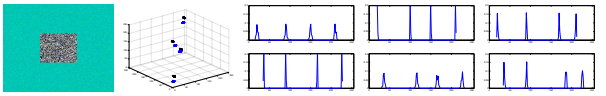
(d) Test image (e) Initialization (f) BHAT, 1-D EMD and VEMD (from left to right)

Figure 1. Comparison with synthetic image #1 in which the foreground and background have identical marginal distributions. By zooming in the reference image (a) one can observe the foreground–the small rectangular region located at the center. The test image (d) is identical to the reference image. 1-D EMD fails to handle this case while BHAT and VEMD succeeds.

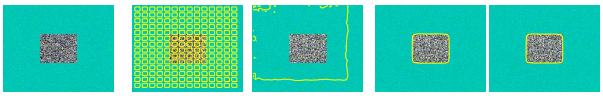
2) Synthetic image #2: In this synthetic image as shown in Fig. 2(a), the small region at the image center is the foreground. In the reference image, the gray-level distributions of three channels are independent, each following univariate Gaussian of various means and small variances. In the test image 2(d), the background distribution is identical to that in the reference image; the foreground distribution is similar to those in the reference image but have distinct means. Hence, the foreground distribution in the test image can be approximately seen as a shifted version of that in the reference one. Fig. 2(b) shows the scatter plots of the foregrounds in the reference (black) and test (blue) images; their respective marginal distributions are shown in Fig. 2(c). Fig. 2(f) presents the convergence results of BHAT, 1-D EMD and VEMD (from left to right) starting from the initialization contour 2(e). The BHAT flow which is based on bin-wise, Bhattacharyya distance, fail to handle this case. However, the 1-D EMD and VEMD flows which are based cross-bin metrics converge to the true boundary.

B. Comparisons in Real Images

We test the performance of three methods vs illumination changes and varying noise level. The reference zebra



(a) Reference image (b) Scatter plot (c) Marginal distributions of foreground in the reference (top) and test image (bottom)



(d) Test image (e) Initialization (f) BHAT, 1-D EMD and VEMD (from left to right)

Figure 2. Comparison with synthetic image #2. The reference image (a) and test image (d) have identical background distributions but different foreground ones. The foreground distribution in (d) can be approximately seen as a shifted version of that in (a). 1-D EMD and VEMD converges to the true boundary but BHAT does not.

image and cheetah image are shown in Fig. 3, from which the reference foreground and background distributions are obtained. Initializations of contours are similar to those in the previous section.

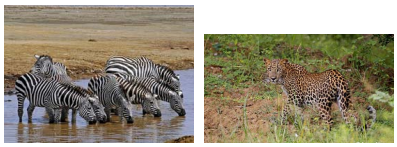


Figure 3. Reference zebra image and cheetah image.

We randomly change the illumination by scaling each gray-level of three color channels as follows:

$$I_S(x,y) = \gamma_S \cdot I_S(x,y), \quad S = \{R,G,B\},$$

where $I_S(x,y)$ denotes the gray-value of channel S at point (x,y), γ_S denotes a random number uniformly distributed on the interval $[0.7, 1.3]$. As to the noise, we corrupt every pixel through the formula

$$I_S(x,y) = I_S(x,y) + \lambda_S(x,y), \quad S = \{R,G,B\},$$

where $\lambda_S(x,y)$ has Gaussian distribution with zero mean and variance δ^2.

Fig. 4 shows the segmentation results of three gradient flows in the zebra images. As seen in Fig. 4(a), when illumination changes, the results of BHAT are very poor in that the contour almost always converges to nothing. In contrast, 1-D EMD and VEMD are more robust to illumination changes: the zebras are well segmented from the background. It can also be seen that the segmentation results of VEMD are more accurate than 1-D EMD. Fig. 4(b) shows the segmentation results in the presence of noise. DEMD accurately delineates the boundaries of zebras despite increasing noise levels. The results of 1-D EMD is inferior to those of EMD but is still satisfied. Invariably, BHAT does not work–the contour converges to nothing in all noisy images.

(a) Illumination changes. Top row: BHAT; middle row: 1-D EMD; bottom row: VEMD

(b) Noise pollution (variance of Gaussian increases from left to right). Top row: BHAT; middle row: 1-D EMD; bottom row: VEMD

Figure 4. Comparison of segmentation results (yellow curves) in zebra images. Note that the image with no yellow curves superposed indicates that the contour converges to nothing on that image.

(a) Illumination changes. Top row: BHAT; middle row: 1-D EMD; bottom row: VEMD

(b) Noise pollution (variance of Gaussian noise increases from left to right). Top row: BHAT; middle row: 1D-EMD; bottom row: VEMD

Figure 5. Comparison of segmentation results (yellow curves) in cheetah images. Note that the image with no yellow curves superposed indicates that the contour converges to nothing on that image.

The segmentation results in the cheetah images are shown in Fig. 5. Fig. 5(a) shows the results of three gradient flows against illumination variation. The VEMD successfully outlines the object boundary; the 1-D EMD has poor results in the cheetah images which contain illumination changes; the BHAT converges to nothing in all images but the second and third ones. Fig. 5(b) presents convergence of three gradient flows vs increasing noise levels: the BHAT exhibits very bad results; the 1-D EMD gets worse as the noise level increases; VEMD well segments the cheetah from the background in all five noisy images.

In view of the experiments above, we conclude that the bin-wise metrics such as Bhattacharyya distance is not suitable for segmentation of images that contains large illumination changes and noise level. The 1-D EMD shows relatively good segmentation results in the zebra images, but its performance deteriorates greatly in the cheetah images. The reason may be that the gray-level correlations of three channels in the zebra images are small. As such the marginal distributions is adequate to characterize their statistic in the zebra images. However, the probability distributions of the foreground and/or background in the cheetah images are complex so that only the joint distributions of three channels are sufficient for statistics modeling. In all cases, VEMD demonstrates strong robustness to illumination changes and noise.

