# Robust Dynamical Model for Simultaneous Registration and Segmentation in a Variational Framework: A Bayesian Approach

Pratim Ghosh, Mehmet 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

We introduce a dynamical model for simultaneous registration and segmentation in a variational framework for image sequences, where the dynamics is incorporated using a Bayesian formulation. A linear stochastic equation relating the tracked object (or a region of interest) is first derived under the assumption that the successive images in the sequence are related by a dense and possibly non-linear displacement field. This derivation allows for the use of a computationally efficient and recursive implementation of the Bayesian formulation in this framework. The contour of the tracked object returned by the dynamical model is not only close to the previously detected shape but is also consistent with the temporal statistics of the tracked object. The performance of the proposed approach is evaluated on real image sequences. It is shown that, with respect to a variety of error metrics such as F-measure, mean absolute deviation and Hausdorff distance, the proposed approach outperforms the state-of-the art approach without the dynamical model.

### 1. Introduction

We consider the problem of simultaneous registration/segmentation in time sequence imagery. A variational framework for this problem is introduced and a solution based on a generic dynamical model is derived. The proposed solution is quite powerful and is applicable in a wide range of applications, ranging from medical imaging to surveillance videos.

Assume that we are given a sequence of images  $I(\mathbf{x},t)|_{t=1,2,...,T}$  consisting of an object that is undergoing motion and/or other deformations. We also assume that the initial condition of the object of interest is known *a priori*. Application examples include Magnetic Resonance Imaging (MRI) sequences of the heart where the goal is to track the endocardium (see Fig. 5), or tracking a vehicle (see

Fig. 6) in a surveillance scenario. The main problem is to segment and track the object of interest from the cluttered background in the presence of noise, motion, and shape deformations/variations. In such cases, segmentation methods that do not take into account shape information or those that are primarily driven by the image data (e.g., edge information) are not quite effective. More recently, statistical shape modeling is used to address some of the above issues. The initial works on statistical shape modeling were developed using the active shape model (ASM) [6], active appearance model (AAM) [5] and the the deformable shape model [12]. The ASM and AAM methods were based on the statistical distribution of the shape boundary points. In [13], the authors combined ASM and AAM in a multistage framework to increase the robustness of the shape matching process to noise and clutter. The authors in [17] introduced a new deformable shape model of the segmenting curve for medical applications. The model is obtained by applying the principal component analysis on a collection of signed distance functions which represent the training shapes. In a recent work [21], the authors developed a subject specific dynamic model to segment the left ventricle from cardiac MR images. Here, the authors considered both the intersubject (variations of the shapes among different patients) and the intra-subject variabilities (dynamics of the heart). Multi-linear PCA (principal component analysis) and ICA (independent component analysis) were employed to learn different variabilities. They demonstrated improvement in performance compared to the traditional segmentation approaches.

All of the above mentioned methods do not explicitly account for the registration while segmenting an object of interest from image sequences. A promising emerging approach in this context is combined registration and segmentation, with registration helping the segmentation task which in turn is used to improve the alignment with the next frame in the time sequence.

Consider the registration part. Assume two consecutive frames from the sequence  $I(\mathbf{x},t): \Omega \to \mathbb{R}^2$  and  $I(\mathbf{x},t-1):$ 



Figure 1. (a) Block diagram of the proposed framework. (b) A close up view of the *Prior* block. The decision box (diamond shaped) clearly shows the computation of two different priors. Here SP and DP stand for the static and the dynamical prior respectively. See Sec 3 for detailed explanation of the prediction and the update blocks.

 $\Omega \to \mathbb{R}^2$ . The displacement vector field  $\mathbf{u}(\mathbf{x}, t-1, t)$ :  $\Omega \to \mathbb{R}^2$  defines the mapping between the frames such that  $I(\mathbf{x}, t) = I(\mathbf{T}(\mathbf{x}), t-1)$  where  $\mathbf{T}(\mathbf{x}) = \mathbf{x} - \mathbf{u}$ . For the rest of the discussion we assume that  $\mathbf{u}(\mathbf{x}, t-1, t) \equiv \mathbf{u}(\mathbf{x}) \equiv \mathbf{u}$ unless mentioned otherwise. Now the intensity based functional for non linear registration can be written as:

$$C_{reg}(\mathbf{u}) = \underbrace{\frac{1}{2} \int_{\Omega} (I(\mathbf{x}, t) - I(\mathbf{T}(\mathbf{x}), t-1))^2 d\mathbf{x}}_{C_{reg}} + \alpha \underbrace{\frac{1}{2} \int_{\Omega} trace(\nabla \mathbf{u} \, \nabla \mathbf{u}^T) d\mathbf{x}}_{C_{reg}^2}$$
(1)

where  $(.)^T$  represents the transpose operation,  $C_{reg}^1$  is the data term,  $C_{reg}^2$  is the regularizer for the vector field, and  $\alpha$  controls the smoothness of the derived vector field **u**. This non-linear registration model [4] can handle larger shape variations than any rigid registration model like the one used in [17].

