Dynamic Cone-Beam Reconstruction Using a Variational Level Set Formulation

Andreas Keil, Jakob Vogel, Günter Lauritsch, and Nassir Navab

***Abstract*—Tomographic reconstruction from cone-beam X-ray data is only solved for static objects, e. g. the abdomen. In cardiac imaging, a rotational angiography sequence takes approx. 5 s and therefore spans several heart beats. Since such an acquisition scheme forces a trade-off between consistency of the scene and reasonable angular spacings between camera positions, standard reconstruction techniques fail at recovering the $3\mathrm{D}+t$ scene.**

We propose a new reconstruction framework based on variational level sets including a new energy term for symbolic reconstruction and incorporating the motion into the level set formalism. The resulting simultaneous estimation of shape and motion proves feasible in the presented experiments. Since the proposed formulation offers a great flexibility in incorporating other data terms, it could be of interest for other reconstruction settings as well.

***Index Terms*—Cardiac imaging, Cone-beam CT, Dynamic reconstruction, Rotational angiography.**

I. Introduction

The clinical motivation for providing a $3\mathrm{D}(+t)$ reconstruction of the coronary arteries from rotational angiography data is to provide the physician with intra-interventional volumetric data. Currently, patients with acute coronary syndrome either get a conventional CT (for a definite rule-out) or are directly sent to the catheter lab where diagnosis and intervention are performed at once using a C-arm system. In the former case, the physician may obtain a 3D reconstruction which is not intra-interventional whereas in the latter case, there are only series of 2D X-rays available for diagnosis and navigation.

Bringing the two worlds together requires a reconstruction from calibrated angiographic projections which can be obtained during a rotational run ($\approx 190°$) of the C-arm around the patient (see Fig. 1). Such a run takes about 4 s to 5 s. The resulting inconsistent projection data inhibits 3D reconstruction. This is the reason why a simultaneous estimation of shape and motion is needed in order to compensate for the heart motion during the reconstruction of the shape.

The ill-posedness of a direct tomographic $4\mathrm{D}$ reconstruction suggests to seek a symbolic or binary reconstruction first and then use the recovered motion for a later tomographic reconstruction. Such a symbolic reconstruction is performed on the contrasted coronary arteries. After enhancing these tubular structures with some kind of "vesselness" filter (like the ones presented in [1] or [2]) in 2D, they are suitable features, covering the motion in the relevant area around the patient's heart.

Please note that this work is an extension of [3], allowing a wider range of motions (periodic deformable vs. global rigid) and requiring different evaluation measures.

A. Keil, J. Vogel, and N. Navab are affiliated with Computer Aided Medical Procedures (CAMP), Fakultät für Informatik, Technische Universität München, Germany and G. Lauritsch is with Siemens AG, Healthcare Sector, Forchheim, Germany. The corresponding author is Andreas Keil, Fakultät für Informatik, I-16, Technische Universität München, Boltzmannstr. 3, 85748 Garching, Germany. E-Mail: andreas.keil@cs.tum.edu.

This work was funded by Siemens AG, Forchheim, Germany.

II. Related Work

To the authors' knowledge, all previous work on cardiac cone-beam CT makes strong use of the assumption that the heart motion can be grouped into several phases (usually defined by a percentage value between two adjacent R-peaks). Within such a phase (e. g. $10\% - 20\%$), the heart is assumed to re-position exactly. This permits a retrospective gating using the simultaneously recorded ECG signals. Based on this, Blondel et al. [4], Hansis et al. [5], and Movassaghi et al. [6] mostly rely on epipolar geometry and triangulation. Temporally distant but spatially consistent projections (yielding a wider baseline) are used to reconstruct 3D points and track them over time. Using traditional computed tomography solutions (like filtered back projection [7] or algebraic reconstruction [8], [9]) Prümmer et al. [10] and Schäfer et al. [11] both perform phase-wise tomographic reconstructions. These phase-wise reconstructions can then be fused if the motion between cardiac phases is somehow known. Of these two, [11] focuses on the motion-compensated FDK-reconstruction algorithm, assuming a known motion field, whereas [10] also proposes to do multiple sweeps for acquiring enough projection data.

