

Robust Simultaneous Registration and Segmentation with Sparse Error Reconstruction

Pratim Ghosh, *Student Member, IEEE*, and B.S. Manjunath, *Fellow, IEEE*

Abstract—We introduce a fast and efficient variational framework for Simultaneous Registration and Segmentation (SRS) applicable to a wide variety of image sequences. We demonstrate that a dense correspondence map (between consecutive frames) can be reconstructed correctly even in the presence of partial occlusion, shading, and reflections. The errors are efficiently handled by exploiting their sparse nature. In addition, the segmentation functional is reformulated using a dual Rudin-Osher-Fatemi (ROF) model for fast implementation. Moreover, nonparametric shape prior terms that are suited for this dual-ROF model are proposed. The efficacy of the proposed method is validated with extensive experiments on both indoor, outdoor natural and biological image sequences, demonstrating the higher accuracy and efficiency compared to various state-of-the-art methods.

Index Terms—Segmentation, registration, tracking, optimization

1 INTRODUCTION

TRACKING objects in image sequences is a fundamental problem in computer vision. Error-free tracking is essential for various vision applications such as surveillance, security systems, and medical/biological image analysis. However, data uncertainty presents significant challenges for reliable object tracking. This includes, for example, illumination changes across consecutive frames, rigid/nonrigid deformation of the object-of-interest, partial occlusion, and data corruption.

Numerous frameworks have been proposed in recent years for robust tracking handling the above-mentioned problems. For example, deformable shape models [2], Active Shape Models (ASMs) [3], and active appearance models (AAMs) [4] capture the statistics of shape and appearance variations using training examples. In [5], ASM and AAM are combined in a multiscale fashion to achieve robustness against noise and clutter. Tsai et al. [6] apply Principal Component Analysis (PCA) to model the variability in the training shapes represented using Signed Distance Functions (SDFs). Zhu et al. [7] propose a subject specific dynamic model designed for medical application using multilinear PCA. However, the performance of these methods heavily depends on the training data, which may not encompass all possible scenarios.

Recently, graph-based algorithms have also been proposed for tracking deformable objects. For example, Ishikawa and Jermyn [8] developed a polynomial time algorithm to extract similar objects from a set of time sequence images.

Schoenemann and Cremers [9] proposed a real-time solution for tracking implemented on a GPU based on finding a minimum cost cyclic path in the product space spanned by the template shape and the given image. The cost is derived both from the image data and the shape prior term. In [10], the previous solution was extended to incorporate the edge information, which provides robustness against illumination changes. In another work [11], Schoenemann and Cremers introduced a motion layer decomposition framework which is solved using both discrete and continuous optimization. To be specific, the layer regularization and continuity are imposed using graph cuts (the discrete part) and motion estimation is performed using standard optical flow (the continuous part) method in this framework. The proposed method not only separates various independently moving layers from an input video, but also yields super resolved version of them. The super resolution is achieved using total variation deblurring. Moreover, it is shown to be robust against occlusion. However, it is unclear how this algorithm may be adapted, particularly for tracking objects with known shape statistics. There exist discriminative methods [12], [13] which cast visual tracking as a two-class classification problem. In these methods, the classifiers are updated dynamically to account for appearance changes and partial occlusion of the object-of-interest. Approaches such as [14], [15] select the target as the one which minimizes the projection error in the space spanned by observed (tracked results from previous frames) and trivial templates (with one nonzero element).

All of the above-mentioned methods seem to disregard rich dense flow information between consecutive frames in time sequence images. One way to include this information is to first compute a correspondence map (registration) between successive frames. We can then compute the correlated segmentation of the video frames to solve the tracking problem. It has been recently shown in [16], [17], and [18] that tracking performance can be further improved by solving the registration and the segmentation problem simultaneously. In [19], Ehrhardt et al. propose such a

- P. Ghosh is with Microsoft Corporation, One Microsoft Way, Redmond, WA 98052-6399. E-mail: pratim@ece.ucsb.edu.
- B.S. Manjunath is with the Department of Electrical and Computer Engineering, University of California Santa Barbara, Santa Barbara, CA 93106. E-mail: manj@ece.ucsb.edu.

Manuscript received 6 May 2011; revised 28 Jan. 2012; accepted 18 Apr. 2012; published online 27 Apr. 2012.

Recommended for acceptance by T. Cootes.

For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-2011-05-0296.

Digital Object Identifier no. 10.1109/TPAMI.2012.103.

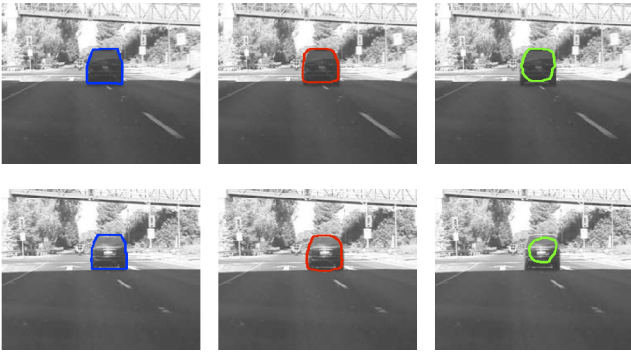


Fig. 1. Examples demonstrating the capability of the proposed method to handle drastic illumination changes using robust correspondence map and sparse error estimation. Two sample frames from an image sequence are shown row-wise. Blue contours show the ground-truth (first column). The obtained results using the proposed method (second column) and the method in [1] (third column) are presented using red and green contours, respectively.

Simultaneous Registration and Segmentation (SRS) in the variational framework. The performance is improved in [1] with the help of a dynamical prior term (SRS+DP). The same authors recently generalized the SRS+DP framework, Generalized SRS (GSRS), which can handle shading, reflections, and illumination changes. In this current work, we extend GSRS [20] using a robust estimation framework. We also propose modifications to the GSRS algorithm for fast and efficient implementation. The effectiveness of the proposed method is illustrated in Fig. 1.

The key contributions of this paper are:

- Design of a fast variational formulation for simultaneous registration and segmentation applicable to time sequence imagery. This consists of two steps:
 - Reconstruction of sparse error (due to partial occlusion, shading, reflections) between consecutive frames using a regularized variant of the L^1 norm.
 - Reformulation of the segmentation functional using the dual Rudin-Osher-Fatemi (ROF) model for fast and efficient implementation.
- The dynamical shape prior term is represented nonparametrically in a way that does not assume any prior distribution of training shapes and can handle an infinite degree of curve variations.
- The new definition allows the incorporation of the *combination term*¹ in an effortless manner with the developed fast segmentation formulation.

This paper is organized as follows: Section 2 explains the new robust estimation framework for simultaneous registration and segmentation. The concepts relating to sparse error reconstruction and the shape-based dual formulation are presented in Sections 2.1.1 and 2.2.1, respectively. Section 3 describes a new combination term which is easily incorporated into the segmentation framework. The experimental results on natural and biological image sequences are provided in Section 4. We discuss future extensions and conclude in Section 5.

1. This is composed of both the segmentation and the registration part.

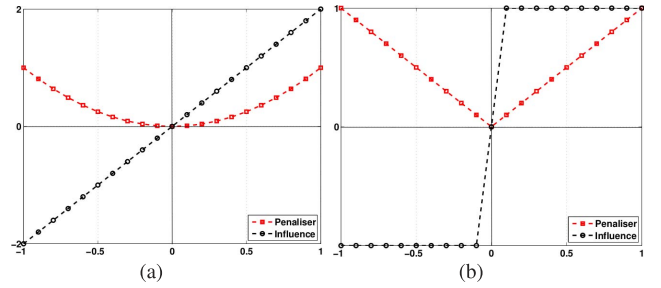


Fig. 2. (a) Quadratic penalizer $\rho = x^2$ and its influence function $\psi = \rho' = 2x$. (b) Regularized L^1 penalizer $\rho = \sqrt{x^2 + \epsilon^2}$, with $\epsilon = 0.001$, and its influence function $\psi = \rho' = \frac{x}{\sqrt{x^2 + \epsilon^2}}$. Note that the influence function saturates very quickly for the latter case.

2 SIMULTANEOUS REGISTRATION AND SEGMENTATION

Consider $I(t) : \Omega(\subseteq \mathbb{R}^2) \rightarrow \mathbb{R}$ and $I(t-1) : \Omega \rightarrow \mathbb{R}$ are the two 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 $\phi^o(\mathbf{x}, t) : \Omega \rightarrow \mathbb{R}$. In this work, we assume that the initial contour of the object of interest is known a priori.

