

# Dynamic Cone Beam Reconstruction Using a New Level Set Formulation

Andreas Keil<sup>1</sup>, Jakob Vogel<sup>1</sup>, Günter Lauritsch<sup>2</sup>, and Nassir Navab<sup>1</sup> \*

<sup>1</sup> Computer Aided Medical Procedures, TU München, Germany, [keila@cs.tum.edu](mailto:keila@cs.tum.edu)

<sup>2</sup> Siemens AG, Healthcare Sector, Forchheim, Germany

**Abstract.** This paper addresses an approach toward tomographic reconstruction from rotational angiography data as it is generated by C-arms in cardiac imaging. Since the rotational acquisition scheme forces a trade-off between consistency of the scene and reasonable baselines, most existing reconstruction techniques fail at recovering the  $3D + t$  scene. We propose a new reconstruction framework based on variational level sets including a new data term for symbolic reconstruction as well as a novel incorporation of 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 as well as hard or soft constraints, it allows an adaption to a wider range of problems and could be of interest to other reconstruction settings as well.

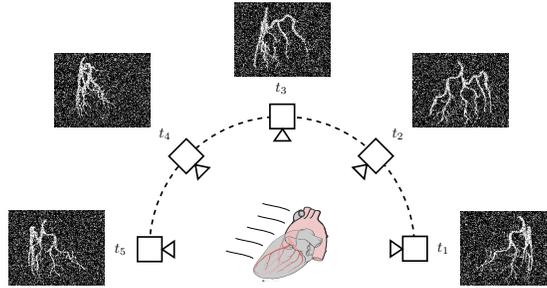
## 1 Introduction

The clinical motivation for providing a  $3D(+t)$  reconstruction of the coronary arteries from rotational angiography data is to provide the physician with intra-interventional 3D data. Currently, patients with chest pain and other symptoms for a cardiac infarction 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 ( $190^\circ$ ) of the C-arm around the patient (see Fig. 1). Such a run takes about 4 s to 5 s which is a hard limit for technical as well as security reasons. Therefore, a human heart beats about 4 to 7 times during imaging. The resulting inconsistent projection data inhibits 3D reconstruction. This is the reason why a simultaneous reconstruction of shape and motion is needed in order to compensate for the heart motion during the reconstruction of the shape.

---

\* 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. This work was funded by Siemens AG, Healthcare Sector, Forchheim, Germany.



**Fig. 1.** 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/>)

The ill-posedness of a direct tomographic 4D reconstruction suggests to seek a symbolic/binary reconstruction first and then use the recovered motion for a later tomographic reconstruction. Such a symbolic reconstruction is performed on the coronaries since they are contrasted and cover the motion in the relevant area around the patient’s heart.

## 2 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 to the same state in each of the phase’s images. This permits a retrospective gating using the simultaneously recorded ECG signals. Based on this, Blondel et al. [1], Hansis et al. [2], and Movassaghi et al. [3] 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 [4] or algebraic reconstruction [5, 6]) Prümmer et al. [7] and Schäfer et al. [8] perform phase-wise tomographic reconstructions. These phase-wise reconstructions can then be fused if the motion between cardiac phases is somehow known. [8] focuses on the motion-compensated FDK-reconstruction algorithm assuming a known motion field whereas [7] also proposes to do multiple sweeps for acquiring enough projection data.

For the following reasons we propose a level set framework for symbolic reconstruction instead of using tomographic- or triangulation-based methods: Due to the bad image quality of contrasted angiographic X-ray projections, an algorithm not explicitly using correspondences but just a soft coupling of points in 3D space would be desirable. Although healthy hearts beat in a more or less regular manner, assuming exact re-positioning and perfectly periodic ECG signals is a quite strong requirement. This is in particular problematic for patients with pathologies like congenital cardiac defects or a prior bypass-surgery. To this end,

a soft coupling in time domain could also prove to be advantageous. Level sets have the further advantage of being able to handle the unknown tree structure.

