Simultaneous Correspondence and Non-Rigid 3D Reconstruction of the Coronary Tree from Single X-ray Images

Eduard Serradell[1]   Adriana Romero[2,4]   Rubén Leta[3]   Carlo Gatta[2,4]   Francesc Moreno-Noguer[1]

[1]Institut de Robòtica i Informàtica Industrial, CSIC-UPC, Barcelona, Spain
[2]Centre de Visió per Computador, UAB, Barcelona, Spain
[3]Hospital de la Santa Creu i Sant Pau, Barcelona, Spain
[4]Departament de Matemàtica Aplicada i Anàlisi, UB, Barcelona, Spain

{eserradell,fmoreno}@iri.upc.edu, aromero@cvc.uab.es, rleta@hsp.santpau.es, carlo.gatta@ub.edu

Abstract

We present a novel approach to simultaneously reconstruct the 3D structure of a non-rigid coronary tree and estimate point correspondences between an input X-ray image and a reference 3D shape. At the core of our approach lies an optimization scheme that iteratively fits a generative 3D model of increasing complexity and guides the matching process. As a result, and in contrast to existing approaches that assume rigidity or quasi-rigidity of the structure, our method is able to retrieve large non-linear deformations even when the input data is corrupted by the presence of noise and partial occlusions. We extensively evaluate our approach under synthetic and real data and demonstrate a remarkable improvement compared to state-of-the-art.

1. Introduction

Recovering the 3D structure of a non-rigid coronary tree from single X-ray images is a highly ambiguous problem since many different 3D configurations can virtually have the same projection. As shown in Fig. 1 the problem becomes even more challenging because X-ray images are often affected by a series of artifacts such as noise, blurring, partial occlusions and vessel discontinuity. Thus, solving this problem requires from prior knowledge about the type of deformations the structure can undergo.

Standard approaches within medical imaging assume a reference 3D scan of the tree is known and that deformations in the input image are negligible. This reduces the shape recovering task to a rigid 3D-to-2D registration [9, 11, 16]. There exist a recent attempt of addressing the non-rigidity nature of the problem, although it has only been shown effective for relatively small deformations [8].

We may find other related areas in computer vision that essentially solve the same problem but in a different context, for instance, the techniques for 3D non-rigid surface
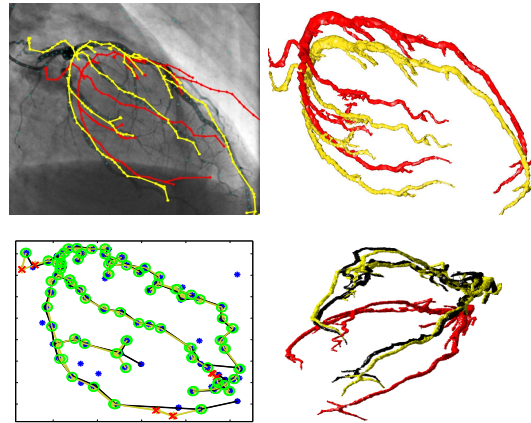

Figure 1. **Recovering the structure of a non-rigid coronary tree.** **Top:** Given an input X-ray image (left), and a reference structure (shown in red at the top-right image) we are able to retrieve the 3D configuration of the coronary tree in the input image (yellow). **Bottom:** Results on synthetic data, for which we know the ground truth and allows us to evaluate the method under noise, occlusions (blue dots), and different levels of deformation. The left figure depicts the 2D results and the figure on the right represents the ground truth (black), the prior (red) and our solution (yellow). Note that even when the prior significantly differs from the ground truth, our approach yields very accurate results.

reconstruction [5, 14, 17] and articulated human pose estimation from monocular images [1, 18, 21, 27]. In these approaches, though, it is often easy to obtain large amounts of training data and build detailed parametric models for specific deformations or 2D-to-3D mappings that directly link 2D observations with 3D configurations. Unfortunately, producing these detailed models and mappings is beyond our possibilities, because X-ray images are harmful for the patient and, besides one single reference 3D scan of the coronary tree, no further prior knowledge can be used.

We therefore propose a novel approach that, given solely

one single X-ray image and a reference 3D configuration, simultaneously recovers the 3D structure of the coronary tree in the input image and establishes matches with the reference shape. As shown in Fig. 1, our method can recover the 3D structure in the input image even when it highly differs from that of the reference configuration. In addition we can handle large amounts of noise and occlusions.