2.1 Registration Model

Given two consecutive image frames, the aim is to find a displacement vector field $\mathbf{u} : \Omega \rightarrow \mathbb{R}^2$, establishing dense correspondence between them. This vector field $\mathbf{u}(\mathbf{x}) = [u(\mathbf{x}), v(\mathbf{x})]$ is obtained historically by maximizing the posterior distribution $\mathcal{P}(\mathbf{u}|I(t), I(t-1))$ (see, for example, [21], [22] and references therein):

$$\begin{aligned} \mathbf{u}^* &= \underset{\mathbf{u}}{\operatorname{argmax}} \mathcal{P}(\mathbf{u}|I(t), I(t-1)) \\ &\propto \underset{\mathbf{u}}{\operatorname{argmax}} \mathcal{P}(I(t), I(t-1)|\mathbf{u})\mathcal{P}(\mathbf{u}). \end{aligned} \quad (1)$$

With certain simplifying assumptions we can obtain the corresponding energy functional of (1) as

$$\begin{aligned} \min_{\mathbf{u}} E_R(\mathbf{u}; I(t-1), I(t)) \\ = \frac{1}{2} \int_{\Omega} [(I(\mathbf{T}(\mathbf{x}), t) - I(\mathbf{x}, t-1))^2 + \alpha_u \operatorname{trace}(\nabla \mathbf{u} \nabla \mathbf{u}^T)] d\mathbf{x}, \end{aligned} \quad (2)$$

which is the classical formulation of Horn and Schunck [23] (HS). Here, $(\cdot)^T$ represents the transpose operation, $\mathbf{T}(\mathbf{x}) = \mathbf{x} + \mathbf{u}(\mathbf{x})$, and α_u controls the smoothness of the derived vector field. The classical HS model (2) uses quadratic penalizers, $\rho(x^2) = x^2$, for both the data and smoothness terms to compute the flow field. The influence function, ψ (i.e., ρ'), for quadratic penalizer is linear and unbounded (see Fig. 2). Thus, the estimation process is severely influenced [21], [24] by the outlying measurements. Also the quadratic penalizer does not favor discontinuities [25] in the computed flow field. Consequently, the focus has moved to designing robust penalizers for data and smoothness terms. See [26] and references therein for detailed explanation.

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

In summary, it has been demonstrated that L^1 minimization [25] is more suitable than the quadratic penalizer for computing the displacement vector fields. Brox et al. [26], Sun et al. [24], and Papenberg et al. [27] have also used a differentiable and regularized variant of L^1 norm $\rho(x^2) = \sqrt{x^2 + \epsilon^2}$ (with $\epsilon = 0.001$), for computational reasons. This is also the most robust convex penalizer known in the optical flow community. Various schemes have been proposed along with the robust penalizer to improve the performance on the Middlebury optical flow benchmark dataset [28]. These include Rudin-Osher-Fatemi structure-texture decomposition [25] of the input sequence, coarse-to-fine estimation of the flow field, median filtering for outlier removal [25], etc. Effort [26] has also been made to incorporate the median filtering heuristic into the optical flow framework in a principled way. The effectiveness of ROF decomposition is elaborated in [25]. The textural part, $I_T(\alpha, \mathbf{x}) = I(\mathbf{x}) - \alpha I_S(\mathbf{x})$, of the image is demonstrated to be invariant to illumination changes, more specifically, robust to shadow and shading reflection artifacts. However, this is a preprocessing step and may not be generalizable enough for handling gross pixel corruptions or occlusions. In this work, we simultaneously estimate both misalignment (dense correspondence map) and corruption in the image sequence. *We assume the corruption due to partial occlusion, shading, reflections, etc., is sparse in nature.*

2.1.1 Simultaneous Modeling of Sparse Corruption and Dense Correspondence Map

Consider $G: \Omega \rightarrow \mathbb{R}$ models the sparse corruptions between successive time frames. We can rewrite (1) by incorporating G as follows:

$$\begin{aligned} \langle \mathbf{u}^*, G^* \rangle &= \arg\max_{\mathbf{u}, G} \mathcal{P}(\mathbf{u}, G | I(t), I(t-1)) \\ &\propto \arg\max_{\mathbf{u}, G} \mathcal{P}(I(t), I(t-1) | \mathbf{u}, G) \mathcal{P}(\mathbf{u}, G) \\ &= \arg\max_{\mathbf{u}, G} \mathcal{P}(I(t), I(t-1) | \mathbf{u}, G) \mathcal{P}(\mathbf{u}) \mathcal{P}(G). \end{aligned} \quad (3)$$

We assume \mathbf{u} and G are independent of each other in the last step. Maximizing the expression in (3) is equivalent to minimizing the following energy functional under certain simplifying assumptions:

$$\begin{aligned} \min_{\mathbf{u}, G} E_{RG}(\mathbf{u}, G; I(t-1), I(t)) \\ = \int_{\Omega} \rho_u(|I(\mathbf{T}(\mathbf{x}), t) - I(\mathbf{x}, t-1) - G|^2) d\mathbf{x} \\ + \alpha_u \int_{\Omega} \rho_u(|\nabla u|^2 + |\nabla v|^2) d\mathbf{x} \\ + \alpha_{g_1} \int_{\Omega} \rho_u(|G|^2) d\mathbf{x} + \alpha_{g_2} \int_{\Omega} \rho_u(|\nabla G|^2) d\mathbf{x}, \end{aligned} \quad (4)$$

where $\rho_u(y^2) = \sqrt{y^2 + \epsilon^2}$, α_{g_1} and α_{g_2} control the sparsity and the smoothness in the estimated corruption. The full derivation of E_{RG} and the related assumptions are presented in Appendix A, which can be found in the Computer Society Digital Library at <http://doi.ieeecomputersociety.org/10.1109/TPAMI.2012.103>. The Euler Lagrange equations that need to be satisfied for \mathbf{u} are

$$\begin{aligned} \overbrace{\rho'_u(I_d^2).I_d I_x - \alpha_u \text{div}(\rho'_u(|\nabla u|^2 + |\nabla v|^2) \nabla u)}^{c_u} &= 0, \\ \overbrace{\rho'_u(I_d^2).I_d I_y - \alpha_u \text{div}(\rho'_u(|\nabla u|^2 + |\nabla v|^2) \nabla v)}^{c_v} &= 0, \end{aligned} \quad (5)$$

where

$$\begin{aligned} I_d &= I(\mathbf{T}(\mathbf{x}), t) - I(\mathbf{x}, t-1) - G, \\ I_x &= \partial_x I(\mathbf{T}(\mathbf{x}), t), \\ I_y &= \partial_y I(\mathbf{T}(\mathbf{x}), t), \end{aligned} \quad (6)$$

Similarly, the Euler Lagrange equation that needs to be satisfied for G is

$$-\rho'_u(I_d^2).I_d + \alpha_{g_1} \rho'_u(|G|^2).G - \alpha_{g_2} \text{div}(\rho'_u(|\nabla G|^2) \nabla G) = 0. \quad (7)$$

The solution of the system of equations in (5) and (7) is challenging since these are highly nonlinear. We adapted the approach from [24], [27] to solve these nonlinear equations. More specifically, \mathbf{u} and G are both solved iteratively in multiple scales using fixed point scheme. We implement a fully implicit scheme for the smoothing terms (e.g., the term associated with α_u and α_{g_2}) and a semi-implicit scheme for the data terms (e.g., the first term in (4) and the term associated with α_{g_1}). The procedure for simultaneous estimation of \mathbf{u} and G is described in Appendix B, which is available in the online supplemental material. The derivative terms, such as I_x and I_y , are computed as the combined version of $\nabla I(\mathbf{T}(\mathbf{x}), t)$ and $\nabla I(\mathbf{x}, t-1)$ for better accuracy. Also the term corresponding to warping, $I(\mathbf{T}(\mathbf{x}), t)$, is obtained using bicubic interpolation. Examples of sparse error reconstruction are demonstrated in Figs. 3 and 4. In the former example, the target image is artificially modified, whereas the error is naturally introduced during imaging in the latter.

2.2 Segmentation Model

The segmentation for the current frame t , based on a dynamic prior term $\hat{\phi}^-(t)$, can be computed by maximizing the conditional distribution $\mathcal{P}(\phi^o(t) | I(t), \hat{\phi}^-(t))$:

$$\begin{aligned} \phi^* &= \arg\max_{\phi^o(t)} \mathcal{P}(\phi^o(t) | I(t), \hat{\phi}^-(t)) \\ &\propto \arg\max_{\phi^o(t)} \mathcal{P}(I(t) | \phi^o(t), \hat{\phi}^-(t)) \mathcal{P}(\phi^o(t) | \hat{\phi}^-(t)) \\ &\propto \arg\max_{\phi^o(t)} \mathcal{P}(I(t) | \phi^o(t)) \mathcal{P}(\hat{\phi}^-(t) | \phi^o(t)) \mathcal{P}(\phi^o(t)). \end{aligned} \quad (8)$$

Here, we assume that $I(t)$ is conditionally independent of $\hat{\phi}^-(t)$ given $\phi^o(t)$. The dynamical prior term [1] maximizes the a posteriori distribution given all the past observations $\phi^o(1), \phi^o(2), \dots, \phi^o(t-1)$. We would like to note that our dynamical prior formulation is motivated by the work of Cremers [29]. However, the approach in [29] does not employ the rich flow information between consecutive image frames for the computation of the prior. If we assume that the tracking process is first order Markov, the dynamical prior term can be efficiently computed using only $\mathbf{u}(\mathbf{x}, t-2, t-1)$ and the previous observation $\phi(\mathbf{x}, t-1)$ as

$$\hat{\phi}^-(\mathbf{x}, t) = \hat{\phi}(\mathbf{x}, t-1) - \mathbf{u}^T(\mathbf{x}, t-2, t-1) \nabla \hat{\phi}(\mathbf{x}, t-1). \quad (9)$$

