NIH-PA Author Manuscript

Image Segmentation Using Active Contours Driven by the Bhattacharyya Gradient Flow

Oleg Michailovich [Member, IEEE],
School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA. He is currently with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N3L 3G1 Canada (olegm@uwaterloo.ca)

Yogesh Rathi, and
School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA (yogesh.rathi@gatech.edu)

Allen Tannenbaum [Member, IEEE]
School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA and also with the Department of Electrical and Computer Engineering, The Technion —Israel Institute of Technology, Haifa, Israel (tannenba@ece.gatech.edu)

Abstract

This paper addresses the problem of image segmentation by means of active contours, whose evolution is driven by the gradient flow derived from an energy functional that is based on the Bhattacharyya distance. In particular, given the values of a photometric variable (or of a set thereof), which is to be used for classifying the image pixels, the active contours are designed to converge to the shape that results in *maximal* discrepancy between the empirical distributions of the photometric variable inside and outside of the contours. The above discrepancy is measured by means of the Bhattacharyya distance that proves to be an extremely useful tool for solving the problem at hand. The proposed methodology can be viewed as a generalization of the segmentation methods, in which active contours maximize the difference between a finite number of empirical moments of the "inside" and "outside" distributions. Furthermore, it is shown that the proposed methodology is very versatile and flexible in the sense that it allows one to easily accommodate a diversity of the image features based on which the segmentation should be performed. As an additional contribution, a method for automatically adjusting the smoothness properties of the empirical distributions is proposed. Such a procedure is crucial in situations when the number of data samples (supporting a certain segmentation class) varies considerably in the course of the evolution of the active contour. In this case, the smoothness properties of the empirical distributions have to be properly adjusted to avoid either over- or underestimation artifacts. Finally, a number of relevant segmentation results are demonstrated and some further research directions are discussed.

Index Terms

Active contours; Bhattacharyya distance; image segmentation; kernel density estimation

I. INTRODUCTION

The quintessential goal of image segmentation is to partition the image domain into a number of (mutually exclusive) subdomains over which certain properties of the image appear to be homogeneous. The homogeneity, however, turns out to be a somewhat vague notion when applied for characterizing localized regions of natural images, and, therefore, it should be used with a precaution whenever the problems that involve such images are dealt with. Thereupon, it is usually more convenient to look at image segmentation from a more practical perspective and think of it as a way to decompose an image into a number of its fragments, each of which can be associated with a distinct *class*. The latter is usually distinguished on a semantic basis and assumed to be either an *object* or a *background*. Thus, for example, the object may be associated with a diseased organ in medical imaging [1], [2], an intruder in surveillance video [3], [4], a moving part of a machine in robotics [5], [6], a maneuvering vehicle in traffic control [7], [8], or a target in navigation and military applications [9], [10].

A multitude of diverse image segmentation methods have been proposed over the last few decades. In spite of this diversity, however, the majority of these methods seem to follow a similar algorithmic pattern. The latter involves making hypotheses regarding the structure and properties of the image to be segmented, defining a set of the image *features* based on which segmentation classes are discriminated, and finally applying a *decision threshold* in either explicit or implicit manner [11]. Thus, for example, an image feature (or a collection thereof) can be considered as a random variable described by a set of conditional likelihood functions. Consequently, the Bayesian decision theory [12] could be used to determine the *Bayesian* decision threshold as a *minimizer* of the posterior probability of misclassification error.

Although the above segmentation approach is by no means generic for all types of segmentation methods proposed hitherto, it espouses an "ideology" that seems to be common for many problems of this kind. In particular, independently of whether the segmentation is based on a local [13], [14] or global [15], [16] analysis, whether it utilizes deformable contours [17], [18] or polygons [19], [20], the most successful segmentation methodology will be the one that minimizes the *probability of misclassification error*. Hence, this probability appears to constitute a universal criterion, based on which image segmentation methods may be designed.

Unfortunately, in many practical settings, little is usually known about the statistics of segmentation classes, and, as a result, an explicit definition of the probability of misclassification error is either very complicated or even impossible. In such situations, it is tempting to find alternative criteria, which are *simpler* to formulate, and based on which one could perform the segmentation with accuracy comparable to that of the methods based on *explicitly* minimizing the error probability.

It is worthwhile noting that, long before the problem of segmenting digital images has become an indispensable part of modern image processing and computer vision, the quest for the aforementioned criteria was undertaken in the fields of communication and radars [21]. Similar to the case of segmentation/ classification, suitable criteria were sought there to be used as substitutes for the *probability of detection error*. In this connection, it was found that some useful choices for such criteria can be based on the notion of *a distance* between probability distributions [21]. The latter, for instance, could be the distributions of observations in a binary hypothesis testing, in which case it can be shown that the further *apart* one can make these distributions, the smaller will be the probability of mistaking one for the other [22].

In order to define a distance between probability distributions, a number of information-theoretic measures can be used [23]. Some standard choices (which are also involved in basic results of information theory) include the Fisher ratio, the Kullback–Leibler (KL) divergence [24], [25], and the Bhattacharyya distance [22] (which is a special case of a more general distance introduced by Chernoff [26]).

When comparing the above distances in application to the problem of signal selection, it was observed in [21] that, in a number of practically important cases, the Bhattacharyya distance turns out to give better results as compared to the KL divergence. Furthermore, there is also a "technical" advantage of using the Bhattacharyya distance for its having a particularly simple analytical form. Specifically, the Bhattacharyya distance between two probability densities $p_1(z)$ and $p_2(z)$, with $z \in \mathbb{R}^N$, is defined as $-\log B$, where B is the Bhattacharyya coefficient given by

$$B = \int_{\mathbb{R}^N} \sqrt{p_1(z)p_2(z)} dz. \quad (1)$$

It is interesting to note that the functions $\sqrt{p_i(z)}$ (with $i \in \{1,2\}$) belong to the unit sphere of $\mathbb{L}^2(\mathbb{R}^N)$, and, thus, B in (1) can be thought of as a *direction cosine* between two points on the sphere. For this reason, the values of B are always confined within the interval [0, 1].

Lately, the exceptional properties of the Bhattacharyya distance have shown its worth for applications in computer vision. Thus, for example, in [27], it was shown that for detection, location, or segmentation algorithms based on the maximum likelihood (ML) or on the minimum-description length principle, and for realistic cases where the object and background statistical parameters are unknown, the Bhattacharyya distance can characterize the image difficulty for a large family of probability laws. Particularly, it was shown that, for a wide spectrum of distributions under test, the average number of misclassified pixels is a monotonously decreasing, bijective function of the Bhattacharyya distance between the probability densities of the object and of the background.

Later on, the Bhattacharyya distance was employed in [28] for image segmentation and tracking. In this study, tracked objects were characterized by model distributions over some photometric variable. Consequently, a tracked object was identified as the region of image whose *interior* generates a *sample* distribution which most closely matches the model distribution. As a possible measure of the match, the Bhattacharyya coefficient B was used in [28] with apparent success.

This paper takes the idea of [28] a few steps further by extending its applicability to the cases when the model distributions of tracked objects cannot be defined beforehand. In particular, the segmentation method described in this paper identifies the object as the image region whose interior generates a sample distribution that *maximizes* the distance to the sample distribution generated by corresponding exterior. In order to assess the above distance, we use the Bhattacharyya coefficient B as given by (1).[1]

It should be noted that virtually all practically important probability distributions can be uniquely represented by (either finite or infinite) sets of their *moments*. In such cases, the discrepancy between two (or more) distributions could be "translated" into the discrepancy

---

[1]Since log is a strictly monotone, increasing function, maximizing the Bhattacharyya distance is equivalent to minimizing the corresponding coefficient. For this reason, we prefer using the coefficient instead of the distance, as it leads to simpler analytical expressions.

between their corresponding moments, and comparing the empirical distributions of a photometric variable inside and outside of the "object region" could be superseded by comparing the corresponding empirical moments. Thus, for example, in the seminal work by Chan and Vese [29], the object of interest is identified as the image region, whose interior and exterior differ the most in terms of their corresponding mean intensities. In the case when both the object and the background have (approximately) identical mean values, second-order moments can be involved into an analogous procedure [30]. Obviously, the chain of statistical moments can be extended unlimitedly, progressively increasing the "discrimination power" of resulting segmentation algorithms on one hand and complicating their implementation on the other. Needless to say, working with the moments also raises the dilemma of determining their number and orders which would be appropriate for specific data at hand. On the other hand, the method proposed in this paper is free of the above limitations, since it takes into consideration the entire shapes of probability distributions, thereby being capable of "sensing" the discrepancy between virtually infinite number of empirical moments.

Apart from providing the explicit formulas necessary for implementing the proposed segmentation method, the current study extends its contribution in an additional direction. Since the proposed method works with empirical distributions (which are functions of the data samples, whose number varies in time as the segmentation converges), the issue of estimation consistency and its preservation during the iterations should not be overseen. Consequently, in this paper we also introduce a computationally efficient procedure which allows one to automatically control and adjust the smoothness properties of the estimated distributions.

In concluding this introductory section, it should be mentioned that the proposed segmentation method exploits the technique of active contours that has become very popular over the last few decades [31]. This methodology is based on the utilization of deformable contours which conform to various object shapes. The contour deformation is typically driven by a *gradient flow* stemming from minimization of an energy functional, which can be dependent on either local (e.g., gradient field) or global (e.g., mean intensity) properties of the image. In the latter case, the resulting active contours are referred to as *region-based*, and this is the group of methods to which the approach reported in this paper belongs.

The paper is organized as follows. Section II briefly revises some fundamental aspects of the level-set framework and introduces the Bhattacharyya energy functional, as well as the related gradient flow. The statistical assumptions underpinning the proposed approach are discussed in Section III. Some possible choices of the discriminative features are summarized here, as well. Section IV deals with the problem of automatically controlling the smoothness of empirical distributions and describes a simple method to perform this control. Section V discusses a number of possible ways of extending the results of the preceding sections to multiobject scenarios. Several key results of our experimental study are demonstrated in Section VI, while Section VII provides examples of practical cases in which the proposed methodology can be advantageous over some existing techniques. Section VIII finalizes the paper with a discussion and conclusions.

II. BHATTACHARYYA FLOW

A. Level-Set Representation of Active Contours

In order to facilitate the discussion, we confine the derivations below to the case of two classes (i.e., when the problem to be dealt with is that of segmenting an object of interest out of its background), followed by describing some possible ways to extent the proposed methodology to multiobject scenarios.

In the two-class case, the segmentation problem is reduced to the problem of partitioning the domain of definition $\Omega \subset \mathbb{R}^2$ of an image $I(x)$ (with $x \in \Omega$) into two mutually exclusive and complementary subsets Ω_- and Ω_+. These subsets can be represented by their respective characteristic functions χ_- and χ_+, which can, in turn, be defined by means of a *level-set function* $\varphi(x)$: $\Omega \to \mathbb{R}$ in the following manner. Let $\mathscr{H}$ be the Heaviside function defined in the standard way as

$$\mathscr{H}(\tau) = \begin{cases} 1, & \text{if } \tau \geq 0 \\ 0, & \text{if } \tau < 0. \end{cases} \quad (2)$$

Then, one can define $\chi_-(x) = \mathscr{H}(-\varphi(x))$ and $\chi_+(x) = \mathscr{H}(\varphi(x))$, with $x \in \Omega$.

Given a level-set function $\varphi(x)$, its *zero level set* $\{x | \varphi(x) \equiv 0, x \in \Omega\}$ is used to *implicitly* represent a curve—*active contour*—embedded into Ω. For the sake of concreteness, we associate the subset Ω_- with the support of the object of interest, Ω_+ while is associated with the support of corresponding background. In this case, the objective of active-contour-based image segmentation is, given an initialization $\varphi_0(x)$, to construct a *convergent* sequence of level-set functions $\varphi_t(x)$ $\{\varphi_t(x)\}_{t>0}$ (with $\varphi_t(x)|_{t=0} = \varphi_0(x)$ such that the zero level-set of $\varphi_\infty(x)$ coincides with the boundary of the object of interest.

The above sequence of level-set functions can be conveniently constructed using the variational framework [32]. Specifically, this sequence can be defined by means of a *gradient flow* that minimizes the value of a properly defined cost functional [31]. In the case of the present study the latter is derived in the following way. First, the image to be segmented $I(x)$ is transformed into a *vector-valued* image of its local features $\mathbf{J}(x)$.[2] Note that the feature image $\mathbf{J}(x)$ ascribes to every pixel of $I(x)$ an N-tuple of its associated features, and, hence, it can be formally represented as a map from Ω to $\mathbb{R}^N$. Subsequently, given a level-set function $\varphi(x)$, the following two quantities are computed:

$$P_-(z|\phi(x)) = \frac{\int_\Omega K_-(z - \mathbf{J}(x))\chi_-(x)dx}{\int_\Omega \chi_-(x)dx} = \frac{\int_\Omega K_-(z - \mathbf{J}(x))\mathscr{H}(-\phi(x))dx}{\int_\Omega \mathscr{H}(-\phi(x))dx} \quad (3)$$

and

$$P_+(z|\phi(x)) = \frac{\int_\Omega K_+(z - \mathbf{J}(x))\chi_+(x)dx}{\int_\Omega \chi_+(x)dx} = \frac{\int_\Omega K_+(z - \mathbf{J}(x))\mathscr{H}(\phi(x))dx}{\int_\Omega \mathscr{H}(\phi(x))dx} \quad (4)$$

where $z \in \mathbf{R}^N$, and $K_-(z)$ and $K_+(z)$ are two scalar-valued functions with either compact or effectively compact supports (e.g., Gaussian densities). Provided that the *kernels* $K_-(z)$ and $K_+(z)$ are normalized to have unit integrals with respect to the feature vector z, *viz.* $\int_{\mathbf{R}^N} K_-(z)\,dz = \int_{\mathbf{R}^N} K_+(z)\,dz = 1$, the functions $P_-(z|\varphi(x))$ and $P_+(z|\varphi(x))$ given by (3) and (4) are nothing else but *kernel-based estimates* of the probability density functions (pdfs) of the image features observed over the subdomains Ω_- and Ω_+, respectively [33], [34].