The key contributions that make this possible are twofold. First we use a generative model that progressively increases its complexity and allows a coarse to fine fitting while 3D-to-2D matches are estimated. Second, we take advantage of a recursive parameterization of the coronary tree that introduces dependencies between all the nodes of the tree, and diffuses the local constraints to the whole structure. Both the parameterization and the generative model are then integrated within a Kalman-based optimization framework. In the results section we will show that the overall methodology has significant advantages when compared to state-of-the art approaches.

2. Related Work

Recovering the 3D structure of the coronary tree from single vascular images involves dealing with many different issues. Besides the inherent ambiguity of the monocular non-rigid reconstruction, the problem is further accentuated due to the presence of noise in the images and partial occlusions between different branches of the tree. This complexity has been traditionally alleviated by considering the vascular system as a rigid structure [9, 11, 16] and using multiple views [26, 29]. To the best of our knowledge, [8] is the only approach in the medical imaging literature that considers the non-rigid nature of the problem. They introduce 3D priors and inextensibility constraints into a steepest descent scheme to solve for the shape. Yet, their optimization procedure is only effective under relatively simple deformations as those occurring in the liver artery.

On the other hand, our approach has similarities with the techniques to reconstruct non-rigid 3D surfaces and estimate articulated pose from monocular images. Among the former, it has been shown that 3D shape can be retrieved by imposing local inextensibility and constraints introduced by a set of 3D-to-2D correspondences between the input image and a reference shape [5, 14, 17, 19]. In essence we will also use the same kind of assumptions, although we will need from additional constraints since in our context the 3D-to-2D correspondences are unknown and have to be resolved simultaneously with the shape. Moreover, local distance constraints are much less restrictive when dealing with points linked through a tree-like structure than when dealing with neighboring points on a surface. In addition, many of these approaches impose strong shape priors based on previously acquired training data [17, 19] while in our approach accurate training data is hard to obtain and we have to rely on very weak shape priors.

Since the coronary tree may be regarded as an articulated structure, one might think in applying the techniques of articulated pose estimation to our problem [1, 21, 27]. These approaches rely on large amounts of data for learning a mapping from 2D image observations to 3D poses, and have the advantage of not requiring to solve the 2D-to-3D correspondence problem. Yet, as said above, while obtaining sufficient training data is feasible for applications such as human pose estimation [2, 23], it becomes prohibitive in our framework, as the number of X-ray images that may be captured for each patient is limited.

Recent works suggest introducing similar constraints as those used for non-rigid shape recovery into the formulation of articulated pose estimation problems [18, 22, 24]. This allows fitting more detailed parametric 3D models [22] and reducing the dependency of articulated pose estimation techniques on the training data [18]. However, reducing the dependency on training data has the drawback of increasing the sensitivity to artifacts into the input data.

Drawing particular inspiration on these approaches, our method also combines tools from the techniques for articulated pose estimation and shape recovery. However, in order to tackle problems with much larger amounts of image noise and occlusions, we propose using a generative 3D model that progressively increases its complexity and adaptability while establishing correspondences and detecting and rejecting outlier points. In addition, we represent the articulated structure using a recursive parametrization that, as we will show in the results section, yields remarkable improved results when compared against [18].

3. Algorithm Overview

The focus of this paper is on retrieving the 3D structure of the coronary tree while establishing 3D-to-2D point correspondences between an input X-ray image and a reference 3D scan. However, the overall algorithm requires additional tasks as we next detail:

1. Feature extraction: Our input data is an X-ray image and a volumetric 3D Computed Tomography (CT) scan. In a preprocessing step we segment the vessel regions on both sets of data and extract points of interest.