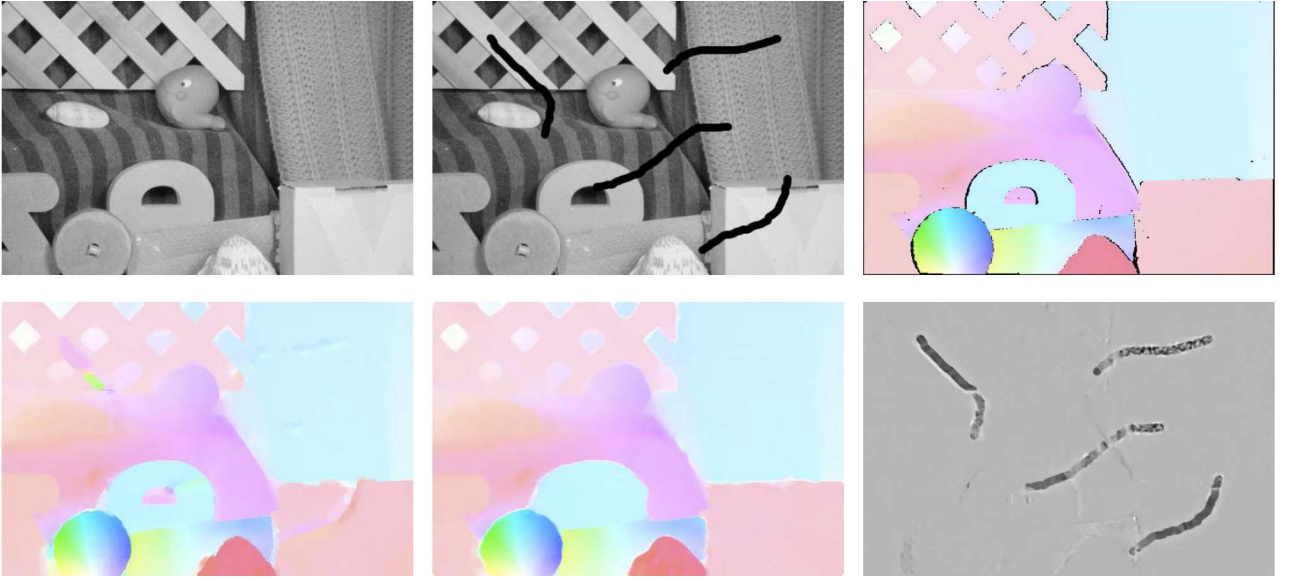


Fig. 3. This figure demonstrates the effectiveness of incorporating the sparse error term into the optical flow formulation. The top row shows the source image, the target image, and the groundtruth flow of the rubber-whale sequence from the Middlebury optical flow benchmark. Note that the target image is modified by thick black lines. The first figure in the second row shows the state-of-the-art flow [26] (with average-end-point-error 0.110). The last two images show the computed flow (with average-end-point-error 0.102), and reconstructed sparse error term $G(x, y)$ using the proposed method. Standard color coding is used to present the optical flow.

In this, the modeling and the observation errors obey the Additive White Gaussian Noise model (AWGN). See [1] for detailed explanation. As shown in [19], [1], and [20], (8) can be implemented using the standard Active Contours Without Edges (ACWEs) [30] model:

$$\begin{aligned} \min_{\phi^o} E_S(\phi^o(t); I(t), \hat{\phi}^-(t)) \\ = \alpha_{s_1} \int_{\Omega} \ln p(I|\theta_2) + H_{\epsilon}(\phi^o(\mathbf{x}, t)) \ln \frac{p(I|\theta_2)}{p(I|\theta_1)} d\mathbf{x} \\ + \alpha_{s_2} \frac{1}{2} \int_{\Omega} |\phi^o(\mathbf{x}, t) - \hat{\phi}^-(\mathbf{x}, t)|^2 d\mathbf{x} \\ + \int_{\Omega} |\nabla H_{\epsilon}(\phi^o(\mathbf{x}, t))| d\mathbf{x}. \end{aligned} \quad (10)$$

Here, α_{s_1} and α_{s_2} are the positive constants, θ_1, θ_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 model in (10) has a number of shortcomings.

- For simplicity we assume $\alpha_{s_2} = 0$ for this discussion. Then the flow minimizing energy (i.e., standard ACWE flow) corresponding to (10) is

$$\partial_{\tau} \phi^o = H'_{\epsilon}(\phi^o) \left(\alpha_{s_1} \ln \frac{p(I|\theta_1)}{p(I|\theta_2)} + \nabla \cdot \left(\frac{\nabla \phi^o}{|\nabla \phi^o|} \right) \right),$$

where τ represents the evolution of the respective function. Here, H_{ϵ} is chosen to be noncompactly supported and a smoothed version of the Heaviside function. The steady state solution of the above flow is thus equivalent to

$$\partial_{\tau} \phi^o = \left(\alpha_{s_1} \ln \frac{p(I|\theta_1)}{p(I|\theta_2)} + \nabla \cdot \left(\frac{\nabla \phi^o}{|\nabla \phi^o|} \right) \right),$$

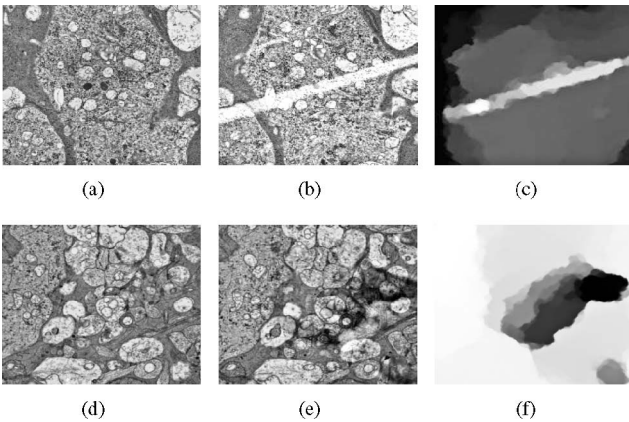


Fig. 4. This figure demonstrates more examples of sparse error reconstruction. (a), (b) and (d), (e) are two different consecutive z-slices from a Transmission Electron Micrograph stack. Errors are naturally introduced due to slice damage while imaging. The corresponding sparse errors are shown in (c) (black pixels ≈ 0) and (f) (white pixels ≈ 0), respectively. TEM data is courtesy of the Marc Laboratory, Moran Eye Center, University of Utah.

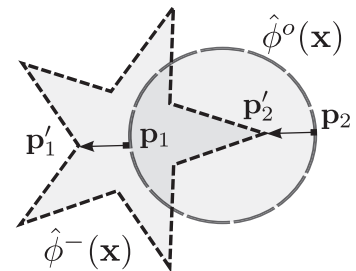


Fig. 5. Explanation of the distance measure defined in (12). Here the star and the circular shapes represent the zero level-sets of the dynamic prior term and the current contour position, respectively. The function evaluates the sum of squared euclidean distances of the points on $\phi^o(\mathbf{x})$ to the closest point on $\hat{\phi}^-(\mathbf{x})$. For example, the distances between $p_1 \rightarrow p'_1$ and $p_2 \rightarrow p'_2$ are computed in this process.

TABLE 1
The Overall Performance of Various Approaches

Approaches	Measure	Car1	Car2	Person1	Person2	Person3	Face1	Face2
MIL	f_m	0.53	0.76	0.18	0.16	0.56	0.86	0.86
	p_c	0.3	0.8	0.2	0.0	0.4	1.0	0.9
ET	f_m	0.25	0.12	0.12	0.19	0.62	0.77	0.64
	p_c	0.0	0.0	0.1	0.0	0.5	0.9	0.4
IVT	f_m	0.77	0.64	0.14	0.17	0.62	0.91	0.87
	p_c	1.0	0.6	0.2	0.1	0.5	1.0	1.0
MS	f_m	0.30	0.72	0.13	0.19	0.23	0.58	0.62
	p_c	0.2	0.7	0.1	0.7	0.1	0.4	0.4
RSRS	f_m	0.91	0.86	0.86	0.84	0.82	0.84	0.82
	p_c	1.0	1.0	1.0	1.0	1.0	1.0	0.9
RSRS\G	f_m	0.72	0.65	0.78	0.50	0.60	0.70	0.72
	p_c	0.8	0.6	0.9	0.5	0.7	0.6	0.6
SRS+DP	f_m	0.21	0.12	0.17	0.28	0.23	0.29	0.37
	p_c	0.1	0.1	0.1	0.2	0.1	0.1	0.1
SRS	f_m	0.2	0.12	0.16	0.25	0.14	0.29	0.3
	p_c	0.1	0.1	0.1	0.1	0.0	0.1	0.1

The F-measure (f_m) and the PASCAL-criterion (p_c) are evaluated over 150, 100, 230, 130, 500, 400, and 400 frames for the Car1, Car2, Person1, Person2, Person3, Face1, and Face2 sequences, respectively.

which corresponds to the following energy:

$$E = \int_{\Omega} \alpha_{s_1} \ln \left(\frac{p(I|\theta_2)}{p(I|\theta_1)} \right) \phi^o d\mathbf{x} + \int_{\Omega} |\nabla \phi^o| d\mathbf{x}. \quad (11)$$

This functional (as well as the standard ACWE model) does not have a minimizer [31] since it is homogeneous of degree one in ϕ^o .

- Next, the shape prior term associated with α_{s_2} (see (10)) compares two shapes (the prior and the current shape) based on their signed distance representation. The dissimilarity measured in this way strongly depends on the choice of domain Ω [32]. To this end, the dissimilarity between same two shapes changes by simply extending the domain of integration.

In this paper, we propose a formulation which addresses the above-mentioned drawbacks.

2.2.1 A New Shape-Based Dual Formulation of the Segmentation Functional

The formulation is developed based on [33], where Bresson et al. proposed a fast numerical scheme to globally minimize a standard variational model for segmentation. The scheme is based on the dual formulation of the Total Variation (TV) norm. To the best of our knowledge, this is the first attempt to include the shape prior term with such a dual formulation.