Although we will present novel data terms and a new space-time coupling, we still want to point the reader to the following works which we share some ideas with: Yoon et al. [9] perform a CT-like reconstruction from X-ray data using multiphase level sets. This work enables the reconstruction of piece-wise constant tissue from very few projections but does not deal with motion. Rathi et al. [10] and Cremers et al. [11] perform deformable tracking on 2D images using active contours which is related to our time-coupling.

Additionally, there is a lot of related work on 3D reconstruction from optical images using level sets, graph cuts, or voxel occupancy techniques. For the sake of brevity, we do not delve into this field but just want to mention Franco et al. [12] who give a nice derivation and solution to the problem of 3D reconstruction from probabilistic silhouette images in a *synchronized* multi-view environment.

### 3 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. The main theoretical contributions of this paper are the development of energy terms, fitting a level set function to the given image data, and its usage with a dynamic<sup>3</sup> level set function.

#### 3.1 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 objects using a level set function

$$\Phi_0 : \begin{cases} \mathbb{R}^3 \rightarrow \mathbb{R} \\ \mathbf{x}_0 \mapsto \Phi_0(\mathbf{x}_0) \end{cases} \quad (1)$$

on some reference domain with coordinates  $\mathbf{x}_0$  and the convention  $\Phi(\mathbf{x}_0) < 0$  for “inside” or reconstructed points. In order to establish a temporal relationship of the reconstruction frames, this level set function is made dynamic by introducing a warping transformation  $\varphi$ . Similar to what was presented in [13], this transformation maps points from location  $\mathbf{x}$  at time  $t$  to coordinates  $\mathbf{x}_0$  in the reference frame where the shape is reconstructed using one single level set function  $\Phi_0$ . For the experiments presented in this paper, we modeled a rigid motion over time using 6 temporal B-splines (with 10 degrees of freedom each) for the time-dependent rotation matrix  $\mathbf{R}$  and translation vector  $\mathbf{T}$  yielding coefficients  $\alpha \in \mathbb{R}^{6 \times 10}$ :

$$\varphi : \begin{cases} \mathbb{R}^3 \times \mathbb{R} \times \mathbb{R}^{6 \times 10} \rightarrow \mathbb{R}^3 \\ (\mathbf{x}, t, \alpha) \mapsto \mathbf{R}(t, \alpha) \cdot \mathbf{x} + \mathbf{T}(t, \alpha) \end{cases} \quad (2)$$

<sup>3</sup> “Dynamic” in this context means the deformation of the level set function over the real time variable  $t$  (as opposed to an evolution of the level set function over artificial time during iterations).

Note that the reference frame is arbitrary and not fixed to any point in time. This way, we avoid any bias toward a specific time.

Putting together equations (1) and (2), we obtain the dynamic level set function  $\Phi : \mathbb{R}^3 \times \mathbb{R} \times \mathbb{R}^{6 \times 10} \rightarrow \mathbb{R}$

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

Note that one could also choose to directly model a 4D level set function  $\Phi(\mathbf{x}, t)$ . But using a dynamic 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.
- Memory requirements are much lower compared to a 4D grid if  $\boldsymbol{\varphi}$  is parametrized.

### 3.2 Reconstruction Energies

Having built a model for the shape and motion to be reconstructed, we now proceed to setting up an energy functional that fits the reconstruction parameters  $\Phi_0$  and  $\boldsymbol{\alpha}$  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 [14] but taking into account the projective character of the imaging device:

