# Generalized Simultaneous Registration and Segmentation

Pratim Ghosh, Emre Sargin and B.S. Manjunath
Department of Electrical and Computer Engineering
University of California, Santa Barbara, CA 93106-9560
{pratim,msargin,manj}@ece.ucsb.edu

## Abstract

*Simultaneous registration and segmentation (SRS) provides a powerful framework for tracking an object of interest in an image sequence. The state-of-the-art SRS-based tracking methods assume that the illumination is maintained constant across consecutive frames. However, this assumption does not hold in many natural image sequences due to dynamic light source and shadows. We propose a generalized model for SRS-based tracking in this paper to account for non-uniform additive illumination changes. More specifically, we introduce two new terms in the SRS energy functional which address the above mentioned problem. The first term couples the shape-based cue and intensity-based cue to establish a correspondence between them. The second term compensates for the illumination change which is complementary to the first term. We demonstrate that the proposed SRS energy functional yields superior performance over the state-of-the-art SRS-based methods for various indoor and outdoor image sequences.*

## 1. Introduction

We consider the problem of object segmentation and tracking in time sequence imagery. For this purpose, a generalization of the existing simultaneous registration and segmentation technique is derived and an effective variational optimization framework is introduced. The proposed generalization makes the tracking robust to varying appearance (intensity) as well as non-rigid deformations of the object shape.

Assume that the image sequence $I(\mathbf{x}, t)|_{t=1,2,\ldots,T}$ consists of an object of interest undergoing motion and/or other deformations. We also consider the variation of the object appearance over time possibly due to additive non-uniform illumination changes, inter-reflections and scattering which are frequently encountered in a practical object tracking scenario. In previous approaches, the possible variations in the object appearance and shape are statistically modeled based on the training examples, *e.g*., deformable shape model [14], active shape model (ASM) [8], active appearance model (AAM) [7], the multi-scale combination [15] of ASM and AAM to increase the robustness against noise and clutter. Similarly, Tsai *et al*. [22] developed a parametric level-set method for shape based image segmentation. Zhu *et al*. [27] proposed a subject specific dynamic model for medical application based on multi-linear principal component analysis (MPCA). MPCA is used to learn the different modes of variation in the object of interest. However, the performance of these methods suffers from abrupt changes in the object appearance which is not covered in the training set. Recently, Schoenemann and Cremers [20] proposed a combinatorial solution to track deformable objects. The algorithm is based on finding a minimum cost cyclic path in the product space spanned by the template shape and the given image. The cost of a cyclic path is computed from the image data as well as from the template shape. A variant of this combinatorial solution was also proposed in [19]. This method uses edge information and hence is robust to illumination changes - the flipside being that it is computationally very expensive. There also exist polynomial time graph algorithms [12] for extracting similar objects from multiple images (time-sequence) simultaneously which do not consider any shape information.

Another set of approaches [26, 17, 24, 16] have been developed recently which explicitly take into account the registration problem between consecutive time instances for segmenting the object of interest over time. It has been recently demonstrated [1, 23, 18] that the performance can be improved significantly by simultaneously solving the segmentation and registration problems. Ehrhardt *et al*. [10] proposed such simultaneous registration and segmentation (SRS) in the variational framework. More recently Ghosh *et al*. [11] (SRS+DP) demonstrated that the performance can be further improved by a dynamical prior (DP) term. Though the state-of-the-art SRS approaches are indeed promising in terms of segmentation and tracking, they do not explicitly account for frequently occurring illumination changes across consecutive time frames. Our main

contribution in this paper is to propose a framework that accounts for such illumination variations.

The key components of the proposed approach are:

- A generalized model is introduced for the level-set based simultaneous registration and segmentation framework which is robust to non-uniform additive illumination changes over time.
- The model is derived in a principled way by formulating the SRS problem in a maximum *a posteriori* (MAP) framework.
- The proposed model tightly couples multiple cues for establishing correspondence and also compensates for the illumination changes over consecutive views.

The paper is organized as follows. Section 2 briefly summarizes the main components of a recently proposed simultaneous registration and segmentation algorithm (SRS+DP) [11]. In Section 3 we introduce a maximum *a posteriori* formulation (MAP) which generalizes the concept of SRS+DP. A new energy functional is derived from the MAP formulation in Section 4. In Section 5 we present some qualitative and quantitative evaluation of the proposed approach. We conclude in Section 6.

## 2. Review of the Previous approach

Let $I(t) : \Omega \rightarrow \mathbb{R}$ and $I(t-1) : \Omega \rightarrow \mathbb{R}$ denote the images at two consecutive time instances $t$ and $t-1$. The shape of the tracked object of interest at any arbitrary time $t$ is embedded in the level-set function $\phi^o(\mathbf{x}, t) : \Omega \rightarrow \mathbb{R}$[1]. We assume that any kind of relation/correspodence between consecutive (say, $t$ and $t-1$) frames can be defined using a displacement vector field $\mathbf{u}(\mathbf{x}, t-1, t)$. In this we also assume that the initial contour of the object of interest is known *a priori*. Below we describe the driving equations for the segmentation and registration modules in state-of-the-art SRS framework. Interested readers are referred to [9, 13, 11, 10] for detailed explanation.