Let us consider a dynamic shape prior term, $d_p^2 : \Omega \rightarrow \mathbf{R}^+$, as

$$d_p^2(\phi^o(\mathbf{x}); \hat{\phi}^-(\mathbf{x})) = \int_{\Omega} \rho_p(|\hat{\phi}^-(\mathbf{x})|^2) |\nabla H(\phi^o)| d\mathbf{x}, \quad (12)$$

where $\rho_p(\cdot)$ is a robust estimator. Here, we assume $\rho_p(y^2) = y^2$. Thus, the model in (12) integrates the square of the signed distance representation of the dynamical prior term along the zero level-set of the current contour position. Consider Fig. 5 for an explanation. Alternatively, the distance computed in (12) is equal to $|\hat{\phi}^-(\mathbf{p}_i') - \phi^o(\mathbf{p}_i)|^2$ in a discrete setting. A similar distance measure is employed in [34]. However, we would like to note that Bresson et al. [34] use a principal component analysis-based parametric shape representation for their formulation. The parameters

are the mean shape and a few principal components of a population of training shapes represented as signed distance functions. As pointed out in [35], this formulation has a number of limitations. First, the shape distribution is assumed to be Gaussian, which is generally not true for complex shape variations. Second, a linear combination of SDFs (for example, the mean shape) does not produce an SDF in general. In contrast, we use a nonparametric model (see (12)) for shape representation. This representation trivially avoids the above-mentioned drawbacks of the parametric shape representation. It can also handle infinite degree curve variations as opposed to parametric models, which lead to better performance. Moreover, the objective function in [34] is optimized using the standard gradient descent procedure which is generally slow [33] in nature.

We now explain a fast numerical scheme based on dual formulation of the TV-norm incorporating the shape prior term.

Consider the following constrained minimization problem:

$$\begin{aligned} \min_{0 \leq \psi^o \leq 1} E_{DS}(\psi^o(\mathbf{x}); I(t), d_p^2) \\ = \underbrace{\int_{\Omega} d_p^2 |\nabla \psi^o(\mathbf{x})| d\mathbf{x}}_{\text{TV}_{d_p^2}(\psi^o)} + \alpha_{s_1} \underbrace{\int_{\Omega} \ln \left(\frac{p(I|\theta_2)}{p(I|\theta_1)} \right) \psi^o(\mathbf{x}) d\mathbf{x}}_{\eta(\mathbf{x}, \theta_1, \theta_2)}. \end{aligned} \quad (13)$$

The above expression is similar³ to (11) (assuming $d_p^2 \equiv 1, \forall \mathbf{x}$). Here the solution is constrained to lie between 0 and 1 since the unconstrained problem does not have a stationary solution as mentioned before. The first term in (13) measures the weighted total variation norm of the function ψ^o . It can be rigorously shown that (see [31], [33] and references therein) (13) is equivalent to solving the following unconstrained problem:

$$\begin{aligned} \min_{\psi^o} E_{DS}(\psi^o(\mathbf{x}); I(t), d_p^2) \\ = \text{TV}_{d_p^2}(\psi^o) + \int_{\Omega} [\alpha_{s_1} \eta(\mathbf{x}, \theta_1, \theta_2) \psi^o + \hat{\alpha} \zeta(\psi^o)] d\mathbf{x}, \end{aligned} \quad (14)$$

3. Please note that ϕ^o and ψ^o represent the same observed contour in different ways.

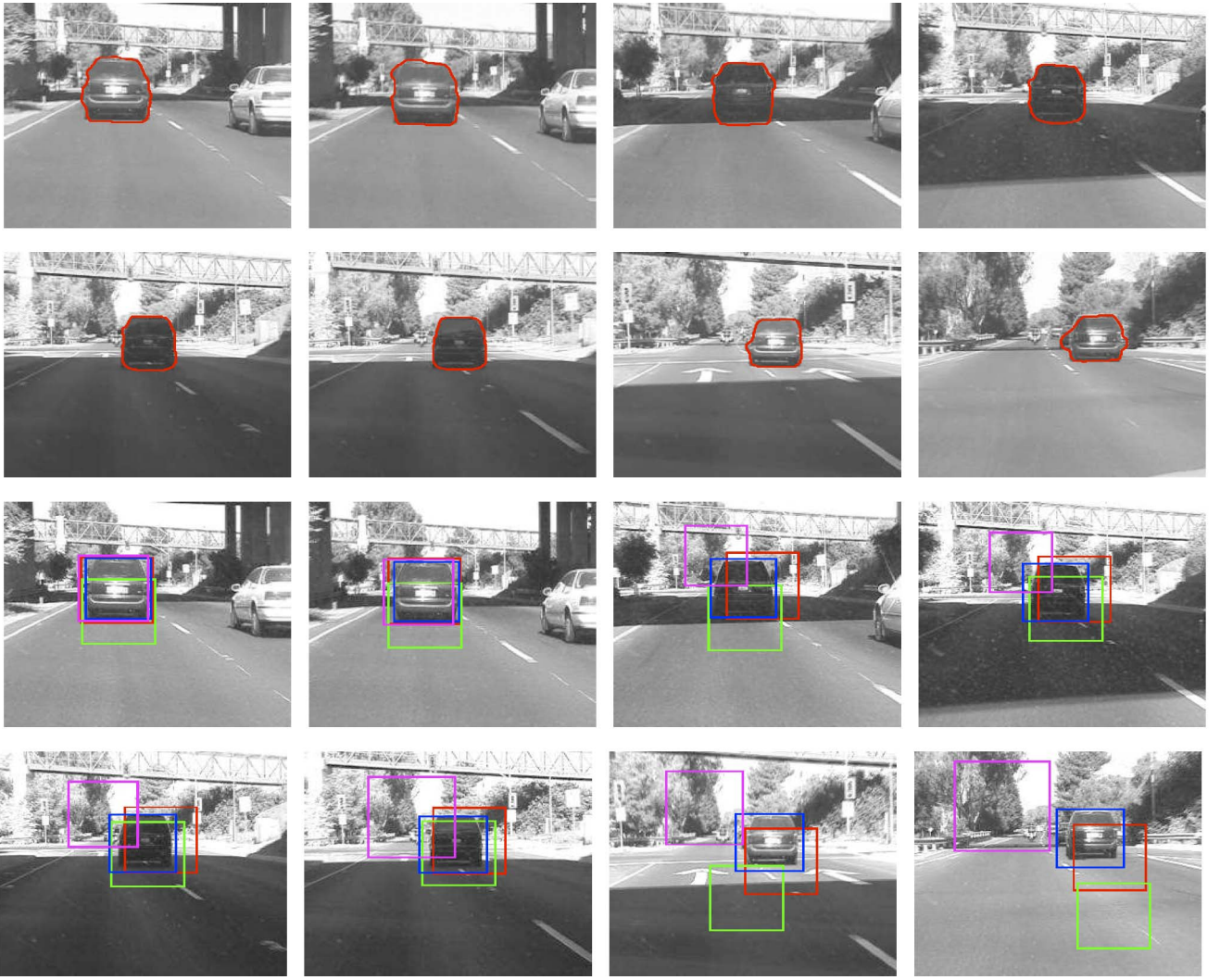


Fig. 6. Sample visual results from the *Car1* sequence. The first two rows present the results (in red contour) for the proposed tracker on eight sample frames. The last two rows show the results corresponding to various other trackers (*MIL* in red, *ET* in green, *IVT* in blue, and *MS* in pink contour) for the same sample frames.

where $\zeta(y) := \max\{0, 2|y - \frac{1}{2}| - 1\}$ and $\hat{\alpha} > \frac{\alpha_{s1}}{2} \|\eta(\mathbf{x})\|_{L^\infty(\Omega)}$. We can also rewrite $\zeta(y)$ as

$$\zeta(y) = \begin{cases} -2y, & \text{if } y < 0, \\ 2(y - 1), & \text{if } y > 1, \\ 0, & \text{otherwise.} \end{cases}$$

As can be seen, the two energies in (13) and (14) agree for $\{\psi^o \in L^\infty(\Omega) : 0 \leq \psi^o(\mathbf{x}) \leq 1 \forall \mathbf{x}\}$. It is thus sufficient to prove that any minimizer of (14) satisfies the constraint $0 \leq \psi^o \leq 1$. A proof is provided in Appendix C, which is available in the online supplemental material. The variational formulation (14) can now be solved using a convex regularity [36] term:

$$\begin{aligned} \min_{\psi^o, \varphi} E_{DS}(\psi^o, \varphi; I(t), d_p^2) \\ = \text{TV}_{d_p^2}(\psi^o) + \frac{1}{2\beta} \|\psi^o - \varphi\|^2 \\ + \int_{\Omega} [\alpha_{s1} \eta(\mathbf{x}, \theta_1, \theta_2) \varphi + \alpha \zeta(\varphi)] d\mathbf{x}, \end{aligned} \quad (15)$$

where $\beta > 0$ is a small constant so that φ is a close approximation of ψ^o . Note that (15) is a convex minimization problem which can be solved by iteratively updating ψ^o and φ in an alternative manner:

1. Solve for ψ^o with fixed φ :

$$\min_{\psi^o} \left\{ \text{TV}_{d_p^2}(\psi^o) + \frac{1}{2\beta} \|\psi^o - \varphi\|^2 \right\}. \quad (16)$$

2. Then solve for φ with fixed ψ^o :

$$\begin{aligned} \min_{\varphi} \left\{ \frac{1}{2\beta} \|\psi^o - \varphi\|^2 \right. \\ \left. + \int_{\Omega} [\alpha_{s1} \eta(\mathbf{x}, \theta_1, \theta_2) \varphi + \alpha \zeta(\varphi)] d\mathbf{x} \right\}. \end{aligned} \quad (17)$$