In order to avoid making strong (and often not valid) assumptions about data consistency, we propose a variational level set framework for symbolic reconstruction instead of using tomographic- or triangulation-based methods. This enables

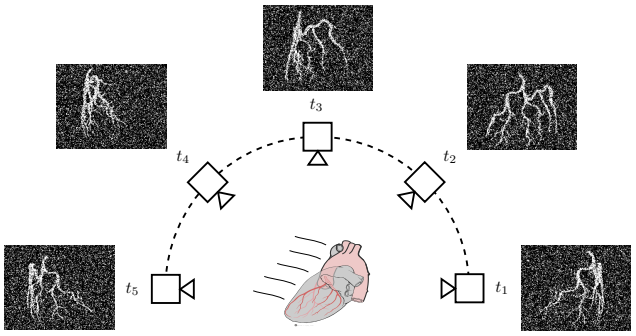

Fig. 1. The problem setting: A rotational angiography of a dynamic scene. (Image is derived from work by Patrick J. Lynch, medical illustrator; C. Carl Jaffe, MD, cardiologist. http://creativecommons.org/licenses/by/2.5/)

a soft coupling in space and time, thereby having the potential to yield more robust algorithms.

Although we employ a novel model and energy formulation, we still want to point the reader to the following works which we share some ideas with: Yoon et al. [12] perform a CT-like reconstruction from X-ray data using multiphase level sets. This enables the reconstruction of piece-wise constant tissue from very few projections but does not deal with motion. Rathi et al. [13] and Cremers et al. [14] perform deformable tracking on 2D images using active contours which is related to our time-coupling.

Additionally, there is a lot of related work in the computer vision community on 3D reconstruction from optical images using level sets, graph cuts, or voxel occupancy grids. For the sake of brevity, we do not delve into this field but just want to mention Franco et al. [15] who present a nice solution to the problem of 3D reconstruction from probabilistic silhouette images in a *synchronized* multi-view environment.

III. Methods

Having laid out our motivation for developing a level set framework (offering the desired soft coupling) for symbolic reconstruction, we now proceed to its modeling and implementation.

A. Dynamic Level Sets

Since we seek to obtain a symbolic or binary reconstruction of our 3D scene over time, we have chosen to model the "inside" and "outside" of reconstructed coronaries using a level set function

$$\Phi_0 : \left\{ \begin{array}{ccc} \mathbb{R}^3 & \rightarrow & \mathbb{R} \\ \boldsymbol{x}_0 & \mapsto & \Phi_0(\boldsymbol{x}_0) \end{array} \right. \tag{1}$$

on some reference domain with coordinates $\boldsymbol{x}_0$ and the convention $\Phi_0(\boldsymbol{x}_0) < 0$ for "inside" or "reconstructed vessel points". (Please refer to [16] for a short introduction to variational level set methods.) In order to establish a temporal relationship of the reconstruction frames, this level set function is made dynamic by introducing a parametric warping transformation $\boldsymbol{\varphi}$. This transformation maps points from location $\boldsymbol{x}$ at time t to coordinates $\boldsymbol{x}_0$ in the reference frame. Note that this reference frame (where the shape is reconstructed using one single level set function Φ_0) is arbitrary and not fixed to any point in time. This way, we avoid any bias toward a specific point in time. For the experiments presented in this paper, we model the motion as

$$\boldsymbol{\varphi}(\boldsymbol{x}, t, \boldsymbol{\alpha}) = \boldsymbol{R}(t, \boldsymbol{\alpha}) \cdot \boldsymbol{x} + \boldsymbol{T}(t, \boldsymbol{\alpha}) + \boldsymbol{u}(\boldsymbol{x}, t, \boldsymbol{\alpha}) \ , \tag{2}$$

where the rotation matrix $\boldsymbol{R}$ and the translation vector $\boldsymbol{T}$ together represent a dynamic rigid motion and $\boldsymbol{u}$ is a deformable motion. The former is modeled using 6 temporal B-splines with 12 degrees of freedom each, the latter is a tensor-product spline with $5^3 \cdot 12$ knots in $\mathbb{R}^3$, totaling to $\boldsymbol{\alpha} \in \mathbb{R}^{4572}$. Both motion parts are assumed to be periodic at this stage. However, there is no restriction when modeling the scene motion and future work will be aimed at including global, non-periodic motions, too.

