Guiding Image Segmentation on the Fly: Interactive Segmentation From a Feedback Control Perspective

Liangjia Zhu, Peter Karasev, Ivan Kolesov, Romeil Sandhu, *Member, IEEE*, and Allen Tannenbaum, *Fellow, IEEE*

***Abstract*—Image segmentation is a fundamental problem in computational vision and medical imaging. Designing a generic automated method that works for various objects and imaging modalities is a formidable task. Instead of proposing a new specific segmentation algorithm, we present a general design principle on how to integrate user interactions from the perspective of feedback control theory. Lyapunov stability analysis is employed to design and analyze an interactive segmentation system. Then, stabilization conditions are derived to guide the algorithm design. Finally, the effectiveness and robustness of the proposed method are demonstrated.**

***Index Terms*—Dynamical system, evolutionary process, feedback control, hybrid systems, interactive image segmentation.**

I. Introduction

THE problem of image segmentation has been an active research field over the past several decades and remains as a very challenging task. Although user (human) knowledge can recognize and partition an image into necessary regions, current automated algorithms fail to capture such boundaries on a consistent basis over wide ranging image modalities, i.e., there exists no "universal" segmentation algorithm. This said, the issue of effectively integrating user prior knowledge into a segmentation design is a driving principle behind existing state-of-the-art methods. These application-driven methodologies utilized prior models to aid in the segmentation process [1]. However, these methods remain automated and suffer from the same tacit issues for which they were designed to overcome. As a result, users directly participate in the segmentation loop in various image segmentation systems [2]. Given such input, the research community has classified this area of research as *interactive segmentation* [3]–[5]. Typically, one starts by initializing a method and then iteratively modifies the intermediate results until the algorithm obtains a satisfactory result. This process loop can be alternatively viewed as a feedback control process [6], whereby the user's editing action is a controller, the visualization/monitoring is an observer, and the changes of the intermediate results drive the system dynamics. Once the segmentation is complete, the system converges to an expected result. In this manner, the user automatically fills the "information" gap between an imperfect model to that of a desired segmentation. However, much of the work along this line of research does not explicitly model the user's role from the systems and control perspective. That is, user inputs are used *passively* without consideration to its contribution to the stability or convergence of the overall system. To the best of our knowledge, there are very few attempts to model segmentation from the perspective of feedback control.

Feedback principles have been used in the literature either based on empirical rules, such as boundary consistency [7] and connectivity of foreground pixels [8], or by modeling user contribution as a weighted term in an objective function [9], [10] to *close* the segmentation process. The important property of stability for a control system was not touched upon in these works. In our previous work [11], we formulated an interactive image segmentation methodology as a form of feedback control of a given partial differential equation (PDE) system and then derived its stability conditions. The method was chosen to be the classical region-based active contours for single-object segmentation. One major advantage of this formulation is that although the user does not know a perfect model/method for a segmentation task (which in fact rarely exists), one can start from a classic model and alter the segmentation in a principled manner. We can then quantitatively design and evaluate the performance of a control-based segmentation system.

Manuscript received April 17, 2017; revised August 21, 2017; accepted November 20, 2017. Date of publication January 11, 2018; date of current version September 25, 2018. This work was supported by the Air Force Research Laboratory under Grant FA9550-15-1-0045 and Grant FA9550-17-1-0435, by the Army Research Office under Grant W911NF-17-1-049, by the National Center for Research Resources under Grant P41-RR-013218, by the National Institute of Biomedical Imaging and Bioengineering under Grant P41-EB-015902, by National Cancer Institute under Grant 1U24CA18092401A1, by National Institute on Aging under Grant R01 AG053991, and by a grant from the Breast Cancer Research Foundation. Recommended by Associate Editor J.-i. Imura. *(Corresponding author: Liangjia Zhu.)*

L. Zhu, I. Kolesov, R. Sandhu, and A. Tannenbaum are with the Department of Computer Science and Applied Mathematics/Statistics, Stony Brook University, Stony Brook, NY 11794 USA (e-mail: liangjia.zhu@stonybrook.edu; ivan.kolesov@stonybrook.edu; romeil.sandhu@stonybrook.edu; allen.tannenbaum@stonybrook.edu).

P. Karasev was the Georgia Institute of Technology, Atlanta, GA 30332 USA. He is now with Agilent Technologies Inc., Santa Clara, CA 95051 USA (e-mail: pkarasev@gatech.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2018.2792328

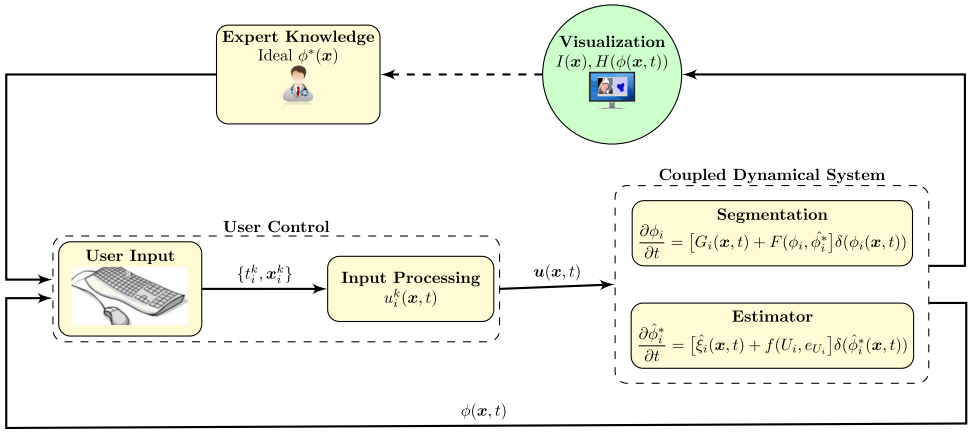


Fig. 1. Diagram of the control-based segmentation framework. The feedback compensates for deficiencies in automatic segmentation by utilizing the expert's knowledge.

In this paper and with regard to our previous work [11], we expand our methodology to generalized cases of evolution-based segmentation methods in addition to supporting multiple-object segmentation. Fig. 1 shows the diagram of the proposed framework. A user incorporates their prior knowledge to generate corrections as input to the closed-loop system. The segmentation boundary evolution and explicit estimate of the user's ideal segmentation are updated within an inner loop.

The main contributions of this work are the following.

1) We present a framework on how to design an interactive segmentation system from a feedback control perspective. Stabilization conditions for an interactive segmentation system are derived yielding a tangible framework for algorithm design and analysis. A key feature is that we are bridging image segmentation with control theory. As such, we are able to leverage new developments to design and analyze more sophisticated image segmentation systems. For example and in this work, a hybrid systems framework is adopted to model user input so as to provide a more practical approach than the continuous case derived in [11].
2) We present a relaxed framework with regard to the core algorithm (i.e., the only requirement is the ability to be described as a dynamical system). In short, this can be applied to multiobject segmentation, handles both region- and distance-based metrics, and supports scalar and vector images.
3) We present a framework capable of analyzing the behavior of some existing interactive segmentation methods, providing directions for further improvements.

We should note and emphasize that the proposed controlled segmentation framework offers control theoretics to enhance existing methods as opposed to replacing the method itself.

This paper is organized as follows. Section II reviews automated and interactive segmentation methods as well as basics from feedback control theory. Section III gives the formulation of an automated segmentation as a dynamical system for both region- and distance-based metrics. Section IV provides the control laws to stabilize the dynamical system for generic cases and shows specific examples on how to realize these control laws. Section V reports segmentation results followed by conclusion in Section VI.

II. Literature Review

A. (Semi)Automatic Image Segmentation

Variational image segmentation or active contour models are one class of algorithms commonly used in automated image segmentation. An underlying assumption employed in these algorithms is that the optimization of an energy functional defined over image features leads to an expected partition of the image [12]. Examples of commonly used image features are edges [13], regional statistics [14], and their combinations [15]. Multiobject segmentation has been addressed either by adding constraints to penalize violating areas [16] or by introducing competing components for adjacent regions [17], [18]. The shortest-distance-based image segmentation methods have been proposed along this line [19]. Image classification/clustering can also be modeled by using a variational framework [20]. Comprehensive introduction and review of variational segmentation methods can be found in [21] and [22] and references therein.

Alternative ways of representation, such as based on graph or clustering, have been widely used in image segmentation as well. See [23] and [24] and references therein for comprehensive reviews.

B. Interactive Image Segmentation

Automatic segmentation is attractive, but user intervention is inevitable in some critical tasks, especially for medical images [25]. The limitations of automatic or semiautomatic segmentation methods are well known and have been extended to integrate user interactions. Depending on the focus of these methods, the interactive segmentation methods are roughly summarized as follows.

1) User Interface: Some classic automatic segmentation methods have been combined with a powerful visualization and user editing system to achieve interactive segmentation. This is

popular in some medical image segmentation systems [2], [26], where classic methods such as region growing, active contours, and graph cuts are adopted as core techniques. Common features shared by these methods are efficient user interaction, easiness of manipulation, especially for 3-D volumes.

2) Segmentation Accuracy: Much of the research focus has placed emphasis on improving the segmentation accuracy per user input. Typically, image segmentation is associated with the optimization of an objective function. Thus, advances have been made either by improving classic objective functions such as active contours [10], graph cuts [27], and geodesic distance [28] or by proposing new formulations to interpret user input such as conditional random field [29] and random walker [5]. One of notable advances in this direction is the reformulation of some variational algorithms, which are subject to local minima, into convex optimization framework with global optimal solutions [30], [31].

3) User Input Modeling and Accumulation: How to understand the meaning of user input to better equip a segmentation algorithm has recently received attention in the literature. User input has been modeled as non-Euclidean kernels [32], combined with the shape constraint [33], and categorized [34] to reduce user efforts in interaction. Since user interaction is not an isolated and passive action, learning-based methods have been proposed to model the temporal/historic knowledge of user input to increase efficiency [35], [36].

4) System Design: With regard to the system design, recent work has illuminated the *causality* and overall performance of a system. User input is an integral part of the system, which is combined with the current and previous states to determine the next segmentation. Most of these algorithms are rooted either in variational [11], [37], [38] or graph-based segmentations [3], [4].

In addition to these four categories of research, graphics processing unit computing has also been increasingly utilized to develop real-time interactive segmentation systems [9], [31].

C. Feedback Control Theory

The principle of feedback is most often used in a control system. The basic idea is to use the difference between the signal to be controlled and a desired reference signal to determine system actions. *We should note that feedback control ought not to be confused with simply taking feedback in a controlled system—there is subtle but distinct difference to unambiguously indicate when the performance is guaranteed* [6], [39]. The second requirement can be considered as studying the stability of a system around a given state or an equilibrium point, referred to as *stabilization* of a system.