The functional in (16) is the well-known total variation-based image denoising model of Rudin, Osher, and Fatemi. This is subsequently solved using the dual model of the functional. The solution of (17) is simply computed pointwise since it is independent of any spatial derivatives of the

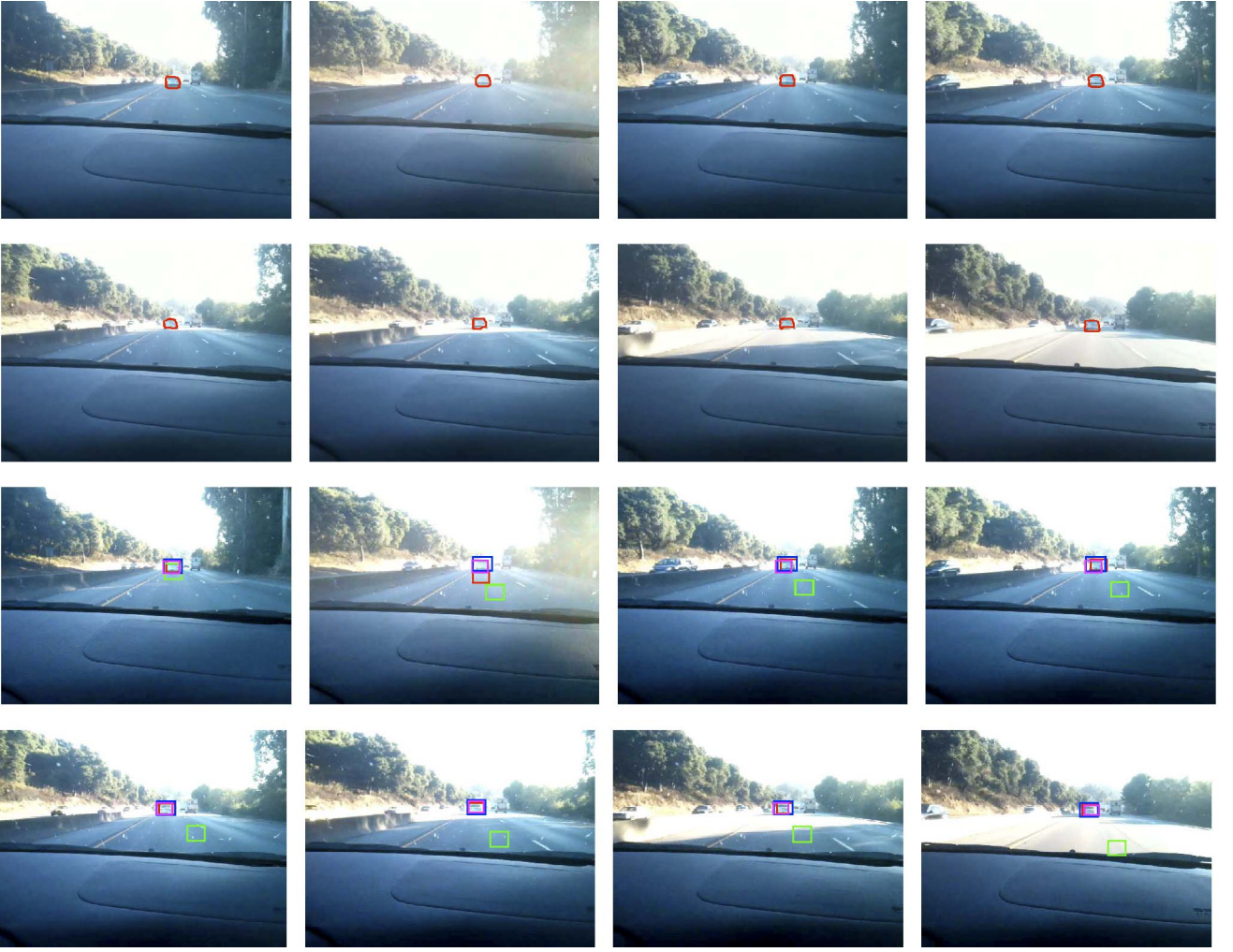


Fig. 7. Sample visual results from the *Car2* sequence. The first two rows present the results (in red contour) for the proposed tracker on eight sample frames. The last two rows show the results corresponding to various other trackers (*MIL* in red, *ET* in green, *IVT* in blue, and *MS* in pink contour) for the same sample frames.

evaluated function. Refer to Appendix D, which is available in the online supplemental material, for the update rules.

3 COMBINING REGISTRATION AND SEGMENTATION FUNCTIONAL

A commonly used objective function [19], [1], [20] in this regard is the one which establishes the correspondence between the target ($\phi^o(t)$) and the reference ($\phi^o(t-1)$) level-set functions using a mapping function $\mathbf{u}(\mathbf{x}, t-1, t)$. Mathematically we can write

$$\begin{aligned} \min_{\mathbf{u}, \phi^o} E_C(\mathbf{u}, \phi^o(t); \phi^o(t-1)) \\ = \frac{1}{2} \int_{\Omega} N_{\epsilon'}(\mathbf{x}) (\phi^o(\mathbf{x}, t) - \phi^o(\mathbf{T}(\mathbf{x}), t-1))^2 d\mathbf{x}, \end{aligned} \quad (18)$$

where

$$N_{\epsilon'}(\mathbf{x}) = \begin{cases} 1, & \text{if } |\phi^o(\mathbf{x}, t)| < \epsilon' \vee |\phi^o(\mathbf{x}, t-1)| < \epsilon'; \\ 0, & \text{otherwise.} \end{cases}$$

The functional E_C tries to register isocontours within a narrow band of the zero level sets of the source and target

shapes. The width of the narrow band is chosen by the parameter ϵ' . The choice of ϵ' is crucial since it determines the scale of objects to be registered. See [37] for a detailed discussion. In contrast, we propose a functional which does not require the scale information a priori. The new combination term is proposed (similarly to (12)) as

$$\begin{aligned} \min_{\mathbf{u}, \phi^o(t)} E_{NC}(\mathbf{u}, \phi^o(t); \phi^o(t-1)) &= d_c^2 \\ &= \int_{\Omega} \rho_c(|\phi^o(\mathbf{T}(\mathbf{x}), t-1)|^2) |\nabla H(\phi^o(\mathbf{x}, t))| d\mathbf{x}, \end{aligned} \quad (19)$$

where $\rho_c(y^2) = y^2$.

To summarize, the proposed method simultaneously estimates \mathbf{u} , G , and $\phi^o(t)$ for the current frame at time t in an alternate fashion. The combination term, see (19), modifies the update rule for the estimated functions. Recalling (5), the new Euler Lagrange equations that need to be satisfied for \mathbf{u} can be written as

$$\begin{aligned} \rho'_c(\phi_w^2) \phi_w \phi_x |\nabla H(\phi^o(t))| + \mathcal{C}_u &= 0, \\ \rho'_c(\phi_w^2) \phi_w \phi_y |\nabla H(\phi^o(t))| + \mathcal{C}_v &= 0, \end{aligned} \quad (20)$$



Fig. 8. Sample visual results from the *Person1* sequence. The first two rows present the results (in red contour) for the proposed tracker on eight sample frames. The last two rows show the results corresponding to various other trackers (*MIL* in red, *ET* in green, *IVT* in blue, and *MS* in pink contour) for the same sample frames.

where

$$\phi_w = \phi^o(\mathbf{T}(\mathbf{x}), t-1), \quad \phi_x = \partial_x \phi_w, \quad \phi_y = \partial_y \phi_w. \quad (21)$$

The update rule for the sparse error, $G(x, y)$, remains the same since it is independent of the combination term E_{NC} . Due to the combination term, there is also a minor modification to the segmentation problem in (16):

$$\min_{\psi^o} \left\{ \text{TV}_{d_p+d_e}(\psi^o) + \frac{1}{2\beta} \|\psi^o - \varphi\|^2 \right\}. \quad (22)$$

4 EXPERIMENTAL RESULTS

Experiments on natural image sequences. We present results on various indoor and outdoor sequences. We use seven different image sequences, namely, *Car1*,⁴ *Car2*, *Person1*, *Person2*,⁵ *Person3*,⁶ *Face1*, and *Face2*,⁷ to evaluate

4. This sequence is downloaded from <http://www.cs.toronto.edu/~dross/ivt/>.

5. *Person1* and *Person2* are both downloaded from <http://groups.inf.ed.ac.uk/vision/CAVIAR/CAVIARDATA1/>.

6. http://scallops.ece.ucsb.edu/mediawiki/index.php/Main_Page.