The core idea of the preset approach is quite intuitive and it is based on the assumption that, for a properly selected subset of image features, the "overlap" between the informational contents of the object and of the background has to be minimal. In other words, if one thinks of the active contour as of a *discriminator* that separates the image pixels into two subsets,

---

[2] In the section that follows, we will elaborate on some possible choices of the feature space. Meanwhile, in order to clarify the meaning standing behind of this operation, suffice it to note that $\mathbf{J}$ could be, for example, the image $I(x)$ itself or the vector-valued image of its partial derivatives.

then the optimal contour should minimize the *mutual information* between these subsets. It is worthwhile noting that, for the case at hand, minimizing the mutual information is equivalent to maximizing the KL divergence between the pdfs associated with the "inside" and "outside" subsets of pixels. For the reasons discussed below, however, instead of the divergence, we propose to maximize the Bhattacharyya distance between the pdfs. Specifically, the optimal active contour $\varphi^*(z)$ is defined as

$$\phi^*(x) = \arg\inf_{\phi(z)} \left\{ \tilde{B}(\phi(x)) \right\} \quad (5)$$

where

$$\tilde{B}(\phi(x)) = \int_{z \in \mathbb{R}^N} \sqrt{P_-(z \mid \phi(x)) P_+(z \mid \phi(x))} \, dz \quad (6)$$

with $P_-(z|\varphi(x))$ and $P_+(z|\varphi(x))$ being given by the (3) and (4), correspondingly.

B. Gradient Flow

In order to contrive a numerical scheme for minimizing (6), its first variation should be computed first. The first variation of $\tilde{B}(\varphi(x))$ (with respect to $\varphi(x)$) can be easily shown to be given by

$$\frac{\delta \tilde{B}(\phi(x))}{\delta \phi(x)} = \frac{1}{2} \int_{z \in \mathbb{R}^N} \left(\frac{\partial P_-(z|\phi(x))}{\partial \phi(x)} \sqrt{\frac{P_+(z|\phi(x))}{P_-(z|\phi(x))}} + \frac{\partial P_-(z|\phi(x))}{\partial \phi(x)} \sqrt{\frac{P_-(z|\phi(x))}{P_+(z|\phi(x))}} \right) dz. \quad (7)$$

Differentiating (3) and (4) with respect to $\varphi(x)$, one obtains

$$\frac{\partial P_-(z|\phi(x))}{\partial \phi(x)} = \frac{\delta(\phi(x))}{A_-} \left(P_-(z|\phi(x)) - K_-(z) - \mathbf{J}(x) \right) \quad (8)$$

and

$$\frac{\partial P_+(z|\phi(x))}{\partial \phi(x)} = \frac{\delta(\phi(x))}{A_+} \left(K_+(z - \mathbf{J}(x)) - P_+(z|\phi(x)) \right) \quad (9)$$

where $\delta(\cdot)$ is the delta function, and A_- and A_+ are the areas of Ω_- and Ω_+ given by $\int_\Omega \chi_-(x) \, dx$ and $\int_\Omega \chi_+(x) \, dx$, respectively.

By substituting (8) and (9) into (7) and combining the corresponding terms, one can arrive at

$$\frac{\delta \tilde{B}(\phi(x))}{\delta \phi(x)} = \delta(\phi(x)) V(x) \quad (10)$$

where

$$V(x) = \frac{1}{2} \tilde{B}(\phi(x))(A_-^{-1} - A_+^{-1}) + \frac{1}{2} \int_{z \in \mathbb{R}^N} K_+(z - \mathbf{J}(x)) \frac{1}{A_+} \sqrt{\frac{P_-(z|\phi(x))}{P_+(z|\phi(x))}} dz - \frac{1}{2} \int_{z \in \mathbb{R}^N} K_-(z - \mathbf{J}(x)) \frac{1}{A_-} \sqrt{\frac{P_+(z|\phi(x))}{P_-(z|\phi(x))}} dz. \quad (11)$$

Assuming the same kernel $K(z)$ is used for computing the last two terms in (11), i.e., $K(z) = K_-(z) = K_+(z)$, the latter can be further simplified to the following form:

$$V(x) = \frac{1}{2}\tilde{B}(\phi(x))(A_-^{-1} - A_+^{-1}) + \frac{1}{2}\int_{z \in \mathbb{R}^N} K(z - \mathbf{J}(x))L(z|\phi(x))dz \quad (12)$$

where

$$L(z|\phi(x)) = \frac{1}{A_+}\sqrt{\frac{P_-(z|\phi(x))}{P_+(z|\phi(x))}} - \frac{1}{A_-}\sqrt{\frac{P_+(z|\phi(x))}{P_-(z|\phi(x))}}. \quad (13)$$

Finally, introducing an artificial *time* parameter *t*, the gradient flow of that minimizes (6) is given by

$$\phi_t(x) = -\frac{\delta\tilde{B}(\phi(x))}{\delta\phi(x)} = -\delta\phi(x)V(x) \quad (14)$$

where the subscript denotes the corresponding partial derivative, and *V*(*x*) is defined as given by either (11) or (12).

From the viewpoint of statistical estimation, the cost function (6) can be thought of as accounting for the fidelity of estimation of the optimal level-set function to observed features of *I*(*x*). However, this cost function does not take into consideration some plausible properties of the optimal solution, and, as a result, minimizing (6) alone could be too sensitive to measurement noises and/or errors in the data. In order to alleviate this sensitivity, one can attempt to filter out the spectral components of the solution which belong to the noise subspace. For the case at hand, one can *regularize* the solution via constraining the length of the active contour, in which case the optimal level-set function $\varphi^*(x)$ is given by

$$\phi^*(x) = \arg\inf_{\phi(x)} \left\{ \tilde{B}(\phi(x)) + \alpha\int_\Omega \|\nabla\mathscr{H}(\phi(x))\|dx \right\} \quad (15)$$

where ∇ denotes the operator of gradient, $\|\cdot\|$ is the Euclidean norm, and $\alpha > 0$ is a regularization constant, which controls the compromise between fidelity and stability. The gradient flow associated with minimizing the cost functional in (15) can be shown to be equal to

$$\phi_t(x) = \delta(\phi(x))(\alpha\kappa - V(x)) \quad (16)$$

where κ is the curvature of the active contour given by $\kappa = -\text{div}\{\nabla\varphi(x)/(\|\nabla\varphi(x)\|)\}$. The form of (15) suggests that the gradient flow (16) will converge to a (local) solution of minimal curvature, which results in the maximal distance between the empirical distributions $P_-(z)$ and $P_+(z)$ as measured by the Bhattacharyya coefficient (1). All the segmentation results reported in the present study have been obtained via numerically solving (16) with the value of α set to be equal to 1.

Finally, it should be noted that the gradient flow (16) would change the level-set function $\varphi(x)$ over a set of *zero measure*, if the formal delta function $\varphi(\cdot)$ was exploited in computations. In order to overcome this "technical" difficulty, it is common to extend the numerical support of the level-set evolution via replacing the delta function by its smoother version $\delta_\Delta(\cdot)$ that could be defined as given by, e.g., [29]

$$\delta_\Delta(\phi(x)) = \begin{cases} \frac{1}{2\Delta}(1+\cos(\frac{\pi\phi(x)}{\Delta})) & |\phi(x)| \leq \Delta \\ 0, & \text{otherwise.} \end{cases} \quad (17)$$

Note that the support of δ_Δ is finite and it is controlled though the user-defined parameter Δ.

III. STATICAL FRAMEWORK

A. Assumptions and Limitations

The proposed model for evolving the active contours is based of the notion of discrepancy between the empirical densities $P_-(z)$ and $P_+(z)$ of the feature vector z associated with the object and background classes, respectively. We note that using pdfs for estimating the optimal shape of the active contour is also supported by the well-known *likelihood principle* [35], which states that in making either inferences or decisions about a parameter of interest after data is observed, all relevant experimental information is contained in the likelihood function of the observed data.

However, before we turn to describing some additional aspects related to solving (15) by means of the gradient flow (16), the statistical assumptions, based on which the solution is computed, should be underlined. First of all, it has been assumed that, for the case of two segmentation classes, the samples of $I(x)$ can be characterized by the two conditional densities $p(z|z \in \Omega_-)$ and $p(z|z \in \Omega_+)$, which describe the distribution of the feature vector z in the object and in the background class, respectively. Moreover, it has been assumed that the pixels of the feature image $\mathbf{J}(x)$ that pertain to the object class are independent and identically distributed (i.i.d.) copies of z obeying $p(z|z \in \Omega_-)$, whereas the samples of $\mathbf{J}(x)$ representing the background are i.i.d. copies of z distributed according to $p(z|z \in \Omega_+)$.

An obvious disadvantage of using the above statistical assumptions is the fact that they do not allow one to take into consideration the dependency structure between the samples of $I(x)$ (or between those of $\mathbf{J}(x)$, for that matter). It goes without saying that ignoring the interpixel dependency is a simplification, which has been mainly used to make the proposed segmentation method feasible from the viewpoint of its practical implementation. Undoubtedly, the precision of the method could be improved, if the above dependency was accounted for by means of, e.g., the theory of Markov fields [36]. However, such a more rigorous modeling would have necessitated the estimation of the dependency structure, thereby substantially increasing the computational cost of the overall segmentation process. Hence, in a certain sense, assuming the interpixel independency "trades" the model precision for the practicability of the resulting numerical scheme.

It should also be noted that the kernel density estimates (3) and (4) exploit the assumption of *ergodicity*, which allows one to estimate the densities based on "spatial" averaging rather than on ensemble averaging. Assuming the ergodicity could also be viewed as a limitation, though somewhat less critical one as compared to the assumption of interpixel independency.

Finally, a few comments should be made regarding the form of the *velocity $V(x)$* as given by (12). One can see that the latter consists of two terms. The first of these terms is independent of the spatial coordinate x, and it results in either increment or decrement of the *mean value* of the level-set function $\varphi(x)$ by a constant amount depending on the ratio between the areas A_- and A_+. On the other hand, the second component of the velocity is coordinate dependent, and it can be viewed as a *smoothed* version of the function $L(x)$ defined by (13), whose form deserves a special attention. Specifically, $L(x)$ is defined as a difference between

the square roots of the *likelihood ratios* $P_-(z)/P_+(z)$ and $P_+(z)/P_-(z)$ weighted by corresponding areas. This fact immediately connects the proposed curve-evolution model to the *hypothesis testing* by means of likelihood ratio tests [37]. As a matter of fact, at each iteration, the gradient flow (16) "performs" a likelihood ratio test so as to alter the contour's shape by the forces, which make it include into its interior the image pixels, which are *likely* to belong to the object class. The existence of such an interpretation, as well as its connection to the classical theory of hypothesis testing, further supports the reasonability of the proposed variational framework.

B. Examples of Discriminative Features

In this section, a number of possible definitions of the feature vector *z* are discussed. The set of examples given below is by no means complete, but it rather represents the features most frequently used in practice.

Formally, the transition from the data image $I(x):\Omega \to \mathbb{R}$ to the vector-valued image $\mathbf{J}(x):\Omega \to \mathbb{R}^N$ of its features can be descried by a transformation $\mathcal{W}\cdot\}$ applied to $I(x)$, *viz.* $\mathbf{J}(x) = \mathcal{W} I(x)\}$. Hence, the question of selecting a useful set of image features is essentially equivalent to the question of defining a proper transformation $\mathcal{W}$ Perhaps, the simplest choice here is to define the feature image $\mathbf{J}(x)$ to be identical to $I(x)$, which corresponds to the case of $\mathcal{W}$ being identity and $N = 1$.[3] In this case, the features used for the classification are the gray levels of $I(x)$, and the resulting segmentation procedure is essentially *histogram based* [38].

Although the above choice has proven useful in numerous practical settings, it is definitely not the best possible for the cases when both object and background have similar intensity patterns. In this case, it seems reasonable to take advantage of a relative displacement of these patterns with respect to each other via transforming $I(x)$ to the space of its partial derivatives. This transformation is performed by setting $\mathcal{W} \equiv \nabla$, in which case the feature space becomes two-dimenational, i.e., $N = 2$.

As a next logical step, one can smooth the above gradient $\nabla I(x)$ using a set of low-pass filters with progressively decreasing bandwidths. This construction brings us directly to the possibility to define $\mathcal{W}$ to be a *wavelet transform* [39]. Note that, in this case, each pixel of the resulting $\mathbf{J}(x)$ carries information on *multiresolution* (partial) derivatives of $I(x)$. It should be noted that using the wavelet coefficients as discriminative features for image segmentation has long been successfully used in numerous applications [40].

The dependency structure between the partial derivatives of $I(x)$ can be captured by the *structural tensor* defined as given by

$$\mathbf{T}(x) = \begin{pmatrix} (\partial_{x_1} I(x))^2 & \partial_{x_1} I(x)\partial_{x_2} I(x) \\ \partial_{x_1} I(x)\partial_{x_2} I(x) & (\partial_{x_2} I(x))^2 \end{pmatrix} \quad (18)$$

with $x = (x_1, x_2)$, and ∂_{x_1} and and ∂_{x_2} denoting the corresponding partial derivatives. In this case, the feature space is 3-D, as for each x, $\mathbf{J}(x)$ is defined to be equal to $[(\partial_{x_1} I(x))^2, \partial_{x_1} I(x) \partial_{x_2} I(x), \partial_{x_2} I(x))^2]$. Note, however, that this choice of the feature space should be treated with precaution, since the latter is no more a linear space. This is because the structural tensor (18) is positive definite, and, therefore, the pixels of $\mathbf{J}(x)$ defined as above are no more elements of an Euclidean space, but rather of a nonlinear manifold. Fortunately, the

---

[3]It should be noted that, even though in the definitions above the data image $I(x)$ is scalar-valued, it should not be necessarily so. $I(x)$ may be a color image, as well.

availability of kernel-based methods for estimating probability densities defined over nonlinear manifolds makes it possible to apply our approach in this situation, as well [41].