Similarly one can write down a cost functional for the segmentation part as (from Chan et al. [3]):

$$C_{cv}(\phi^{o}(\mathbf{x},t)) = \overbrace{\int_{\Omega} \ln p(I|\theta_{2}) + H_{\epsilon}(\phi^{o}(\mathbf{x},t)) \ln \frac{p(I|\theta_{2})}{p(I|\theta_{1})} d\mathbf{x}}^{C_{cv}^{1}} + \nu \underbrace{\int_{\Omega} |\nabla H_{\epsilon}(\phi^{o}(\mathbf{x},t))| d\mathbf{x}}_{C_{cv}^{2}}$$
(2)

where  $C_{cv}^1$  is the data term,  $C_{cv}^2$  is the length term,  $\phi^o(\mathbf{x}, t)$ :  $\Omega \to \mathbb{R}$  is the level-set function,  $H_\epsilon$  is the regularized Heav-

iside function, and  $\theta_1$ ,  $\theta_2$  parameterize the object and the background pdfs.

Recently, it is shown that the segmentation result can be significantly improved by performing the registration and segmentation simultaneously since the solutions to the two parts (i.e. the registration and the segmentation parts) above are inter-related. To date different variants have been developed in this context. For instance, the work in [20] developed an active contours framework, where the registration was formulated using only the affine registration model. Also see the work in [14] and references therein for simultaneous registration and segmentation (SRS) using other approaches. In [19], the authors developed a variational formulation for registration guided image segmentation where the solutions for both problems were computed simultaneously. A variant of this technique is developed in [15], where the segmentation is obtained with respect to the deformation. Another set of techniques [18, 1] have been developed for SRS by computing the deformation field only around the object of interest.

## 2. Simultaneous Registration and Segmentation

Here we briefly review a recent work [8] on simultaneous registration and segmentation (SRS) framework based on which the proposed approach is developed. Consider Fig. 1. Here I(t-1) and I(t) denote successive image frames in a video sequence at times t-1 and t respectively. It is assumed that the segmentation ( $\phi^o(\mathbf{x}, t-1)$ ) of the reference image  $I(\mathbf{x}, t-1)$  is known to us. Based on the fact that **u** maps the pixel from  $I(\mathbf{x}, t-1)$  to  $I(\mathbf{x}, t)$ , the authors in [8] proposed a relation between the reference (at t-1) and the target (at



Figure 2. Example of a sequence of images. The first two (i.e. A and B) and the last two (i.e. B and C) image patches are examples of good and bad candidates for the computation of displacement field respectively. See the sudden background change from B to C (highlighted rectangular regions) near the rear end of the car edge, which violates the brightness constancy assumption for the computation of  $\mathbf{u}$  in those regions.

t) level-set functions as follows:

$$\phi^{o}(\mathbf{x},t) = \phi^{o}(\mathbf{T}(\mathbf{x}),t-1)$$
(3)

The above formula couples the segmentation and the registration process into a single objective. In [8], the functional corresponding to the objective in Eq. (3) was formulated as:

$$C_{comb}(\phi^{o}(\mathbf{x},t),\mathbf{u}) = \frac{1}{2} \int_{\Omega} N_{\epsilon}(\phi^{o}(\mathbf{x},t),\phi^{o}(\mathbf{x},t-1))(\phi^{o}(\mathbf{x},t)) - \underbrace{\phi^{o}(\mathbf{T}(\mathbf{x}),t-1)}_{\text{prior}})^{2} d\mathbf{x}$$
(4)

where  $N_{\epsilon}$  is a binary function which selects the pixels in a  $\epsilon$  neighborhood of the shapes (i.e.  $\phi^{o}(\mathbf{x}, t)$  and  $\phi^{o}(\mathbf{x}, t-1)$ ) under consideration for the optimization. Essentially, the functional in Eq. (4) tries to keep the target segmentation close to the deformed reference segmentation result.