A common technique used in system stabilization is by analyzing the Lyapunov function of a system [40], [41]. The basic idea is to check whether the Laypunov function is dissipative along all possible trajectories of a dynamical system around the equilibrium point, rather than solving the system equation directly. It has been applied to stabilize systems driven by single [42] as well as coupled [43] PDEs. Lyapunov analysis has been extended to derive stability conditions for time-delayed systems [44] and discrete-time systems [45] and to synchronize two chaotic systems with impulsive signals [46].

Continuous dynamical systems with discrete events have been widely studied in impulsive control [47] and hybrid systems [48] because of their effectiveness, efficiency, and flexibility to address challenging stabilization problems in dynamical systems.

The existence of feedback has long been observed and utilized in modeling the human vision system [49]. However, only very few attempts have been made to employ the feedback principle to design an image processing system. Early research add feedback in image segmentation either based on empirical rules [7], [8] or by adding user contribution as a weighted term [9], [10], without addressing the issue of stability for these closed-loop systems. For example, the formulations in [9] and [10] can be rewritten as

$$\frac{\partial \phi}{\partial t} = \delta(\phi)(G(\phi) + \lambda L(\xi)) \tag{1}$$

where $\delta(\phi)$ is the derivative of the regularized Heaviside function, which is the level-set function, $G(\phi)$ represents contribution from image content, $L(\xi)$ models online labeling error ξ from user input, and λ is a *global* constant scalar that balances the user's influence to the segmentation. Refer to (4) and Section IV-B for details. The value of λ is usually determined empirically, rather than automatically adjusted to image content as in the proposed framework. In our previous work [11], interactive image segmentation is formulated as controlling region-based active contours, where the control law is derived rigorously by using the Laypunov stability theorem.

III. Automatic Segmentation as an Open-Loop System

A large class of segmentation algorithms can be considered as evolutionary processes, where these regions (or their boundaries) "compete" with each other and evolve based on certain quantifiable criteria [50]. Examples include region growing/competition, classical active contour models, and distance-based segmentation. Typically, the evolutionary process can be described by a dynamical system, driven by the optimization of certain energy functionals. As an example, the level-set formulation of active contours has been extensively employed; see [51] and [52] and the references therein. We note that level sets were proposed in [53] and subsequently developed and applied in many works.

Let $I : \Omega \to \mathbb{R}^n$ be an image defined on $\Omega \in \mathbb{R}^m$, where $m \geq 2$ and $n \geq 1$. Image segmentation consists of determining a partition $\mathcal{R} = \{\Omega_i\}, i = 1, \ldots, N$ such that $\Omega = \cup_{i=1}^{N} \Omega_i$ and $\Omega_i \cap \Omega_j = \emptyset$ if $i \neq j$ [50], [54].

Each region $\Omega_i(\boldsymbol{x})$ is an open space represented by the zero-level set of a smooth function $\phi_i : \Omega \to \mathbb{R}$, such that

$$\begin{aligned} \phi_i(\boldsymbol{x}) &> 0, \quad \text{for} \quad \boldsymbol{x} \quad \text{inside} \quad \Omega_i \\ \phi_i(\boldsymbol{x}) &= 0, \quad \text{for} \quad \boldsymbol{x} \in \partial\Omega_i \\ \phi_i(\boldsymbol{x}) &< 0, \quad \text{for} \quad \boldsymbol{x} \quad \text{outside} \quad \Omega_i. \end{aligned} \tag{2}$$

The regularized Heaviside function

$$H(\phi(\boldsymbol{x})) = \begin{cases} 1, & \text{if } \phi > \epsilon \\ 0, & \text{if } \phi < -\epsilon \\ \frac{1}{2}\left(1 + \frac{\phi}{\epsilon} + \frac{1}{\pi}\sin\left(\frac{\pi\phi}{\epsilon}\right)\right), & \text{otherwise} \end{cases} \tag{3}$$

is used to indicate the exterior and interior regions such that its derivative denoted by $\delta(\phi(\boldsymbol{x}))$ is well defined [52]. With a slight abuse of notation, these are denoted by $H(\phi)$ and $\delta(\phi)$ for simplicity hereinafter. Then, the dynamical system describing the evolution is defined as follows:

$$\begin{aligned} \frac{\partial \phi_i}{\partial t} &= G_i(\boldsymbol{x}, t)\delta(\phi_i) \\ \phi_i(\boldsymbol{x}, 0) &= \phi_i^0(\boldsymbol{x}) \end{aligned} \tag{4}$$

where $G_i : \Omega \times \mathbb{R}^+ \to \mathbb{R}$ represents the intrinsic dynamics that describes the evolution of the region $\Omega_i(\boldsymbol{x})$, and t is an artificial time that keeps track of the evolution process from an initial state $\phi_i^0(\boldsymbol{x})$. Without loss of generality, each $G_i(\boldsymbol{x}, t)$ may be decomposed into two *competing* components as

$$G_i(\boldsymbol{x}, t) = -\left[g_i(\boldsymbol{x}, t) - g_i^c(\boldsymbol{x}, t)\right] \tag{5}$$

where $g_i : \Omega_i \times \mathbb{R}^+ \to \mathbb{R}$ represents the contribution from $\Omega_i(\boldsymbol{x})$ and $g_i^c : \Omega_i^c \times \mathbb{R}^+ \to \mathbb{R}$ is that from all other $\Omega_j(\boldsymbol{x})$ for $j \neq i$. A negative sign is used in front of (5) due to the definition of H for the interior region. This way of decomposition ensures that no overlaps or gaps are developed between different regions during the evolution process, which has been used for modeling multiple active contours in different region-based algorithms [17], [18], [20]. On the other hand, segmentation algorithms that are based on clustering pixels according to their "distances" to given seed points naturally fit into this formulation, since image-based distance to given points can be implemented using the level-set formulation [19]. Therefore, the presented framework works for *region-based active contour models* and *distance-based clustering*.

A. Region-Based Active Contour Models

The function $g_i(\boldsymbol{x}, t)$ may be defined via the statistics of $I(\boldsymbol{x})$ inside the region $\Omega_i(\boldsymbol{x})$. A simple example of such a statistical measure is given by

$$g_i(\boldsymbol{x}, t) = \left[I(\boldsymbol{x}) - \mu_i(t)\right]^2 \delta(\phi_i) \tag{6}$$

where $I(\boldsymbol{x}) \in \mathbb{R}$ and $\mu_i(t)$ is the average value of $I(\boldsymbol{x})$ inside the region Ω_i at time t [14], [55], and

$$g_i^c(\boldsymbol{x}, t) = \min_{j \neq i} g_j(\boldsymbol{x}, t) \quad \text{for each } \boldsymbol{x} \in \Omega_i. \tag{7}$$

Other region-based energies [15], [18], [56] may be employed in a similar way.

B. Distance-Based Clustering

Let $\Omega_{i,\boldsymbol{x}}$ be a set of points inside a region Ω_i, which is referred to as "seed points" hereafter. The distance between a point $\boldsymbol{x} \in \Omega$ and the seed region is defined as

$$d(\boldsymbol{x}, \Omega_{i,\boldsymbol{x}}) = \min_{\boldsymbol{y} \in \Omega_{i,\boldsymbol{x}}} \min_{C \in \theta(\boldsymbol{x}, \boldsymbol{y})} \int_0^1 g_\gamma(C(p)) \|C'(p)\| dp \tag{8}$$

where $\theta(\boldsymbol{x}, \boldsymbol{y})$ is the family of all curves connecting points $\boldsymbol{x}$ and $\boldsymbol{y}$ on the image, and $p \in [0, 1]$ is the parameterization of a specific path $C : [0, 1] \to \mathbb{R}^m$ weighted by an image-based function $g_\gamma : \mathbb{R}^m \to \mathbb{R}^+$. The distance $d(\boldsymbol{x}, \Omega_{i,\boldsymbol{x}})$ may be computed using the level-set formulation as well by interpreting it as a front propagation problem with an image-dependent distance measure $1/g_\gamma$, where $g_\gamma(I) = 1 + \|\nabla I\|_2^2$ [19]. After computing the distance from a point to each seed region, the point is assigned to the closest region.

With a slight abuse of notation, let $d_i(\boldsymbol{x}, t)$, $d_{i^c}(\boldsymbol{x}, t)$, and $d_{\min}(\boldsymbol{x}, t)$ be the distance between point $\boldsymbol{x}$ and region Ω_i, the shortest distance between the point $\boldsymbol{x}$ and any regions other than $\Omega_i(\boldsymbol{x}, t)$, and the shortest distance between the point and all regions, respectively. An example of natural dynamics acting on ϕ_i is defined as

$$g_i(\boldsymbol{x}, t) = \begin{cases} g_\gamma(I), & \text{if } d_i(\boldsymbol{x}, t) \neq d_{\min}(\boldsymbol{x}, t) \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

and

$$g_i^c(\boldsymbol{x}, t) = \begin{cases} g_\gamma(I), & \text{if } d_{i^c}(\boldsymbol{x}, t) \neq d_{\min}(\boldsymbol{x}, t) \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

Equations (9) and (10) are used in the evolution of (4) and (5). This formulation is essentially a clustering process based on the shortest distance from a point to all regions.

Fig. 2 shows an example of automatic segmentation by using the models described in this section. With simple user initializations, either circular regions or scribbles, these two classical algorithms capture the majority of the objects, while missing some details. In particular, the region-based active contour has both evolution leakage and unreached regions, while the distance-based clustering fails to mark correct boundaries, where seed regions were not properly specified. Though sophisticated methods [28], [30], [31] may achieve better segmentation results with the same initializations, our main focus is how to improve these classical methods by integrating user inputs from the perspective of feedback control.

IV. Interactive Segmentation as Feedback Control

Suppose the user has an ideal segmentation of the image domain into regions in mind: $\{\phi_i^*(\boldsymbol{x})\}$ for $i = 1, \ldots, N$. Then, the goal is to design a feedback control system

$$\begin{aligned} \frac{\partial \phi_i}{\partial t} &= \left[G_i(\boldsymbol{x}, t) + F(\phi_i, \phi_i^*)\right]\delta(\phi_i) \\ \phi_i(\boldsymbol{x}, 0) &= \phi_i^0(\boldsymbol{x}) \end{aligned} \tag{11}$$

such that $\lim_{t \to \infty} \phi_i(\boldsymbol{x}, t) \to \phi_i^*(\boldsymbol{x})$ for $i = 1, \ldots, N$, where $F(\phi_i, \phi_i^*)$ is the control law to be defined below.