### 2.1. Segmentation module

The segmentation for the current frame (at time $t$) can be computed by maximizing the following *a posteriori* probability:

$$\underset{\phi^o(t)}{\operatorname{argmax}}\, \mathcal{P}(\phi^o(t)|I(t), \hat{\phi}^-(t)) \tag{1}$$

where $\hat{\phi}^-(t)$ is the dynamical prior at time $t$. In [11] the authors proposed an probabilistic formulation to compute $\hat{\phi}^-(t)$ which was shown to be consistent with the temporal statistics of the tracked object. To this end $\hat{\phi}^-(t)$ is designed in such a way so as to maximize the *a posteriori* probability given all the past observations $\phi^o(1), \phi^o(2)..., \phi^o(t-1)$ for the segmentation of the current frame. This problem of computing $\hat{\phi}^-(t)$ was solved by formulating a linear stochastic equation:

$$\phi(\mathbf{x}, t) = \phi(\mathbf{x}, t\text{–}1) - \mathbf{u}^T(\mathbf{x}, t\text{–}2, t\text{–}1)\nabla\phi(\mathbf{x}, t\text{–}1) + w \tag{2}$$

and a observation model:

$$\phi^o(\mathbf{x}, t) = \phi(\mathbf{x}, t) + v \tag{3}$$

where $w$ (with pdf $\mathcal{P}(w) \sim \mathcal{N}(0, Q)$) and $v$ (with pdf $\mathcal{P}(v) \sim \mathcal{N}(0, R)$) represent the modeling and the observation errors, respectively. Using Eq. (2) and (3) the dynamical prior can be computed as:

$$\hat{\phi}^-(\mathbf{x}, t) = \hat{\phi}(\mathbf{x}, t\text{–}1) - \mathbf{u}^T(\mathbf{x}, t\text{–}2, t\text{–}1)\nabla\hat{\phi}(\mathbf{x}, t\text{–}1). \tag{4}$$

Assuming $I(t)$ and $\hat{\phi}^-(t)$ are conditionally independent given $\phi(t)$ and are also mutually independent we can write:

$$\mathcal{P}(\phi^o(t)|I(t), \hat{\phi}^-(t)) \propto \mathcal{P}(I(t)|\phi^o(t))\mathcal{P}(\hat{\phi}^-(t)|\phi^o(t))\mathcal{P}(\phi^o(t)) \tag{5}$$

It can be shown that maximizing the expression in Eq. (5) is equivalent to minimizing the following energy function under certain simplifying assumptions:

$$\begin{aligned} E(\phi^o(t); I(t), \hat{\phi}^-(t)) = & \\ & \int_\Omega \ln p(I|\theta_2) + H_\epsilon(\phi^o(\mathbf{x}, t)) \ln \frac{p(I|\theta_2)}{p(I|\theta_1)} d\mathbf{x} \\ & + \beta \frac{1}{2} \int_\Omega |\phi^o(\mathbf{x}, t) - \hat{\phi}^-(\mathbf{x}, t)|^2 d\mathbf{x} \\ & + \nu \int_\Omega |\nabla H_\epsilon(\phi^o(\mathbf{x}, t))| d\mathbf{x} \end{aligned} \tag{6}$$

where $\nu$ and $\beta$ are the positive constants, $\theta_1$, $\theta_2$ parameterize the object and the background pdfs, and $H_\epsilon(z) = \frac{1}{2}[1 + \frac{2}{\pi}\arctan(\frac{z}{\epsilon})]$ is the regularized Heaviside function. The term associated with $\nu$ is similar to the term used in [4] which penalize the length of the curve represented by the zero level-set of $\phi^o(t)$.

### 2.2. Registration module

The displacement vector field $\mathbf{u}$ is used to establish the correspondence between two consecutive frames using multiple cues, for example, image intensity and shape based cues.

**Intensity based registration:** The displacement vector field $\mathbf{u}(t-1, t)$ relating intensity cue for establishing the correspondence can be computed by maximizing the *a posteriori* probability $\mathcal{P}(\mathbf{u}|I(t), I(t-1))$. Applying the Bayes rule we can write:

$$\mathcal{P}(\mathbf{u}|I(t), I(t-1)) \propto \mathcal{P}(I(t), I(t-1)|\mathbf{u})\mathcal{P}(\mathbf{u}) \tag{7}$$

[1] For the rest of the discussion we assume that $\phi(\mathbf{x}, t) \equiv \phi(t)$ unless mentioned otherwise.

With certain simplifying assumptions the maximization problem in Eq. (7) can be reduced to minimization of the following energy functional:

$$E(\mathbf{u}; I(t-1), I(t)) = \frac{1}{2}\int_{\Omega}(I(\mathbf{x},t) - I(\mathbf{T}(\mathbf{x}),t-1))^2 d\mathbf{x} + \alpha\frac{1}{2}\int_{\Omega}\text{trace}(\nabla\mathbf{u}\,\nabla\mathbf{u}^T)d\mathbf{x} \quad (8)$$

where $(.)^T$ represents the transpose operation, $\mathbf{T}(\mathbf{x}) = \mathbf{x} - \mathbf{u}(\mathbf{x})$, and $\alpha$ controls the smoothness of the derived vector field.

**Shape based registration:** Similarly one can write down the expression $\mathbf{u}^*$ for the vector field $\mathbf{u}$ which relates the shape based cue (represented as level-set functions) from two consecutive frames:

$$\mathbf{u}^* = \underset{\mathbf{u}}{\text{argmax}}\,\mathcal{P}(\mathbf{u}|\hat{\phi}(t-1), \phi^o(t)) \quad (9)$$

where $\hat{\phi}(t-1)$ is the final estimated shape at $t-1$. An equivalent minimization problem to Eq. (9) can be written [10] as:

$$E(\mathbf{u}; \hat{\phi}(t-1), \phi^o(t)) = \frac{1}{2}\int_{\Omega} N_\epsilon(\phi^o(t), \hat{\phi}(t-1))(\phi^o(\mathbf{x},t) - \hat{\phi}(\mathbf{T}(\mathbf{x}),t-1))^2 d\mathbf{x} \quad (10)$$

where $N_\epsilon$ is a binary function which confines the optimization around the $\epsilon$ neighborhood of the shapes (i.e. $\phi^o(\mathbf{x},t), \hat{\phi}(\mathbf{x},t-1)$) under consideration. This cost function penalizes the deviation of the target shape ($\phi^o(\mathbf{x},t)$) from the transformed reference shape ($\hat{\phi}(\mathbf{x},t-1)$). The formula in Eq. (10) also combines the concept of registration and segmentation into a single objective term.