The prior term in Eq. (4) depends on the previous shape  $(\phi^{o}(\mathbf{x}, t-1))$ , and the displacement vector field  $\mathbf{u}(\mathbf{x}, t-1, t)$ , which is being simultaneously estimated in the current curve evolution process (see Eq. (4)). A common source of error in the estimation of u is when the background (especially near the object edges) does not remain constant across successive frames. In optical flow literature [10], this assumption is often referred to as the brightness constancy assumption where it is assumed that the following relation holds:  $I(\mathbf{x}, t) = I(\mathbf{x} + \Delta \mathbf{x}, t + \Delta \mathbf{t})$ , for small enough values of  $\Delta t$ . Consider, for example the three rectangular regions of interest A, B and C in Fig 2. A, B, and C represent the corresponding regions in these time sequence images. The brightness constancy assumption for the computation of displacement field holds good for A and B but not for B and C (due to the white spot at the rear end of the car). Thus, it is difficult to ensure an absolutely error-free computation of u between B and C in practical experimental settings. It is to be emphasized that this estimation error occurs for u even when the input  $\phi^o(\mathbf{x}, t-1)$  is error-free due to the problem with the background constancy assumption. Hence, in the SRS framework [8], an error may be introduced in the



Figure 3. The first and second rows present the segmentation result (black contour) using the proposed approach and the previous work in [8] respectively. The displayed results zoom into the rear part of the three car images shown in Fig. 2. The results for three other successive frames are also shown to describe the nature of the segmentation error for SRS without any dynamic prior. The previous work in [8] shows accumulation of segmentation error in the successive frames.

computation of  $\phi^o(\mathbf{x}, t)$  due to the estimation error in **u**. From the given formulation of the SRS framework [8], it is not obvious whether it is possible to compensate and hence avoid the accumulation of such errors. See for example, the second row of the Fig. 3, where this segmentation error (for the previous framework [8]) accumulates over successive frames even though it was absent in the first two frames.

To overcome this estimation error for  $\phi^o(\mathbf{x}, t)$ , we employ a dynamic prior term  $\phi^{-}(t)$  which maximizes 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 dynamic prior term avoids the dependence on the current estimate  $\mathbf{u}(\mathbf{x}, t-1, t)$  in the estimation of  $\phi^{o}(\mathbf{x},t)$ , thus providing robustness against noise/clutter and other disturbances. We show later that if the underlying process is assumed to be first order Markov, the proposed dynamical prior  $\phi^{-}(t)$  can be efficiently computed using only  $\mathbf{u}(t-2,t-1)$  and  $\phi^o(t-1)$  - an implicit assumption here is that the past estimates of  $\mathbf{u}(t-2, t-1)$  and  $\phi^{o}(1), \phi^{o}(2), \dots, \phi^{o}(t-1)$  are error-free. This error-free assumption is also made in the SRS framework in [8]. As we will demonstrate in the following sections, with the dynamic prior  $\phi^{-}(\mathbf{x},t)$ , our proposed approach overcomes this SRS problem in an effective way. See Fig. 3 (the first row) for some examples. The key components of our method are:

- A generic dynamical model using Bayesian formulation for the level-set based simultaneous segmentation and non-linear registration is proposed.
- Our dynamical model is unsupervised, as we do not require any prior learning of the shape variation in a temporal sequence or between different sequences of same kind.
- We demonstrate that the Bayesian problem reduces to solving a linear stochastic equation after certain simplifying assumptions on the underlying model of SRS

and thus can be solved in a computationally efficient way.

The development and the implementation of the proposed dynamical model is described in Section 3. In Section 4 we present quantitative and qualitative evaluation of the proposed approach. We conclude with discussions in Section 5.

# 3. Model-based Probabilistic Tracking for Segmentation

In this section, we introduce a Bayesian formulation for incorporating the dynamics of the object into SRS framework and a computationally efficient approach is devised for the implementation purpose.