7. Both *Face1* and *Face2* sequences can be found in http://vision.ucsd.edu/~bbabenko/project_miltrack.shtml.

the performance of the proposed tracking framework. We compare our results with various state-of-the-art trackers, for example, Simultaneous Registration and Segmentation tracking [19] and its variant (*SRS+DP*) [1], Mean Shift (*MS*) tracking [38], Multiple Instance Learning (*MIL*) tracking [13], Incremental Visual (*IVT*) tracking [39], Ensemble tracking (*ET*) [12]. *Car1*, *Face1*, and *Face2* are the gray scale image sequences and the rest of them are composed of 24-bit color images. We also demonstrate the performance of the proposed tracker without robustly estimating \mathbf{u} and G . We refer to the proposed tracker and its variant as Robust-SRS (*RSRS*) and *RSRS*\G in the subsequent discussion, respectively. The overall performance is demonstrated in Table 1. We compute the F-measure, f_m , and the PASCAL-criterion, p_c (with 50 percent overlap), to capture the accuracy of different approaches. f_m is defined as

$$f_m = \frac{2PR}{P+R}, \quad \text{where } P = \frac{|\Omega_g \cap \Omega_a|}{\Omega_a}, \quad \text{and } R = \frac{|\Omega_g \cap \Omega_a|}{\Omega_g}$$

where $\Omega_g (\subseteq \mathbb{R}^2)$ and $\Omega_a (\subseteq \mathbb{R}^2)$ are the two regions describing the object-of-interest in *groundtruth* and *automatic* segmentation in the same time frame and $|\cdot|$ represents the cardinality of a given set. Similarly, p_c is defined as

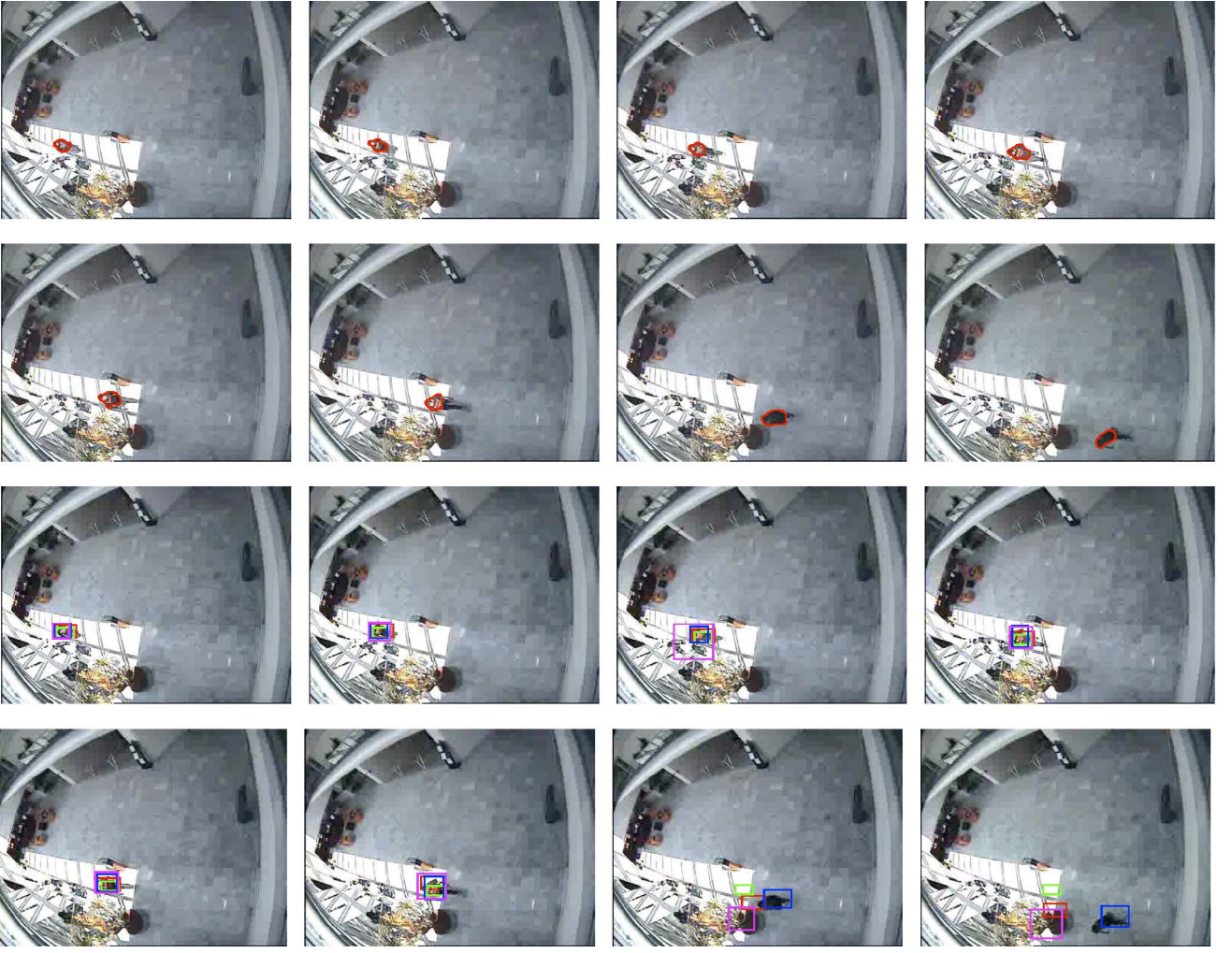


Fig. 9. Sample visual results from the *Person2* sequence. The first two rows present the results (in red contour) for the proposed tracker on eight sample frames. The last two rows show the results corresponding to various other trackers (*MIL* in red, *ET* in green, *IVT* in blue, and *MS* in pink contour) for the same sample frames.

$$p_c(x) = \begin{cases} 1, & \text{if } x \geq 0.5; \\ 0, & \text{otherwise.} \end{cases} \quad \text{where } x = \frac{|\Omega_o \cap \Omega_a|}{|\Omega_o \cup \Omega_a|},$$

The object-of-interest in the *Car1* sequence undergoes pronounced nonuniform illumination changes as it passes under a bridge. Some sample tracking results are presented in Fig. 6. Most of the state-of-the-art trackers, except *IVT*, lose the target fully or partially due to illumination changes. Similarly, in *Car2* (see Fig. 7), the vehicle goes through several shadow and bright regions. The overall performance (other than *ET*) is better for this sequence than the previous one, the reason being that the relative object movement with respect to the camera is small compared to *Car1* sequence.

In the *Person1* sequence (see Fig. 8), the tracked object gets occluded partially by a similar looking object (both man and woman are wearing black coats). It can be seen that all state-of-the-art trackers wrongly latch onto the occluding object (the man in this sequence) in the rest of the frames. Rich motion information helps the proposed tracker to attach to the tracked object until the end of sequence. The result corresponding to the *Person2* sequence is presented in Fig. 9. In this sequence, most of the state-of-the-art trackers

lose the object when it switches from bright to a darker region. The proposed tracker again performs satisfactorily on this sequence.

The fifth sequence, *Person3* (see Fig. 10), is an indoor sequence. The object is walking in the corridor passing through several bright and dark regions. There is severe reflection from the floor and the wall. It is also to be noted that the object undergoes large scale changes as it goes away from the camera. *MS* and *IVT* produce undesired results for this sequence. Although *MIL* and *ET* can track the object, they do not locate it well due to drastic scale changes. The proposed approach tracks the target very well throughout the whole sequence, even in the presence of illumination and scale changes.

In the *Face1* and *Face2* image sequences, the objective is to track and segment frontal human face over time. Some sample visual results for these two sequences are presented in Figs. 11 and 12, respectively. The main challenge here is to cope with the severe occlusions and minor illumination changes. We would like to note that the shape deformation or the relative displacement with respect to the camera of the faces is negligible compared to other image sequences. To summarize, the proposed method is seen to perform reasonably well even in the presence of severe occlusions.