We would like to note that the displacement vector field $\mathbf{u}(\mathbf{x}, t-1, t)$ independently relates the shape and image intensity based cues for establishing correspondence. The intensity based correspondence in the registration module is effective only if the corresponding pixels have equal gray values. Unfortunately, this assumption is commonly violated, *e.g.*, when brightness constancy assumption ($I(\mathbf{x},t) = I(\mathbf{x}+\Delta\mathbf{x}, t+\Delta t)$) does not hold, and when constant/linear illumination change across consecutive views. The authors in [11] proposed a solution with the use of a dynamical prior term to address the problem of correspondence when brightness constancy assumption is violated. However, the framework of SRS+DP [11] still lacks a model which can compensate for the constant/linear illumination change or other intensity based disturbances. We can partially address the correspondence problem in presence illumination change across consecutive views by a functional as used in [3, 2] :

$$E(\mathbf{u}; I(t-1), I(t)) = \frac{1}{2}\int_{\Omega}\Big(|\nabla I(\mathbf{x},t) + - \nabla I(\mathbf{T}(\mathbf{x}),t-1)|\Big)^2 d\mathbf{x} \quad (11)$$

where $\nabla = (\partial x, \partial y)^T$ denotes the spatial gradients. The functional in Eq. (11) (obtained from gradient constancy [3] assumption) allows some small variations in the grey value. Nevertheless, this functional is clearly not sufficient when the illumination changes non-uniformly across consecutive views. To demonstrate this we perform a simple test. Consider Fig. 1. The background of a circular object (top left corner) is modified by adding a spatially varying function as shown in the top right corner. The object intensity is chosen in such a way that it does not saturate across time. Thus we also allow some variations in the object appearance across time. The last two rows compares the performance of the proposed approach with the state-of-the-art technique. We next develop the proposed approach systematically from the concept of simultaneous registration and segmentation which is robust to non-uniform illumination change.

## 3. Generalized Simultaneous Registration and Segmentation

In this section we first introduce a maximum *a posteriori* (MAP) framework that generalizes the concept of simultaneous registration and segmentation. Secondly, we show that the SRS+DP framework can be obtained from this MAP formulation with certain simplifying assumptions. Thirdly, we point out the limitations of such assumptions, especially when there is drastic illumination change across consecutive views. Finally, a novel approximation is introduced which is demonstrated to be quite effective for this purpose.

Consider the following maximum *a posteriori* problem:

$$\underset{\mathbf{u},\phi^o(t)}{\text{argmax}}\,\mathcal{P}(\mathbf{u}, \phi^o(t)|I(t), I(t-1), \hat{\phi}(t-1), \hat{\phi}^-(t)). \quad (12)$$

Decomposing the above probability expression we obtain:

$$\begin{aligned}\mathcal{P}\big(\mathbf{u}, \phi^o(t)\big|I(t), I(t-1), \hat{\phi}(t-1), \hat{\phi}^-(t)\big) &= \mathcal{P}\big(\mathbf{u}\big|\phi^o(t), I(t), I(t-1), \hat{\phi}(t-1), \hat{\phi}^-(t)\big)\\ &\quad \mathcal{P}\big(\phi^o(t)\big|I(t), I(t-1), \phi(t-1), \hat{\phi}^-(t)\big)\\ &= \underbrace{\mathcal{P}\big(\mathbf{u}\big|\phi^o(t), I(t), I(t-1), \hat{\phi}(t-1)\big)}_{\mathcal{C}_1}\\ &\quad \underbrace{\mathcal{P}\big(\phi^o(t)\big|I(t), \hat{\phi}^-(t)\big)}_{\mathcal{C}_2}.\end{aligned} \quad (13)$$

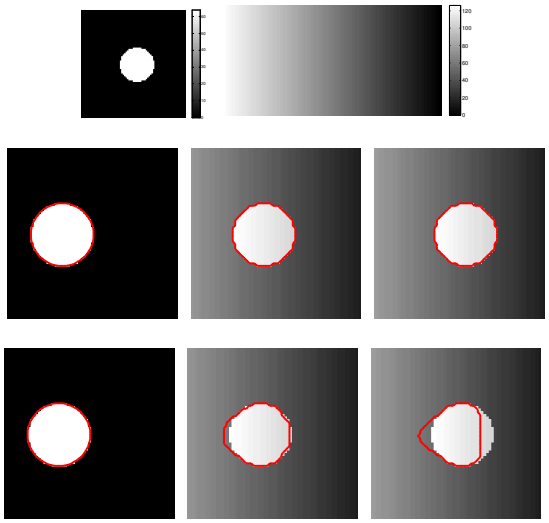

Figure 1. Example of circular object-of-interest (top-left), and spatially varying background (top-right). The intensity of the object is 64 and the background is varying linearly from 128 to 0 (from left to right). A synthetic sequence consists of three images (last two rows). The second row presents the segmentation/tracking result (red contour) for the proposed approach and the third row shows the same for the state-of-the-art technique [11]. The state-of-the-art work can not segment/track the object reliably due to non-uniform illumination change across consecutive time frames. The contour is specified for the first frame (*i.e.* the first column in second and third rows) surrounding the circular object for both the methods. In these images the background is actually used as shown in the top right corner. However, we focus in the vicinity of the object for better visualization.