A dynamic level set function is now obtained by using the motion model (2) to warp the shape model (1), yielding the dynamic level set function

$$\Phi(\boldsymbol{x}, t, \boldsymbol{\alpha}) = \Phi_0\big(\boldsymbol{\varphi}(\boldsymbol{x}, t, \boldsymbol{\alpha})\big) \ . \tag{3}$$

Note that one could also directly model a 4D level set function $\Phi(\boldsymbol{x}, t)$. But using a parametric warping function $\boldsymbol{\varphi}$ has several advantages:

- The shape reconstruction is implicitly regularized over time, since there is only one shape model.
- The motion can be recovered directly, simplifying its later use in a tomographic reconstruction as well as enabling a direct motion regularization (as opposed to imposing temporal soft constraints on Φ).
- Memory requirements for separate shape and motion models are much lower compared to a 4D Φ grid if the motion is parametrized.

B. Reconstruction Energies

Having modeled the shape and the motion to be optimized, we now describe the energy functional that fits the reconstruction parameters Φ_0 (implicitly representing the shape) and $\boldsymbol{\alpha}$ (representing the motion) to the given L projection images I_l acquired at times t_l, $1 \leq l \leq L$. The projection images' pixels are assumed to contain intensity values in $[0, 1]$, corresponding to the probability that the associated ray hit a vessel. Imposing penalties on false positive and false negative reconstructed points in space works in a manner similar to what was first presented by Chan and Vese [16] in the segmentation domain but taking into account the projective character of the imaging device:

Let V be the reconstruction volume, $\boldsymbol{P}_l : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ the projection operator for frame l, and H the Heaviside step function (or rather a mollified version of it, see [16] for examples). The false positive term is then

$$\begin{aligned} E_{\text{FP}}(\Phi_0, \boldsymbol{\alpha}) = \sum_{l=1}^{L} \int_V S_{\text{FP}}\Big(I_l\big(\boldsymbol{P}_l(\boldsymbol{x})\big)\Big) \cdot \\ \cdot \big[1 - H\big(\Phi_0(\boldsymbol{\varphi}(\boldsymbol{x}, t_l, \boldsymbol{\alpha}))\big)\big] \cdot \big[1 - I_l\big(\boldsymbol{P}_l(\boldsymbol{x})\big)\big] \ \mathrm{d}\boldsymbol{x} \ , \end{aligned} \tag{4}$$

where $S_{\text{FP}}(i) = H\left(\frac{1}{2} - i\right)$ is a switching function, enabling the false positive penalty for low intensities/probabilities $i \in [0, \frac{1}{2}]$ only. In this formula, the first two factors filter out the false (1st factor) positive (2nd factor) reconstructions, whereas the 3rd factor weights the penalty. This way, reconstructed points are penalized every time they are hit by a "non-vessel ray".

Penalizing false negatives works in a similar way. However, the big difference is that we cannot accumulate penalties in volume space. Due to the images being probabilistic projections, we may impose a false negative penalty if, and only if, no object is reconstructed along the whole ray corresponding to a high intensity pixel.[1] Thus, whole rays have to be

[1]Note that another approach would be to focus on a point in space and impose a false negative penalty iff *all* projected intensities enforce an object. However, this would favor "empty" reconstructions due to the initially inconsistent data.

considered instead of single points:

$$E_{\text{FN}}(\Phi_0, \boldsymbol{\alpha}) = \sum_{l=1}^{L} \int_A S_{\text{FN}}\big(I_l(\boldsymbol{p})\big) \cdot \cdot H\left(\min_{\boldsymbol{x} \in X_l(\boldsymbol{p})} \Phi_0\big(\boldsymbol{\varphi}(\boldsymbol{x}, t_l, \boldsymbol{\alpha})\big)\right) \cdot I_l(\boldsymbol{p}) \ \mathrm{d}\boldsymbol{p} \quad (5)$$

Here, $A \subset \mathbb{R}^2$ is the projection image space, $X_l(\boldsymbol{p})$ is the set of volume points corresponding to pixel $\boldsymbol{p}$ in image l, and $S_{\text{FN}}(i) = H\left(i - \frac{1}{2}\right)$ is the switching function enabling the term for falsely reconstructed points only. The three factors here are responsible for selecting pixels which indicate a vessel to be reconstructed on the ray to pixel $\boldsymbol{p}$ (1st factor), selecting rays where all Φ values are positive, i. e. there is no object reconstructed (2nd factor), and adding a weighted penalty (3rd factor), respectively.