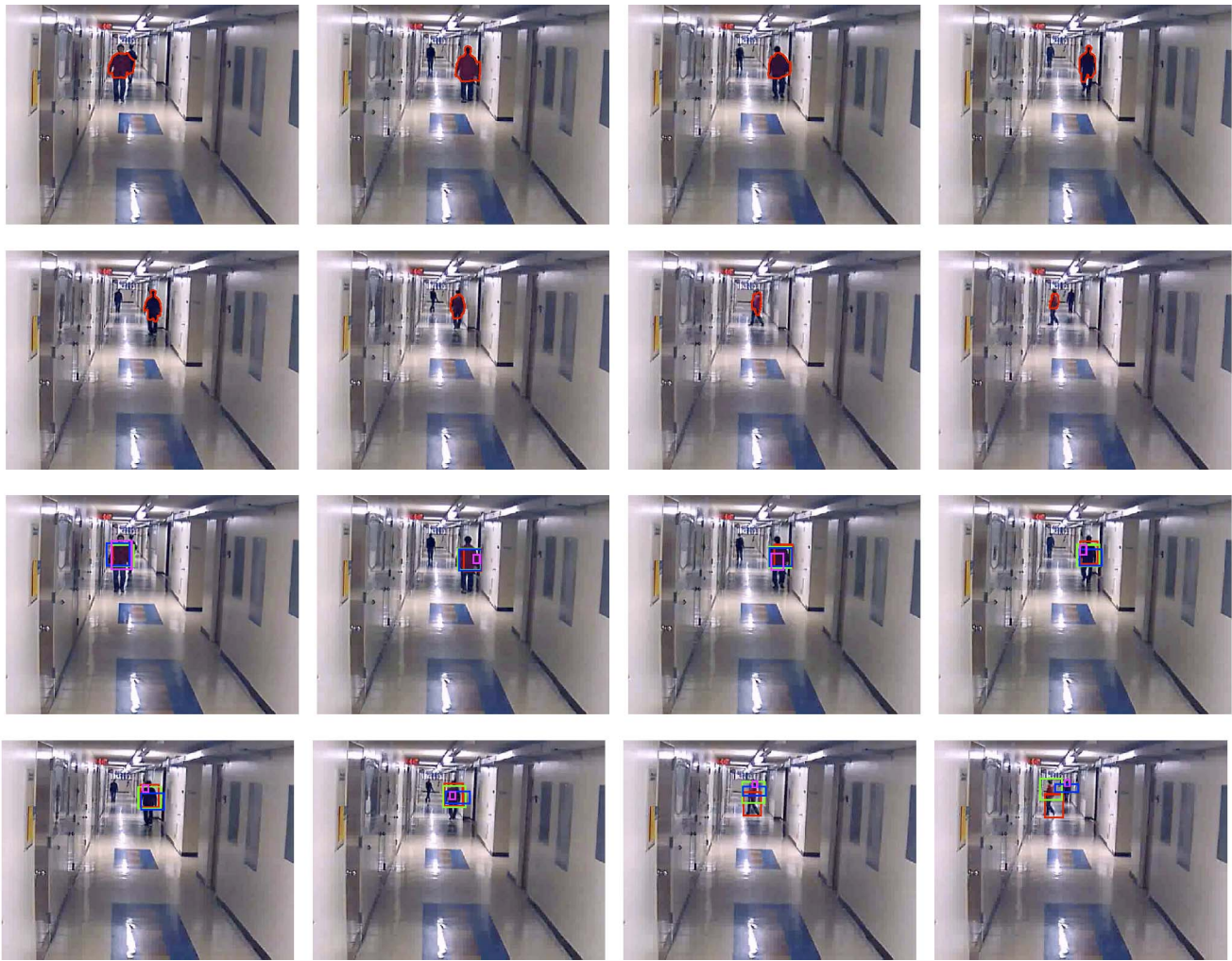


Fig. 10. Sample visual results from the *Person3* sequence. The first two rows present the results (in red contour) for the proposed tracker on eight sample frames. The last two rows show the results corresponding to various other trackers (*MIL* in red, *ET* in green, *IVT* in blue, and *MS* in pink contour) for the same sample frames.

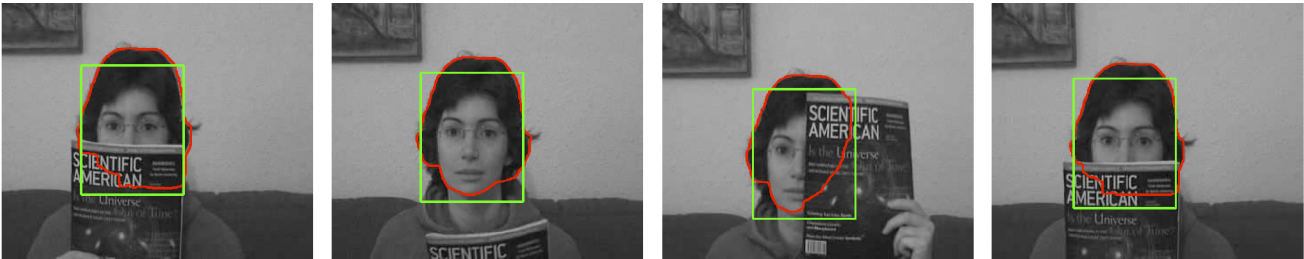


Fig. 11. Sample tracking results from the *Face1* sequence. The ground truth and the proposed tracking output (*RSRS*) are shown using green and red color contours, respectively.

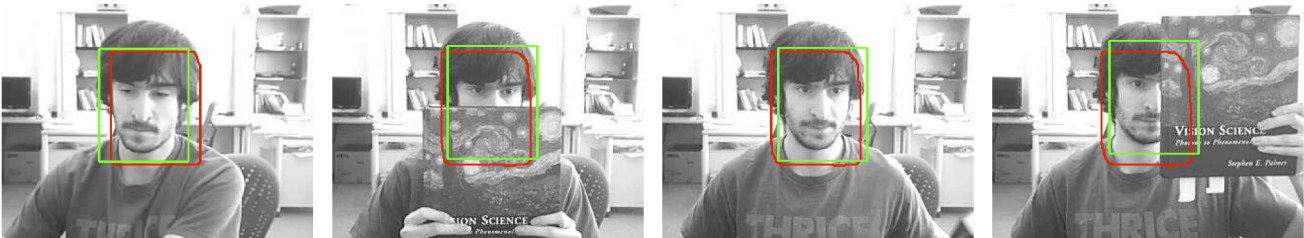


Fig. 12. Sample tracking results from the *Face2* sequence. The ground truth and the proposed tracking output (*RSRS*) are shown using green and red color contours, respectively.

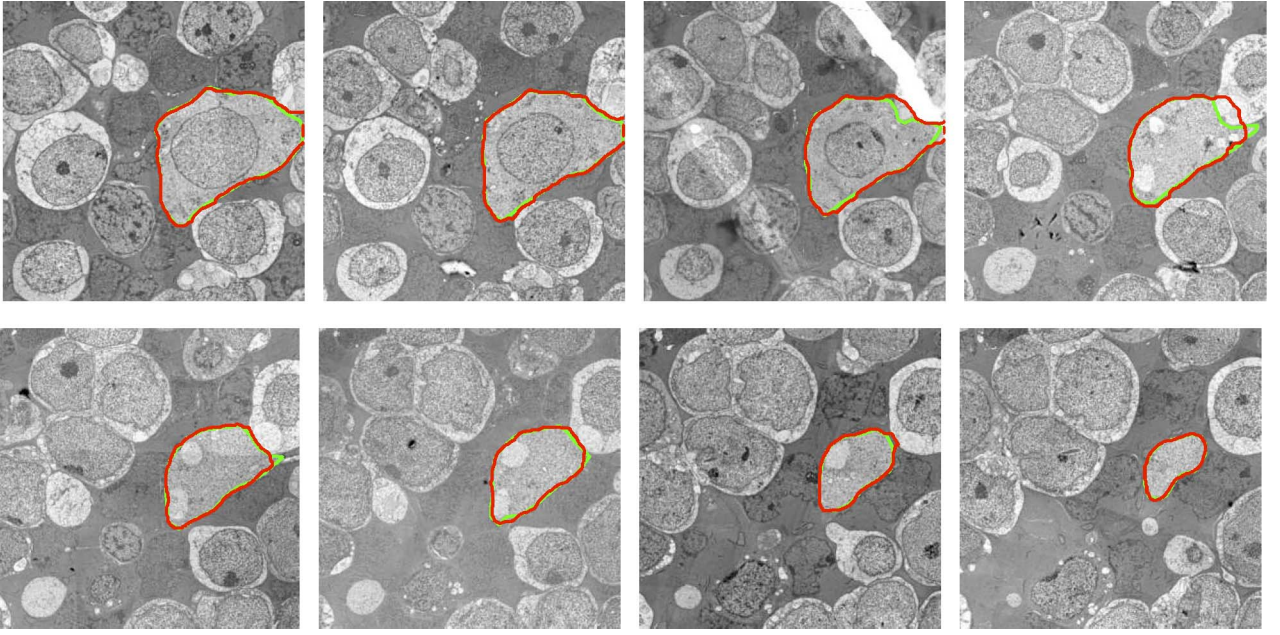


Fig. 13. Sample tracking results on electron micrograph z -slices. The groundtruth and obtained results are presented using green and red contours, respectively.

Experiments on transmission electron micrograph (TEM) z -slices. We also study the effectiveness of the proposed approach on TEM z -slices. Note that the tracking is performed over depth (along z direction) for this dataset. In this experiment, a neuronal structure is tracked over 70 z -slices. See Fig. 13 for the results. This dataset is extremely challenging due to noisy gradients and poor data terms. Also, some of the intermediate frames are missing because of complete data corruption. Note that the proposed approach tracks the structure very well throughout even in the presence of large nonlinear deformation.

Parameter setting. We keep all the parameters approximately constant throughout our experiments. We have three different coefficients from E_{RG} (see (4)), e.g., α_u , α_{g_1} , and α_{g_2} are empirically set to 8×10^{-4} , 9×10^{-5} , and 3×10^{-3} , respectively. The weight for the dynamical prior term d_p^2 and the combination term d_c^2 are both approximately set to 2×10^{-5} . The coefficient α_{s_1} corresponding to (15) is selected randomly from the range $[0.02, 0.05]$.

We first extract a simple gray scale/RGB-color histogram (for natural images) or texture information (for TEM z -slices) of the foreground and background from a few frames. These are then used to obtain the likelihood values $\eta(\mathbf{x})$ for the rest of the frames. In the current implementation, we also use the edge information of the image. The edge information is easily incorporated by minimizing $\text{TV}_{g+d_p^2+d_c^2}$ instead of $\text{TV}_{d_p^2+d_c^2}$ in (22). Here, $g = \frac{1}{1+\gamma|\nabla I_\sigma|^2}$ is the inverse of the smoothed gradient magnitude function. In our experiments, we keep $\gamma = 600$. The proposed tracking framework is implemented in MATLAB. In our implementation we update \mathbf{u} , ϕ^o , and G twice for each time frame. It takes about 3 seconds for processing each frame of size 200×200 on a 6 GB, 2.67 GHz Intel Xeon CPU. Note that the majority of the processing time is spent computing the updates for \mathbf{u} and G . In our current

approach, ϕ^o is computed much faster (≈ 10 times) than in [1] and [20].

5 CONCLUSION

In this paper, we propose a robust variational framework for level-set-based simultaneous registration and segmentation applied to natural and biological image sequences. The developed formulation is demonstrated to be robust against partial occlusion, drastic illumination changes, and data corruption. Various deformations of the object of interest are seamlessly tackled by robustly computing a dense correspondence map between consecutive frames. For this, we simultaneously estimate a sparse error between successive frames. Moreover, we introduce a fast and efficient shape prior-based dual ROF model for the segmentation counterpart.