To obtain $\mathcal{C}_1$ we assume that $\mathbf{u}$ is independent of $\hat{\phi}^-(t)$ given $\phi^o(t)$. We also assume that $\phi^o(t)$ is independent of the previous image $I(t-1)$ and is independent of $\hat{\phi}(t-1)$ given the dynamical prior $\hat{\phi}^-(t)$ to get $\mathcal{C}_2$. The dependency graph between different variables is depicted in Fig. 2 for better understanding.

Note that, $\mathcal{C}_2$ is associated with the energy functional $E(\phi^o(t); I(t), \hat{\phi}^-(t))$ (Eq. (6)). However, maximization of $\mathcal{C}_1$ via a single energy functional is intractable due to multiple conditional dependencies. Accordingly, based on certain assumptions, one can approximate $\mathcal{C}_1$ with the product of probabilities conditioned on fewer variables. If we assume the intensity based cue i.e., $I(t)$ and $I(t-1)$ are conditionally independent of the shape based cue i.e., $\phi^o(t)$ and $\hat{\phi}(t-1)$, given the flow field $\mathbf{u}$, and are also mutually independent, then we can approximate the $\mathcal{C}_1$ term as:

$$\mathcal{C}_1 \approx \mathcal{P}\big(\mathbf{u}\big|I(t), I(t-1)\big)\mathcal{P}\big(\mathbf{u}\big|\phi^o(t), \hat{\phi}(t-1)\big) \qquad (14)$$

Note that the above approximation corresponds to the energy functionals $E(\mathbf{u}; I(t-1), I(t))$ (from Eq. (8)) and $E(\mathbf{u}; \hat{\phi}(t-1), \phi^o(t))$ (from Eq. (10)) in the SRS+DP framework. In this regard, we would like to mention:

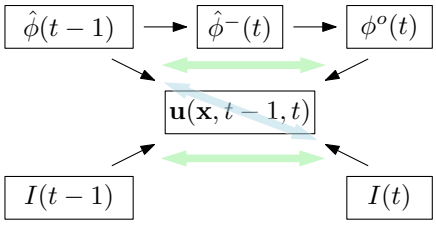


Figure 2. Dependency graph model in a typical SRS framework. The light green double arrows correspond to the terms which are considered in SRS+DP. The proposed approach makes use of the other cross-term which is shown by the light blue double arrow.

- Image intensity based correspondence (*i.e.*, $\mathcal{P}(\mathbf{u}|I(t), I(t-1)))$ may be highly corrupted due to non uniform illumination change.
- Although, the second term (*i.e.*, $\mathcal{P}(\mathbf{u}\big|\phi^o(t), \hat{\phi}(t-1))$ is independent of the illumination, it may introduce instability since $\mathbf{u}$ is conditioned on another optimization variable $\phi^o(t)$.

Thus, the approximation adopted in SRS+DP may not be relevant in certain cases. Alternatively, one can obtain a new approximation:

$$\mathcal{C}_1 \approx \mathcal{P}\big(\mathbf{u}\big|I(t), \hat{\phi}(t-1)\big)\mathcal{P}\big(\mathbf{u}\big|\phi^o(t), I(t-1)\big), \qquad (15)$$

if the pair $I(t)$ and $\hat{\phi}(t-1)$ is assumed to be conditionally independent of the pair $I(t-1)$ and $\phi^o(t)$ given $\mathbf{u}$ and they are also assumed to be mutually independent. The effectiveness of this cross connection between intensity and shape based cues for establishing correspondence is elaborated in Section 4.

In a general set up, one can sequentially optimize for $\mathbf{u}$ by alternating the approximations for $\mathcal{C}_1$ based on the two assumptions (Eq. (14) and Eq. (15)). However, for computational reasons, we opt for the batch update which combines the two approximations into a single objective function. Furthermore, we drop the $\mathcal{P}\big(\mathbf{u}\big|\phi^o(t), I(t-1)\big)$ term which may introduce instability since $\mathbf{u}$ is conditioned on another optimization variable $\phi^o(t)$. The final approximation can be summarized as:

$$\mathcal{C}_1 \approx \overbrace{\mathcal{P}\big(\mathbf{u}\big|I(t), I(t-1)\big)}^{\mathcal{C}_1^1} \underbrace{\mathcal{P}\big(\mathbf{u}\big|\phi^o(t), \hat{\phi}(t-1)\big)}_{\mathcal{C}_1^2} \underbrace{\mathcal{P}\big(\mathbf{u}\big|I(t), \hat{\phi}(t-1)\big)}_{\mathcal{C}_1^3}. \qquad (16)$$

We next explain the formulation and the properties of the new term $\mathcal{C}_1^3$. An important modification for the term $\mathcal{C}_1^1$ is also proposed which is complementary to $\mathcal{C}_1^3$.