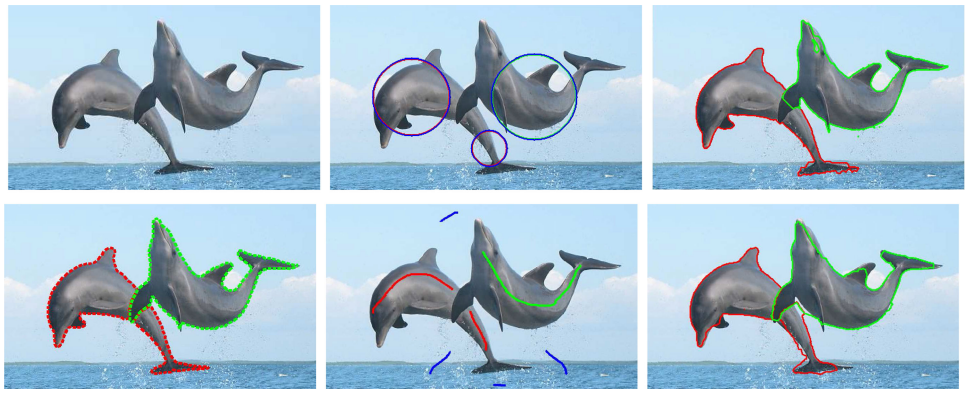


Fig. 2. Example of automatic segmentation. First column: source image and manual segmentation. Second column: initialization for region-based active contour (top) and distance-based clustering (bottom). Last column: automatic segmentation results.

A. Existence of a Regulatory Control

Define $\xi := \{\xi_1, \ldots, \xi_N\}$ and $V(\xi, t)$ as the pointwise and total labeling error, respectively:

$$\xi_i(\boldsymbol{x}, t) = H(\phi_i) - H(\phi_i^*) \tag{12}$$

$$V(\xi, t) = \frac{1}{2}\sum_{i=1}^{N}\int_\Omega \xi_i^2(\boldsymbol{x}, t)d\boldsymbol{x}. \tag{13}$$

If ϕ_i^*s are given and $V(\xi, t) \in C^1$, the determination of $F(\cdot, \cdot)$ is straightforward by applying the Lyapunov's direct method for stabilization. The control signal F uses the bounds of the image-dependent term $G_i(\boldsymbol{x}, t)$

$$g_M(\boldsymbol{x}) := \sup_{\forall t \in \mathbb{R}^+, i=1,\ldots,N} \{|G_i(\boldsymbol{x}, t)|\}. \tag{14}$$

Lemma 1: For the image bound defined in (14), there exists at least one label, $i \in \{1, \ldots, N\}$, such that $G_i(\boldsymbol{x}, t) < g_M(\boldsymbol{x})$ for a point on the zero-level set of ϕ_i.

Proof: By contradiction, otherwise, we have $G_i(\boldsymbol{x}, t) \geq g_M(\boldsymbol{x}), \forall i = 1, \ldots, N$. This may only hold when $G_i(\boldsymbol{x}, t) = g_M(\boldsymbol{x}), \forall i = 1, \ldots, N$, which implies points along the boundaries of adjacent regions will move in different directions. This contradicts the condition of no overlap between any two regions. ■

Theorem IV.1: Notation as above. The control law

$$F(\phi_i, \phi_i^*) = \alpha_i^2(\boldsymbol{x}, t)\xi_i(\boldsymbol{x}, t), \tag{15}$$

where $\alpha_i^2(\boldsymbol{x}, t) \geq g_M(\boldsymbol{x})$, asymptotically stabilizes the system (11) from $\{\phi_i(\boldsymbol{x}, t)\}$ to $\{\phi_i^*(\boldsymbol{x})\}, i = 1, \ldots, N$, for ϵ defined in (3) sufficiently small.

Furthermore, if ξ_i is large in the sense of

$$\int_\Omega \delta^2(\phi_i)\xi_i^2(\boldsymbol{x}, t)d\boldsymbol{x} \geq \rho\int_\Omega \xi_i^2(\boldsymbol{x}, t)d\boldsymbol{x}, i = 1, \ldots, N \tag{16}$$

for a given constant $\rho > 0$, then there exists $\nu > 0$ such that the control law exponentially stabilizes the system with a convergence rate of $e^{-\nu t}$.

Proof: As H is differentiable and continuous of ϵ, so does the ξ_i. Note that $\xi \to \{-1, 0, 1\}$ as $\epsilon \to 0$, we first look at the case $\xi \in \{-1, 0, 1\}$. Take $V(\xi, t)$ as the candidate Lyapunov function and differentiate it with respect t. Take $V(\xi, t)$ as the candidate Lyapunov function and differentiate it with respect to t

$$\begin{aligned}
\frac{dV(\xi, t)}{dt} &= \sum_{i=1}^{N}\int_\Omega \xi_i\frac{\partial \xi_i}{\partial t}d\boldsymbol{x} \\
&= \sum_{i=1}^{N}\int_\Omega \xi_i\delta(\phi_i)\frac{\partial \phi_i}{\partial t}d\boldsymbol{x} \\
&= \sum_{i=1}^{N}\int_\Omega \delta^2(\phi_i)[G_i\xi_i - \alpha_i^2\xi_i^2]d\boldsymbol{x} \\
&\leq \sum_{i=1}^{N}\int_\Omega \delta^2(\phi_i)[|G_i||\xi_i|^2 - \alpha_i^2\xi_i^2]d\boldsymbol{x} \\
&= \sum_{i=1}^{N}\int_\Omega \delta^2(\phi_i)\xi_i^2[|G_i| - \alpha_i^2]d\boldsymbol{x} \\
&< \sum_{i=1}^{N}\int_\Omega \delta^2(\phi_i)\xi_i^2[g_M - \alpha_i^2]d\boldsymbol{x} \\
&\leq 0.
\end{aligned} \tag{17}$$

The first inequality is from the fact that $\xi_i \in \{-1, 0, 1\}$. The last inequality holds because of the Lemma 1. Therefore, $dV(\xi, t)/dt$ is negative definite under the given condition due to the continuity of ξ_i on ϵ.

Furthermore, if the condition (16) is satisfied, from (17), we have

$$\begin{aligned}
\frac{dV(\xi, t)}{dt} &\leq \sum_{i=1}^{N}\int_\Omega \delta^2(\phi_i)\xi_i^2[|G_i| - \alpha_i^2]d\boldsymbol{x} \\
&= -\sum_{i=1}^{N}\int_\Omega \delta^2(\phi_i)\xi_i^2[\alpha_i^2 - |G_i|]d\boldsymbol{x} \\
&\triangleq -\sum_{i=1}^{N}\int_\Omega \delta^2(\phi_i)\xi_i^2\nu_i d\boldsymbol{x} \\
&= -\sum_{i=1}^{N}\tilde{\nu}_i\int_\Omega \delta^2(\phi_i)\xi_i^2 d\boldsymbol{x}
\end{aligned}$$

$$= -\tilde{\nu}\sum_{i=1}^{N}\int_{\Omega}\delta^2(\phi_i)\xi_i^2 d\boldsymbol{x}$$
$$\leq -2\tilde{\nu}\rho V(\xi,t) \tag{18}$$

where $\tilde{\nu}\in\left(0,\max_{\{\boldsymbol{x},i\}}\left\{\alpha_i^2-|G_i|\right\}\right)$. The mean value theorem is applied to move the ν_i out of the integral. Note that $V(\xi,t)$ is in fact defined as $0.5\|\xi\|_2^2$. Therefore, the system is exponentially stable with the convergence rate of $\nu=2\tilde{\nu}\rho$ under the given condition. ■

Remark 1: This theorem gives a sufficient condition for the existence of a control law that stabilizes the dynamical system to a given desired steady-state value. Intuitively, the control law defines a localized input required to change the intrinsic dynamics of the autonomous system.

B. Label-Error Estimation

In practice, $\{\phi_i^*\}$ is not given or completely available beforehand. User input based on current segmentation is utilized to predict/estimate the ideal segmentation on the fly.

1) User Input Processing: User interaction is modeled as a binary decision as to whether a given location is correctly labeled as inside or outside the expected segmentation. Let $L=\{1,\ldots,N\}$ be the set of labels corresponding to regions $\Omega_i(\boldsymbol{x})$. User inputs are properly applied to the image to capture segmentation errors. Afterwards, the effect of user input is propagated.

Let $\Omega_{t_i^k}$ be the kth user input applied to the ith label at time t_i^k. The user input is modeled as a discrete event

$$u_i^k(\boldsymbol{x},t_i^k)=\begin{cases}q, & \text{if} \quad \boldsymbol{x}\in\Omega_{t_i^k}\\ p, & \text{if} \quad \boldsymbol{x}\in\Omega\backslash\Omega_{t_i^k}\end{cases} \tag{19}$$

where $q,p\in\mathbb{R}$ are constants that model the instant effect of the user input.

2) Accumulation of User Input: Let $\boldsymbol{u}(\boldsymbol{x},t)$ denote the set of all user input effects at time t. For a given label i, let $\boldsymbol{u}_i$ denote the accumulated effect from all $u_i^k(\boldsymbol{x},t)$ and $\boldsymbol{u}_i^c$ be the effect all $u_j^k(\boldsymbol{x},t)$, $j\neq i$, $j=1,\ldots,N$ and $k=1,2,\ldots$. The total effect of user input for label i, denoted by $U_i(\boldsymbol{x},t)$, is given as a function of $\boldsymbol{u}_i(\boldsymbol{x},t)$ and $\boldsymbol{u}_i^c(\boldsymbol{x},t)$. As the explicit form of user input is metric dependent, see (38) and (40) for detailed definitions.

3) Label-Error Estimation: Let $\{\hat{\phi}_i^*\}$ be an estimate of $\{\phi_i^*\}$ and define the error terms by

$$\hat{\xi}_i(\boldsymbol{x},t):=H(\phi_i)-H(\hat{\phi}_i^*)$$
$$e_{U_i}(\boldsymbol{x},t):=H(\hat{\phi}_i^*)-H(U_i). \tag{20}$$

The feedback in (11) will use the estimate $\{\hat{\phi}_i^*\}$:

$$\frac{\partial\phi_i}{\partial t}=[G_i(\boldsymbol{x},t)+F(\phi_i,\hat{\phi}_i^*)]\delta(\phi_i)$$
$$\phi_i(\boldsymbol{x},0)=\phi_i^0(\boldsymbol{x}). \tag{21}$$

The estimator is an observer-like system driven by accumulated user input U_i with an error term e_{U_i} given by

$$\frac{\partial\hat{\phi}_i^*}{\partial t}=\left[\hat{\xi}_i+f\left(U_i,e_{U_i}\right)\right]\delta(\hat{\phi}_i^*)$$
$$\hat{\phi}_i^*(\boldsymbol{x},0)=\phi_i^0(\boldsymbol{x}) \tag{22}$$

where $f\left(U_i,e_{U_i}\right)$ is a tuning function such that the total labeling error $V(\hat{\xi},t):=E(t)+\hat{V}(\hat{\xi},t)$ has a negative semidefinite derivative. Here

$$E(t):=\frac{1}{2}\sum_{i=1}^{N}\int_{\Omega}|U_i|e_{U_i}^2 d\boldsymbol{x} \quad \text{(estimator versus user input)} \tag{23}$$

$$\hat{V}(\hat{\xi},t):=\frac{1}{2}\sum_{i=1}^{N}\int_{\Omega}\hat{\xi}_i^2 d\boldsymbol{x} \quad \text{(estimator versus visualization)} \tag{24}$$

with $\hat{\xi}=\cup_{i=1}^{N}\hat{\xi}_i(\boldsymbol{x},t)$.