We are currently studying the effect of imperfect initializations on tracking performance. Also, since the motion and the sparse error estimation procedure is computationally demanding, we would like to explore incremental estimation, which can further increase the speed of the proposed approach. In this work, we address tracking and segmentation of a single object. Future work also includes extension to simultaneous tracking of multiple objects or structures.

ACKNOWLEDGMENTS

This work is supported by the Center for Bioimage Informatics under grants NSF-III 0808772 and NSF-OIA 0941717.

REFERENCES

- [1] P. Ghosh, M.E. Sargin, and B.S. Manjunath, "Robust Dynamical Model for Simultaneous Registration and Segmentation in a Variational Framework: A Bayesian Approach," *Proc. 12th IEEE Int'l Conf. Computer Vision*, Oct. 2009.

- [2] M. Leventon, W. Grimson, and O. Faugeras, "Statistical Shape Influence in Geodesic Active Contours," *Proc. Fifth IEEE EMBS Int'l Summer School on Biomedical Imaging*, June 2002.
- [3] T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham, "Active Shape Models: Their Training and Application," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38-59, Jan. 1995.
- [4] T. Cootes et al., "Active Appearance Models," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681-685, June 2001.
- [5] S.C. Mitchell et al., "Multistage Hybrid Active Appearance Model Matching: Segmentation of Left and Right Ventricles in Cardiac MR Images," *IEEE Trans. Medical Imaging*, vol. 20, no. 5, pp. 415-423, May 2001.
- [6] A. Tsai et al., "A Shape-Based Approach to the Segmentation of Medical Imagery Using Level Sets," *IEEE Trans. Medical Imaging*, vol. 22, no. 2, pp. 137-154, Feb. 2003.
- [7] 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," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1-8, June 2008.
- [8] H. Ishikawa and I. Jermy, "Region Extraction from Multiple Images," *Proc. Eighth IEEE Int'l Conf. Computer Vision*, pp. 509-516, 2001.
- [9] T. Schoenemann and D. Cremers, "Globally Optimal Shape-Based Tracking in Real-Time," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, June 2008.
- [10] T. Schoenemann and D. Cremers, "A Combinatorial Solution for Model-Based Image Segmentation and Real-Time Tracking," *IEEE Trans. Pattern Analysis Machine Intelligence*, vol. 32, no. 7, pp. 1153-1164, July 2010.
- [11] T. Schoenemann and D. Cremers, "High Resolution Motion Layer Decomposition Using Dual-Space Graph Cuts," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1-7, 2008.
- [12] S. Avidan, "Ensemble Tracking," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 2005.
- [13] B. Babenko, M. Yang, and S. Belongie, "Visual Tracking with Online Multiple Instance Learning," *IEEE Pattern Analysis and Machine Intelligence*, vol. 32, no. 7, pp. 1153-1164, July 2009.
- [14] X. Mei and H. Ling, "Robust Visual Tracking Using L1 Minimization," *Proc. IEEE Int'l Conf. Computer Vision*, pp. 1-8, 2009.
- [15] X. Mei, H. Ling, and Y. Wu, "Minimum Error Bounded Efficient L1 Tracker with Occlusion Detection," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2011.
- [16] 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," *Medical Image Computing and Computer-Assisted Intervention*, vol. 3749, pp. 286-293, 2005.
- [17] G. Unal and G. Slabaugh, "Coupled PDES for Non-Rigid Registration and Segmentation," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 168-175, June 2005.
- [18] K. Saddi, C. Chef'd'hotel, M. Rousson, and F. Chriet, "Region-Based Segmentation via Non-Rigid Template Matching," *Proc. IEEE Int'l Conf. Computer Vision*, pp. 1-7, Oct. 2007.
- [19] J. Ehrhardt, A. Richberg, and H. Handels, "A Variational Approach for Combined Segmentation and Estimation of Respiratory Motion in Temporal Image Sequences," *Proc. IEEE Int'l Conf. Computer Vision*, pp. 1-7, Oct. 2007.
- [20] P. Ghosh, M.E. Sargin, and B.S. Manjunath, "Generalized Simultaneous Registration and Segmentation," *Proc. IEEE Int'l Conf. Computer Vision*, June 2010.
- [21] M. Black and P. Anandan, "The Robust Estimation of Multiple Motions: Parametric and Piecewise-Smooth Flow Fields," *Computer Vision and Image Understanding*, vol. 63, no. 1, pp. 75-104, 1996.
- [22] D. Cremers and S. Soatto, "Motion Competition: A Variational Approach to Piecewise Parametric Motion Segmentation," *Int'l J. Computer Vision*, vol. 62, no. 3, pp. 249-265, 2005.
- [23] B. Horn and B. Schunck, "Determining Optical Flow," *Computer Vision*, vol. 17, pp. 185-203, 1981.
- [24] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High Accuracy Optical Flow Estimation Based on a Theory for Warping," *Proc. European Conf. Computer Vision*, pp. 25-36, 2004.
- [25] A. Wedel, T. Pock, C. Zach, H. Bischof, and D. Cremers, "An Improved Algorithm for TV-L1 Optical Flow," *Statistical and Geometrical Approaches to Visual Motion Analysis*, vol. 5064, pp. 23-45, 2009.
- [26] D. Sun, S. Roth, T. Darmstadt, and M. Black, "Secrets of Optical Flow Estimation and Their Principles," *Proc. IEEE Int'l Conf. Computer Vision*, June 2010.
- [27] N. Papenberg, A. Bruhn, T. Brox, S. Didas, and J. Weickert, "Highly Accurate Optic Flow Computation with Theoretically Justified Warping," *Int'l J. Computer Vision*, vol. 67, no. 2, pp. 141-158, 2006.
- [28] S. Baker, S. Roth, D. Scharstein, M. Black, J. Lewis, and R. Szeliski, "A Database and Evaluation Methodology for Optical Flow," *Proc. IEEE Int'l Conf. Computer Vision*, 2007.
- [29] D. Cremers, "Dynamical Statistical Shape Priors for Level set Based Tracking," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 8, pp. 1262-1273, Aug. 2006.
- [30] T. Chan and L. Vese, "Active Contours without Edges," *IEEE Trans. Image Processing*, vol. 10, no. 2, pp. 266-277, Feb. 2001.
- [31] T. Chan, S. Esedoglu, and M. Nikolova, "Algorithms for Finding Global Minimizers of Image Segmentation and Denoising Models," *SIAM J. Applied Math.*, vol. 66, no. 5, pp. 1632-1648, 2006.
- [32] D. Cremers and S. Soatto, "A Pseudo-Distance for Shape Priors in Level Set Segmentation," *Proc. Second IEEE Workshop Variational, Geometric and Level Set Methods in Computer Vision*, pp. 169-176, 2003.
- [33] X. Bresson, S. Esedoglu, P. Vanderghenst, J. Thiran, and S. Osher, "Fast Global Minimization of the Active Contour/Snake Model," *J. Math. Imaging and Vision*, vol. 28, no. 2, pp. 151-167, 2007.
- [34] X. Bresson, P. Vanderghenst, and J. Thiran, "A Variational Model for Object Segmentation Using Boundary Information and Shape Prior Driven by the Mumford-Shah Functional," *Int'l J. Computer Vision*, vol. 68, no. 2, pp. 145-162, 2006.
- [35] D. Cremers, S. Osher, and S. Soatto, "Kernel Density Estimation and Intrinsic Alignment for Shape Priors in Level Set Segmentation," *Int'l J. Computer Vision*, vol. 69, no. 3, pp. 335-351, 2006.
- [36] J. Aujol, G. Gilboa, T. Chan, and S. Osher, "Structure-Texture Image Decomposition—Modeling, Algorithms, and Parameter Selection," *Int'l J. Computer Vision*, vol. 67, no. 1, pp. 111-136, 2006.
- [37] M. Taron, N. Paragios, and M. Jolly, "Registration with Uncertainties and Statistical Modeling of Shapes with Variable Metric Kernels," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 31, no. 1, pp. 99-113, Jan. 2008.
- [38] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-Based Object Tracking," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 5, pp. 564-575, May 2003.
- [39] D. Ross, J. Lim, R. Lin, and M. Yang, "Incremental Learning for Robust Visual Tracking," *Int'l J. Computer Vision*, vol. 77, no. 1, pp. 125-141, 2008.



Pratim Ghosh received the BE degree in electrical engineering from Jadavpur University, Kolkata, India, in 2004, and the ME degree in system science and automation (with distinction) from the Indian Institute of Science, Bangalore, India, in 2006, and the PhD degree in electrical and computer engineering from the University of California, Santa Barbara, in 2011. He interned at Janelia Farm, Howard Hughes Medical Institute, in the summer of 2009. His research interests span the areas of computer vision and pattern recognition, with particular emphasis on variational methods for image segmentation. He is a student member of the IEEE.



B.S. Manjunath received the MS degree in systems science and automation from the Indian Institute of Science and the PhD degree in electrical engineering from the University of Southern California. He is working as a professor of electrical and computer engineering at the University of California, Santa Barbara. His research interests include image processing and computer vision, and he has published more than 250 peer reviewed papers on image segmentation, registration, texture analysis, content-based image and video retrieval, image steganography and forensics, camera networks, and bio-image analysis. He is also a coinventor on 24 US/international patents. He was an associate editor of the *IEEE Transactions on Pattern Analysis and Machine Intelligence*, among others. He is a fellow of the IEEE.