## 4. New Functional

Consider the following functional (corresponding to the term $\mathcal{C}_1^3$) which connects the intensity based cue with the

shape based cue to obtain a robust estimate for $\mathbf{u}$:

$$E(\mathbf{u};\hat{\phi}(t-1),I(t))=\int_{\Omega} g_{I(t)}(\mathbf{x})\big|\nabla H_{\epsilon}(\hat{\phi}(\mathbf{T}(\mathbf{x}),t-1))\big|d\mathbf{x} \tag{17}$$

where $g_{I(t)}(\mathbf{x}) : [0,+\infty[\rightarrow \mathbb{R}^{+}$ is a strictly decreasing function computed from $I(t)$ representing the object of interest in the image. The second term inside the absolute sign ($|.|$) projects the previous segmentation map into the current frame $I(t)$. Thus the functional tries to deform the previous segmentation in such a way so as to get a better alignment with the object representation in the current frame. Here we assume the object is represented using a simple edge indicator function, i.e., $g_{I(t)} = \frac{1}{1+|\nabla I_{\sigma}(t)|}$, where $I_{\sigma}(t)$ is the Gaussian smoothed image. Thus $g$ varies inversely with the edge strength. The other choice for $g$, possibly better but computationally intensive, is computing the object representation in a discriminative way [5] using conditional random field. The advantage of using $g$ is that we do not rely only on the relative information provided by the image intensities which can be corrupted due to illumination change or due to other disturbances. In contrast, we can reliably estimate $\mathbf{u}$ using $\mathcal{C}_1^3$ as long as the object is visible in the current frame. The energy is non-convex and can be optimized locally. The Euler-Lagrange of Eq. (17) provides the update equations for $\mathbf{u}(\mathbf{x})$:

$$\frac{\partial \mathbf{u}}{\partial \tau} = -\nabla \cdot \left( g_{I(t)}(\mathbf{x}) \frac{\nabla \hat{\phi}(\mathbf{T}(\mathbf{x}))}{|\nabla \hat{\phi}(\mathbf{T}(\mathbf{x}))|} \right) \delta_{\epsilon}\big(\hat{\phi}(\mathbf{T}(\mathbf{x}))\big) \nabla \hat{\phi}(\mathbf{T}(\mathbf{x})) \tag{18}$$

where $\partial_x$ and $\partial_y$ are the partial derivatives along $x$ and $y$ directions respectively, and $\delta_{\epsilon}(z) = \frac{1}{\pi}\frac{\epsilon}{\epsilon^2+z^2}$ is the regularized dirac delta function. In Eq. (18) we drop the time argument from $\hat{\phi}$ for notational simplicity.

Apart from the advantage indicated above the term $\mathcal{C}_1^3$ is also effective when there is no illumination change. However, the presence of intensity-wise correspondence (the term $\mathcal{C}_1^1$ ) can yield false negatives if there is non-uniform illumination change. To cope with this we modify the energy term corresponding to $\mathcal{C}_1^1$ as:

$$\begin{aligned} E^{n}(\mathbf{u};&I(t-1),I(t)) = \\ &\frac{1}{2}\int_{\Omega}\Big(I(\mathbf{x},t) - I(\mathbf{T}(\mathbf{x}),t-1) + \mathcal{G}(\mathbf{x})\Big)^2 d\mathbf{x} \\ &+ \alpha\frac{1}{2}\int_{\Omega}\mathrm{trace}(\nabla\mathbf{u}\,\nabla\mathbf{u}^{T})d\mathbf{x} + \gamma\frac{1}{2}\int_{\Omega}|\nabla\mathcal{G}|^2 d\mathbf{x} \end{aligned} \tag{19}$$

where $\alpha$, $\gamma$ are the proportionality constants and $\mathcal{G}$ compensates for the illumination change, inter-reflection *etc*. across consecutive views. We would like to mention that there exists a considerable body of literature for recovering the reflectance properties in a real scene. A commonly used parametric form in this regard is bidirectional reflectance distribution function (BRDF) [25]. Instead, we employ a non-parametric function $\mathcal{G}$ to compensate for the illumination changes which does not assume any particular configuration regarding the moving object w.r.t. the light source. The overall optimization framework is presented in Algorithm 1. Note that a gradient descent scheme is employed for optimizing different variables in the SRS framework. As a result, the algorithm is prone to find local minima. To avoid this, we first compute an initial guess for $\mathbf{u}$ using a multi-resolution registration technique. It is then used to initialize $\mathbf{u}$ in Algorithm 1. It is to be noted that the performance of a gradient descent scheme depends on good initialization in general. We employ Crank-Nicolson scheme [21] (accurate of order (2,2)) to implement the diffusion equations as depicted in step 4 and step 6 of Algorithm 1. The step 5 in Alg. 1 can be implemented using standard Chan-Vese [4] formulation. The proposed algorithm adds a negligible amount of computational complexity to SRS+DP, which is linear in terms of the size of the image.

**Algorithm 1** Algorithm for simultaneous estimation of $\mathbf{u}$ and $\phi^o(t)$.

1: Consider we have $\hat{\phi}(t-1)$, $I(t-1)$, and $I(t)$.
2: Initialize $\mathbf{u}$ and $\phi^o(t)$ and $\mathcal{G}$.
3: **while** $|\phi^o(t,\tau) - \phi^o(t,\tau-1)| \geq \Delta$ **do**
4: Optimize over $E = E^n(\mathbf{u};I(t-1),I(t)) + E(\mathbf{u};\hat{\phi}(t-1),\phi^o(t)) + E(\mathbf{u};\hat{\phi}(t-1),I(t))$ ($\mathbf{u}$ related functionals).
5: Optimize over $E = E(\phi^o(t);I(t),\hat{\phi}^{-}(t)) + E(\mathbf{u};\hat{\phi}(t-1),I(t))$ ($\phi^o(t)$ related functionals).
6: Optimize over $E = E^n(\mathbf{u};I(t-1),I(t))$ ($\mathcal{G}$ related functionals).
7: Update $\mathbf{u}$ and $\phi^o(t)$ and $\mathcal{G}$.
8: **end while**

## 5. Experimental Results

We choose four different time sequences to evaluate the performance of the proposed approach. All of the sequences are 100 frames long. We use 20 frames (out of 100 frames) for tuning the parameters for each method. The remaining 80 frames are used for testing purpose. The first two sequences $\mathcal{D}_1$ and $\mathcal{D}_2$[2] are obtained from a stationary surveillance camera and the objective is to segment and track a given vehicle over a certain period of time. The sequence $\mathcal{D}_1$ (see Fig. 3) has pronounced non-uniform illumination change over consecutive time frames since the vehicle is

[2]Both sequences are downloaded from `http://i21www.ira.uka.de`.