Let  $V$  be the reconstruction volume,  $\mathbf{P}_l : \mathbb{R}^3 \rightarrow \mathbb{R}^2$  the projection operator for frame  $l$ , and  $H$  the Heaviside step function (or rather a mollified version of it, see [14] for examples). The false positive term then is

$$E_{\text{FP}}(\Phi_0, \boldsymbol{\alpha}) = \sum_{l=1}^L \int_V S_{\text{FP}}(I_l(\mathbf{P}_l(\mathbf{x}))) \cdot [1 - H(\Phi_0(\boldsymbol{\varphi}(\mathbf{x}, t_l, \boldsymbol{\alpha})))] \cdot [1 - I_l(\mathbf{P}_l(\mathbf{x}))] \, d\mathbf{x} , \quad (4)$$

where  $S_{\text{FP}}(i) = H(\frac{1}{2} - i)$  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 (1<sup>st</sup> factor) positive (2<sup>nd</sup> factor) reconstructions, whereas the 3<sup>rd</sup> factor is a weighted 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 only impose a false negative penalty if, and only if, no object is reconstructed along the whole ray corresponding to a high intensity pixel.<sup>4</sup> Thus, whole rays have to be considered instead of single

<sup>4</sup> 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.

points:

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

Here,  $A \subset \mathbb{R}^2$  is the projection image space,  $X_l(\mathbf{p})$  is the set of volume points corresponding to pixel  $\mathbf{p}$  in image  $l$ , and  $S_{\text{FN}}(i) = H(i - \frac{1}{2})$  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  $\mathbf{p}$  (1<sup>st</sup> factor), selecting rays where all  $\Phi$  values are positive, i. e. there is no object reconstructed (2<sup>nd</sup> factor), and adding a weighted penalty (3<sup>rd</sup> factor), respectively.

The two data terms seem to be of very different type. This is remedied by either appropriately weighting them or reformulating the false negative term to a volume integral using the coarea formula.

### 3.3 Regularization

In terms of regularization we only need to care about shape regularization at this point 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

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

for penalizing the level set surface and thereby favoring reconstructions with low surface curvatures.

### 3.4 Implementation

Optimizing the system

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

resulting from putting together the terms (4)–(6), is rather complex as two sets of parameters must be computed simultaneously, namely the shape model  $\Phi_0$  and the deformation parameters  $\alpha$ . The former is minimized using the variational derivative of  $\frac{\delta E}{\delta \Phi_0}$ , the latter by calculating the gradient  $\nabla_{\alpha} E$ . 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  $\Phi_0$  and  $\alpha$ .

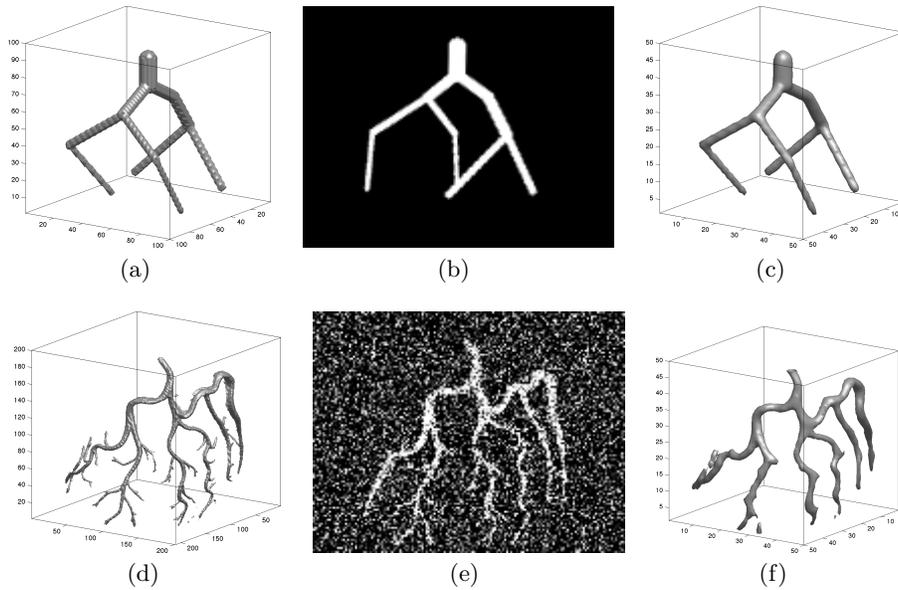
The most demanding issue to solve is the computation of  $E_{\text{FN}}$  and its derivative. It involves ray casting (customized to using  $3 \times 4$  projection matrices and applying the estimated motion for every sample point) for computing the minimum contained in the equation’s second factor. Updates to  $\Phi_0$  have to be applied

at the sample points (which are in general not at grid locations of  $\Phi_0$ ) and thus be “backward-interpolated”.