Another interesting choice of $\mathbf{J}(x)$ can be followed in the scenarios, in which $I(x)$ appears as an element of a sequence of tracking images. In this case, $\mathbf{J}(x) : \Omega \to \mathbf{R}^2$ can be defined to be the *vector field of local displacements* of the gray levels of $I(x)$. Specifically, let $\partial_t I(x)$ denote the temporal (partial) derivative of $I(x)$. Let also $G(x)$ be a 2-D column vector with its first and second components equal to $\partial_{x_1} I(x) \partial_t I(x)$ and $\partial_{x_2} I(x) \partial_t I(x)$, respectively. Then, the *gray-level constancy* constraint [42] can be shown to result in the following least square solution for the local displacement field $\mathbf{v}(x)$:

$$\mathbf{v}(x) = - \mathbf{T}^{-1}(x)G(x), \forall x \in \Omega \quad (19)$$

with $\mathbf{T}(x)$ being the structure tensor given by (18).[4] Consequently, in the case when the motion of the tracked object is independent of that of its background, the feature image can be defined by setting $\mathbf{J}(x) \equiv \mathbf{v}(x)$. We note that this choice seems to be reasonable for many tracking scenarios, where the background motion is either negligible or associated with the ego-motion of camera, which rarely correlates with the dynamics of tracked objects. It is also worthwhile noting that *motion-based* segmentation [43] represents an independent field of research, which embraces many powerful segmentation techniques. Thus, the method presented in this study becomes a specific instance of this class of approaches, when $\mathbf{J}(x)$ is related to motion-based features of $I(x)$.

The local moments of $I(x)$ [44], multiresolution versions thereof [45], and the local fractal dimension [46] are among many other image features, which could be used for the segmentation. Note that a combinational use of all the features mentioned above is also possible. Thus, for example, by letting the feature vector $z = (z_1, z_2, z_3)$ (whose specific realization at x is given by $\mathbf{J}(x)$) be composed of the intensity (z_1) and local velocity components (z_2, z_3), one can perform the segmentation based on *both* gray-scale and motion information, thereby bridging the gap between the segmentation approaches, which utilize these features separately. It should be noted that in this case, it is reasonable to assume that the intensity is independent of the motion. Consequently, the pdf of z can be factorized $p(z) = p(z_1, z_2, z_3) = p(z_1) p(z_2, z_3)$ as so as to reduce the computational complexity of segmentation.

IV. UPDATING THE EMPIRICAL DISTRIBUTIONS

A. Kernel-Based Estimation

In the current study, the conditional densities $p(z|z \in \Omega_-)$ and $p(z|z \in \Omega_+)$ are estimated using the kernel estimation method [33], [34]. In this section, we consider the subject of properly defining the kernel's *bandwidth*, which should be consistent with the data size for the estimates to be reliable [47]. Before we start, it should be noted that the continuously-defined form of the kernel estimates in (3) and (4) was used merely for the convenience of derivation of the gradient flow (16). In what follows, however, we resort to the discrete form of these estimates, as it is more appropriate to the practical case, where data images are represented by their samples over Ω.

In the discrete setting, the kernel density estimation amounts to approximating the (unknown) pdf $p(z)$ of an *N*-dimensional random vector z by

---

[4]Note that, strictly speaking, the tensor $\mathbf{T}(x)$ has rank one, and, hence, it is not invertible. To overcome this technical difficulty, it is common to average the values of the tensor over a small (e.g., 3×3) neighborhood of x.

$$f(z)=\frac{1}{n}\sum_{i=1}^{n}K_\sigma(z-z^i) \quad (20)$$

where $\left\{z^i\right\}_{i=1}^{n}$ are n independent realizations of z. In (20), the kernel $K_\sigma(z)$ is parameterized by a vector σ and has a unit integral, i.e., $\int_{\mathbb{R}^N}K_\sigma(z)\,dz=1$.

A number of possible definitions of the kernel $K_\sigma(z)$ are possible [34], among which the most frequent one is to define to be a *Gaussian pdf*, and this is the choice followed in the current study. Moreover, to facilitate the numerical implementation of the density estimation, we use a separable (isotropic) form of the Gaussian pdf, in which case $K_\sigma(z)$ is defined as

$$K_\sigma(z)=(2\pi)^{-N/2}\prod_{k=1}^{N}\sigma_k^{-1}\exp\left\{-z_k^2/2\sigma_k^2\right\} \quad (21)$$

where $z_1, z_2,\ldots,z_N$, are the coordinates of z, and the standard deviations $\sigma=\left\{\sigma_k\right\}_{k=1}^{N}$ control the extension of $K_\sigma(z)$ along corresponding directions. It should be noted that the separability of the kernel in (21) does not imply the separability of the estimate $f(z)$ that is now given by

$$f(z)=\frac{1}{n}\sum_{i=1}^{n}\prod_{k=1}^{N}K_{\sigma_k}(z_k-z_k^i) \quad (22)$$

with $K_{\sigma_k}(z_k)=\left(\sqrt{2\pi\sigma_k^2}\right)^{-1}\exp\left\{-z_k^2/2\sigma_k^2\right\}$.

The kernel method of density estimation has proven valuable in numerous applications. It is well known, however, that effective use of this method requires proper choice of the *bandwidth* parameters σ_k. When insufficient smoothing is done (i.e., the bandwidth parameters are too small), the resulting density estimate is too rough and contains spurious features that are artifacts of the sampling process. On the other hand, when excessive smoothing is done, important features of the underlying structure are smoothed away. Consequently, to optimize the accuracy of the estimates in (3) and (4), the bandwidth parameters $\left\{\sigma^k\right\}_{k=1}^{N}$ should be properly defined. This can be done using the procedure we describe in the next section.

B. Defining the Bandwidth

In order to avoid unnecessary mathematical abstraction, we confine the discussion below to the case of $N = 1$. Note that such a reduction of dimensionality can by no means be considered as a limitation, as the kernel $K_\sigma(z)$ [as given by (21)] is separable. Consequently, the results derived below can be straightforwardly applied to each 1-D component of $K_\sigma(z)$ *independently*.

In the scalar case, the kernel estimate of $p(z)$ is given by $f(z)=n^{-1}\sum_{i=1}^{n}K_\sigma(z-z^i)$, with $\left\{z^i\right\}_{i=1}^{n}$ being n observations of z and $K_\sigma(z)$ being a scaled version of the normal density $K(z)=\left(\sqrt{2\pi}\right)^{-1}\exp\left\{-z^2/2\right\}$, *viz.* $K_\sigma(z) = \sigma^{-1}K(z/\sigma)$. In this case, a standard way to determine the optimal value of the bandwidth parameter is via minimization of the

asymptotic mean integrated square error between the original pdf $p(z)$ and its estimate $f(z)$. This optimal value can be shown to be given by [47]

$$\sigma^* = \left(\frac{\int K^2(z)dz}{n \|p''(z)\|_2^2 \left(\int z^2 K(z)dz \right)^2} \right)^{1/5} \quad (23)$$

where $\|p''(z)\|_2$ denotes the $\mathbb{L}_2$-norm of the second derivative of $p(z)$. Substituting the Gaussian kernel in (23) results in

$$\sigma^* = \left[2\sqrt{\pi} \|p''(z)\|_2^2 \right]^{-1/5} n^{-1/5}. \quad (24)$$

An exact computation of the above optimal value is obviously impossible, since it requires knowing the derivative of an unavailable quantity, i.e., of $p(z)$. In order to overcome this difficulty, it is standard to use an approximation of the above derivative instead of its exact value. Constructing approximations of this kind has long been an intense research topic in statistics, where all such methodologies proposed so far can now be categorized into the *first* and the *second* generation methods [47]. Although, in general, the second generation methods are more accurate, their implementation often requires iterative solution of nonlinear equations, which makes these methods difficult to integrate into the segmentation procedure under consideration. On the other hand, a considerable gain in computational efficiency (on account of insignificant loss in accuracy) can be achieved by exploiting the first generation methods. One of the methods of this kind (which seems to be first proposed in [48]) is based on assuming the unknown $p(z)$ to be Gaussian. In this case, the resulting optimal bandwidth is given by

$$\sigma^* = (4/3)^{1/5} \sigma_0 n^{-1/5} \approx \sigma_0 n^{-1/5} \quad (25)$$

where σ_0^2 is the second moment of $p(z)$, which can be effortlessly estimated as the sample variance of $\left\{ z^i \right\}_{i=1}^n$. Although assuming $p(z)$ to be Gaussian may be an oversimplification, this assumption has been observed to work quite satisfactory in numerous cases of practical interest. For this reason, in the present study, we choose to define the optimal bandwidth according to (25).

In the case of two segmentation classes, the active contour divides the image domain into two subdomains. As a result, the total number of pixels n of $I(x)$ can be expressed as the sum of n_- pixels belonging to the object class and n_+ pixels belonging to the class of background. Moreover, as the shape of the active contour constantly changes in the course of its evolution, so do the numbers n_- and n_+. As a result, the optimal bandwidths σ_- and σ_+ corresponding to the object and its background, respectively, become functions of the iteration time t, as they are given by

$$\sigma_-^*(t) = \sigma_0^- n_-^{-1/5}(t) \quad \sigma_+^*(t) = \sigma_0^+ n_+^{-1/5}(t) \quad (26)$$

where σ_0^- and σ_0^+ are the standard deviations associated with the conditional densities $p(z|z \in \Omega_-)$ and $p(z|z \in \Omega_+)$, respectively. In order to estimate the constant factors σ_0^- and σ_0^+, one can first classify the image features using some simple"course-level" algorithm, followed by computing the standard deviations of the classes thus obtained. Alternatively, one can estimate these constants by means of the *expectation maximization* algorithm assuming a Gaussian mixture model [49] for the (unconditional) likelihood of x. Yet even simpler

method for computing σ_0^- and σ_0^+ would be to set them both equal to the standard deviation of the entire (i.e., unclassified) data set. Though trivial, the latter approach has been observed to work reliably in practice, and, for this reason, it was used to derive all the results reported in the experimental part of this paper.

The variability of the optimal bandwidths in (26) creates the need for constantly recomputing the kernel density estimates (3) and (4). It goes without saying that such "re-estimations" are very undesirable from the practical point of view, as they could considerably increase the overall computational load. Thus, it is tempting to find a way to update the density estimates in a computationally efficient manner. A possible solution for this problem is proposed next.

C. Bandwidth Adjustment via Isotropic Diffusion

To simplify the presentation, let us consider first the problem of estimating the pdf $p(z)$ of a random variable z given a *time-varying* set of its independent observations $\left\{z^i\right\}_{i=1}^{n(t)}$. Suppose that, at time t_1, the set $\left\{z^i\right\}_{i=1}^{n(t_1)}$ consists of $n(t_1)$ observations, and, hence, the kernel estimate of $p(z)$ is given by

$$f(z|t_1)=\frac{1}{n(t_1)}\sum_{i=1}^{n(t_1)}K_{\sigma(t_1)}(z-z^i) \quad (27)$$

where $\sigma(t_1)$ is the optimal bandwidth corresponding to $n(t_1)$ and $K_{\sigma(t_1)}(z)$ is an appropriately scaled Gaussian kernel. Subsequently, suppose that at a *latter* time t_2, some of the observations are *excluded* from $\left\{z^i\right\}_{i=1}^{n(t_1)}$. Let the subset of the indices of these "excluded" observations be denoted by $Y \subset \{1, 2,...,n(t_1)\}$. Then, the density estimate $f(z|t_1)$ can be updated according to

$$f(z|t_1) \mapsto f(z|t_2)=\frac{n(t_1)f(z|t_1) - \sum_{i\in Y}K_{\sigma(t_1)}(z-z^i)}{n(t_2)} \quad (28)$$

where $n(t_2) = n(t_1) - \|Y\|$ with $\|Y\|$ being the size of Y.[5] It should be noted that computing (28) is "cheap," as normally $Y \ll n(t_1)$. Denoting the new set of observations by $\left\{z^i\right\}_{i=1}^{n(t2)}$, the updated density estimate (28) can be alternatively represented as given by

$$f(z|t_2)=\frac{1}{n(t_2)}\sum_{i=1}^{n(t_2)}K_{\sigma(t_1)}(z-z^i). \quad (29)$$

The above estimate could be considered as optimal, if the kernel $K_{\sigma(t_1)}$ in (29) was replaced by $K_{\sigma(t_2)}$, with $\sigma(t_2)$ being the optimal bandwidth corresponding to $n(t_2)$. Consequently, in order to restore the optimality of (29), the kernel $K_{\sigma(t_1)}$ in (29) should be "substituted" by $K_{\sigma(t_2)}$. In the current paper, we propose to do this via the process of *isotropic diffusion* [50]. Specifically, $K_{\sigma}(t_1)$ is transformed into $K_{\sigma}(t_2)$ by means of the following diffusion equation:

---

[5]If the set $\left\{z^i\right\}_{i=1}^{n(t_1)}$ had "gained" some observations, the sign in (28) should have been changed to plus and $n(t_2)$ would have been equal to $n(t_1) + \|Y\|$, with Y being the indices of the "gained" observations.