In addition to stabilizing $V(\hat{\xi},t)$, the control proposed in Theorem IV.2 is designed to achieve a useful qualitative behavior. When the user is satisfied with the agreement between ϕ_i and their ideal ϕ_i^*, it is assumed that U_i remains constant; either the user never needed to apply a correction near $\boldsymbol{x}$ or has otherwise stopped adding more inputs. In this case, $\hat{\phi}_i^*$ should follow ϕ_i. Conversely, when U_i grows due to persistent user input, $\hat{\phi}_i^*$ is to become increasingly driven toward U_i irrespective of agreement between $\hat{\phi}_i^*$ and ϕ_i. Subsequently, $\hat{\phi}_i^*$ should pull ϕ_i along due to the coupling term $F(\phi_i,\hat{\phi}_i^*)$ [see (21)] in the closed-loop dynamics of ϕ_i.

Theorem IV.2: Notation as in Theorem IV.1. Let $f(U_i,e_{U_i})=-|U_i|e_{U_i}$ and consequently

$$\frac{\partial\hat{\phi}_i^*}{\partial t}=\left[\hat{\xi}_i-|U_i|e_{U_i}\right]\delta(\hat{\phi}_i^*). \tag{25}$$

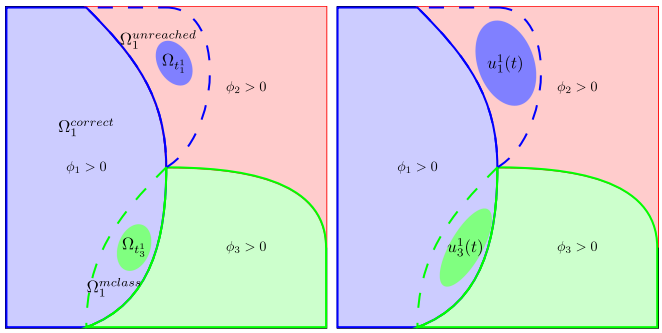
Assume that user input has stopped (U_i remains constant). Then, $V(\hat{\xi},t)$ has a negative-semidefinite derivative, namely

$$V'(\hat{\xi},t)\leq-\sum_{i=1}^{N}\int_{\Omega}\delta^2(\hat{\phi}_i^*)\left[\hat{\xi}_i-|U_i|e_{U_i}\right]^2 d\boldsymbol{x}. \tag{26}$$

Proof: Computing the time derivative $V'(\hat{\xi},t)=E'(t)+\hat{V}'(\hat{\xi},t)$, we obtain

$$E'(t)=\sum_{i=1}^{N}\int_{\Omega}|U_i|e_{U_i}\frac{\partial e_{U_i}}{\partial t}d\boldsymbol{x}$$
$$=\sum_{i=1}^{N}\int_{\Omega}|U_i|e_{U_i}\left(\delta(\hat{\phi}_i^*)\frac{\partial\hat{\phi}_i^*}{\partial t}\right)d\boldsymbol{x} \tag{27}$$

$$\hat{V}'(\hat{\xi},t)=\sum_{i=1}^{N}\int_{\Omega}\hat{\xi}_i\frac{\partial\hat{\xi}_i}{\partial t}d\boldsymbol{x}$$
$$=\sum_{i=1}^{N}\int_{\Omega}\hat{\xi}_i\left[\delta(\phi_i)\frac{\partial\phi_i}{\partial t}-\delta(\hat{\phi}_i^*)\frac{\partial\hat{\phi}_i^*}{\partial t}\right]d\boldsymbol{x}. \tag{28}$$



(a) (b)

Fig. 3. Example of segmentation errors with (a) user inputs and (b) propagation of the user input effect.

Substituting for $\frac{\partial \phi_i}{\partial t}$ and $\frac{\partial \hat{\phi}_i^*}{\partial t}$, we obtain

$$E'(t) = \sum_{i=1}^{N} \int_\Omega \delta^2(\hat{\phi}_i^*)[-|U_i|^2 e_{U_i}^2 + |U_i| e_{U_i} \hat{\xi}_i] d\boldsymbol{x} \quad (29)$$

$$\hat{V}'(\hat{\xi}, t) = \sum_{i=1}^{N} \int_\Omega \delta^2(\phi_i)[G_i \hat{\xi}_i - \alpha_i^2 \hat{\xi}_i^2] d\boldsymbol{x} - \sum_{i=1}^{N} \int_\Omega \delta^2(\hat{\phi}_i^*)[\hat{\xi}_i^2 - |U_i| e_{U_i} \hat{\xi}_i] d\boldsymbol{x}. \quad (30)$$

Adding (29) and (30) and combining the $\delta^2(\hat{\phi}_i^*)$ terms, we obtain

$$E'(t) + \hat{V}'(\hat{\xi}, t) = \sum_{i=1}^{N} \int_\Omega \delta^2(\phi_i)[G_i \hat{\xi}_i - \alpha_i^2 \hat{\xi}_i^2] d\boldsymbol{x} - \sum_{i=1}^{N} \int_\Omega \delta^2(\hat{\phi}_i^*)[\hat{\xi}_i - |U_i| e_{U_i}]^2 d\boldsymbol{x}. \quad (31)$$

When α_i satisfies Theorem IV.1, it follows that

$$V'(\hat{\xi}, t) \leq -\sum_{i=1}^{N} \int_\Omega \delta^2(\hat{\phi}_i^*) \left[\hat{\xi}_i - |U_i| e_{U_i}\right]^2 d\boldsymbol{x}. \quad (32)$$

■

C. Hybrid System: For Efficient User Interactions

At a given time t, the regions of current and ideal segmentation for a given label i, denoted by $\Omega_i(\boldsymbol{x})$ and $\Omega_i^*(\boldsymbol{x})$, respectively, can be represented as

$$\Omega_i(\boldsymbol{x}) = \Omega_i^{\text{correct}}(\boldsymbol{x}) \cup \Omega_i^{\text{mclass}}(\boldsymbol{x})$$
$$\Omega_i^*(\boldsymbol{x}) = \Omega_i^{\text{correct}}(\boldsymbol{x}) \cup \Omega_i^{\text{unreached}}(\boldsymbol{x}) \quad (33)$$

where $\Omega_i^{\text{correct}}(\boldsymbol{x})$ is the correct segmentation, $\Omega_i^{\text{mclass}}(\boldsymbol{x})$ is the region misclassified as the label i, and $\Omega_i^{\text{unreached}}(\boldsymbol{x})$ is the region not reached by the label i. Note that $\Omega_i^{\text{mclass}}(\boldsymbol{x})$ is also an unreached region of other labels. See Fig. 3, where $\Omega_1^{\text{mclass}}(\boldsymbol{x}) = \Omega_3^{\text{unreached}}(\boldsymbol{x})$ and user inputs, $\Omega_{t_1^1}$ and $\Omega_{t_3^1}$, are focused on the unreached regions and then propagated. User inputs are applied as a sequence of discrete events to the coupled dynamical system to form a hybrid system, rather than waiting for their effect to reach current fronts as in [11].

Let $P : \Omega \to \mathbb{R}$ be such that $H(P(\Omega_{t_i^k})) = H(U_i(\Omega_{\boldsymbol{x}_i^k}))$ for a user input applied at $\Omega_{t_i^k}$ and t_i^k. Then, a hybrid system is formed as the coupled system (21), (25) with a sequence of discrete events at time $t_i^k, k = 1, \ldots,$

$$\Delta\hat{\phi}_i^*(\Omega_{t_i^k}, t_i^{k+}) = P(\Omega_{t_i^k}) - \hat{\phi}_i^*(\Omega_{t_i^k}, t_i^k)$$
$$\Delta\phi_i(\Omega_{t_i^k}, t_i^{k+}) = \hat{\phi}_i^*(\Omega_{t_i^k}, t_i^{k+}) - \phi_i(\Omega_{t_i^k}, t_i^k). \quad (34)$$

One simple example of P is $P = \text{sign}(U_i)$.

Assumption 1: We assume that the size of each unreached region $\Omega_i^{\text{unreached}}(\boldsymbol{x})$ can be driven sufficiently small (relative to the level of image resolution) by a *finite* sequence of user inputs $\{\Omega_{t_i^k}\}$ at time t_i^k, $k = 1, \ldots, K_i$, where K_i is the number of user inputs for label i.

Remark 2: The rationale for this assumption is based on the fact that since $\Omega_i^{\text{unreached}}(\boldsymbol{x})$ is compact, then it has a *finite subcover*, consisting of a finite set $\{\omega_i^k\}$ with $\overline{\omega_i^k} = H(u_i^k(t))$ from the user inputs u_i^k defined in (19). In the extreme case, the user may manually accomplish the segmentation task with a finite number of user interactions. Thus, it is reasonable to assume the existence of a "finite" number of user inputs required for a desired segmentation. This is of course always true in practice.

Corollary IV.3: Consider the coupled system (21), (25) with Assumption 1. Then, the hybrid system (34) has the same stability properties as the dynamical system defined by (21) and (25) and, thus, converges to the ideal segmentation.

Proof: Since Theorem IV.2 and Assumption 1 are satisfied, one has that $V'(\hat{\xi}, t) \leq 0$ for almost all $t \in \mathbb{R}^+$. In addition, there are only a finite number of discrete events. Thus, the bound of $V'(\hat{\xi}, t)$ in (26) still holds for $t > t_K$, where t_K is the last instant of impulse. Moreover, Assumption 1 ensures that one will get the desired ideal segmentation. ■

Remark 3: The gist of the above result is that if the number of discrete inputs is finite, then the asymptotic property of the derivative of the total error $V'(\hat{\xi}, t)$ will not be affected by the discrete impulses. Therefore, the stability of the coupled system as defined by (21) and (25) does not change with the finite number of impulses (discrete inputs).

Remark 4: Applying user input immediately to the segmentation process has been empirically used in many interactive segmentation algorithms [3], [4], [28]. This corollary sheds light on the rationale of this commonly used strategy from the perspective of hybrid systems.

D. Examples of Control-Based Segmentation Methods

Two representative methods are discussed in the following section.