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, optimized using the OpenMP framework. Even though GPUs appear to be a natural choice for ray casting, their bad support for arbitrary writes disqualifies them for this algorithm.

## 4 Experiments and Discussion

We tested our method using synthetic and phantom data. The “synthetic” data was created by modeling tubes of considerable diameter clearly visible in the projection images (see Fig. 2(a) and (b)) while the “phantom” data was physically built, scanned, reconstructed (without motion) and segmented. It contains thin vessels of just 1 or 2 voxels diameter as visible in Fig. 2(d). In both cases, we used  $3 \times 4$  projection matrices, obtained from the calibration of a real stationary C-arm, to generate synthetic views of the data. During the virtual image acquisition process we applied a dynamic but rigid motion with realistic intervals and amplitudes. An image enhancement step as necessary in the real applica-



**Fig. 2.** Two examples of the imaging and reconstruction process. Top row: “Synthetic” data without noise. Bottom row: “Phantom” data with 50 % noise. From left to right: Ground truth models, exemplary projection, and the final reconstruction. Note that the projections do not show a static setting, but a snapshot of a moving artery tree.

tion could be omitted grace to the use of symbolic ground truth data. Instead, we randomly added Gaussian noise (with zero mean and standard deviations of 25 % and 50 % of the full intensity range) to the projection images in order to test the proposed algorithm’s sensitivity to noise. Sample projections with different magnitudes of noise are shown in Fig. 2.

As a result of these steps, we obtained a series of projection images of the moving artery tree, and their corresponding projection matrices. In order to speed up testing, we worked on rather coarse data using 48 projections at  $155 \times 120$  pixels each (compared to 200-400 images at  $620 \times 480$  pixels each in a real setting). The reconstruction volume  $V$  covered a cube of size  $(15\text{cm})^3$ , discretized as grid of  $50^3$  voxels.

All experiments have been run on high-performance hardware, including quad-core and 24-core equipment. The execution time depends on several factors such as noise and complexity of both motion and image content. For the 24-core machine, an average execution time of roughly 5 min for 100 iterations has been attained, after which the result was stable.

In order to compute a meaningful error measure, we collected all spatial points  $\mathbf{X}_{k=1..K}$  corresponding to vessel voxels in the ground truth data. Afterwards, for each moment  $t_l$ , we warped these points using both the ground truth motion  $\widehat{\mathbf{R}}(t_l)$ ,  $\widehat{\mathbf{T}}(t_l)$  and the reconstructed motion  $\mathbf{R}(t_l, \boldsymbol{\alpha})$ ,  $\mathbf{T}(t_l, \boldsymbol{\alpha})$  and computed the reconstruction error

$$\left\| [\widehat{\mathbf{R}}(t_l) \cdot \mathbf{X}_k + \widehat{\mathbf{T}}(t_l)] - [\mathbf{R}(t_l, \boldsymbol{\alpha}) \cdot \mathbf{X}_k + \mathbf{T}(t_l, \boldsymbol{\alpha})] \right\|_2 \quad (8)$$

for every point  $X_k$  at every moment  $t_l$ .

A comparison of errors for two data sets and three noise levels is given in Table 1. Obviously, the algorithm supports a fair amount of noise. Still, low-noise images (25 %) compare best to the segmented vessel images to be used in the final application. The phantom data set performs worse than the synthetic model. However, this problem can most likely be traced to the coarse resolution we used despite the fine structures of this data set (see Fig. 2, bottom row). Nevertheless, the motion was usually still well estimated, in these cases. The average errors for reasonably posed problems with max. 25 % noise does not exceed 1 mm (and thus is sub-voxel accurate) even though we downsampled all data by factor four in space and time compared to the real setting.