Let  $\phi(\mathbf{X}, t)$  be a set of level-set function samples  $\{\phi(\mathbf{x}, t)\}$ , obtained by uniformly sampling the spatial domain  $\Omega$  ( $\mathbf{x} \in \Omega$ ). Then, the maximum *a posteriori* estimate of  $\phi(\mathbf{X}, t)$  given all the observations up to time *t* can be formulated as:

$$\hat{\phi}(\mathbf{X},t) = \operatorname*{argmax}_{\phi(\mathbf{X},t)} \mathcal{P}(\phi(\mathbf{X},t) | \phi^o(\mathbf{X},1\!:\!t))$$
(5)

where  $\phi^o(\mathbf{X}, 1 : t)$  represents the set of observations  $\{\phi^o(\mathbf{X}, 1), \phi^o(\mathbf{X}, 2), ..., \phi^o(\mathbf{X}, t)\}$  obtained till the current time t. The computation of observations from images is described later in this section. For computational efficiency, we will now assume that each element of the embedding function,  $\phi(\mathbf{x}, t)$ , is independent of (a)  $\phi(\mathbf{X}, t) \setminus \phi(\mathbf{x}, t)$ , and (b)  $\phi^o(\mathbf{X}, 1:t) \setminus \phi^o(\mathbf{x}, 1:t)$ , given  $\phi(\mathbf{x}, t-1)$  for both cases. Here,  $\setminus$  represents the set difference operation. Therefore, the right hand side of Eq. (5) reads:

$$\mathcal{P}(\phi(\mathbf{X},t)|\phi^{o}(\mathbf{X},1:t)) = \prod_{\mathbf{x}\in\mathbf{X}} \mathcal{P}(\phi(\mathbf{x},t)|\phi^{o}(\mathbf{x},1:t))$$
(6)

Thus the maximization problem in (5) reduces to maximizing each elements as follows:

$$\hat{\phi}(\mathbf{x},t) = \operatorname*{argmax}_{\phi(\mathbf{x},t)} \mathcal{P}(\phi(\mathbf{x},t) | \phi^{o}(\mathbf{x},1\!:\!t))$$
(7)

Next consider the underlying model of SRS in Eq. (3) between t and t+1 time instances:

$$\phi(\mathbf{x},t+1) = \phi(\mathbf{x} - \mathbf{u}(\mathbf{x},t,t+1),t)$$
(8)

Applying Taylor series expansion on both sides and truncating the higher order terms, we obtain:

$$\phi(\mathbf{x},t) + \phi_t(\mathbf{x},t) \approx \phi(\mathbf{x},t) - \mathbf{u}^T(\mathbf{x},t,t+1)\nabla\phi(\mathbf{x},t)$$
  

$$\phi(\mathbf{x},t+1) - \phi(\mathbf{x},t) \approx -\mathbf{u}^T(\mathbf{x},t,t+1)\nabla\phi(\mathbf{x},t)$$
  

$$\phi(\mathbf{x},t+1) \approx \phi(\mathbf{x},t) - \mathbf{u}^T(\mathbf{x},t,t+1)\nabla\phi(\mathbf{x},t)$$
(9)

where  $\phi_t$  and  $\nabla \phi$  represent the temporal and the spatial derivatives respectively, and  $(.)^T$  represents the transpose

operation. From Eq. (9), we can obtain a linear stochastic equation for the SRS model provided in Eq. (8) as:

$$\phi(\mathbf{x},t) = \phi(\mathbf{x},t-1) - \mathbf{u}^{T}(\mathbf{x},t-2,t-1)\nabla\phi(\mathbf{x},t-1) + w$$
(10)

where w represents a certain error term with distribution  $\mathcal{P}(w) \sim \mathcal{N}(0, Q)$ . We approximate  $\mathbf{u}(\mathbf{x}, t-1, t)$  by  $\mathbf{u}(\mathbf{x}, t-2, t-1)$  assuming a continuous motion of the object of interest. The error involved in this approximation and the Taylor series expansion error in Eq. (9) is modeled using an Additive White Gaussian Model (AWGN).

The observed variable is related to its original variable as:

$$\phi^{o}(\mathbf{x},t) = \phi(\mathbf{x},t) + v \tag{11}$$

where v is the observation error with pdf  $\mathcal{P}(v) \sim \mathcal{N}(0, R)$ . Based on the model described by Eq. (10) and (11),  $\hat{\phi}(\mathbf{x}, t)$  (from Eq. (7)) can be recursively estimated as described below [11].