(a) Frame 10 (b) Frame 20 (c) Frame 30 (d) Frame 40

(e) Frame 50 (f) Frame 60 (g) Frame 70 (h) Frame 80

Figure 3. This figure demonstrates some example visual results on the dataset $\mathcal{D}_1$. The corresponding frame numbers are written below. The obtained results using GSRS, SRS+DP, and SRS are shown in green, red and black contours respectively.

going through several shadow and bright regions. In $\mathcal{D}_2$ (see the first row in Fig. 4) the visibility is poor due to heavy snowfall and the inter-reflection (between the vehicle and snow) makes the segmentation and tracking quite challenging. The third sequence $\mathcal{D}_3$[3] is an indoor sequence of a plastic bottle (see the second row in Fig. 4) with drastic illumination changes where the motion is due to camera movement. In the case of sequence $\mathcal{D}_4$ (see the third row in Fig. 4) both the camera and the vehicle (*i.e.* object-of-interest) are moving at varying speed. Thus the object motion in $\mathcal{D}_4$ is not smooth like the other sequences considered for the experiment. Moreover, the illumination condition changes drastically in the later half of the sequence. We compare with three other state-of-the-art approaches namely, SRS [10], SRS+DP [11] and a blob based tracker [6] (BT). In the subsequent discussion we refer the proposed approach as GSRS.

We use F-measure, an area based error metric, to capture the performance of different approaches. Consider $\Omega_g$ ($\subseteq \mathbb{R}^2$) and $\Omega_a$ ($\subseteq \mathbb{R}^2$) be the two regions describing the object-of-interest in *groundtruth* and *automatic* segmentation in the same time frame. The F-measure is defined as:

$$F = \frac{2\mathcal{P}\mathcal{R}}{\mathcal{P}+\mathcal{R}}, \text{ where } \mathcal{P} = \frac{|\Omega_g \cap \Omega_a|}{\Omega_a}, \text{ and } \mathcal{R} = \frac{|\Omega_g \cap \Omega_a|}{\Omega_g}$$

[3]The bottle data can be found in `http://www-cvpr.iai.uni-bonn.de/data/bottledata.zip`.

where $\mathcal{P}$ is the precision and $\mathcal{R}$ is the recall; and $|.|$ represents the cardinality of a given set. Fig. 3 shows the results on the dataset $\mathcal{D}_1$. The proposed approach segments and tracks the vehicle desirably in all the time frames even in presence of severe illumination changes. Fig. 4 demonstrates some visual results on the three other sequences. We present detailed quantitative results in Table 1 for different algorithms and for different datasets. The F-measure is computed after every 10 frames. As can be seen, the proposed approach outperforms the state-of-the-art techniques in almost all time instances. We submit a video corresponding to the sequence $\mathcal{D}_4$ as supplementary material. It demonstrates the segmentation/tracking result in red contour for the proposed scheme (GSRS).

## 6. Conclusion

We introduce a generalized model for the level-set based simultaneous registration and segmentation framework applied to time sequence imagery. The proposed model is derived in a principled way by formulating a MAP problem. We show that the multiple cues, *e.g.*, shape and intensity, can be strongly coupled for establishing better correspondence. Also we explicitly account for the additive non-uniform illumination changes across consecutive views through a spatially varying term. Our method is demonstrated to be robust against various frequently occurring

Table 1. A comprehensive overview of the performance of different SRS methods over different datasets. Each sequence has 80 time frames. The F-measure is computed after every 10 frames.

| Datasets | Methods | F-measure at frame number- | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $10^{th}$ | $20^{th}$ | $30^{th}$ | $40^{th}$ | $50^{th}$ | $60^{th}$ | $70^{th}$ | $80^{th}$ |
| $\mathcal{D}_1$ | GSRS | 0.9324 | 0.9364 | 0.9175 | 0.9098 | 0.9179 | 0.8937 | 0.8552 | 0.8712 |
| | SRS+DP | 0.9187 | 0.9167 | 0.8881 | 0.8476 | 0.8427 | 0.8279 | 0.8146 | 0.8105 |
| | SRS | 0.8392 | 0.6789 | 0.4718 | 0.1068 | -lost tracking- | | | |
| | BT | 0.4038 | 0.4317 | 0.4985 | 0.4165 | -lost tracking- | | | |
| $\mathcal{D}_2$ | GSRS | 0.8749 | 0.8829 | 0.8404 | 0.8660 | 0.8638 | 0.8502 | 0.8651 | 0.9018 |
| | SRS+DP | 0.8636 | 0.8362 | 0.7910 | 0.8170 | 0.7954 | 0.7970 | 0.8389 | 0.8883 |
| | SRS | 0.7726 | 0.8314 | 0.7643 | 0.7827 | 0.8048 | 0.8020 | 0.7729 | 0.7820 |
| | BT | 0.6978 | 0.6859 | 0.7815 | 0.8030 | 0.6397 | 0.6630 | 0.7734 | 0.8094 |
| $\mathcal{D}_3$ | GSRS | 0.9558 | 0.9222 | 0.8930 | 0.8739 | 0.8917 | 0.8822 | 0.8349 | 0.8713 |
| | SRS+DP | 0.9433 | 0.9179 | 0.8665 | 0.8318 | 0.8777 | 0.8594 | 0.7643 | 0.7631 |
| | SRS | 0.7520 | 0.8287 | 0.7090 | 0.4949 | 0.4113 | 0.6891 | 0.6025 | 0.3864 |
| | BT | 0.3488 | 0.1432 | 0.0306 | 0.0349 | 0.0602 | 0.0620 | 0.1270 | 0.1060 |
| $\mathcal{D}_4$ | GSRS | 0.9451 | 0.9350 | 0.7263 | 0.8202 | 0.8793 | 0.8557 | 0.7517 | 0.8651 |
| | SRS+DP | 0.8218 | 0.6635 | 0.6688 | 0.8711 | 0.8209 | 0.7826 | 0.6940 | 0.7869 |
| | SRS | 0.8065 | 0.7701 | 0.6943 | 0.5817 | 0.3158 | 0.1314 | 0.0607 | 0.0538 |
| | BT | 0.6805 | 0.8167 | 0.7192 | 0.6093 | 0.4797 | 0.2812 | 0.2624 | 0.1193 |