## 5 Conclusion

The method presented in this paper is a promising alternative to other cardiac cone beam reconstruction procedures. Its major benefit is that it does not depend on hard constraints such as perfect ECG signals (although they may be included as a soft constraint at a later point) or an exact re-positioning of cardiac anatomy between heart beats.

However, in its present form, the motion description does not yet cover all possible motions that one encounters in the clinical setting. Future work will

**Table 1.** Comparison of the reconstruction errors. The two data sets “Synthetic” and “Phantom” were reconstructed at three different noise levels. All errors are given in mm and have been evaluated for a series of 5-10 experiments.

Data Set	Noise	Mean	St.D.	Max.	Med.
Synthetic	0%	0.54	0.30	2.19	0.47
Synthetic	25%	0.68	0.36	3.14	0.60
Synthetic	50%	2.36	2.53	11.73	1.18
Phantom	0%	0.91	0.48	4.41	0.82
Phantom	25%	0.88	0.46	4.37	0.81
Phantom	50%	4.15	2.70	9.79	3.86

Data Set	Noise	Mean	St.D.	Max.	Med.
Synthetic	0 – 50%	1.20	1.70	11.73	0.64
Phantom	0 – 50%	1.98	2.22	9.79	1.03
both	0%	0.81	0.47	4.41	0.72
both	25%	0.83	0.45	4.37	0.75
both	50%	3.68	2.77	11.73	3.07

thus aim at more universal descriptions providing more degrees of freedom, such as affine transformations and fully deformable models. Especially in the latter case, application-specific soft constraints (e.g. relating ECG and motion) will most likely become necessary.

## References

1. Blondel, C. et al.: Reconstruction of coronary arteries from a single rotational X-ray projection sequence. *IEEE TMI* **25**(5) (2006) 653–663
2. Hansis, E. et al.: Projection-based motion compensation for gated coronary artery reconstruction from rotational x-ray angiograms. *Phys. Med. Biol.* **53** (2008)
3. Movassaghi, B. et al.: 3D reconstruction of coronary stents in vivo based on motion compensated X-ray angiograms. In: *MICCAI*. Volume 4191 of LNCS. (2006)
4. Feldkamp, L.A., Davis, L.C., Kress, J.W.: Practical cone-beam algorithm. *J. Opt. Soc. Am. A* **1**(6) (1984) 612–619
5. Gordon, R. et al.: Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography. *J. Theor. Biol.* **29**(3) (1970)
6. Andersen, A.H., Kak, A.C.: Simultaneous algebraic reconstruction technique (SART): A superior implementation of the ART algorithm. *US Imag.* **6**(1) (1984)
7. Prümmer, M. et al.: Cardiac C-arm CT: Efficient motion correction for 4D-FBP. In: *NSS and MIC*, Springer (2006) 1–20
8. Schäfer, D. et al.: Motion-compensated and gated cone beam filtered back-projection for 3-D rotational X-ray angiography. *IEEE TMI* **25**(7) (2006) 898–906
9. Yoon, S., Pineda, A.R., Fahrig, R.: Level set reconstruction for sparse angularly sampled data. In: *NSS and MIC*. Volume 6. (2006) 3420–3423
10. Rathi, Y. et al.: Particle filtering for geometric active contours with application to tracking moving and deforming objects. In: *CVPR*. Volume 2., IEEE (2005) 2–9
11. Cremers, D.: Dynamical statistical shape priors for level set-based tracking. *IEEE Trans. PAMI* **28**(8) (2006) 1262–1273
12. Franco, J.S., Boyer, E.: Fusion of multi-view silhouette cues using a space occupancy grid. In: *ICCV*. Volume 2. (2005) 1747–1753
13. Blume, M., Keil, A., Navab, N., Rafecas, M.: Blind motion compensation for positron-emission-tomography. In: *SPIE Med. Imag. Proc. of the SPIE* (2009)
14. Chan, T.F., Vese, L.A.: Active contours without edges. *IEEE TIP* **10**(2) (2001)