$$u_\tau(z, \tau) = u_{zz}(z, \tau),\, u(z, \tau{=}0) = u_0(z) \quad (30)$$

where τ denotes the *diffusion* time (which should not be confused with the "iteration" time *t*). The closed form solution to (30) can be readily shown to be given by

$$u(z, \tau) = u_0(z) * K_{\sqrt{2\tau}}(z) \quad (31)$$

where $K_\sigma(z)$ denotes, as before, the Gaussian density with its standard deviation equal to σ, and * stands for the operator of convolution. Moreover, the *semigroup* property of the diffusion implies that

$$u(z, \tau_2) = u_0(z) * K_{\sqrt{2\tau_2}}(z) = u(z, \tau_1) * K_{\sqrt{2\Delta\tau}}(z) = u_0(z) * K_{\sqrt{2\tau_1}}(z) * K_{\sqrt{2\Delta\tau}}(z) \quad (32)$$

where $\Delta\tau = \tau_2 - \tau_1$. Consequently, by setting $u_0(z) = \delta(z)$, we obtain

$$K_{\sqrt{2\tau_1}}(z) * K_{\sqrt{2\Delta\tau}}(z) = K_{\sqrt{2\tau_2}}(z). \quad (33)$$

The above expression suggests that the kernel $K_{\sigma(t_1)}$ can be transformed into $K_{\sigma(t_2)}$ by *diffusing* the former for the period of time equal to $\Delta\tau = (\sigma^2(t_2) - \sigma^2(t_1))/2$. As a result, assuming that the optimal bandwidth depends on data size according to (25), the time interval necessary for the above diffusive density update is given by

$$\Delta\tau = \frac{\sigma_0}{2}\big(n(t_2)^{-2/5} - n(t_1)^{-2/5}\big). \quad (34)$$

Needless to say, the diffusion process can be applied directly to (29), since the latter is formed as a *linear combination* of mutually shifted versions of $K_{\sigma(t_1)}$. Thus, the entire process of updating this estimate consists of two steps. First, provided that the difference $\Delta n = n(t_2) - n(t_1)$ has exceeded a predefined threshold, the estimate is recomputed according to (28) (or using an analogous formula for the case when $n(t)$ increases). Subsequently, the resulting density is subjected to the process of diffusion for the time duration of $\Delta\tau$ defined by (34). We note that, in practice, the diffusion can be implemented either via direct convolution with $K_{\sqrt{2\Delta\tau}}$ or by means of solving a discrete approximation to (30).

D. Forward versus Backward Diffusion

In the case when the number of data samples *decreases*, the update procedure described above leads to the well-posed operation of *direct* (forward) diffusion. In this case, however, when this number *increases*, the time interval in (34) becomes negative, and, as a result, the diffusion needs to be performed in the *backward* direction in time. Unfortunately, such an *inverse* diffusion is well known to be an ill-conditioned, unstable operation.

The above instability makes it impossible to "run" the inverse diffusion for relatively long time intervals. However, the fact that it usually takes some time for the inverse diffusion to blow-up can be used to "regularize" the process of density estimation in the following way. When the sample size increases, one can propagate the diffusion in the backward direction till the first signs of instability start to show up. At this moment, the diffusion is terminated and the kernel density estimate is re-estimated by mean of, e.g., the *fast Gauss transform* [51]. In practice, however, it was observed that such re-estimations almost never have to be done. This is because, the optimal bandwidth given by (25) is proportional to $n^{-1/5}$, which is a very slow-varying function of *n* for relatively large values of the latter (see Fig. 1). As a

result, for $n > 10^3$, substantial variations in the number of data points result in only *negligible* variations of corresponding bandwidth. It should be noted, however, that the situation changes cardinally when, for example, the active contour converges on a small target represented by relatively small number of image samples. In this case, the variability of $n^{-1/5}$ is significant, and, as a result, the optimal bandwidth is very "sensitive" to variations of n. Thus, for the case of decreasing n, updating the smoothness properties of kernel density estimates becomes crucial. Fortunately, it can be done via the well-posed and computationally efficient process of forward diffusion as described in the preceding section.

Finally, we note that, in the current study, the velocity $V(x)$ in (16) was computed using the simplified expression (12), which utilizes a single kernel function $K(z)$ [as opposed to $K_-(z)$ and $K_+(z)$ in (11)]. This simplification, however, is allowed by the fact that the second component in (12) is always estimated using a *fixed* number of data samples, which is defined by the support of $\delta_\Delta(x)$. In summary, the proposed algorithm can be summarized in the form of the pseudo-code shown in Table I.

V. MULTIOBJECT CASE

In order to extend the applicability of the proposed method to the case when images contain more than one object of interest, its multiclass version should be addressed next. In particular, in such a scenario, the image domain Ω is considered to be a union of M (mutually exclusive) subdomains $\left\{\Omega^k\right\}_{k=1}^{M}$, each of which is associated with a corresponding (conditional) pdf $p_k(z)$. Consequently, the first step to be done is to generalize the definition of the Bhattacharyya coefficient (1) to the case of M densities. Such a natural generalization is known as the *average* Bhattacharyya coefficient and it is given by [21]

$$B_M = \frac{2}{M(M-1)} \sum_{j>i}^{M} \int_{\mathbb{R}^N} \sqrt{p_i(z)p_j(z)}\, dz \quad (35)$$

with $i, j = 1, 2, \ldots, N$. The above coefficient represents a *cumulative* measure of discrepancy between all the possible pairs of the densities under consideration. Note that the normalization constant $2/(M(M-1)$ guarantees that the coefficient (35) takes its values in the interval [0, 1].

The additivity of the construction in (35) makes it trivial to derive the corresponding gradient flow. In fact, each additive term of the cost functional can be differentiated independently so that the resulting gradient flow is given as a *sum* of the gradient flows related to the additive components in (35). In order to complete the multiclass formulation of the Bhattacharyya flow, however, a few words need to be said regarding the definition of the level-set function.

Obviously, in the multiclass case, using only *one* level-set function would be insufficient to solve the problem at hand. Consequently, we follow the multiphase segmentation formulation of [52] that uses p level-set functions $\left\{\phi^i(x)\right\}_{i=1}^{p}$, which are capable of segmenting the image $I(x)$ into (up to) 2^p regions. The idea standing behind the above approach is as simple as brilliant. In particular, since one level-set function partitions the image domain into two subdomains, level-set functions can partition into subdomains, each of which is labelled by the *signs* of the level-set functions $\left\{\phi^i(x)\right\}_{i=1}^{p}$ in that subdomain. For example, when $p = 2$, we obtain four subdomains, *viz.* Ω_{--}, Ω_{-+}, Ω_{+-}, and Ω_{++}.[6] For the reason of space limitation, the formulas for the multiphase gradient flow corresponding to the case of (35) are not provided in this paper. However, these formulas can be easily

obtained by "plugging" the results of Section II into the "templates" derived in [52], *mutatis mutandis*.

VI. RESULTS

A. Intensity-Based Segmentation

The experimental study of the present paper consists of three parts, each of which aims at demonstrating different characteristics of the proposed segmentation method. In the current section, the test images are segmented based on their intensities alone. The first example here is the image of Zebra, which is considered to be relatively hard to segment due to the multimodality of the pdf related to the object class. The segmentation results obtained for this image are shown in Fig. 2, the subplots A1–A4 of which depict the initial, two intermediate and the final shape of the active contour, respectively. The corresponding empirical densities of the object and background classes are shown in the lower row of subplots in Fig. 2. We note that the initial position of the contour was chosen to be such that the number of data samples in both segmentation classes were equal. One can see that the algorithm results in a useful segmentation, which well agrees with the true shape of Zebra. It should be noted, however, that the final segmentation ascribes a portion of the shadow near Zebra's hoofs to the object class. It is because that the intensity levels of the shadow are very close to those of the stripes on Zebra's skin. In this case, the intensity information is insufficient to achieve the ideal result.

Our next example is much more challenging than the previous one. This is a synthetic image generated according to the segmentation mask shown in the leftmost subplot of Fig. 3. This image is referred below to as the image of Cat, since its object class resembles the silhouette of a sitting cat. The corresponding data image is shown in the middle subplot of the same figure. To generate this image, its "object" samples were drawn as independent realizations of z_1 obeying the Rayleigh distribution. At the same time, the samples of the corresponding background were drawn as independent realization of another random variable z_2, which was related to z_1 according to $z_2 = 2\mu - z_1$ with μ being the mean value of z_1. It should be noted that the probability densities defined in this way have the same even (central) moment, while their odd moments are identical up to the sign. These densities are shown in the rightmost subplot of Fig. 3.

The segmentation results obtained for the Cat image are demonstrated in Fig. 4. One can see that, even though the cat is virtually indistinguishable from its background to the eye of a human observer, the proposed algorithm is capable of providing a useful segmentation in this case. Note also how well the estimated class-conditional densities agree with the original densities shown in Fig. 3.

B. Segmenting Vector-Valued Features

The object and background classes of the image shown in Subplot A of Fig. 5 have the same intensity distributions, and, hence, their discrimination based on gray-level information alone is by no means possible. It should be noted, however, that the image pattern is defined to be homogeneous all over the image domain, except for a *round* central region (to be understood as the object), which is rotated with respect to its exterior by 45°. In this case, one can take advantage of the relative displacement of the object with respect to the background via defining the feature image $\mathbf{J}(x)$ to be the gradient $\nabla I(x)$ of the test image $I(x)$. The partial derivatives of $I(x)$ in the row and the column directions are shown in Subplots B and C of Fig. 5, respectively, while Subplot D of the figure shows the phase $\angle$

---

[6] It should be noted that, in the case when $2^p > M$, any of the unnecessary $2^p - M$ phases simply remain idle.

[∇*I*(*x*)] of the gradient ∇*I*(*x*). It is interesting to observe that the distributions of the phase values within the object and background classes appear to be very similar. This fact implies the impossibility to segment the object of interest based on the "orientation" information alone. However, using both partial derivatives as discriminative features makes the segmentation easily achievable as shown by Subplots E-H of Fig. 5. One can see that, in this case, the active contour is also capable of correctly identifying the true object shape.

An additional segmentation result is shown by Subplots A-D of Fig. 6. In this case, the vector-valued features are composed of the monochromatic component of the test image of Surfer. Once again, one can see that the active contour succeeds well in finding the correct shape of the object of interest.

C. Combined Intensity-Motion Segmentation

The main objective of our last example is to demonstrate the value of image segmentation based on both "static" and "dynamic" image features. In particular, in this case, the first component of the feature image $\mathbf{J}(x)$ is set to be equal to the image intensity $I(x)$, whereas its two last components are set to be equal to the local displacements (velocities) of the gray levels of $I(x)$ due to the motion of both object and camera.

The subplots A1–A4 of Fig. 7 show the result of segmenting the image of Leopard using the *intensity information only*. Obviously, the similarity between the intensity patterns of Leopard's fur and of the surrounding terrain, implies a similarity between the intensity distributions within the object and background classes. Consequently, the intensity-based segmentation fails to discriminate between the object and background in a satisfactory manner. In this case, the active contour converges on the part of the object that differs the most from the background, while failing to "catch" the whole object. On the other hand, combined intensity-motion segmentation provides quite satisfactory results, which are shown in Subplots B1–B4 of the same figure. Note that the local velocities (displacements) were computed using two successive images of the original sequence according to (19). A subsampled version of this velocity field is depicted in Fig. 8 (where the subsampling has been done for the sake of clarity of visualization). Therefore, we conclude that motion-based features are quite useful in dynamic scenarios, in which image intensity cannot provide sufficient information to perform the image segmentation in a satisfactory way.

VII. COMPARATIVE STUDY

In the preceding section, the viability of the proposed segmentation method was demonstrated experimentally via its successful performance on a number of examples of practical interest. Unfortunately, little can be deduced from these examples as to what advantages are offered by this method as compared to existing segmentation techniques. In particular, the Bhattacharyya distance is only one instance out of a number of possible definitions of distances between probability densities. Perhaps, the most famous among such distances is the KL divergence that could have been used instead of the Bhattacharyya coefficient to derive a level-set evolution [53]. Hence, it is tempting to identify conditions under which the KL divergence would result in an inferior segmentation as compared to the proposed approach.

If the KL divergence had to be used instead of the Bhattacharyya coefficient, then $\tilde{B}(\varphi(x))$ in (15) would have to be replaced by $-D(\tilde{\varphi}(x))$, where the *symmetrized* KL divergence $D(\tilde{\varphi}(x))$ is given by

$$\tilde{D}(\phi(x)) = \frac{1}{2} D_-(\phi(x))$$
$$+ \frac{1}{2} D_+(\phi(x)) \tag{36}$$
$$= \frac{1}{2} \int_{z \in \Re^N} P_-\left(z \middle| \phi(x)\right) \log\left[\frac{P_-(z|\phi(x))}{P_+(z|\phi(x))}\right] dz + \frac{1}{2} \int_{z \in \mathbb{R}^N} P_+\left(z \middle| \phi(x)\right) \log\left[\frac{P_+(z|\phi(x))}{P_-(z|\phi(x))}\right] dz.$$

In this case, the corresponding velocity $V(x)$ (which is to be substituted in the flow (16) can be shown to be equal to

$$V(x) = \frac{1}{2}\left(\frac{D_-(\phi(x)) - 1}{A_-} - \frac{D_+(\phi(x)) - 1}{A_+}\right) + \frac{1}{2}\int_{z \in \mathbb{R}^N} K(z - \mathbf{J}(x)L(z|\phi(x))dz \tag{37}$$

where

$$L(z|\phi(x)) = \frac{1}{A_-}\left(\frac{P_+(z|\phi(x))}{P_-(z|\phi(x))} - \log\left[\frac{P_-(z|\phi(x))}{P_+(z|\phi(x))}\right]\right) - \frac{1}{A_+}\left(\frac{P_-(z|\phi(x))}{P_+(z|\phi(x))} - \log\left[\frac{P_+(z|\phi(x))}{P_-(z|\phi(x))}\right]\right). \tag{38}$$

Apart from its being more complicated to compute as compared with the Bhattacharyya velocity (12)–(13), the KL velocity (37)–(38) has an additional drawback that stems from the properties of the functions involved in its computation. In particular, the logarithm function used in (36)–(38) is known to be very sensitive to variations of its argument in vicinity of relatively small values of the latter. Moreover, the logarithm is undefined at zero, which makes computing the KL velocity susceptible to numerical errors which should be expected when the densities P_- and P_+ approach zero. The above properties of the logarithm are obviously disadvantageous, as they make the KL divergence prone to the errors caused by inaccuracies in estimating the *tails* of probability densities. On the other hand, the square root is a well-defined function in vicinity of zero. Moreover, for relatively small values of its argument, the variability of the square root is considerably smaller than that of the logarithm. Consequently, the Bhattacharyya flow should be much less susceptible to the influence of the inaccuracies mentioned above.