challenges, *e.g*., illumination change, inter-reflection, scattering *etc*.

An area of further improvement can be better initialization of various system parameters while running optimization for different components, *e.g*., the displacement vector field, the level-set evolution and the spatially varying illumination term. Secondly, the motion computation for all the points in image domain (*dense* flow) is generally inefficient and error prone. Instead, we would like to investigate the computation of region specific motion field, which is more suitable for our problem. Currently, we are studying the effect of sequential optimization using different approximations for $\mathcal{C}_1$ on segmentation/tracking performance as discussed in Section 3.

## Acknowledgements

We would like to thank all anonymous reviewers for their valuable comments. This work was supported by the Center for Bioimage Informatics under grants NSF-ITR 0331697 and NSF-III 0808772.

## References

[1] J. An, Y. Chen, F. Huang, D. Wilson, and E. Geiser. A variational pde based level set method for a simultaneous segmentation and non-rigid registration. *LNCS*, 3749:286–293, 2005.

[2] T. Brox, C. Bregler, and J. Malik. Large displacement optical flow. *CVPR, IEEE Int. Conference on*, June 2009.

[3] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. *LNCS*, pages 25–36, 2004.

[4] T. Chan and L. Vese. Active contours without edges. *Image Processing, IEEE Transactions on*, 10(2):266–277, Feb 2001.

[5] D. Cobzas and M. Schmidt. Increased discrimination in level set methods with embedded conditional random fields. *CVPR, IEEE Int. Conference on*, June 2009.

[6] D. Comaniciu, V. Ramesh, and P. Meer. Kernel-based object tracking. *IEEE Trans. on PAMI*, pages 564–577, 2003.

[7] T. Cootes, G. Edwards, C. Taylor, et al. Active appearance models. *IEEE Trans. on PAMI*, 23(6):681–685, 2001.

[8] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models: their training and application. *CVIU*, 61(1):38–59, Jan 1995.

[9] D. Cremers, M. Rousson, and R. Deriche. A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape. *IJCV*, 72(2):195–215, 2007.

[10] J. Ehrhardt, A. Richberg, and H. Handels. A variational approach for combined segmentation and estimation of respiratory motion in temporal image sequences. *Computer Vision, IEEE Int. Conference on*, pages 1–7, Oct. 2007.

[11] P. Ghosh, M. E. Sargin, and B. S. Manjunath. Robust dynamical model for simultaneous registration and segmentation in a variational framework: A bayesian approach. *Computer Vision, IEEE Int. Conference on*, Oct. 2009.

[12] H. Ishikawa and I. Jermyn. Region extraction from multiple images. In *IEEE Int. Conference on Computer Vision*, pages 509–516, 2001.

[13] F. Kschischang, B. Frey, and H. Loeliger. Factor graphs and the sum-product algorithm. *IEEE Trans. on Information theory*, 47(2):498–519, 2001.

[14] M. Leventon, W. Grimson, and O. Faugeras. Statistical shape influence in geodesic active contours. *Biomedical Imaging, IEEE 5th EMBS Int. Summer School on*, June 2002.