**Prediction Model:** The prediction (dynamic prior) for the current time instance t given all the previous observations is obtained by maximizing the following pdf:

$$\hat{\phi}^{-}(\mathbf{x},t) = \operatorname*{argmax}_{\phi(\mathbf{x},t)} \mathcal{P}(\phi(\mathbf{x},t) | \phi^{o}(\mathbf{x},1:t-1))$$
(12)

which simplifies (using Eq. (10) and (11)) to computing:

$$\hat{\phi}^{-}(\mathbf{x},t) = \hat{\phi}(\mathbf{x},t-1) - \mathbf{u}^{T}(\mathbf{x},t-2,t-1)\nabla\hat{\phi}(\mathbf{x},t-1).$$
 (13)

**Observation Model:** The observation is calculated from the combined registration and segmentation framework after including the dynamic prior term into Eq. (2). We can express the dynamic prior term as:

$$\mathcal{P}(\phi^{o}(\mathbf{x},t)|\hat{\phi}(\mathbf{x},1:t-1)) \propto e^{-\frac{1}{2}\left(\beta \int_{\Omega} ||\phi^{o}(\mathbf{x},t)-\hat{\phi}^{-}(\mathbf{x},t)||^{2} d\mathbf{x}\right)}$$
(14)

where  $\beta$  is a positive constant. Now Eq. (2) reads as:

$$C_{cv}(\phi^{o}(\mathbf{x},t)) = C_{cv}^{1}(\phi^{o}(\mathbf{x},t)) + \nu C_{cv}^{2}(\phi^{o}(\mathbf{x},t)) + \beta \underbrace{\frac{1}{2} \int_{\Omega} |\phi^{o}(\mathbf{x},t) - \hat{\phi}^{-}(\mathbf{x},t)|^{2} d\mathbf{x}}_{C_{cv}^{3}} \quad (15)$$

where  $C_{cv}^3$  is the dynamic prior term. The terms  $C_{cv}^1$  and  $C_{cv}^2$  are already defined in Eq. (2).

The complete optimization framework with the dynamic prior term can now be expressed as:

$$C(\mathbf{u}, \phi^{o}(\mathbf{x}, t)) = \gamma_{1}C_{reg}^{1}(\mathbf{u}) + \gamma_{2}C_{reg}^{2}(\mathbf{u}) + \gamma_{3}C_{cv}^{1}(\phi^{o}(\mathbf{x}, t)) + \gamma_{4}C_{cv}^{2}(\phi^{o}(\mathbf{x}, t)) + \gamma_{5}C_{cv}^{3}(\phi^{o}(\mathbf{x}, t)) + \gamma_{6}C_{comb}(\mathbf{u}, \phi^{o}(\mathbf{x}, t))$$
(16)

where  $\gamma_i$ s are the tuning parameters in this framework. The Euler-Lagrange of Eq. (16) provides the update equations

for **u** and the level-set function  $\phi^o(\mathbf{x}, t)$ :

$$\frac{\partial \mathbf{u}(\mathbf{x},\tau)}{\partial \tau} = \gamma_1 \Big( I(\mathbf{T}(\mathbf{x}),t-1) - I(\mathbf{x},t) \Big) \nabla I(\mathbf{x},t-1) |_{\mathbf{T}(\mathbf{x})} 
+ \gamma_2 \Delta \mathbf{u}(\mathbf{x}) + \gamma_6 N_\epsilon (\phi^o(\mathbf{x},t),\hat{\phi}(\mathbf{x},t-1)) \Big( \hat{\phi}(\mathbf{T}(\mathbf{x}),t-1) 
- \phi^o(\mathbf{x},t) \Big) \nabla \hat{\phi}(\mathbf{x},t-1) |_{\mathbf{T}(\mathbf{x})}$$
(17)

and

$$\frac{\partial \phi^{o}(\mathbf{x}, t, \tau)}{\partial \tau} = \delta_{\epsilon}(\phi^{o}(\mathbf{x}, t)) \left(\gamma_{3} \ln \frac{p(I(x)|\theta_{1})}{p(I(x)|\theta_{2})} + \gamma_{4} div(\frac{\nabla \phi^{o}(\mathbf{x}, t)}{|\nabla \phi^{o}(\mathbf{x}, t)|})\right) - \gamma_{5} \left(\phi^{o}(\mathbf{x}, t) - \hat{\phi}^{-}(\mathbf{x}, t)\right) + \gamma_{6} N_{\epsilon}(\phi^{o}(\mathbf{x}, t), \hat{\phi}(\mathbf{x}, t-1)) \left(\phi^{o}(\mathbf{x}, t) - \hat{\phi}(\mathbf{T}(\mathbf{x}), t-1)\right) \right) \tag{18}$$