Although the two data terms differ in their integration domains (volumes vs. imaging planes), they are presented using these intuitive formulas. For computation purposes, the data terms may either be appropriately weighted or one may convert the area integrals in the false negative term to volumetric integrals using the coarea formula.

C. Regularization

In terms of regularization we only need to care about shape regularization since the motion parameters are inherently regularized due to the usage of B-Splines with an appropriate number of knots. For obtaining a smooth shape reconstruction in the reference frame, we use

$$E_{\text{shape}}(\Phi_0) = \int_{V_0} \delta(\Phi_0(\boldsymbol{x}_0)) \cdot \|\nabla \Phi_0(\boldsymbol{x}_0)\| \ \mathrm{d}\boldsymbol{x}_0 \quad (6)$$

for penalizing the level set surface area, thus favoring reconstructions with low surface curvatures.

D. Implementation

Optimizing the system

$$E(\Phi_0, \boldsymbol{\alpha}) = \lambda_{\text{FN}} \cdot E_{\text{FN}}(\Phi_0, \boldsymbol{\alpha}) + \lambda_{\text{FP}} \cdot E_{\text{FP}}(\Phi_0, \boldsymbol{\alpha}) + \lambda_{\text{shape}} \cdot E_{\text{shape}}(\Phi_0) \quad (7)$$

is rather complex as the shape model Φ_0 and the deformation parameters $\boldsymbol{\alpha}$ have to be computed simultaneously. The former is minimized using the variational derivative of $\frac{\delta E}{\delta \Phi_0}$, the latter by calculating the gradient $\nabla_{\boldsymbol{\alpha}} E$. Computing these terms from their analytic forms involves deriving the minimum functional from equation (5), several numerical approximations, and a step size management during gradient descent for Φ_0 and α.

The most demanding issue to solve is the computation of E_{FN} and its derivatives. Computing the minimum functional in the equation's second factor requires a customized ray casting step with a warping of every sample point.

Several approaches to implement such a scheme are possible, including GPU-based methods. After considering aspects related to memory usage and speed of computation, we decided to use a CPU-based procedure, parallelized using OpenMP. Even though GPUs appear to be a natural choice for ray casting, their bad support for "arbitrary writes" disqualifies them for this algorithm.

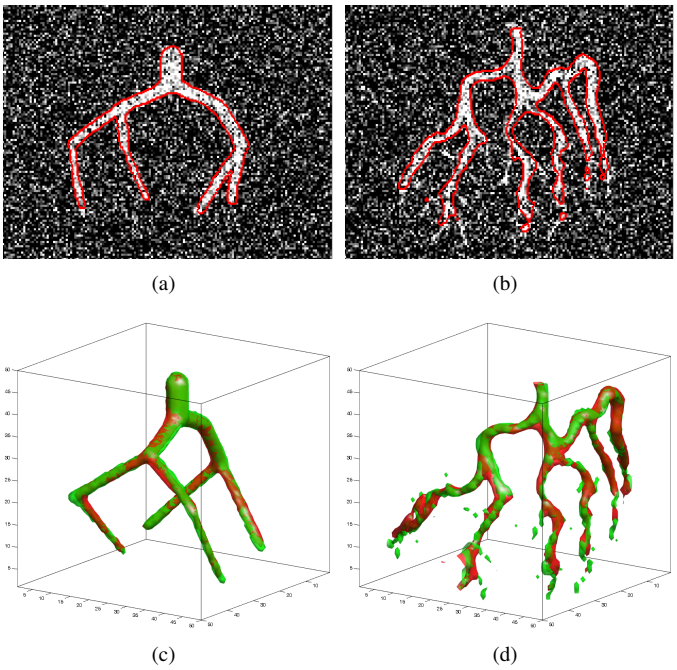

Fig. 2. Exemplary overlays for error evaluation. The noise level for the projections and reconstructions shown is 50 %. Top row: Overlay of reconstructed shape borders (red) on input projection data. Bottom row: 3D overlay of ground truth data (green) and reconstructed shapes (red). The left column shows the "Synthetic" data set whereas the right column shows the "Phantom" data set. Note that the whole setup is dynamic and the printed images can only show a snapshot of the non-rigidly moving artery trees.