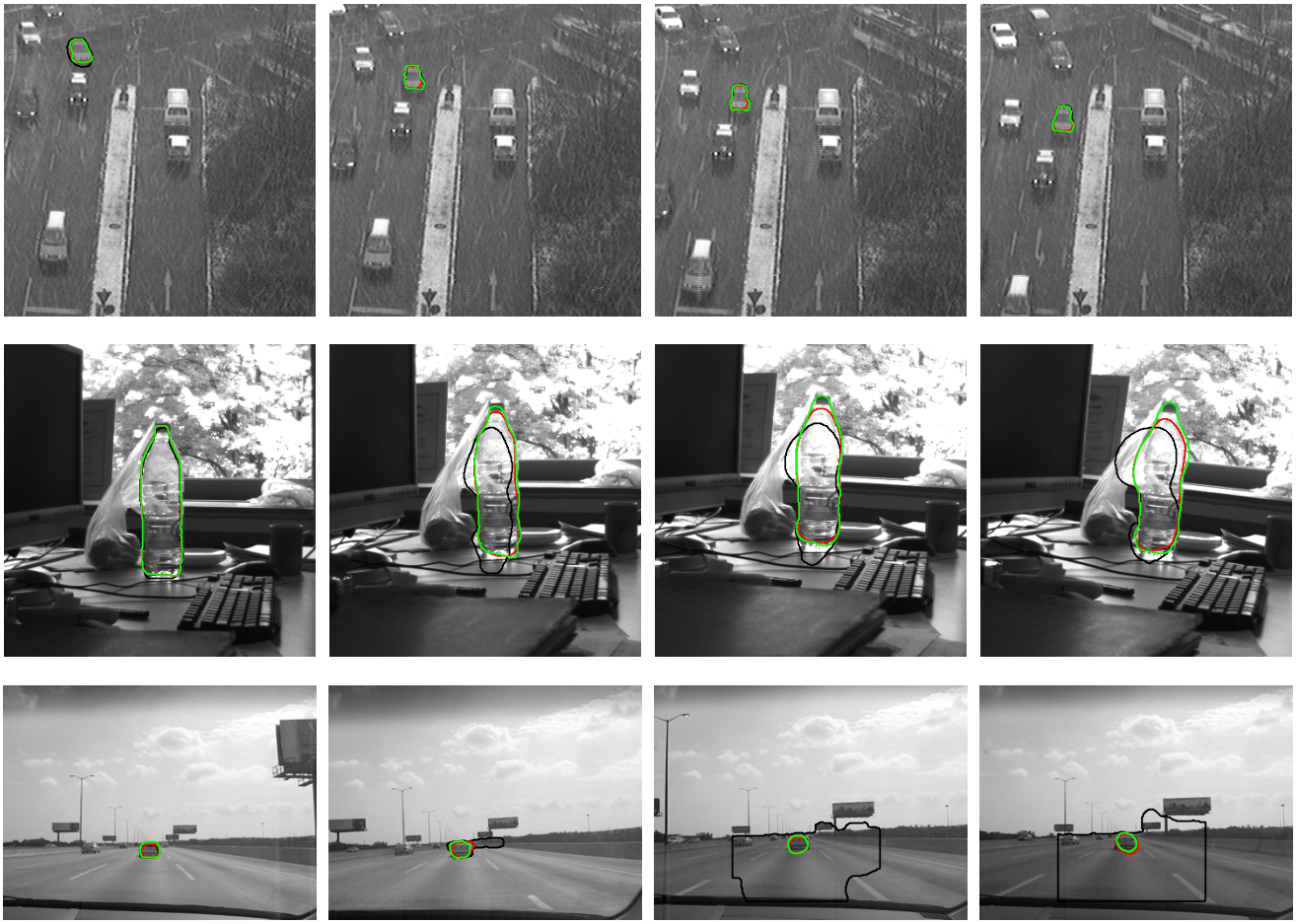

Figure 4. Sample frames and the corresponding results for the datasets $\mathcal{D}_2$ (first-row), $\mathcal{D}_3$ (second-row) and $\mathcal{D}_4$ (third-row) are demonstrated here. In all the plots the obtained results using GSRS, SRS+DP, and SRS are shown in green, red and black contours respectively.


[15] S. C. Mitchell, B. P. Lelieveldt, V. der Geest, et al. Multistage hybrid active appearance model matching: segmentation of left and right ventricles in cardiac mr images. *Medical Imaging, IEEE Trans. on*, 20(5):415–423, May 2001.

[16] N. Paragios and R. Deriche. Geodesic active regions and level set methods for motion estimation and tracking. *Computer Vis. and Image Understanding*, pages 259–282, 2005.

[17] K. Pohl, J. Fisher, J. Levitt, et al. A unifying approach to registration, segmentation, and intensity correction. *Lecture Notes in Computer Science*, 3749:310–318, 2005.

[18] K. Saddi, C. Chefd'hotel, M. Rousson, and F. Cheriet. Region-based segmentation via non-rigid template matching. *Computer Vision, IEEE Int. Conference on*, pages 1–7, Oct. 2007.

[19] T. Schoenemann and D. Cremers. Globally optimal shape-based tracking in real-time. In *IEEE Int. Conference on CVPR*, Anchorage, Alaska, June 2008.

[20] T. Schoenemann and D. Cremers. A combinatorial solution for model-based image segmentation and real-time tracking. *IEEE Trans. PAMI*, 2009.

[21] J. Strikwerda. *Finite difference schemes and partial differential equations*. Society for Industrial Mathematics, 2004.

[22] A. Tsai, J. Yezzi, A., W. Wells, et al. A shape-based approach to the segmentation of medical imagery using level sets. *Medical Imaging, IEEE Trans. on*, 22(2):137–154, Feb 2003.

[23] G. Unal and G. Slabaugh. Coupled pdes for non-rigid registration and segmentation. *CVPR, IEEE Int. Conference on*, 1:168–175, June 2005.

[24] F. Wang and B. Vemuri. Simultaneous registration and segmentation of anatomical structures from brain mri. *Lecture Notes in Computer Science*, 3749:17, 2005.

[25] G. Ward. Measuring and modeling anisotropic reflection. In *Proc. of the 19th annual conf. on Computer graphics and interactive techniques*, pages 265–272, 1992.

[26] A. Yezzi, L. Zllei, and T. Kapur. A variational framework for integrating segmentation and registration through active contours. *Medical Image Analysis*, 7(2):171–185, 2003.

[27] Y. Zhu, X. Papademetris, A. Sinusas, and J. Duncan. Segmentation of left ventricle from 3d cardiac mr image sequences using a subject-specific dynamical model. *CVPR, IEEE Conference on*, pages 1–8, June 2008.