Eq. (17) and (18) are updated simultaneously. Eq. (17) requires solving two diffusion equations, one for each component of **u**. The heat equation in Eq. (17) can be efficiently implemented using Crank-Nicolson scheme [16] which is accurate of order (2,2). It is to be noted that the update equations for **u** and  $\phi^o(\mathbf{x}, t)$  are obtained by the gradient descent scheme which tends to find local, rather than global, minima. Therefore, we first compute an initial guess  $\mathbf{u}_o$  using a multi-resolution non-linear registration technique and then use it to initialize **u** for Eq. (17) and (18).

**Update Model:** Based on the dynamic prior term  $\hat{\phi}^{-}(\mathbf{x},t)$  and  $\phi^{o}(\mathbf{x},t)$ , we can now compute  $\hat{\phi}(\mathbf{x},t)$  as:

$$\hat{\phi}(\mathbf{x},t) = (1-\kappa(t))\hat{\phi}^{-}(\mathbf{x},t) + \kappa(t)\phi^{o}(\mathbf{x},t)$$
(19)

where  $\kappa(t) = \frac{\sigma^{-}(t)}{\sigma^{-}(t)+R}$ . Here,  $\sigma^{-}(t)$  is the covariance of the prediction error and it can be recursively computed by  $\sigma^{-}(t) = \sigma^{-}(t-1) + Q$ , where Q is the variance of w (under the AWGN assumption in Eq. (10)). As can be seen from Eq. (19), the segmentation result  $\hat{\phi}(\mathbf{x},t)$  is a linear combination of the prediction  $\hat{\phi}^{-}(\mathbf{x},t)$  and the observation  $\phi^{o}(\mathbf{x},t)$  (obtained from the SRS framework) and does not solely depend on either of them.

Our formulation for the dynamical model (Eq. (5)) is motivated by the work of Cremers [7], however, our embedding formulation is different. To learn the dynamics, a second order auto regressive model is applied in [7] on the parametric representation of the level-set function. In contrast, we develop the dynamical model directly from the level-set function and thus, the speed and the simplicity of the original Chan Vese model can be exploited in the segmentation process [9]. Another notable difference with [7] is that there is no *separate training phase* in our proposed framework.

Table 1. Comparison of two approaches for the segmentation of cardiac MR images. The performance is obtained over 32 (patients) $\times$ 8 (slices) $\times$ 18 (frames)=4608 images.

Error Metrics	SRS+DP	SRS
F-measure	$0.8828 \pm 0.0576$	$0.8196 \pm 0.1397$
MAD	$0.2458 \pm 0.2019$	$1.1519 \pm 2.0956$
HD	$5.2929 \pm 2.5605$	$11.9556 \pm 13.1597$

#### 4. Experimental Results

To evaluate the performance of the proposed method, we choose two different datasets from two different domains. We choose these datasets to demonstrate the effectiveness of incorporating a dynamic prior in a rapidly changing environment. The problem corresponding to the first dataset [2] is to track the endocardium of the heart (for medical applications) in short axis cardiac MR images. This dataset was constructed by studying 33 subjects. There are 8 z-slices for which the groundtruth is available for all the patients. Thus, we extract results from  $32(\text{patients}) \times 8(\text{slices}) \times 18(\text{frames}) = 4608 \text{ images, since}$ the data from one patient is used to tune the parameters ( $\gamma_i$ s) for both methods. In the second dataset<sup>1</sup> (for surveillance applications), the objective is to track a car in a traffic image sequence. The sequence is 80 frames long. A significant amount of non-linear deformation (corresponding to the endocardium of the heart) is present in the first dataset whereas the second one has substantial amounts of motion with accompanying perspective distortions. In the subsequent discussion, we refer to the prior work as Simultaneous Registration and Segmentation (SRS) and our proposed method as SRS with Dynamic Prior (SRS+DP).