Alternatively to the information-based formulation of [53], the active contours can be propagated using the ML formulation as suggested in [54]. In the notations of this paper, the ML optimal level-set function is sought as a minimizer of the *negative log-likelihood* of observed features that is defined as given by

$$\tilde{L}(\phi(x)) = -\int_\Omega \log[p_-(\mathbf{J}(x))]\mathscr{H}(-\phi(x))dx - \int_\Omega \log[p_+(\mathbf{J}(x))]\mathscr{H}(-\phi(x))dx \tag{39}$$

where $p_-(z)$ and $p_+(z)$ denote *a priori* known probability densities corresponding to the object of interest and its background.

It is interesting to note that, in this case, the velocity $V(x)$ is given by the log-likelihood ratio

$$V(x) = \log\left[\frac{p_-(\mathbf{J}(x))}{p_+(\mathbf{J}(x))}\right]. \tag{40}$$

Thus, for example, in the case when the feature image $\mathbf{J}(x)$ is formed by image intensities (i.e., $\mathbf{J}(x) = I(x)$), and the densities $p_-(z)$ and $p_+(z)$ are Gaussian with their mean values equal to μ_- and μ_+, and their variances equal to σ_-^2 and σ_+^2, respectively, the velocity (40) becomes

$$V(x) = \frac{1}{2}\left[\log(\hat{\sigma}_+^2) - \log(\hat{\sigma}_-^2) + \hat{\sigma}_+^{-2}\left|I(x) - \hat{\mu}_+\right|^2 - \hat{\sigma}_-^{-2}\left|I(x) - \hat{\mu}_-\right|^2\right] \quad (41)$$

where the estimates

$$\hat{\mu}_- = \frac{1}{A_-}\int_\Omega I(x)\mathscr{H}(-\phi(x))dx$$

$$\hat{\mu}_+ = \frac{1}{A_+}\int_\Omega I(x)\mathscr{H}(\phi(x))dx$$

$$\hat{\sigma}_-^2 = \frac{1}{A_-}\int_\Omega |I(x) - \hat{\mu}_-|^2\mathscr{H}(-\phi(x))dx \quad (42)$$

$$\hat{\sigma}_+^2 = \frac{1}{A_+}\int_\Omega |I(x) - \hat{\mu}_+|^2\mathscr{H}(\phi(x))dx$$

are supposed to be recomputed at each iteration.

In addition to its being well founded from a theoretical perspective, the ML formulation is also advantageous in providing a way to predefine an optimal value of the regularization parameter in (15)–(16). Specifically, the *minimum description length* principle allows one to preset this value to be equal to log(8), as in [54]. Unfortunately, this advantage of the ML formulation seems to be counterbalanced by its dependency on *a priori* knowing the class conditional likelihoods $p_-(z)$ and $p_+(z)$, as well as by its being sensitive to inaccuracies in defining the latter. The problematic character of the above properties of the ML approach is demonstrated through the example that follows.

The leftmost subplot of Fig. 9 shows the original segmentation mask (template) that defines the object of interest as a circle over uniform background. The test image shown by the center subplot of the same figure was synthesized according to the above template with both object and background densities defined to be Gaussian pdfs with zero means and variances equal to 1 and 2, respectively. In parallel, another test image was synthesized in similar manner except for the fact that 4% of its (randomly picked) pixels were replaced by independent realizations of a Gaussian random variable with zero mean and variance equal to 16. This image is depicted in the rightmost subplot of Fig. 9. Note that the "substituted" pixels of the latter image can be thought of as *outliers* which are expected to introduce errors in estimation/modeling of the class conditional densities.

The results of intensity-based segmentation of the test images of Fig. 9 are demonstrated in Fig. 10, subplots A1–A3 of which show the final active contours obtained by using the ML velocity (41), KL velocity (37), and the Bhattacharyya velocity (12), respectively. In order to compare these results in a quantitative manner, for each method, the probability of misclassification error was estimated as an *expected* number of misclassified pixels normalized to the total number of pixels, with the expectation being approximated by averaging the results of 200 independent trials. Specifically, the empirical error probabilities for the ML, KL, and Bhattacharyya segmentations were found to be equal to 0.83%, 1.07%, and 0.88%, respectively. One can see that, in this case, the ML segmentation has the lowest error due to its using the correct models for the class conditional likelihoods. On the other hand, the KL segmentation appears to be the worst performer here, because of the numerical errors discussed in the beginning of this section. One can also see that the error of the proposed method closely approaches that of the ML segmentation.

As a next stage, the segmentation methods under consideration were applied to the outlier contaminated image shown in the rightmost subplot of Fig. 9. For this case, typical segmentation results are shown in Subplots B1–B2 of Fig. 10. One can see that the lack of correspondence (caused by the outliers) between the assumed and actual class-conditional likelihoods makes the ML segmentation converge to an incorrect solution (see Subplot B1). The error probability of this segmentation was found to be equal to 10.2%. Moreover, the outliers induce relatively small deviations at the *tails* of the kernel density estimates. These deviations, in turn, are "translated" into sizable errors in estimation of the KL divergence (as well as of the related gradient flow) due to the high sensitivity of the logarithm in vicinity of zero. Consequently, the KL segmentation results in an erroneous solution, as well (Subplot B2). For this case, the error probability was found to be equal to 9.7%. On the other hand, the Bhattacharyya gradient flow succeeds to converge to a useful solution with 0.91% error (as shown by Subplot B3), thereby exhibiting remarkable insensitivity to both model errors and the influence of outliers.

VIII. DISCUSSION AND CONCLUSIONS

In this paper, a method for image segmentation by active contours has been proposed. As in most of the methods using a similar classification mechanism, the derivation of the proposed approach is based on a variational analysis, in which the active contour is driven by the forces stemming from minimization of a cost functional. In the current work, the latter is formulated based on the *discrimination principle* that seems to be intrinsic in the way the human visual system functions [55]. In particular, the cost functional is defined as a measure of dissimilarity between the informational contents of segmentation classes, and as a result, the active contour is forced to converge to the shape that minimizes the "overlap" between these contents.

In the heart of the proposed segmentation method is the notion of a distance between probability densities. In particular, the active contours have been evolved to maximize the Bhattacharyya distance between nonparametric (kernel-based) estimates of the probability densities of segmentation classes. In this case, it would be *conceptually* analogous, if one tried to minimize the mutual information shared by the classes via maximizing the KL divergence between the corresponding probability densities. Yet, in the experimental study of Section VII, it was demonstrated that this alternative criterion may result in the performance inferior to that of the proposed method due to a relatively high sensitivity of the KL divergence to inaccuracies in estimation of the tails of the class conditional densities.

Using the nonparametric estimates of probability densities allows one to apply the proposed segmentation method in the situations when little is known on the distributions of image features within different segmentation classes. However, if the information on distributions was *a priori* available, the segmentation could have been achieved by means of the ML approach of [54], [56], which has an advantage of providing a way to predefine an optimal values of the regularization parameter in (16). Unfortunately, the availability of the above information seems to be rare in practice considering the vast diversity of possible images and their related features. Moreover, the errors in modeling the class conditional densities are capable of substantially degrading the performance of the ML segmentation, as it was demonstrated in Section VII.

As a possible extension, the method proposed in this paper can be easily modified to incorporate shape priors as it is suggested, e.g., in [1]. In this case, the *principal component analysis* (PCA) is used to represent the optimal level-set as an element of a finite-dimensional vector space spanned by the mean value and a predefined (relatively small) number of the principal components which correspond to a given set of training shapes. An

alternative way to impose smoothing constrains on the segmentation can be by using the Sobolev active contours recently proposed in [57].

In this paper, kernel density estimation [33], [34] was employed to compute the class conditional densities in a nonparametric manner. In order to increase the computational efficiency of the estimation, the kernel functions were defined to be *isotropic* Gaussian densities. It is well known, however, that better density estimation is possible using *anisotropic* kernels, as it is shown, e.g., in [58]. Unfortunately, in this case, the computational load should be expected to grow considerably. Moreover, using the isotropic kernels is also advantageous, as it allows further increasing the computational efficiency via updating the kernel bandwidth through the process of isotropic diffusion, as it is shown in Section IV. This scheme would not be possible, if anisotropic kernels were used.

It is also interesting to point out the similarity between the problem of image segmentation by means of active contours and the problem of *blind source separation* [as a specific instance of *independent component analysis* (ICA) [59]]. The latter is a reconstruction problem, in which a number of unknown source signals have to be recovered from measurements of their *algebraic* mixtures. This problem has inspired the proposal of numerous solutions, many of which are based on finding the directions—*independent components*—in the multidimensional space, along which the mixtures (or, better to say, the projections thereof) are *as independent as possible*. In this connection, in order to access the above independency a number of information-based criteria, like the KL divergence, have been intensively used. Returning to the problem of segmentation, one can think of a data image as a *geometric* mixture of sources, i.e., of segmentation classes. In such a case, the active contour acts akin an independent component when trying to separate the image into a number of as independent regions as possible. It is quite interesting to note that the method proposed in the present study employs the concept and tools, which are very much similar to those used in ICA.

Finally, we note that the proposed approach constitutes a natural generalization of the segmentation methods, which classify the image samples based on a finite number of low-order empirical moments. Moreover, the simplicity of the proposed variational formulation allows one to accommodate and use an arbitrary number of diverse image features. Thus, for example, in Section VI-C, images are segmented using both "static" and "dynamic" features. It is also possible to extend the applicability of the method to segmenting the data defined over nonlinear manifolds. This subject well deserves more ample treatment, which defines one of the directions of our future research.

Acknowledgments

REFERENCES

1. Tsai A, Yezzi A, Wells W, Tempany C, Tucker D, Fan A, Grimson WE, Willsky A. A shape-based approach to the segmentation of medical imagery using level sets. IEEE Trans. Med. Imag. 2003 Feb; vol. 22(no. 2):137–154.

2. Grau V, Mewes A, Alcaniz M, Kikinis R, Warfield SK. Improved watershed transform for medical image segmentation using prior information. IEEE Trans. Med. Imag. 2004 Apr; vol. 23(no. 4): 447–458.

3. Stringa E, Regazzoni CS. Real-time video-shot detection for scene surveillance applications. IEEE Trans. Image Process. 2000 Jan; vol. 9(no. 1):69–79. [PubMed: 18255373]

4. Haritaoglu I, Harwood D, Davis LS. $\mathbf{W}^4$: Real-time surveillance of people and their activities. IEEE Trans. Pattern Anal. Mach. Intell. 2000 Aug; vol. 22(no. 8):809–830.

5. Rimey RD, Cohen FS. A maximum-likelihood approach to segmenting range data. IEEE Trans. Robot. Autom. 1988 Jun; vol. 4(no. 3):277–286.

6. Nair D, Aggarwal JK. Moving obstacle detection from a navigating robot. IEEE Trans. Robot. Autom. 1998 Jun; vol. 14(no. 3):404–416.

7. Kato J, Watanabe T, Joga S, Rittscher J, Blake A. An HMM-based segmentation method for traffic monitoring movies. IEEE Trans. Pattern Anal. Mach. Intell. 2002 Sep; vol. 24(no. 9):1291–1296.

8. Mellia M, Meo M, Casetti C. TCP smart framing: A segmentation algorithm to reduce TCP latency. IEEE/ACM Trans. Netw. 2005 Apr; vol. 13(no. 2):316–329.

9. Bhanu B, Holben R. Model-based segmentation of FLIR images. IEEE Trans. Aerosp. Electron. Syst. 1990 Jan; vol. 26(no. 1):2–11.

10. Reed S, Petillot Y, Bell J. An automatic approach to the detection and extraction of mine features in sidescan sonar. IEEE J. Ocean. Eng. 2003 Jan; vol. 28(no. 1):90–105.

11. Zhu SC, Yuille A. Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 1996 Sep; vol. 18(no. 9): 884–900.

12. Duda, RO.; Hart, RE.; Stork, DG. Pattern Recognition. Wiley; New York: 2001.

13. Canny JF. A computational approach to edge detection. IEEE Trans. Pattern Anal. Mach. Intell. 1986 Jun; vol. 8(no. 6):679–698. [PubMed: 21869365]

14. Yezzi A, Kichenassamy S, Kumar A, Olver P, Tannenbaum A. A geometric snake model for segmentation of medical imagery. IEEE Trans. Med. Imag. 1997 Apr; vol. 16(no. 2):199–209.

15. Geman S, Geman D. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. IEEE Trans. Pattern Anal. Mach. Intell. 1984 Nov; vol. PAMI-6(no. 11):721–741. [PubMed: 22499653]

16. Mumford D, Shah J. Optimal approximations by piecewise smooth functions and associated variational problems. Comm. Pure Appl. Math. 1989; vol. 42(no. 5):577–685.

17. Blake, A.; Yuille, A. Active Vision. Cambridge, MA: MIT; 1992.

18. Malladi R, Sethian J, Vemuri B. Shape modeling with front propagation: A level set approach. IEEE Trans. Pattern Anal. Mach. Intell. 1995 Feb; vol. 17(no. 2):158–175.