VI. Conclusion and Future Work

The main contribution of this paper is a novel segmentation method based on the VEMD model. This method can exploit joint probability distributions of multidimensional features. It solved the weakness of the existing 1D EMD segmentation models, namely they are limited to univariate distributions or one-dimensional marginal distributions of multidimensional features. Hence, our method can make full use of the benefits of this cross-bin measure for robust segmentation. The experiments showed that our method is robust to illumination change and noise, outperforming the methods based 1-D EMD and the bin-wise metrics such as Bhattacharyya distance.

Through signature representation, more distinct, high-dimensional features are readily applicable to our method. It is interesting to adopt more powerful feature descriptors in the VEMD to handle segmentation in more challenging situations. In the current work, the VEMD requires the probability distributions of foreground and background be known *a prior*. As in [4], assuming the local distributions in the foreground (resp. background) are similar everywhere, we may develop a different VEMD that do not require the reference distributions. This issue will be studied in the future work.

ACKNOWLEDGMENT

The work was supported by the National Natural Science Foundation of China (60973080, 61170149), Program for New Century Excellent Talents in University (NCET-10-0151), the Fundamental Research Funds for the Central Universities (DUT13RC(3)02), Key Project by Chinese Ministry of Education (210063), and High-level professionals (innovative teams) of Heilongjiang University (Hdtd2010-07).

REFERENCES

[1] D. Cremers, M. Rousson, and R. Deriche, "A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape," *Int. J. of Comput. Vision*, vol. 72, no. 2, pp. 195–215, April 2007.

[2] S. Osher and J. A. Sethian, "Fronts propagating with curvature-dependent speed: algorithms based on hamilton-jacobi formulations," *J. Comput. Phys.*, vol. 79, pp. 12–49, 1988.

[3] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Trans. on Image Processing*, vol. 10, no. 2, pp. 266–277, 2001.

[4] K. Ni, X. Bresson, T. Chan, and S. Esedoglu, "Local histogram based segmentation using the wasserstein distance," *Int. J. of Comput. Vision*, vol. 84, no. 1, pp. 97–111, 2009.

[5] A. Adam, R. Kimmel, and E. Rivlin, "On scene segmentation and histograms-based curve evolution," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, pp. 1708–1714, 2009.

[6] D. Freedman, R. Radke, T. Zhang, Y. Jeong, D. Lovelock, and G. Chen, "Model-based segmentation of medical imagery by matching distributions," *IEEE Trans. on Medical Imaging*, vol. 24, no. 3, pp. 281–292, 2005.

[7] P. Li, "Tensor-sift based Earth Mover's Distance for contour tracking," *J. Math. Imaging Vis.*, vol. 46, no. 1, pp. 44–65, 2013.

[8] S. Jehan-Besson, M. Barlaud, G. Aubert, and O. Faugeras, "Shape gradients for histogram segmentation using active contours," in *Proc. of the Int. Conf. on Computer Vision*, 2003, p. 408.

[9] G. Aubert, M. Barlaud, O. Faugeras, and S. Jehan-Besson, "Image segmentation using active contours: Calculus of variations or shape gradients?" *SIAM Applied Mathematics*, vol. 63, no. 6, pp. 2128–2154, 2003.

[10] D. Freedman, R. J. Radke, T. Zhang, Y. Jeong, and G. T. Y. Chen, "Model-based multi-object segmentation via distribution matching," in *Proc. IEEE Conf. Comp. Vis. Patt. Recog. Workshop*, 2004.

[11] O. Michailovich, Y. Rathi, and A. Tannenbaum, "Image segmentation using active contours driven by the bhattacharyya gradient flow," *IEEE Trans. on Image Processing*, vol. 16, no. 11, pp. 2787–2801, 2007.

[12] Y. Rubner, C. Tomasi, and L. J. Guibas, "The earth mover's distance as a metric for image retrieval," *Int. J. of Computer Vision*, vol. 40, no. 2, pp. 99–121, 2000.

[13] Q. Zhao, Z. Yang, and H. Tao, "Differential earth mover's distance with its applications to visual tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 2, pp. 274–287, 2010.

[14] I. Leichter, "Mean shift trackers with cross-bin metrics," *IEEE Trans. Pattern Anal. Mach. Intell.*, no. 99, p. 1, 2011.

[15] O. Pele and M. Werman, "Fast and robust earth mover's distances," in *Proc. Int. Conf. Comp. Vis.*, 2009, pp. 460–467.

[16] H. Ling and K. Okada, "EMD-L1: An efficient and robust algorithm for comparing histogram-based descriptors," in *Proc. of European. Conf. on Comp. Vis.*, 2006, pp. 330–343.

[17] J. Rabin, G. Peyré, and L. D. Cohen, "Geodesic shape retrieval via optimal mass transport," in *Proc. of European. Conf. on Comp. Vis.*, 2010, pp. 771–784.

[18] R. T. Collins and W. Ge, "CSDD features: Center-surround distribution distance for feature extraction and matching," in *Proc. of European. Conf. on Comp. Vis.*, 2008, pp. 140–153.

[19] D. G. Luenberger and Y. Ye, *Linear and Nonlinear Programming*. Springer-Verlag, 2007.

[20] S. Osher and R. Fedkiw, *Level Set Methods and Dynamic Implicit Surfaces*. Springer-Verlag, 2002.

[21] K. Zhang, H. Song, and L. Zhang, "Active contours driven by local image fitting energy," *Pattern Recognition*, vol. 43, no. 4, pp. 1199–1206, 2010.