We use three different error metrics to capture the performance of our proposed method. They are (a) mean absolute distance (MAD), (b) Hausdorff distance (HD), (c) F-measure. The first two are distance based (i.e. MAD and HD) and the last one (i.e. F-measure) is area based error metric. The MAD captures the global disagreement between two regions (for example, the groundtruth and the obtained result) whereas HD computes the local dissimilarity between them. Let  $\mathcal{M}_g = \{g_1, g_2, ..., g_p\}$  and  $\mathcal{M}_r = \{r_1, r_2, ..., r_q\}$  be two set of points corresponding to the groundthruth (manual) and the obtained (automatic) result respectively. Then, MAD and HD are defined as follows:

$$MAD(\mathcal{M}_g, \mathcal{M}_r) = \frac{1}{2} \left( \frac{1}{p} \sum_{i=1}^p d(g_i, \mathcal{M}_r) + \frac{1}{q} \sum_{j=1}^q d(r_j, \mathcal{M}_g) \right)$$
$$HD(\mathcal{M}_g, \mathcal{M}_r) = \max\left( \max_i d(g_i, \mathcal{M}_r), \max_j d(r_j, \mathcal{M}_g) \right)$$
where  $d(q_i, \mathcal{M}_r) = \min_i |r_i - q_i|$  i.e. the minimum dis-

where  $d(g_i, \mathcal{M}_r) = \min_{r_j \in \mathcal{M}_r} |r_j - g_i|$ , i.e. the minimum dis-

<sup>&</sup>lt;sup>1</sup>The data is downloaded from http://i21www.ira.uka.de

tance from a given point to the other set. Let the regions that circumscribe the set of points  $\mathcal{M}_g$  and  $\mathcal{M}_r$  be denoted by  $\Gamma_g$  and  $\Gamma_r$  respectively. We can compute the F-measure as:

$$F = \frac{2\mathcal{PR}}{\mathcal{P} + \mathcal{R}}, \text{ where } \mathcal{P} = \frac{|\Gamma_g \cap \Gamma_r|}{\Gamma_r}, \text{ and } \mathcal{R} = \frac{|\Gamma_g \cap \Gamma_r|}{\Gamma_g}$$

where  $\mathcal{P}$  is the precision and  $\mathcal{R}$  is the recall; and |.| represents the cardinality of a given set.

For the cardiac MR dataset, the parameters are tuned using a number of sequences (taken from different slices of a particular patient) for each of the two methods separately in the training phase. Then the parameters are kept fixed during the rest (the data corresponding to the remaining 32 patients) of the experiment. Table 1 summarizes the overall performance of the two approaches, where the error values are obtained by averaging over all the patients, slices and frames. We also compare the performance of our proposed method (SRS+DP) with that obtained using SRS along individual frames in Fig. 4. In all the plots, we can see a general trend that the performance is getting worse with time for SRS due to error build up whereas this is clearly not the case for SRS+DP. Figure. 5 presents some visual results on two different image sequences obtained from two different patients.

Figure. 6 summarizes the performance of the two approaches on a traffic sequence. The displayed images are approximately 20 frames apart. We are tracking a car which has approximately the same intensity as the road. SRS shows similar result as the proposed approach (SRS+DP) for the first two images whereas it fails to produce desired results in the last two cases. As explained before (see the discussion in Sec. 2), the rear part of the car is not well segmented in case of SRS due to the lack of any prior knowl-edge regarding its dynamics.

#### **5.** Conclusion

We propose a generic dynamical model applicable to time sequence imagery, which is incorporated into the simultaneous registration and segmentation (SRS) framework. A Bayesian approach is first employed to learn the dynamics of the object of interest. The learning is performed in a computationally efficient way by assuming that the underlying SRS model is first order Markov. Our proposed dynamical model has a number of attractive features. The model is directly developed from the level-set function without constraining it to a lower dimensional parametric space. Also the proposed model acquires the dynamics of the data while the actual segmentation takes place. Thus our method does not require any prior information about the shape variation, i.e. there is no explicit training phase. The proposed method performs better than SRS in presence of noise/clutter and other disturbances. We also demonstrate that the lack of noise robustness of the SRS framework, where the segmentation error accumulates with time, can be circumvented with our dynamical model. The proposed approach introduces a moderate amount of complexity which grows linearly with the number of pixels in the image.

One of the issues with the proposed method is tuning the six parameters in Eq. (16). These need to be set for each class of image sequence data and it was done manually for the two classes used in our experiments. We would like to explore automated tuning given the groundtruth data. Also, we made an assumption regarding the dependency of the level-set function samples while building up the dynamical model (see Eq. (6)) for computational efficiency. Future work includes the analysis of the trade off involved between the segmentation accuracy and the computational efficiency in this assumption.