1) Controlled Region-Based Active Contour Models: The natural dynamics $G_i(\boldsymbol{x}, t)$ described in Section III-A are used. Given a user input as defined in (19) with $p = 1$ and $q = 0$, the effect of which is strengthened by applying a kernel-based method [32] to u_i^k to extend it as

$$u_i^k(\boldsymbol{x}, t_i^{k+}) = h_0(d(\boldsymbol{x}, \Omega_{t_i^k})). \quad (35)$$

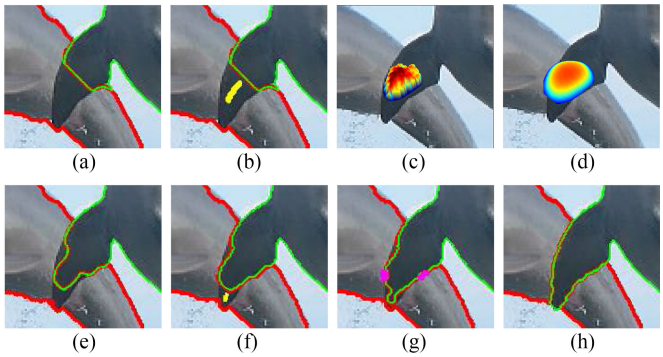


Fig. 4. Example of the user effect to segment flipper using the region-based active contour model. Two dolphins are marked as red (label $L = 1$) and green (label $L = 2$), respectively. (a) Before user interaction. (b) First user input for $L = 2$, $\Omega_{t_2^1}$ at time t_2^1 and (c) extended effect of user input $u_2^1(\boldsymbol{x}, t_2^{1+})$ at time t_2^{1+}. (d) Propagation of the user input effect $\boldsymbol{u}_2(t)$ and (e) the contours at a later time. (f) One more input $\Omega_{t_2^2}$ for $L = 2$. (g) Two inputs for $L = 1$ and (h) the final segmentation.

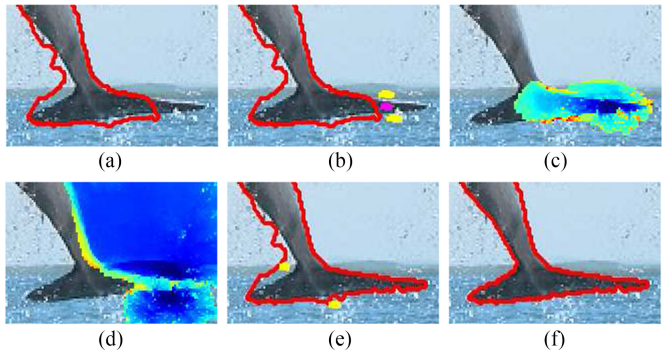


Fig. 5. Example of the user effect to segment flukes using distance-based clustering. (a) Before user interaction. (b) Inputs for flukes (magenta, label $L = 1$) and background (yellow, label $L = 2$). (c) Prorogation of $\boldsymbol{u}_1(\boldsymbol{x}, t)$ from $\Omega_{\boldsymbol{x}_1}$ (dark blue) and (d) $\boldsymbol{u}_2(\boldsymbol{x}, t)$ from $\Omega_{\boldsymbol{x}_2}$ (dark blue). (e) Two inputs for label 2 and (f) the final segmentation.

Here, $d(\boldsymbol{x}, \Omega_{t_i^k})$ is the weighted distance from $\boldsymbol{x}$ to $\Omega_{t_i^k}$ defined in (8), and h_0 is a decreasing function with respect to this distance. In this paper, we use $h_0 = (d_{\max} - d)/(d_{\max} - d_{\min})$, where $d, d_{\max},$ and $d_{\min}$ are the distance and its corresponding extrema, respectively. The value of $d_{\max}$ controls the maximal range the current input affects. The overall effect for label $L = i$ is defined as

$$\boldsymbol{u}_i(\boldsymbol{x}, t) = \sum_k u_i^k(\boldsymbol{x}, t). \tag{36}$$

To propagate and smooth the effect of the user input, a diffusion process is applied to $\boldsymbol{u}_i(\boldsymbol{x}, t)$ as in [11], where

$$\begin{aligned} \frac{\partial \boldsymbol{u}_i}{\partial t} &= \boldsymbol{u}_i + \nabla \cdot \left[H\big((\boldsymbol{u}_i/g_{\mathrm{M}})^2 - 1\big) \nabla \boldsymbol{u}_i \right] \\ \boldsymbol{u}_i(\boldsymbol{x}, 0) &= 0. \end{aligned} \tag{37}$$

The total user input effect is defined as

$$U_i(\boldsymbol{x}, t) := \boldsymbol{u}_i(\boldsymbol{x}, t) - \sum_{j \neq i} \boldsymbol{u}_j(\boldsymbol{x}, t). \tag{38}$$

Then, we have a region-based method system by plugging U_i into (25) of the hybrid system. An example of segmenting the dolphin flipper using the proposed method is shown in Fig. 4.

2) Controlled Distance-Based Clustering: In this example, we use the simple natural dynamics $G_i(\boldsymbol{x}, t)$, as described in Section III-B. More sophisticated schemes such as [28] may be employed. The effect of the user input is defined to have the same metric as the system equations

$$\begin{aligned} u_i^k(\boldsymbol{x}, t) &= \min_{\boldsymbol{y} \in \Omega_{t_i^k}} \min_{C \in \theta(\boldsymbol{x}, \boldsymbol{y})} \int_0^1 g_\gamma(C(l)) \|C'(l)\| dl \\ u_i^k(\Omega_{t_i^k}, t_i^k) &= q, \quad u_i^k(\Omega_{t_i^k}^c, t_i^k) = p \end{aligned} \tag{39}$$

with $q = 0$ and $p = \infty$, where $\theta(\boldsymbol{x}, \boldsymbol{y})$ is the set of all paths connecting points $\boldsymbol{x}$ and $\boldsymbol{y}$, and l is the parameterization of a particular path C weighted by the function g_γ.

The total user input effect at point $\boldsymbol{x}$ is computed as

$$U_i(\boldsymbol{x}, t) := - \min_{u_k \in \boldsymbol{u}_i, \boldsymbol{y} = \boldsymbol{x}} u_k(\boldsymbol{y}, t) + \min_{u_k \in \boldsymbol{u}_i^c, \boldsymbol{y} = \boldsymbol{x}} u_k(\boldsymbol{y}, t). \tag{40}$$

An illustration of segmenting the dolphin flukes using the proposed distance-based clustering algorithm is given in Fig. 5.

As can be seen from these simple examples, the proposed algorithm requires only a small amount of user interactions to correct the segmentations toward ideal boundaries.

E. Implementation

Efficiency is a key factor in determining the performance of an interactive system. Thus, it is desired to have minimal as possible computational loads and memory costs in the implementation of the feedback control system. To this end, inputs to each object are grouped and evolved by using a single unit/array. That is, the state of $\boldsymbol{u}_i(\boldsymbol{x}, t)$ is recorded by using a single array, and there are total of N arrays to keep track of the user input effect. In addition, depending upon the degree of overlaps between regions, independent regions of interest may be specified ahead of time in order to reduce multiple region segmentations into separate individual binary segmentations. This is a common case in medical image segmentation where target structures do not have overlaps (see Fig. 9 for example).

In addition, efficient implementations were employed for the original automatic methods. Specifically, for the *region-based active contour model*, a sparse-level-set implementation [57] was used that keeps track of $\phi(x, t) = 0$ without reinitialization. Furthermore, only five layers, two inside and two outside of the zero-level set, are required. This reduces the computational complexity dramatically as the dimension during evolution is of the same order as the region boundary length/area in the 2-D/3-D case. In the *distance-based clustering* formulation, since G is locally static [see (9)], the evolution reduces to solving a static Hamilton–Jacobi equation [52]. Thus, distance information can be computed locally in a monotonic way for which efficient numerical schemes such as those given in [58] and [59] may be applied. In this case, only one array is required to keep track of all $\boldsymbol{u}_i(\boldsymbol{x}, t), i = 1, \ldots, N$, which makes it possible to achieve

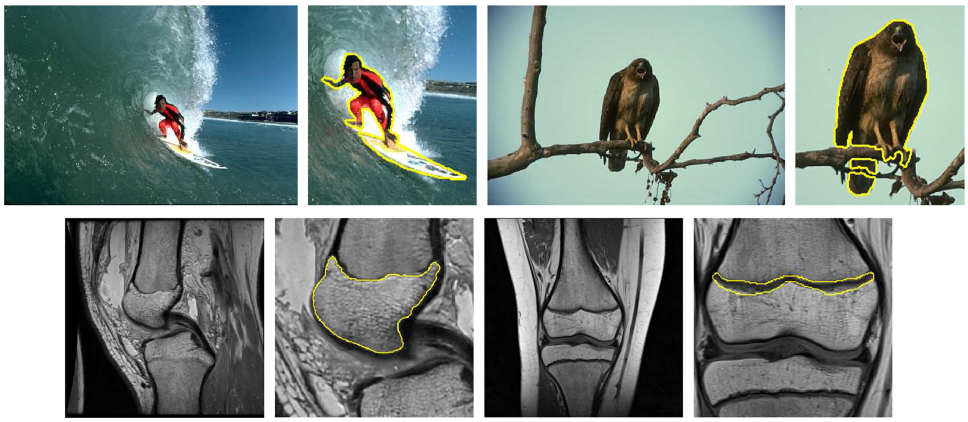


Fig. 6. General (first row) and medical (second row) images used in a quantitative comparison of GrabCut and the proposed algorithm. Manual segmentations are marked in yellow.

real-time interactive segmentation even for 3-D medical images. If no user input is employed, the algorithm has a similar structure as in image segmentation utilizing the fast marching method as originally proposed in [58] and developed and extended in many other works; see [51], [52], and [60] and the references therein. Introducing the control framework enables the algorithm to be nonstatic and, as such, results in increased flexibility.

V. Experimental Results

In this section, we first present examples of the user effect for both *region-based active contours* and *distance-based clustering*. Then, we demonstrate the advantages (and/or disadvantages) of the proposed control framework by comparing it to a popular interactive segmentation method in terms of *user effort* and *predictability*. Next, we present results of applying the proposed framework for segmenting challenging medical images. Finally, we use examples to illustrate the relations between the control-based method and some existing algorithms.

In this section, the localized region-based active contour energy [15] was implemented for the *region-based active contour model*, and a gradient-based distance measure [19] was used for the *distance-based clustering* methodology.

A. Effectiveness of the Proposed Control Framework

1) Selection of Data: Two general images from [61] and two medical images were employed to quantitatively compare the presented methods with the popular GrabCut algorithm [3]. The general images considered are given in the first row of Fig. 6. Specifically, these images were chosen to illustrate strong local contrast and at varying parts of the image (e.g., ambiguous boundaries at the bottom of the bird image). The medical images seen in the second row of Fig. 6 present a different issue—the targeted object (epiphysis/physis) has an intensity profile that is comparable to surrounding objects within the background. In short, the example in Fig. 6 is shown to illustrate and motivate the proposed framework.

2) Quantitative Comparison of User Effort: A location through which the cursor was dragged is defined as an "actuated voxel," and the *total actuated voxels* is a robust indicator of user effort to complete a segmentation.

In this test, the interactive user input via mouse click-and-drag was implemented and measured identically for each algorithm. The extent of the neighborhood around the cursor that marks seed regions in GrabCut were not counted toward the total actuated voxels.