19. Tuceryan M, Jain AK. Texture segmentation using Voronoi polygons. IEEE Trans. Pattern Anal. Mach. Intell. 1990 Feb; vol. 12(no. 2):211–216.

20. Unal G, Krim H, Yezzi A. Fast incorporation of optical flow into active polygons. IEEE Trans. Image Process. 2005 Jun; vol. 14(no. 6):745–759. [PubMed: 15971774]

21. Kailath T. The divergence and Bhattacharyya distance measures in signal selection. IEEE Trans. Commun. Technol. 1967 Feb; vol. COM-15(no. 1):52–60.

22. Bhattacharyya A. On a measure of divergence between two statistical populations defined by their probability distributions. Bull. Calcutta Math. Soc. 1943; vol. 35:99–109.

23. Cover, TM.; Thomas, JA. Elements of Information Theory. New York: Wiley; 1991.

24. Kullback S, Leibler RA. On information and sufficiency. Ann. Math. Statist. 1951; vol. 22:79–86.

25. Kullback, S. Information Theory and Statistics. New York: Wiley; 1959.

26. Chernoff H. A measure of asymptotic efficiency for tests of a hypothesis based on a sum of observations. Ann. Math. Statist. 1952; vol. 23:493–507.

27. Goudail F, Refregier P, Delyon G. Bhattacharyya distance as a contrast parameter for statistical processing of noisy optical images. J. Opt. Soc. Amer. A. 2004 Jul; vol. 21(no. 7):1231–1240.

28. Freedman D, Zhang T. Active contours for tracking distributions. IEEE Trans. Image Process. 2004 Apr; vol. 13(no. 4):518–526. [PubMed: 15376586]

29. Chan T, Vese L. Active contours without edges. IEEE Trans. Image Process. 2001 Feb; vol. 10(no. 2):266–277. [PubMed: 18249617]

30. Yezzi A, Tsai A, Willsky A. A statistical approach to snakes for bimodal and trimodal imagery, in. Proc. ICCV. 1999:898–903.

31. Osher, S.; Paragios, N. Geometric Level Set Methods in Imaging, Vision, and Graphics. New York: Springer; 2003.

32. Gelfand, IM.; Fomin, SV.; Silverman, RA. Calculus of Variations. Englewood Cliffs, NJ: Prentice-Hall; 1975.

33. Silverman, BW. Density Estimation for Statistics and Data Analysis. CRC: Boca Raton, FL; 1986.

34. Simonoff, JS. Smoothing Methods in Statistics. New York: Springer; 1996.

35. Birnbaum A. On the foundations of statistical inference (with discussion). J. Amer. Statist. Assoc. 1962; vol. 57:269–326.

36. Winkler, G. Image Analysis, Random Fields, and Dynamic Monte Carlo Methods: A Mathematical Introduction. New York: Springer-Verlag; 1995.

37. Lehmann, EL. Testing Statistical Hypotheses. New York: Wiley; 1986.

38. Glasbey CA. An analysis of histogram-based thresholding algorithms. Graph. Models and Image Process. 1993; vol. 55:532–537.

39. Mallat S. A theory for multiresolution signal decomposition: The wavelet representation. IEEE Trans. Pattern Anal. Mach. Intell. 1989 Jul; vol. 11(no. 7):674–693.

40. Unser M. Texture classification and segmentation using wavelet frames. IEEE Trans. Image Process. 1995 Apr; vol. 11(no. 4):1549–1560. [PubMed: 18291987]

41. Chikuse, Y. Statistics on Special Manifolds. Springer: New York; 2003.

42. Horn BKP, Schunck BG. Determining optical flow. Artif. Intell. 1981 Aug.vol. 17:185–203.

43. Peleg S, Rom H. Motion based segmentation, in. Proc. 10th Int. Conf. Pattern Recognition. 1990 Jun.vol 1:109–113.

44. Tuceryan M. Moment based texture segmentation. Pattern Recognit. Lett. 1994 Jul.vol. 15:659–668.

45. Sühling M, Arigovindan M, Hunziker P, Unser M. Multiresolution moment filters: Theory and applications. IEEE Trans. Image Process. 2004 Apr; vol. 13(no. 4):484–495. [PubMed: 15376583]

46. Chaudhuri BB, Sarkar N. Texture segmentation using fractal dimension. IEEE Trans. Pattern Anal. Mach. Intell. 1995 Jan; vol. 17(no. 1):72–77.

47. Jones MC, Marron JS, Sheather SJ. A brief survey of bandwidth selection for density estimation. J. Amer. Statist. Assoc. 1996 Mar; vol. 91(no. 433):401–407.

48. Deheuvels P. Estimation nonparamétrique de la densité par histogrammes généralisés. Rev. Statist. Appl. 1977; vol. 25:5–42.

49. Xu L, Jordan MI. On convergence properties of the EM algorithm for Gaussian mixtures. Neural Comput. 1996 Jan; vol. 8(no. 1):129–151.

50. Holman, JP. Heat Transfer. New York: McGraw-Hill; 1981.

51. Greengard L, Strain J. The fast Gauss transform. SIAM J. Sci. Statist. Comput. 1991; vol. 12(no. 1):79–94.

52. Vese LA, Chan TF. A multiphase level set framework for image segmentation using the Mumford and Shah model. Int. J. Comput. Vis. 2002; vol. 50(no. 3):271–293.

53. Kim J, Fisher JW, Yezzi A, Cetin M, Willsky AS. A nonparametric statistical method for image segmentation using information theory and curve evolution. IEEE Trans. Image Process. 2005 Oct; vol. 14(no. 10):1486–1502. [PubMed: 16238055]

54. Martin P, Refregier P, Goudail F, Guerault F. Influence of the noise model on level set active contour segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 2004 Jun; vol. 26(no. 6):799–803. [PubMed: 18579939]

55. Yantis, S. Visual Perception: Essential Readings. Philadelphia, PA: Psych. Press; 2001.

56. Martin P, Refregier P, Goudail F, Guerault F. Nonparametric statistical snake based on the minimum stochastic complexity. IEEE Trans. Image Process. 2006 Sep; vol. 15(no. 9):2762–2770. [PubMed: 16948320]

57. Sundaramoorthi G, Yezzi A, Mennucci A. Sobolev active contours. Int. J. Comput. Vis. 2007 to be published.

58. Breiman L, Meisel W, Purcell E. Variable kernel estimates of multivariate densities. Technometrics. 1977; vol. 19:135–144.

59. Hyvarinen, A.; Karhunen, J.; Oja, E. Independent Component Analysis. New York: Wiley; 2001.

Biographies

Oleg Michailovich (M'02) was born in Saratov, Russia, in 1972. He received the M.Sc. degree in electrical engineering from the Saratov State University in 1994 and the M.Sc. and Ph.D. degrees in biomedical engineering from The Technion—Israel Institute of Technology, Haifa, in 2003.

He is currently with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. His research interests include the application of image processing to various problems of image reconstruction, segmentation, inverse problems, nonparametric estimations, approximation theory, and multiresolution analysis.

Yogesh Rathi received the B.S. degree in electrical engineering and the M.S. degree in mathematics from the Birla Institute of Technology and Science, Pilani, India, in 1997, and the Ph.D. degree in electrical engineering from the Georgia Institute of Technology, Atlanta, in 2006. His Ph.D. dissertation titled "Filtering for Closed Curves" was on tracking highly deformable objects in the presence of noise and clutter.

He is presently a Postdoctoral Fellow in the Psychiatry Neuroimaging Laboratory, Brigham and Women's Hospital, Harvard Medical School, Boston, MA. His research interests include image processing, medical imaging, diffusion tensor imaging, computer vision, and control and machine learning.

Allen Tannenbaum (M'93) was born in New York in 1953. He received the Ph.D. degree in mathematics from Harvard University, Cambridge, MA, in 1976.