### Acknowledgements

We would like to thank all anonymous reviewers for their valuable comments. This work was supported by the NSF Information Technology Research grant # ITR- 0331697.

### References

- 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. *Lecture Notes in Computer Science*, 3749:286–293, 2005.
- [2] A. Andreopoulos and J. K. Tsotsosa. Efficient and generalizable statistical models of shape and appearance for analysis of cardiac mri. *Medical Image Analysis*, 12(3):335–357, June 2008.
- [3] T. Chan and L. Vese. Active contours without edges. Image Processing, IEEE Transactions on, 10(2):266–277, Feb 2001.
- [4] G. Christensen, R. Rabbitt, and M. Miller. Deformable templates using large deformation kinematics. *Image Processing, IEEE Transactions on*, 5(10):1435–1447, Oct 1996.
- [5] T. Cootes, G. Edwards, C. Taylor, et al. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–685, 2001.
- [6] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models: their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, Jan 1995.
- [7] D. Cremers. Dynamical statistical shape priors for level setbased tracking. *Pattern Analysis and Machine Intelligence*, *IEEE Transactions on*, 28(8):1262–1273, Aug. 2006.
- [8] J. Ehrhardt, A. Schmidt-Richberg, and H. Handels. A variational approach for combined segmentation and estimation of respiratory motion in temporal image sequences. *Computer Vision, IEEE 11th International Conference on*, pages 1–7, Oct. 2007.
- [9] B.-W. Hong, E. Prados, S. Soatto, and L. Vese. Shape representation based on integral kernels: Application to image



Figure 4. These plots compare the (a) F-measure, (b) Mean Absolute Deviation, and (c) Hausdorff distance measure obtained from SRS and SRS+DP.



Figure 5. Examples of visual results from cardiac MR data-set. The first row and the third row show the results on few images from two different sequences for the proposed method while the corresponding result for SRS is shown in the second and the last rows. In all the plots the groundtruth and obtained results are shown using red and green contours respectively.



Figure 6. Example of visual results for tracking a car near an intersection. The first and the second rows show the results (on few images) for the proposed method and for the SRS method respectively. In all the plots the groundtruth and computed results are shown using red and green contours respectively.

matching and segmentation. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 1:833– 840, June 2006.

- [10] B. Horn and B. Schunck. Determining optical flow. Computer Vision, 17:185–203, 1981.
- [11] F. Kschischang, B. Frey, and H. Loeliger. Factor graphs and the sum-product algorithm. *IEEE Transactions on information theory*, 47(2):498–519, 2001.
- [12] M. Leventon, W. Grimson, and O. Faugeras. Statistical shape influence in geodesic active contours. *Biomedical Imaging*, 2002. 5th IEEE EMBS International Summer School on, June 2002.
- [13] S. Mitchell, B. Lelieveldt, R. van der Geest, H. Bosch, J. Reiver, and M. Sonka. Multistage hybrid active appearance model matching: segmentation of left and right ventricles in cardiac mr images. *Medical Imaging, IEEE Transactions on*, 20(5):415–423, May 2001.
- [14] K. Pohl, J. Fisher, J. Levitt, M. Shenton, R. Kikinis, W. Grimson, and W. Wells. A unifying approach to registration, segmentation, and intensity correction. *Lecture Notes in Computer Science*, 3749:310–318, 2005.
- [15] K. Saddi, C. Chefd'hotel, M. Rousson, and F. Cheriet. Region-based segmentation via non-rigid template matching. *Computer Vision, IEEE 11th International Conference on*, pages 1–7, Oct. 2007.
- [16] J. Strikwerda. *Finite difference schemes and partial differential equations*. Society for Industrial Mathematics, 2004.
- [17] A. Tsai, J. Yezzi, A., W. Wells, C. Tempany, D. Tucker, A. Fan, W. Grimson, and A. Willsky. A shape-based approach to the segmentation of medical imagery using level

sets. Medical Imaging, IEEE Transactions on, 22(2):137–154, Feb. 2003.

- [18] G. Unal and G. Slabaugh. Coupled pdes for non-rigid registration and segmentation. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 1:168– 175, June 2005.
- [19] F. Wang and B. Vemuri. Simultaneous registration and segmentation of anatomical structures from brain mri. *Lecture Notes in Computer Science*, 3749:17, 2005.
- [20] 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.
- [21] 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. *Computer Vision and Pattern Recognition, IEEE Conference on*, pages 1–8, June 2008.