Three experiments were conducted with different initializations for each image. The actuated pixels after initialization are shown in Fig. 7. At termination, all of the segmentations have greater than 95% overlap with a manually segmented reference. These results show the different characteristics of these algorithms. The proposed algorithm has a lower mean actuated count in all images and tighter clustering in three of them (except the bird image) across repeated segmentations. The wider cluster in the bird image segmentation from the region-based algorithm is reasonable as it has been observed that region-based segmentation methods seem to be more sensitive in segmenting an object with poor local contrast (preventing the bird's tail and claw bleeding through the tree branches) as compared to the distance-based one. The differences of performance are significant for medical images, where the background and the foreground have very similar intensity distribution, since one iteration of the Grabcut can change the segmentation dramatically. On the contrary, the rapid and continuous visual feedback provided by the provided algorithm prevents the developing of a large error, which reduces user's effort in actuating pixels.

3) Comparison of Algorithmic Predictability: Predictability of how the segmentation changes in response to mouse strokes is a criterion for practical ease of use. Quantitatively, the change of segmentation is measured by "reclassified" voxels, of which the assigned labels change between background and foreground. The *predictability* is reflected by looking at the dynamic response between user actuated voxels and reclassified voxels recorded over time as (#newly actuated voxels, #reclassified voxels).

Fig. 8 shows the dynamic response from the experiments described in the previous section. Each mark on the figure corresponds to one iteration when new user input was applied. Linear regression lines are overlaid on the data. All algorithms have a

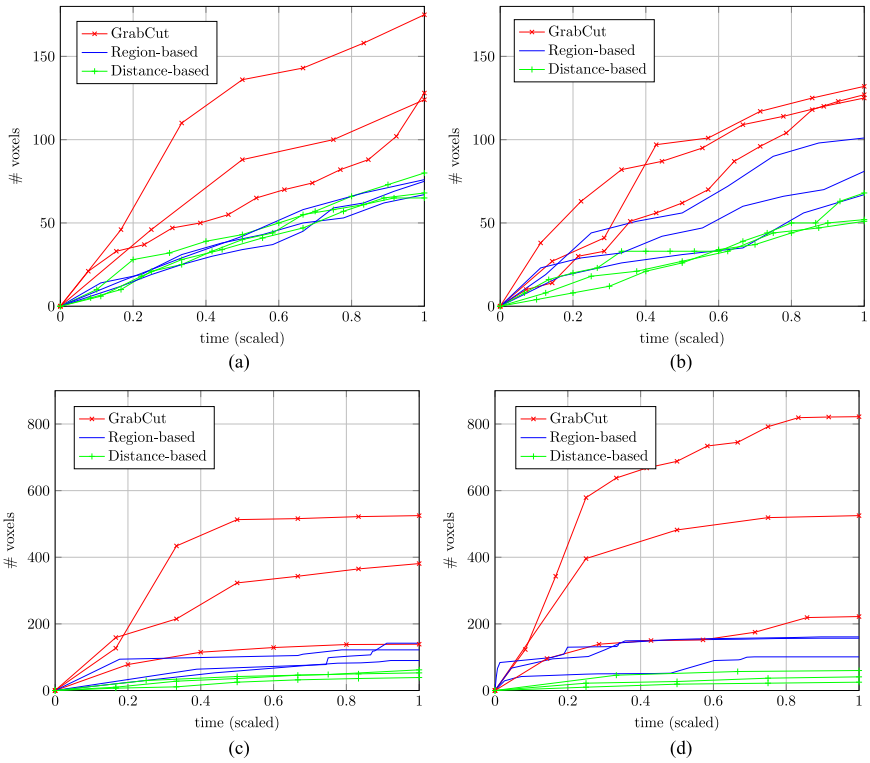


Fig. 7. Comparison of actuated voxels over time, after initialization for (a) surfing, (b) bird, (c) epiphysis, and (d) physis images. The proposed algorithm has a lower mean actuated pixels across repeated segmentations.

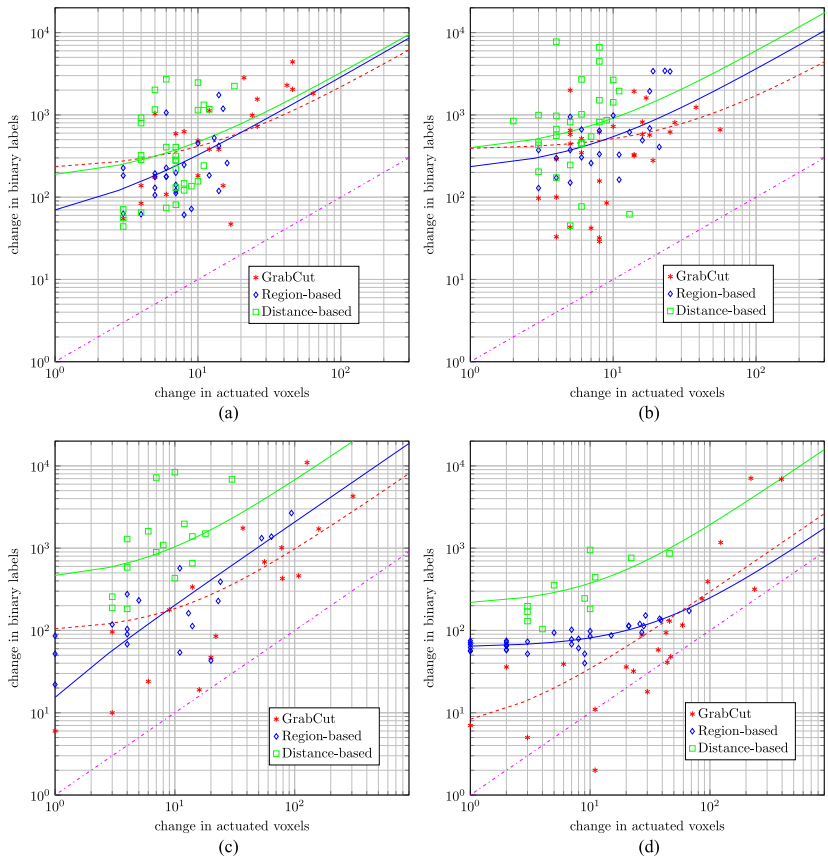


Fig. 8. Comparison of dynamic response to user input; data points and linear fit lines for (a) surfing, (b) bird, (c) epiphysis, and (d) physis images. Points below the dashed pink lines indicate wasted user effort, since more additional voxels were actuated than reclassified.

TABLE I
QUANTITATIVE COMPARISON: MANUAL VERSUS INTERACTIVE APPROACH FOR HEAD–NECK IMAGE SEGMENTATION

	Manual segmentation	Interactive segmentation	Dice		
			User 1	User 2	User 3
Left eye ball	3 min 35 s	2 min	0.85	0.88	0.84
Right eye ball	3 min 25 s	1 min 30 s	0.87	0.94	0.87
Brain stem	9 min 2 s	5 min 30 s	0.86	0.85	0.80
Mandible	29 min 37 s	10 min 15 s	0.81	0.90	0.86

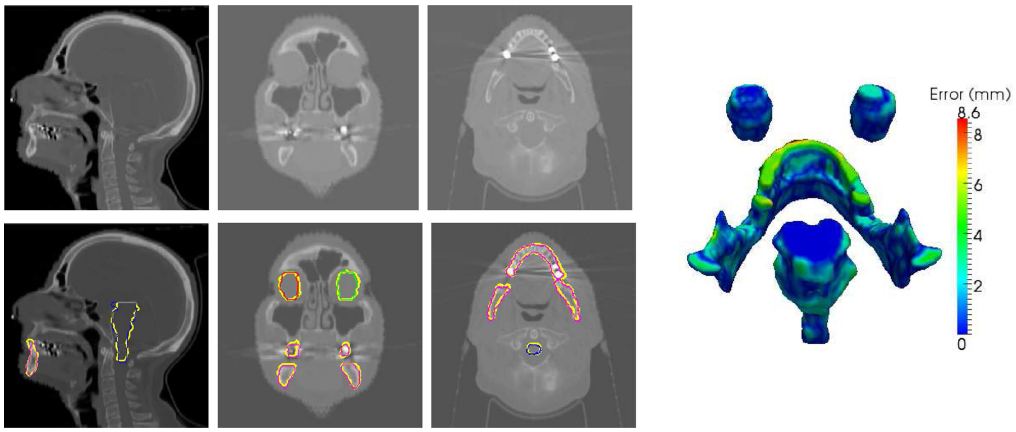


Fig. 9. Example of using the region-based method to segment the left eye (red), right eye (green), brain stem (blue), and mandible (pink), superimposed over manual segmentations (yellow), in axial, coronal, and sagittal views, respectively. The distribution of errors is shown on the right.

very similar dynamic response in the surfing image segmentation [see Fig. 8(a)], since the image has strong local contrast. Differences of predictability become observable in the bird segmentation, where the region-based approach has a tighter distribution along the fitting line because it is less sensitive to the poor-defined boundaries around the bird's tail. The advantages of using the control-based algorithm are shown in the medical image segmentation. Two issues become apparent for the physis segmentation. First, the distribution of GrabCut data points is quite broad; second, some of the GrabCut data points are below the dashed pink line, indicating a waste of user effort, since there are more voxels actuated than reclassified. The dynamic response of GrabCut makes it hard for a user to predict how much change new mouse strokes will cause.

These comparisons do not mean that the proposed algorithm will have better performance than other segmentation algorithms in all cases. On the other hand, we should emphasize that by forming a closed-loop interactive segmentation system, these classical algorithms may be robustified, and certain algorithmic disadvantages may be overcome relative to their open-loop implementations.

B. Application to Medical Image Segmentation

The proposed method was tested on real computed tomography (CT) volumes. In one experiment, four structures, left/right eye ball, brain stem, and mandible, involving in head-neck radiotherapy contouring were segmented from a CT image. The image size is $512 \times 512 \times 146$ voxels. Due to the high similarity of target structures to surrounding tissues and the required precision for the final segmentation result, the proposed region-based method was chosen in this test based on its robustness to these factors. To show the difference of intrauser performance, three users, who were blind to a reference manual segmentation, were involved in this test. The details of this experiment are summarized in Table I. The difference is obvious when examining the times required to perform the segmentations. The speedup is roughly $2\times$ according to Table I. It is also important to notice that the increase in speed becomes more noticeable for large structures that have intricate shapes. The users spent significantly more time outlining the mandible because of its complex boundaries. Additionally, less concentration from the user is required when guided by the interactive method, which further reduces the segmentation time. An example of segmentation is shown in Fig. 9. The largest error is at the mandible because of the scanning artifacts. However, the difference in segmentation accuracy is mainly due to the users understanding of these anatomical structures.