IV. Experiments and Discussion

We test our method using synthetic and phantom data. The "Synthetic" data is created by modeling tubes of considerable diameter clearly visible in the projection images (see Fig. 2(a) and (c)) while the "Phantom" data was physically built, scanned (without motion), reconstructed and segmented. It contains thin vessels which can barely be represented at our current voxel resolution as visible in Fig. 2(d).

In both cases, we use 3×4 projection matrices, obtained from the calibration of a real stationary C-arm, to generate projections of the data. During the simulated image acquisition process we apply a periodic deformable motion with amplitudes of 10 mm for the translations, 10° for the rotations and up to 30 mm for the deformations. An image enhancement step as necessary in the real setting can be omitted grace to the use of symbolic ground truth data. Instead, we add Gaussian noise (with zero mean and standard deviations of 0 % to 50 % of the full intensity range) to the projection images in order to simulate more realistic projections and test the algorithm's sensitivity to noise. Sample projections with 50 % noise are shown in Fig. 2(a) and 2(b).

In order to speed up testing, we work on rather coarse data using 48 projections at 155×120 pixels each (compared to 200-400 images with 620×480 pixels each in a real setting). The reconstruction volume V covers a cube of size $(15\,\text{cm})^3$, discretized as grid of 50^3 voxels (yielding a 3 mm spacing in each dimension).

All experiments have been run on a machine with 24 cores. Execution time depends on several factors such as noise and

TABLE I
COMPARISON OF RECONSTRUCTION ERRORS FOR THE "SYNTHETIC" AND "PHANTOM" DATA SETS AT SIX DIFFERENT NOISE LEVELS EACH.

Data Set	Noise level	R	S_e	S_p
Synthetic	0 %	85.1 %	86.1 %	99.9 %
Synthetic	10 %	84.9 %	84.4 %	99.9 %
Synthetic	20 %	84.6 %	83.5 %	99.9 %
Synthetic	30 %	83.8 %	80.1 %	99.9 %
Synthetic	40 %	83.2 %	80.1 %	99.9 %
Synthetic	50 %	81.3 %	75.9 %	99.9 %
Phantom	0 %	66.7 %	75.2 %	99.6 %
Phantom	10 %	66.6 %	78.0 %	99.6 %
Phantom	20 %	65.0 %	73.8 %	99.6 %
Phantom	30 %	67.0 %	74.2 %	99.6 %
Phantom	40 %	66.3 %	72.8 %	99.6 %
Phantom	50 %	64.7 %	71.7 %	99.6 %

both, motion and scene complexity but is usually below 1 h.

Error measures well-known from the segmentation domain are used for a quantitative evaluation of the symbolic reconstruction: Let T_p, T_n, F_p, and F_n be the true positive, true negative, false positive, and false negative voxel counts for a symbolic reconstruction, resp. Then

$$\text{overlap ratio} \quad R = \frac{2 \cdot T_p}{(T_p + F_n) + (T_p + F_p)} \ , \tag{8}$$

$$\text{sensitivity} \quad S_e = \frac{T_p}{T_p + F_n} \ , \tag{9}$$

$$\text{and specificity} \quad S_p = \frac{T_n}{T_n + F_p} \tag{10}$$

are used to measure the accuracy of the reconstruction.

A comparison of these measures for two data sets and six noise levels is given in TABLE I. The phantom data set performs worse than the synthetic model. The specificity S_p is of limited interest since it mainly correlates with the ratio of vessel voxels to the reconstruction volume's total number of voxels. It always approaches 1 if the reconstruction volume is large enough (T_n is big) and is just given for completeness. The chosen error measures are not really well-suited for evaluating this data with its relatively coarse resolution. Since the vessel structures (esp. of the "Phantom" data set) are barely representable at the given resolution, the voxel counts are also subject to discretization errors. However, a visual evaluation of the recovered motion as in Fig. 2(d) shows that the motion estimation is still very accurate. Unfortunately, a quantitative evaluation of positional errors is quite difficult since there is no reference frame where this error could be measured. A deformable registration (with all its ambiguities) would be required to compute motion errors between reconstruction and ground truth volumes. Previous experiments with rigid motion models (enabling a better evaluation of position errors) in [3], however, have shown that the estimated motion is usually sub-voxel accurate (about 1 mm errors for 1.5 mm voxel spacings). This is even more relevant for subsequent motion-corrected tomographic reconstructions. Comparing the evaluation measures across noise levels shows the very high robustness of the proposed formulation w.r.t. noise.