He has held faculty positions at the Weizmann Institute of Science, Rehovot, Israel; McGill University, Montreal, QC, Canda; ETH, Zurich, Switzerland; The Technion—Israel Institute of Technology, Haifa; the Ben-Gurion University of the Negev, Israel; and the University of Minnesota, Minneapolis. He is presently the Julian Hightower Professor of Electrical and Biomedical Engineering, Georgia Institute of Technology, Atlanta, and Emory University, Atlanta. He has done research in image processing, medical imaging, computer vision, robust control, systems theory, robotics, semiconductor process control, operator theory, functional analysis, cryptography, algebraic geometry, and invariant theory.
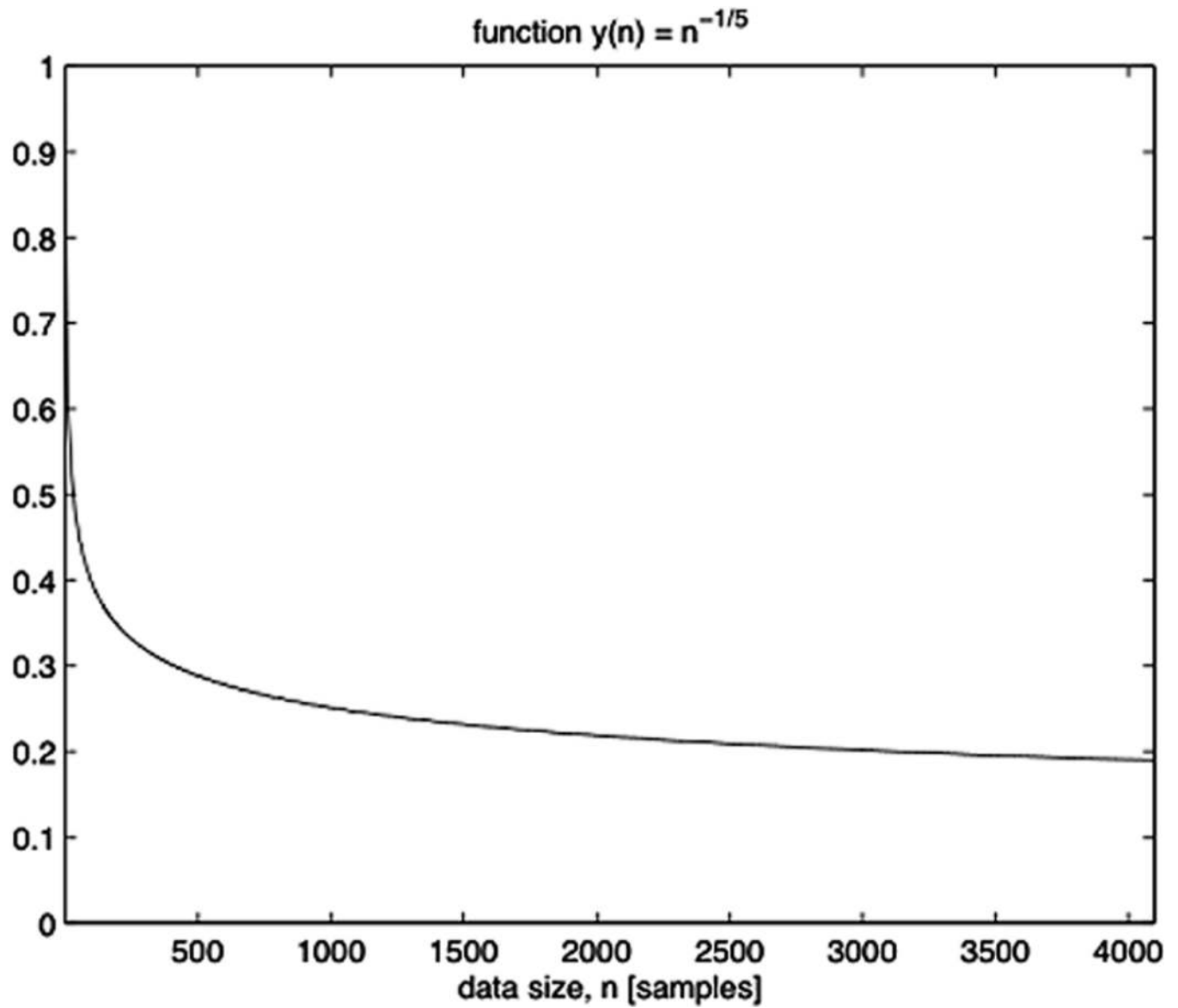
Fig. 1.
Graph of the function $y[n] = n^{-1/5}$.
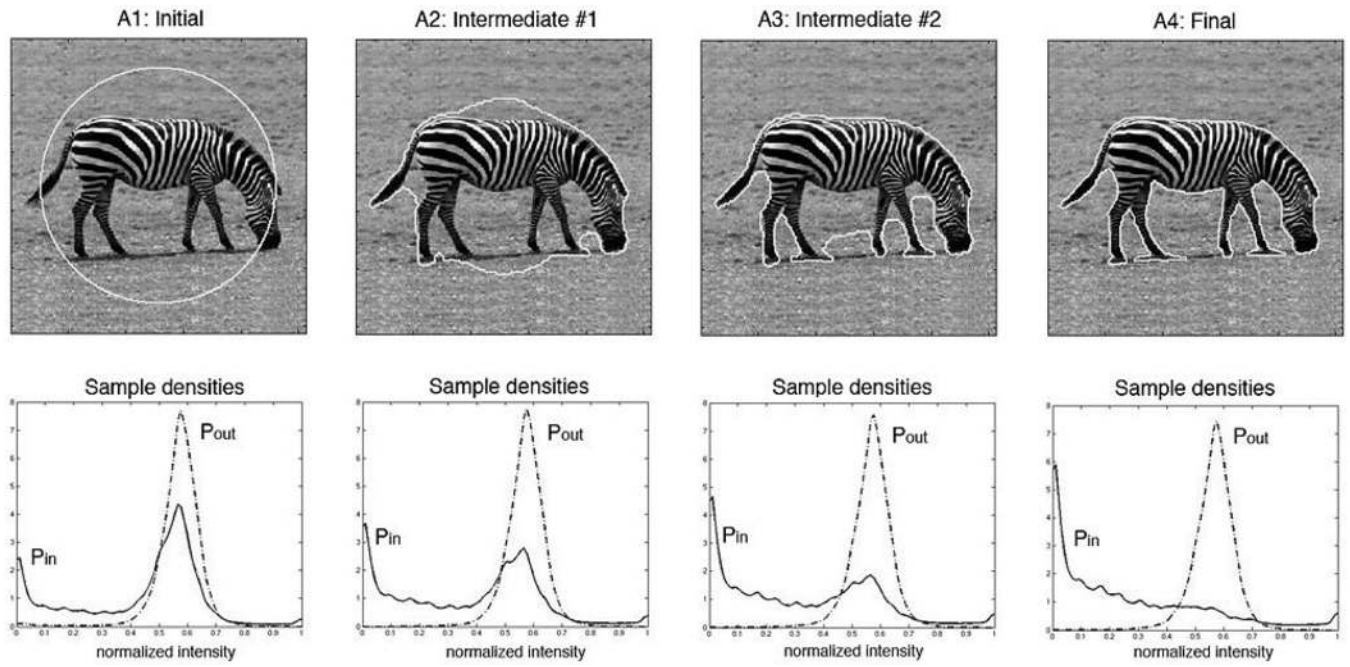
Fig. 2.
(Upper row of subplots) Intensity-based segmentation of Zebra. (Lower row of subplots)
Corresponding empirical densities of Zebra and its background.
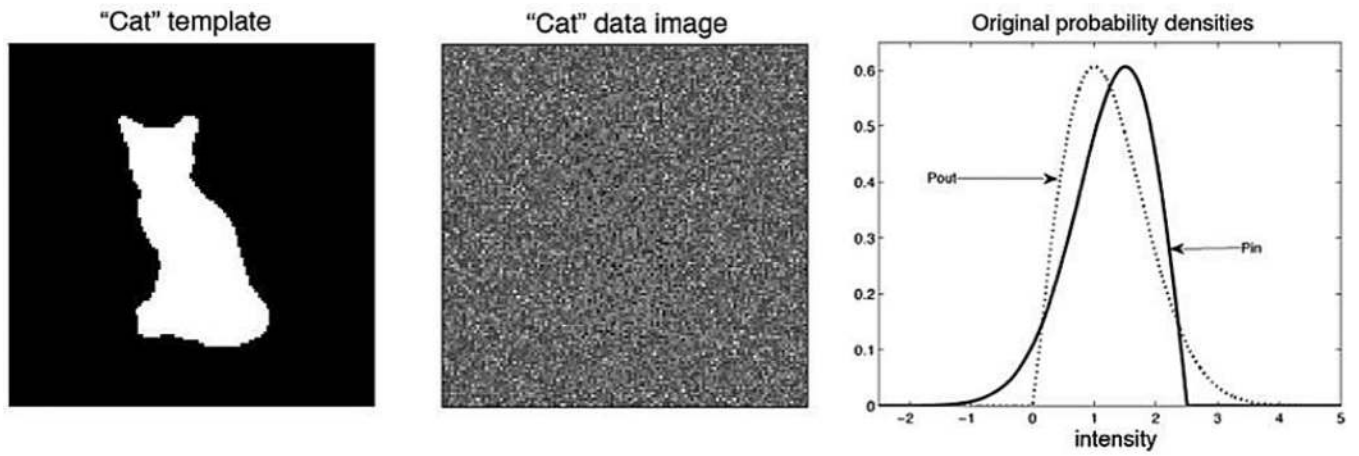
Fig. 3.
(Right) Segmentation mask of Cat. (Middle) Realization of Cat. (Left) Original densities of Cat and its background.
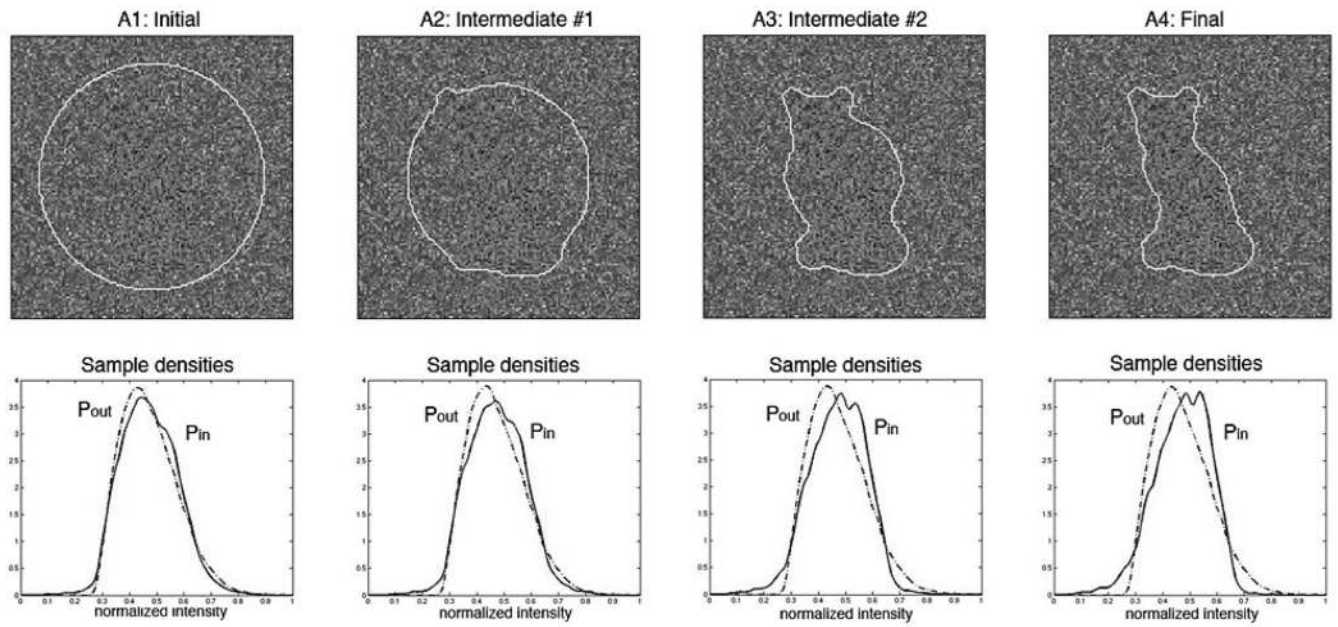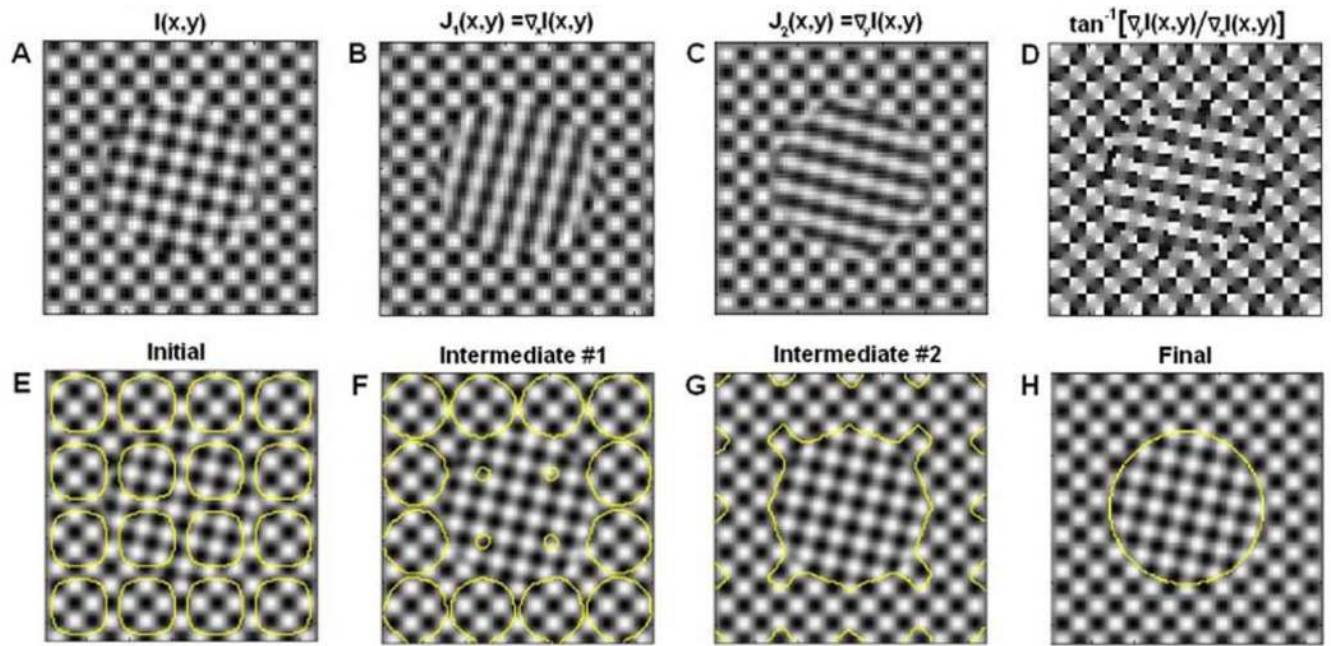
Fig. 4.
(Upper row of subplots) Intensity-based segmentation of Cat. (Lower row of subplots)
Corresponding empirical densities of Cat and its background.

Fig. 5.
(Subplots A–D, left-to-right) Image of rotated pattern to be segmented, the row partial derivatives of the image, the column partial derivative of the image, the phase of the image gradient. (Subplots E–H) Gradient-based segmentation of rotated pattern.
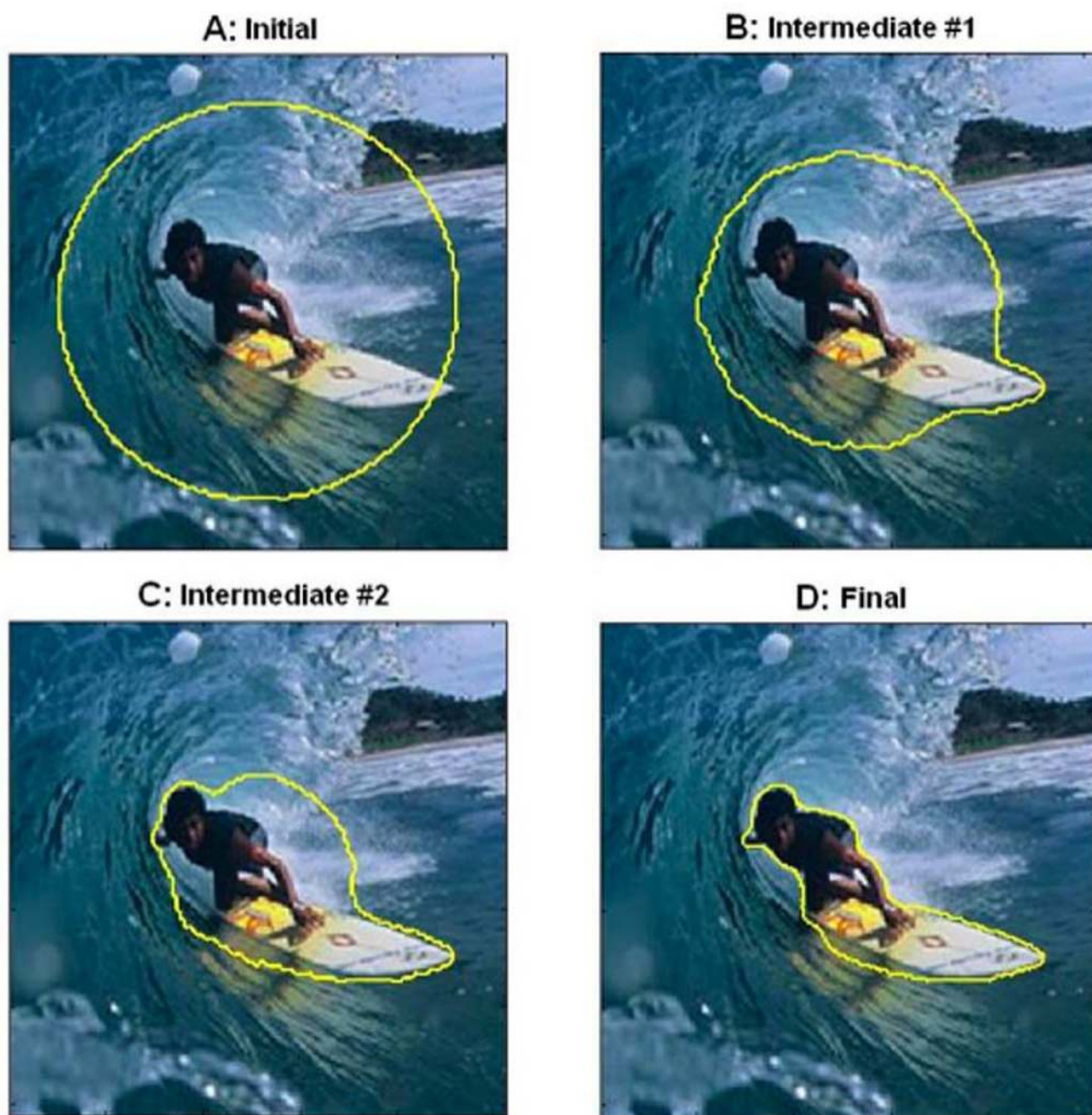
A: Initial