The proposed distance-based method was tested on cardiac chamber segmentation. Three commonly studied chambers in cardiac disease diagnosis, i.e., left/right ventricles and left atrium, were segmented from three cardiac CT images. The size of image slice along the axial direction is all 512×512, with the number of slices ranging from 243 to 292. The performance was summarized in Table II. The proposed method has over $9\times$ speedup on average, while it maintains close to 0.90 Dice coefficient. The segmentation of the right ventricle is harder than the other two due the lack of contrast along

TABLE II
QUANTITATIVE COMPARISON: MANUAL VERSUS INTERACTIVE APPROACH FOR CARDIAC IMAGE SEGMENTATION

	Manual segmentation	Interactive segmentation	Dice		
			Left ventricle	Right ventricle	Left atrium
Case 1	157 min 46 s	18 min 46 s	0.90	0.90	0.89
Case 2	131 min 26 s	13 min 45 s	0.90	0.87	0.92
Case 3	88 min 29 s	13 min 35 s	0.91	0.84	0.93

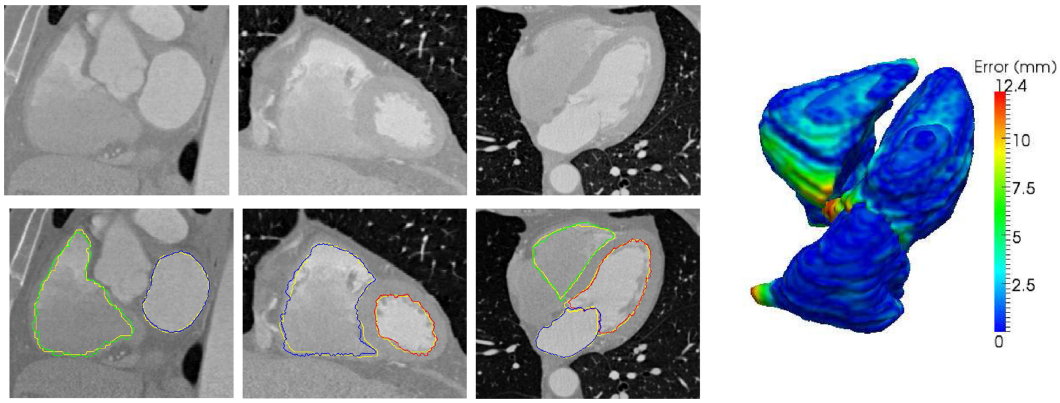


Fig. 10. Example of using the distance-based method to segment the left ventricle (red), right ventricle (green), and left atrium (blue), superimposed over manual segmentations (yellow), in axial, coronal, and sagittal views, respectively. The distribution of errors is shown on the right.

the endocardium and the invisibility of the valves between the right ventricle and atrium. As can be seen from one example segmentation in Fig. 10, large errors are mostly at where chamber boundaries become ambiguous.

Note that all the tests were conducted on desktops with standard CPUs. Indeed, the proposed method has real-time performance for real 3-D medical images, taking into account the factor that the total segmentation time is primarily how long the user takes to evaluate the current segmentation and apply more corrective input.

C. Relation to Existing Interactive Algorithms

The proposed control framework may be employed to close the loop around a number of key existing algorithms. As examples, the formulations in [9] and [10] can be written as (1). These two methods were applied to segment the epiphysis and physis images with different λs, each with three experiments. As shown in Fig. 11, many more user inputs are required if λ is small, while large λ can also increase user's input as it has a similar effect of *excessive input* (see [11]). Note that a large input was required in segmenting the physis using the method [10] because user input is only applied after the automatic segmentation is finished; thus, it cannot prevent large errors from occurring. While it is by no means a definitive comparison, the different characteristics of these algorithms were observed by looking at the minimum average efforts used in segmenting these structures (152, 196, and 120 pixels in epipysis and 169, 404, and 140 in physis, respectively, for [9], [10] and the proposed method).

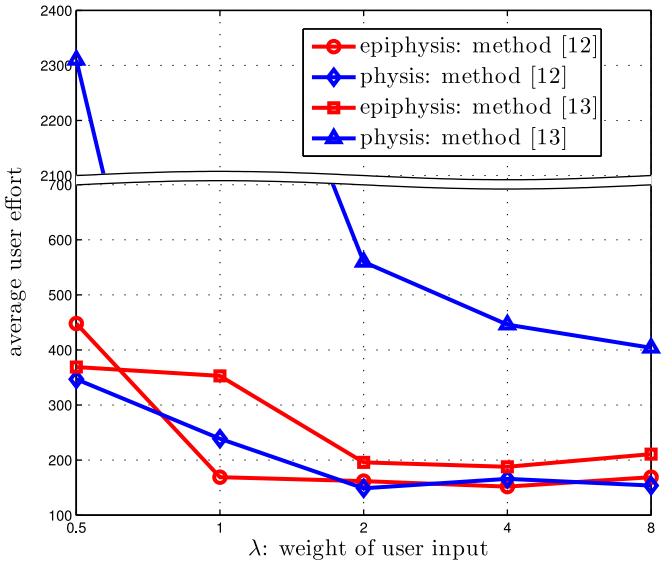


Fig. 11. Average user's effort of methods [12] and [13] with varying λ for the segmentation of epiphysis and physis.

Another feature of the control-based framework is the system's robustness to "noisy" user inputs (see Fig. 12). This is a property inherited from the feedback control design principle that allows admissible input variations [6].

For some distance-based interactive segmentation algorithms such as [28], the type of discrete input described in (34) was implicitly used to model the user interactions. There are other algorithms that may be formulated within the proposed framework. For instance, the algorithm presented in [4] may

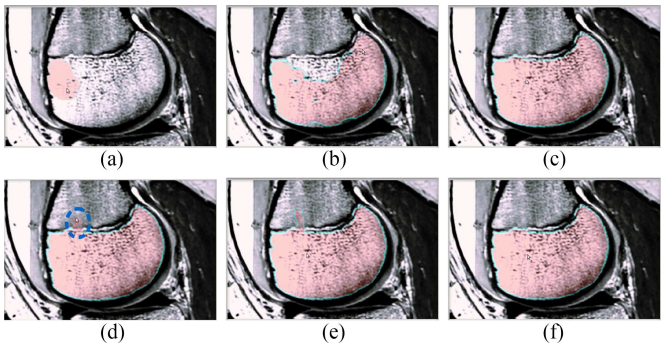


Fig. 12. Robustness of the control-based system with respect to "noisy" user input. (a)–(c) User input for the right segmentation. (d)–(f) System response to "noisy" user input (inside blue region).

be reformulated as a minimal path problem, thereby naturally fitting into the distance-based clustering case.

VI. Conclusion

This paper has presented a systematical way of applying control theory to analyze and design an interactive image segmentation algorithm. As an extension of [11], the new formulation has wider applications and stronger stability conditions. In particular, the proposed method supports both region- and distance-based metrics, handles multiple-object segmentation, and works for both scalar and vector images. The concept of impulsive control was adopted to model user interactions, which justifies the rationale of a widely used empirical strategy from the perspective of feedback control. In addition, conditions for asymptotic and exponential stability are derived, covering different levels of stability requirement.

The experimental results show the effectiveness of adding the proposed control structure to two representative classical methods. Since user's role is seamlessly integrated into a feedback control system, a large error can be prevented from developing and user's effort to corrections is guided by the stability condition. Though the examples used in this paper are based on the level-set formulation, the design principle is generalizable to other interactive segmentation systems that can be described by dynamical systems. It is extensible to discrete systems as well. The focus of this paper is on adopting control theory into image segmentation. There are other factors that determine the performance of an interactive segmentation system. One important topic is informative and efficient visualization to allow effective user interaction [62]. Simple user scribbles or predefined regions were used in the proposed example algorithms. It becomes vital to have more effective user interactions for large-scale 3-D image volumes.

References

[1] T. Heimann and H. Meinzer, "Statistical shape models for 3D medical image segmentation: A review," *Med. Image Anal.*, vol. 13, no. 4, pp. 543–563, 2009.

[2] P. A. Yushkevich *et al.*, "User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability," *Neuroimage*, vol. 31, no. 3, pp. 1116–1128, 2006.

[3] C. Rother, V. Kolmogorov, and A. Blake, ""grabcut": Interactive foreground extraction using iterated graph cuts," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 309–314, Aug. 2004.

[4] V. Vezhnevets and V. Konouchine, ""GrowCut"—Interactive multi-label N-D image segmentation by cellular automata," in *Proc. Int. Conf. Comput. Graph. Vision*, 2005, pp. 150–156.

[5] L. Grady, V. Sun, and J. Williams, "Interactive graph-based segmentation methods in cardiovascular imaging," in *Handbook of Mathematical Models in Computer Vision*, N. Paragios, Y. Chen, and O. Faugeras, Eds. New York, NY, USA: Springer, 2006, pp. 453–469.

[6] J. C. Doyle, B. A. Francis, and A. Tannenbaum, *Feedback Control Theory*, vol. 1. New York, NY, USA: Macmillan, 1992.

[7] B. Bhanu, S. Lee, and J. Ming, "Adaptive image segmentation using a genetic algorithm," *IEEE Trans. Syst., Man, Cybern.*, vol. 25, no. 12, pp. 1543–1567, Dec. 1995.

[8] S. M. Grigorescu, D. Ristic-Durrant, S. K. Vuppala, and A. Gräser, "Closed-loop control in image processing for improvement of object recognition," *IFAC Proc. Vol.*, vol. 41, pp. 5335–5340, 2008.

[9] D. Cremers, O. Fluck, M. Rousson, and S. Aharon, "A probabilistic level set formulation for interactive organ segmentation," *Proc. SPIE*, vol. 6512, Feb. 2007, Art. no. 65120V.

[10] N. Ben-Zadok, T. Riklin-Raviv, and N. Kiryati, "Interactive level set segmentation for image-guided therapy," in *Proc. IEEE Int. Symp. Biomed. Imag.: Nano Macro*, 2009, pp. 1079–1082.

[11] P. Karasev, I. Kolesov, K. D. Fritscher, P. A. Vela, P. Mitchell, and A. Tannenbaum, "Interactive medical image segmentation using PDE control of active contours," *IEEE Trans. Med. Imag.*, vol. 32, no. 11, pp. 2127–2139, Nov. 2013.

[12] D. Mumford and J. Shah, "Optimal approximations by piecewise smooth functions and associated variational problems," *Commun. Pure Appl. Math.*, vol. 42, is. 5, pp. 577–685, Jul. 1989.

[13] S. Kichenassamy, A. Kumar, P. Olver, A. Tannenbaum, and A. Yezzi, "Gradient flows and geometric active contour models," in *Proc. IEEE Int. Conf. Comput. Vision*, 1995, pp. 810–815.

[14] A. Yezzi Jr., A. Tsai, and A. Willsky, "A fully global approach to image segmentation via coupled curve evolution equations," *J. Vis. Commun. Image Represent.*, vol. 13, nos. 1/2, pp. 195–216, 2002.