V. CONCLUSION

The symbolic $3\mathrm{D} + t$ reconstruction presented in this paper is another step towards dynamic reconstruction from rotational cone-beam data. Its major benefit is that it does not depend on hard constraints such as perfect ECG signals or an exact repositioning of cardiac anatomy between heart beats. Thus, the computed motion is more robustly obtained than with other approaches and it may by used for a subsequent, motion-corrected tomographic reconstruction.

The experiments show the framework's ability to obtain a 4D reconstruction from a series of very noisy 2D projections. The types of motion covered by our periodic deformable models are very similar to real cardiac motion. Future work aims at allowing global, non-periodic motions and optimizing the implementation, thus enabling to work with real data.

ACKNOWLEDGMENTS

The authors thank Moritz Blume, Jan Boese, and Martin Brokate for valuable discussions. Credits also go to Tobias Klug and the Chair "LRR" at TUM for providing the high performance hardware and to Christopher Rohkohl for the phantom data.

REFERENCES

[1] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, "Multiscale vessel enhancement filtering," in *Proc. MICCAI*, ser. LNCS, vol. 1496, 1998, pp. 130–137.

[2] T. M. Koller, G. Gerig, G. Székely, and D. Dettwiler, "Multiscale detection of curvilinear structures in 2-D and 3-D image data," in *Proc. ICCV*, 1995, pp. 864–869.

[3] A. Keil, J. Vogel, G. Lauritsch, and N. Navab, "Dynamic cone beam reconstruction using a new level set formulation," in *Proc. MICCAI*, ser. LNCS, 2009.

[4] C. Blondel, G. Malandain, R. Vaillant, and N. Ayache, "Reconstruction of coronary arteries from a single rotational X-ray projection sequence," *IEEE TMI*, vol. 25, no. 5, pp. 653–663, 2006.

[5] E. Hansis, D. Schäfer, O. Dössel, and M. Grass, "Projection-based motion compensation for gated coronary artery reconstruction from rotational x-ray angiograms," *Physics in Medicine and Biol.*, vol. 53, pp. 3807–3820, 2008.

[6] B. Movassaghi et al., "3D reconstruction of coronary stents in vivo based on motion compensated X-ray angiograms," in *Proc. MICCAI*, ser. LNCS, vol. 4191, 2006, pp. 177–184.

[7] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm," *J. Opt. Soc. Am. A*, vol. 1, no. 6, pp. 612–619, 1984.

[8] R. Gordon, R. Bender, and G. T. Herman, "Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography," *J. Theoretical Biol.*, vol. 29, no. 3, pp. 471–481, 1970.

[9] A. H. Andersen and A. C. Kak, "Simultaneous algebraic reconstruction technique (SART): A superior implementation of the ART algorithm," *Ultrasonic Imaging*, vol. 6, no. 1, pp. 81–94, 1984.

[10] M. Prümmer et al., "Cardiac C-arm CT: Efficient motion correction for 4D-FBP," in *Nucl. Science Symp. and Med. Imag. Conf.*, 2006, pp. 1–20.

[11] D. Schäfer, J. Borgert, V. Rasche, and M. Grass, "Motion-compensated and gated cone beam filtered back-projection for 3-D rotational X-ray angiography," *IEEE TMI*, vol. 25, no. 7, pp. 898–906, 2006.

[12] S. Yoon, A. R. Pineda, and R. Fahrig, "Level set reconstruction for sparse angularly sampled data," in *Nucl. Science Symp. and Med. Imag. Conf.*, vol. 6, 2006, pp. 3420–3423.

[13] Y. Rathi, N. Vaswani, A. Tannenbaum, and A. Yezzi, "Particle filtering for geometric active contours with application to tracking moving and deforming objects," in *Proc. CVPR*, vol. 2, 2005, pp. 2–9.

[14] D. Cremers, "Dynamical statistical shape priors for level set-based tracking," *IEEE TPAMI*, vol. 28, no. 8, pp. 1262–1273, 2006.

[15] J.-S. Franco and E. Boyer, "Fusion of multi-view silhouette cues using a space occupancy grid," in *Proc. ICCV*, vol. 2, 2005, pp. 1747–1753.

[16] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Trans. Image Process.*, vol. 10, no. 2, pp. 266–277, 2001.