B: Intermediate #1

C: Intermediate #2

D: Final

Fig. 6.
Color-based segmentation of Surfer.
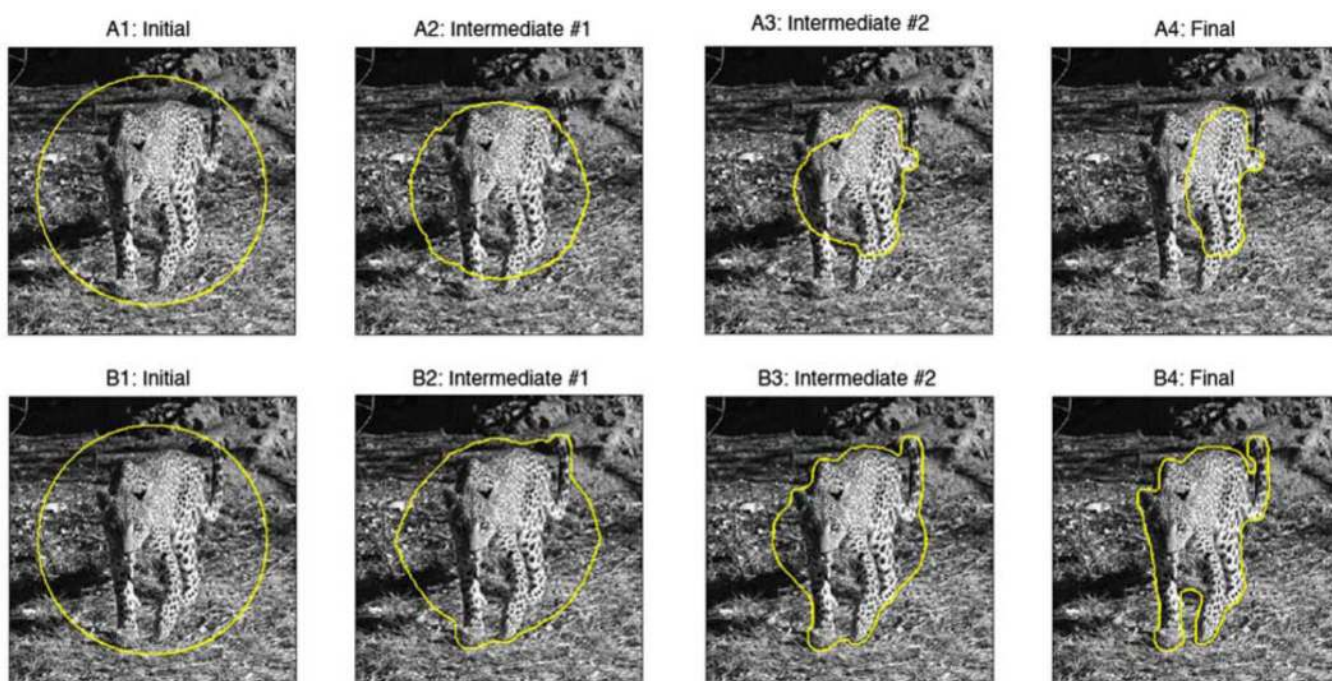
Fig. 7.
(Subplots A1–A4) Intensity-based segmentation of Leopard. (Subplots B1–B4) Joint intensity-motion-based segmentation of Leopard.
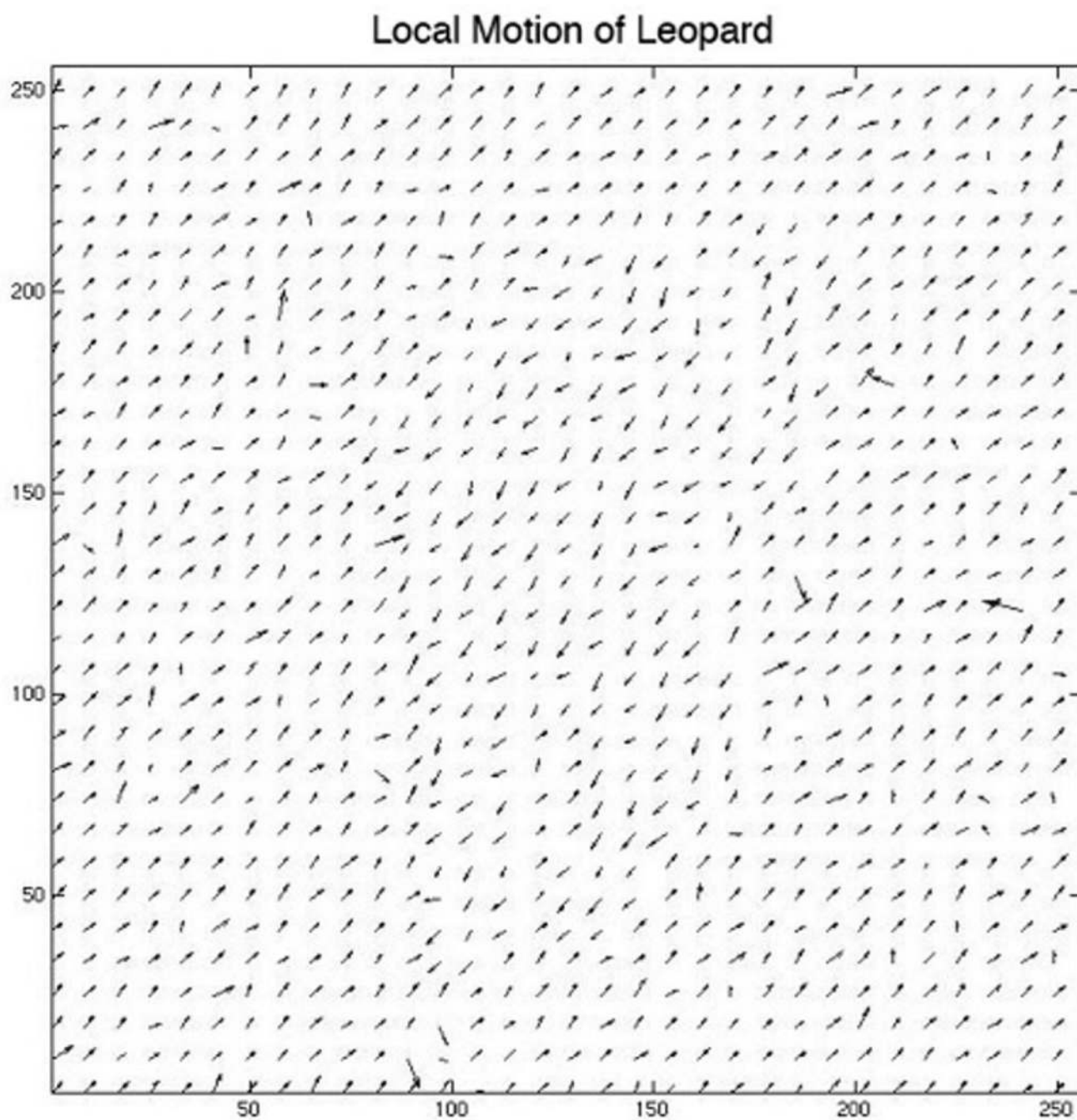
Local Motion of Leopard

Fig. 8.
Local displacement field of Leopard.

Original template **Outliers-free image** **Image with outliers**
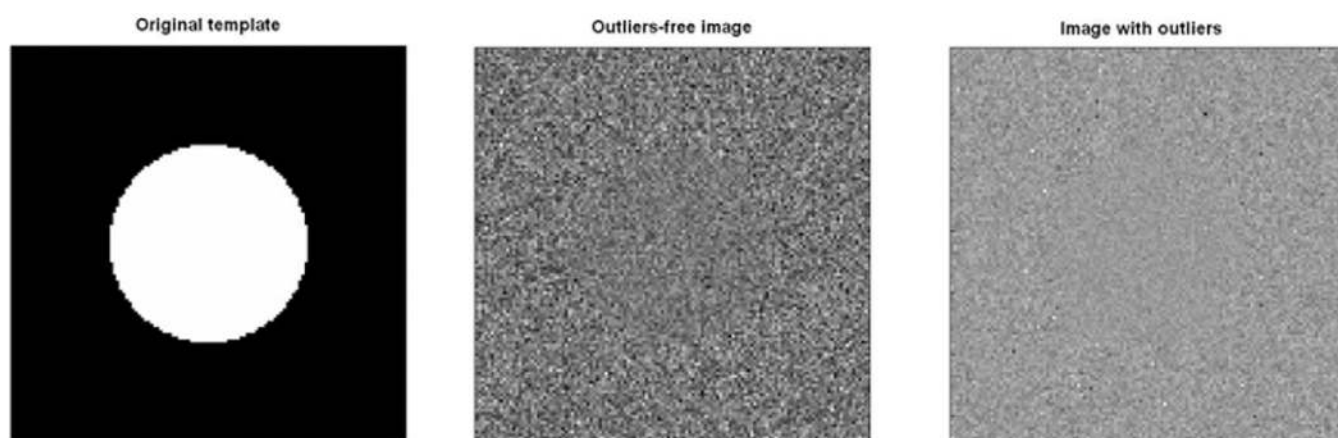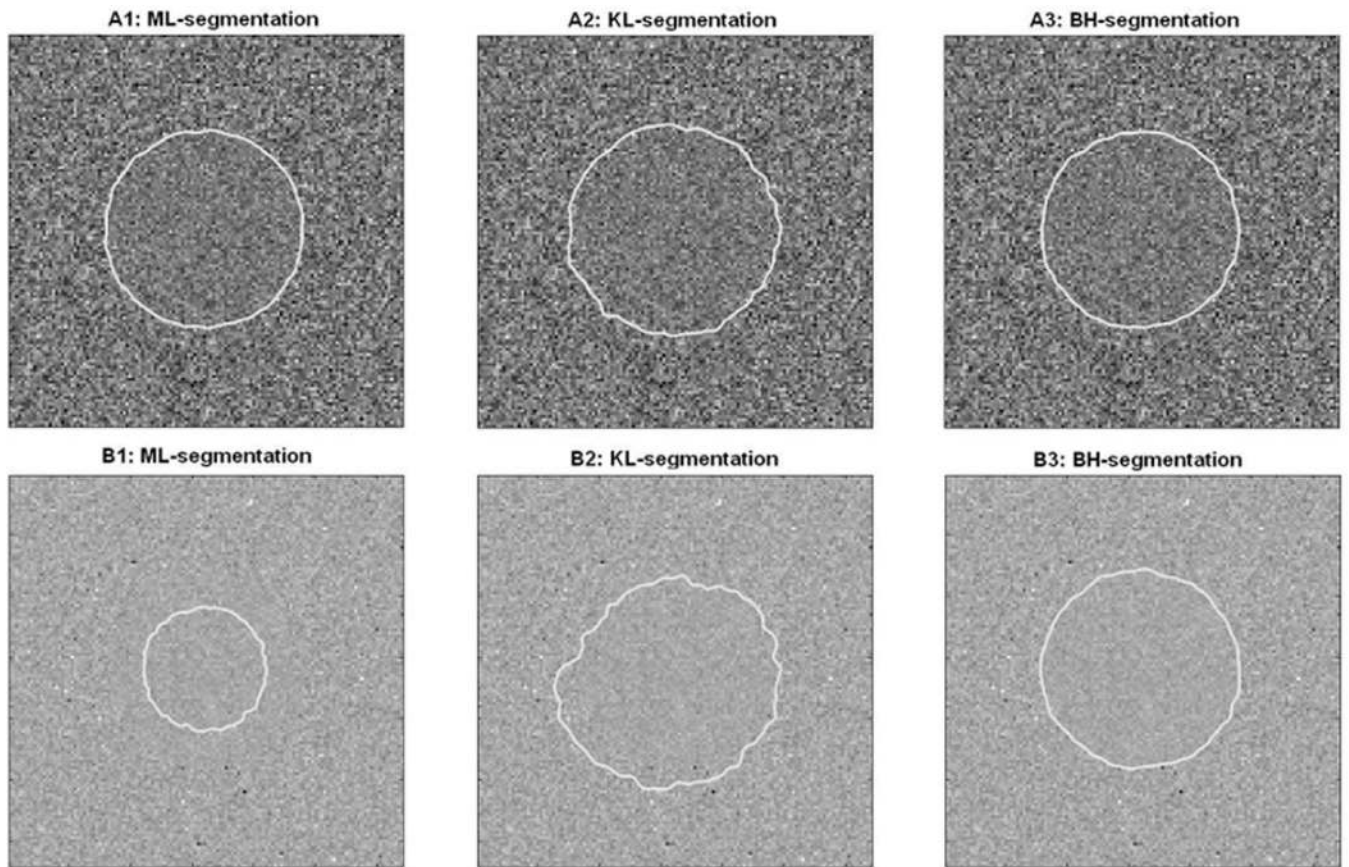

Fig. 9.
(Left) Original template. (Center) Outliers-free image. (Right) Image contaminated by outliers.

Fig. 10.
(Subplots A1–A3) Segmentation of the outliers-free image of Fig. 9 by the active contours maximizing the log-likelihood, KL divergence, and Bhattacharyya distance, respectively. (Subplots B1–B3) Segmentation of the outliers-contaminated image of Fig. 9 by the active contours maximizing the log-likelihood, KL divergence, and Bhattacharyya distance, respectively.

TABLE I

Pseudo-Code of the Proposed Segmentation Method

Set α, Δt

Initiate $\varphi_{t=0}(x)$

Estimate $P_-^{t=0}(z)$ and $P_+^{t=0}(z)$

For $t > 0$ until convergene

- ○ Compute $V(x)$ according to (12)–(13)

- ○ Compute $\kappa(x) = -\mathrm{div}\{\nabla\varphi(\mathrm{x})/\|\nabla\varphi(\mathrm{x})/\|\}$

- ○ Compute $\varphi_{t+1} = \varphi_t + \Delta t\,\delta(\varphi_t(x))\,(\alpha\,\kappa(x) - V(x))$

- ○ Update $P_-^t(z) \rightarrow P_-^{t+1}(z)$ and $P_+^t(z) \rightarrow P_+^{t+1}(z)$ according to (28)

- ○ Diffuse [if necessary] $P_-^{t+1}(z)$ and $P_+^{t+1}(z)$ as described in Section IV-C End

---