[15] S. Lankton and A. Tannenbaum, "Localizing region-based active contours," *IEEE Trans. Image Process.*, vol. 17, no. 11, pp. 2029–2039, Nov. 2008.

[16] H. K. Zhao, T. Chan, B. Merriman, and S. Osher, "A variational level set approach to multiphase motion," *J. Comput. Phys.*, vol. 127, pp. 179–195, 1996.

[17] C. Vázquez, A. Mitiche, and R. Laganière, "Joint multiregion segmentation and parametric estimation of image motion by basis function representation and level set evolution," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 5, pp. 782–793, May 2006.

[18] Y. Gao, R. Kikinis, S. Bouix, M. E. Shenton, and A. Tannenbaum, "A 3D interactive multi-object segmentation tool using local robust statistics driven active contours," *Med. Image Anal.*, vol. 16, no. 6, pp. 1216–1227, 2012.

[19] L. D. Cohen and R. Kimmel, "Global minimum for active contour models: A minimal path approach," *Int. J. Comput. Vision*, vol. 24, no. 1, pp. 57–78, Aug. 1997.

[20] C. Vázquez, A. Mitiche, and I. Ben Ayed, "Image segmentation as regularized clustering: A fully global curve evolution method," in *Proc. Int. Conf. Image Process.*, 2004, pp. 3467–3470.

[21] G. Sapiro, *Geometric Partial Differential Equations and Image Analysis*. New York, NY, USA: Cambridge Univ. Press, 2006.

[22] D. Cremers, M. Rousson, and R. Deriche, "A review of statistical approaches to level set segmentation: Integrating color, texture, motion and shape," *Int. J. Comput. Vision*, vol. 72, pp. 195–215, 2007.

[23] B. Peng, L. Zhang, and D. Zhang, "A survey of graph theoretical approaches to image segmentation," *Pattern Recognit.*, vol. 46, no. 3, pp. 1020–1038, 2013.

[24] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: A review," *ACM Comput. Surveys*, vol. 31, no. 3, pp. 264–323, Sep. 1999.

[25] F. Zhao and X. Xie, "An overview of interactive medical image segmentation," *Ann. BMVA*, vol. 2013, no. 7, pp. 1–22, 2013.

[26] C. J. Armstrong, L. P. Brian, and W. A. Barrett, "Interactive segmentation of image volumes with live surface," *Comput. Graph.*, vol. 31, is. 2, pp. 212–229, 2007.

[27] B. L. Price, B. Morse, and S. Cohen, "Geodesic graph cut for interactive image segmentation," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, Jun. 2010, pp. 3161–3168.

[28] X. Bai and G. Sapiro, "Geodesic matting: A framework for fast interactive image and video segmentation and matting," *Int. J. Comput. Vision*, vol. 82, no. 2, pp. 113–132, 2009.

[29] A. Criminisi, T. Sharp, and A. Blake, "GeoS: Geodesic image segmentation," in *Proc. Eur. Conf. Comput. Vision*, 2008, pp. 99–112.

[30] T. Nguyen, J. Cai, J. Zhang, and J. Zheng, "Robust interactive image segmentation using convex active contours," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3734–3743, Aug. 2012.
[31] C. Nieuwenhuis and D. Cremers, "Spatially varying color distributions for interactive multilabel segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 5, pp. 1234–1247, May 2013.
[32] B. Mory, R. Ardon, A. J. Yezzi, and J. Thiran, "Non-Euclidean image-adaptive radial basis functions for 3D interactive segmentation," in *Proc. IEEE 12th Int. Conf. Comput. Vision*, 2009, pp. 787–794.
[33] V. Gulshan, C. Rother, A. Criminisi, A. Blake, and A. Zisserman, "Geodesic star convexity for interactive image segmentation," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, Jun. 2010, pp. 3129–3136.
[34] W. Yang, J. Cai, J. Zheng, and J. Luo, "User-friendly interactive image segmentation through unified combinatorial user inputs," *IEEE Trans. Imag. Process.*, vol. 19, is. 9, pp. 2470–2479, Sep. 2010.
[35] D. Wang, C. Yan, S. Shan, and X. Chen, "Active learning for interactive segmentation with expected confidence change," in *Computer Vision*, vol. 7724, K. Lee, Y. Matsushita, J. Rehg, and Z. Hu, Eds. Berlin, Germany: Springer, 2013, pp. 790–802.
[36] S. H. Park, I. D. Yun, and S. U. Lee, "Data-driven interactive 3D medical image segmentation based on structured patch model," *Inf. Process. Med. Imag.*, vol. 23, pp. 196–207, 2013.
[37] F. Heckel, O. Konrad, H. K. Hahn, and H. O. Peitgen, "Interactive 3D medical image segmentation with energy-minimizing implicit functions," *Comput. Graph.*, vol. 35, pp. 275–287, 2011.
[38] N. Ben-Zadok, T. R. Raviv, and N. Kiryati, "Interactive level set segmentation for image-guided therapy," in *Proc. IEEE Int. Symp. Biomed. Imag.: From Nano Macro*, 2009, pp. 1079–1082.
[39] P. V. Kokotović, "Joy of feedback: Nonlinear and adaptive, (1991 bode prize lecture)," *Control Syst. Mag.*, vol. 12, pp. 7–17, Jun. 1992.
[40] H. K. Khalil and J. W. Grizzle, *Nonlinear Systems*, vol. 3. Upper Saddle River, NJ, USA: Prentice-Hall, 2002.
[41] F. Clarke, "Lyapunov functions and discontinuous stabilizing feedback," *Annu. Rev. Control*, vol. 35, no. 1, pp. 13–33, 2011.
[42] M. Krstic, "On global stabilization of Burgers' equation by boundary control," *Syst. Control Lett.*, vol. 37, pp. 3123–141, 1999.
[43] F. Di Meglio, R. Vazquez, and M. Krstic, "Stabilization of a system of n+1 coupled first-order hyperbolic linear PDEs with a single boundary input," *IEEE Trans. Automat. Control*, vol. 58, no. 12, pp. 3097–3111, Dec. 2013.
[44] M. Jankovic, "Extension of control Lyapunov functions to time-delay systems," in *Proc. 39th IEEE Decision Control*, 2000, vol. 5, pp. 4403–4408.
[45] M. C. de Oliveira, J. Bernussou, and J. C. Geromel, "A new discrete-time robust stability condition," *Syst. Control Lett.*, vol. 37, no. 4, pp. 261–265, 1999.
[46] A. Khadra, X. Liu, and X. Shen, "Impulsive control and synchronization of spatiotemporal chaos," *Chaos, Solitons Fractals*, vol. 26, is. 2, pp. 615–636, 2005.
[47] T. Yang, *Impulsive Control Theory* (Lecture Notes in Control and Information Sciences), vol. 272. Berlin, Germany: Springer-Verlag, 2001.
[48] P. J. Antsaklis and X. D Koutsoukos, "Hybrid systems: Review and recent progress," in *Software-Enabled Control: Information Technology for Dynamical Systems*, T. Samad and G. Balas, Eds. New York, NY, USA: Wiley, 2003.
[49] S. Grossberg, "Adaptive resonance theory: How a brain learns to consciously attend, learn, and recognize a changing world," *Neural Netw.*, vol. 37, pp. 1–47, 2013.
[50] S. C. Zhu and A. Yuille, "Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 9, pp. 884–900, Sep. 1996.
[51] J. A. Sethian, *Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science on Applied and Computational Mathematics*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, Jun. 1999.
[52] S. Osher and R. P. Fedkiw, *Level Set Methods and Dynamic Implicit Surfaces*, vol. 153. Berlin, Germany: Springer-Verlag, 2003.
[53] A. Dervieux and F. Thomasset, "A finite element method for the simulation of a Rayleigh-Taylor instability," in *Approximation Methods for Navier-Stokes Problems*, R. Rautmann, Ed. Berlin, Germany: Springer, 1980, pp. 145–158.
[54] Rafael C. Gonzalez and Richard E. Woods, *Digital Image Processing*, 3rd ed. Englewood Cliffs, NJ, USA: Prentice-Hall, 2007.
[55] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Trans. Image Process.*, vol. 10, no. 2, pp. 266–277, Feb. 2001.
[56] O. V. Michailovich, Y. Rathi, and A. Tannenbaum, "Image segmentation using active contours driven by the Bhattacharyya gradient flow," *IEEE Trans. Image Process.*, vol. 16, no. 11, pp. 2787–2801, Nov. 2007.
[57] S. T. Whitaker, "A level-set approach to 3D reconstruction from range data," *Int. J. Comput. Vision.*, vol. 29, pp. 203–231, 1998.
[58] J. N. Tsitsiklis, "Efficient algorithms for globally optimal trajectories," *IEEE Trans. Automat. Control*, vol. 40, no. 9, pp. 1528–1538, Sep. 1995.
[59] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numer. Math.*, vol. 1, no. 1, pp. 269–271, 1959.
[60] E. Sifakis and G. Tziritas, "Fast marching techniques for visual grouping," in *Geometric Level Set Methods in Imaging, Vision and Graphics*, S. Osher and N. Paragios, Eds. Berlin, Germany: Springer-Verlag, 2003.
[61] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. 8th IEEE Int. Conf. Comput. Vision*, Jul. 2001, vol. 2, pp. 416–423.
[62] H. Peng, Z. Ruan, F. Long, J. H Simpson, and E. W. Myers, "V3D enables real-time 3D visualization and quantitative analysis of large-scale biological image data sets," *Nature Biotechnol.*, vol. 28, no. 4, pp. 348–353, 2010.

Liangjia Zhu received the Ph.D. degree in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2013.

He is a Postdoctoral Research Fellow with the Department of Computer Science, Stony Brook University, Stony Brook, NY, USA. His research interests include computer vision, medical image analysis, and systems and control.

Peter Karasev received the B.S., M.S., and Ph.D. degrees in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2008, 2010, and 2013, respectively.

He is a Research Engineer with Agilent Technologies Inc., Santa Clara, CA, USA. His current research interests include the application of control-theoretic methods to medical image processing and computer vision.

Ivan Kolesov received the Ph.D. degree in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2013.

He is a Postdoctoral Research Fellow with Stony Brook University, Stony Brook, NY, USA. His research interests include computer vision, medical imaging, and systems and control.

Romeil Sandhu (M'08) received the Ph.D. degree in electrical engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2010.

He is currently with Stony Brook University, Stony Brook, NY, USA. His current research interests include imaging and bioinformatics with a particular emphasis on differential geometry, machine learning, and control theory.

Allen Tannenbaum (F'09) received the Ph.D. degree in mathematics from the Harvard University, in 1976. He is a Distinguished Professor with the Department of Computer Science and Applied Mathematics/Statistics, Stony Brook University, Stony Brook, NY, USA. His research interests include systems and control, signal processing, computer vision, and systems biology.