2. Generative model for the coronary tree: We represent the 3D feature points as a tree, parameterized by the joint angles and distances between consecutive points. Since we cannot explicitly compute deformation modes from training data, we estimate them by performing a Probabilistic Principal Component Analysis (PPCA) over a set of synthetic samples for which we randomized the values of the joint angles. The flexibility of the model will be controlled by the number of components of the PPCA and by the magnitude of the noise used to generate the deformed samples.
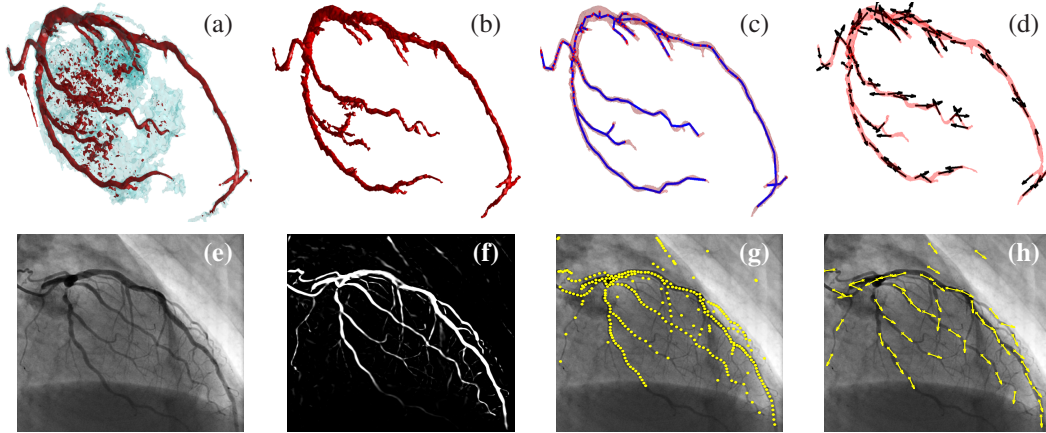
Figure 2. **Feature extraction. Top:** Extraction of the 3D skeleton from the 3D Computed Tomography. (a) Two levels of thresholding. (b) Vessel segmentation. (c) 3D skeletonization. (d) Feature point orientations. **Bottom:** 2D Feature extraction from the X-ray images. (e) Original image. (f) Vesselness segmentation. (g) Extracted feature points (h) Estimated feature orientations.

3. Non-rigid reconstruction and matching: Using a Kalman-based approach, we iteratively solve for the 3D-to-2D correspondences and progressively fit the deformation modes onto the 3D points.

In the following sections we discuss each one of these constituent pieces.

4. Feature Extraction

We next briefly describe the steps we perform to segment the vessel regions and extract the 3D and 2D features from the raw 3D scans and from the X-ray images respectively. Note that this task is by itself an active area of research within medical imaging [4, 7, 20]. Yet, we design relatively simple solutions, partially based on the popular vessel enhancement filter proposed in [6], which although not being free of error, demonstrate the robustness of our algorithm for 3D non-rigid reconstruction.

4.1. 3D Features

The raw data of the volumetric 3D CT scanner is composed of a stack of grayscale image slices. The CT volume is initially segmented using the fuzzy connectedness tree algorithm [28], and as shown in Fig. 2a, a number of regions corresponding to different physiological structures are generated. Fig. 2b shows the vessel-like formations. The vessel centerlines are then accurately detected based on a local steepest gradient ascent of the vesselness measure [6]. Finally, we homogeneously distribute an arbitrary number of nodes along the skeleton, and enforce connection using a minimum spanning tree approach [15](Fig. 2c).

In order to perform the 3D-to-2D matching in a subsequent step, besides using the 3D position, we describe each node based on a unitary vector that indicates the local artery orientation, which is computed from the eigenvectors of the Hessian matrix at the appropriate scale. Although we could

have used more elaborated descriptors, we found the position and orientation to be discriminative enough in our context of 3D heart scans, where despite non-rigid deformations, the global orientation of the heart remains quite stable throughout the cardiac cycle. In Fig. 2d we depict the orientation assigned to each node of the tree.

4.2. 2D Features

A reliable and robust detection of the vessel centerlines in the 2D X-ray images is more challenging than in the 3D case. Occlusions produced by vessel crossings, reduced local contrast in the diaphragm area, discontinuities due to lack of the contrast-agent (or stenosis), are some of the artifacts we may find in these images. Even though a Digital Subtraction Angiography technique can help to reduce some of these problems, its practical use is limited to sequences in which the relative pose of the X-ray equipment does not change with respect to the patient, which is not our case. In addition, while in practice the 3D feature extraction can be performed off-line in a pre-operative stage, 2D features must be computed in real-time and without manual intervention. As a consequence, we will only extract points of interest and will not attempt to solve the connectivity tree of the vascular structure.

The extraction of points of interest is performed by applying the multi-scale vesselness filtering proposed in [6]. We then perform a non-maxima suppression on the vesselness map, followed by a thresholding. For each detected location, a ridge traversal algorithm, similar to the one proposed in [3] is applied, providing the centerlines of arteries. In order to select a specific number of keypoints and ensure they are homogeneously distributed, we further apply a k-means clustering algorithm, were we set k to the number of desired keypoints. Finally, each of the keypoints is again described by the vessel orientation at the point, computed from the eigenvectors of the Hessian matrix. Fig. 2(g,h)

show the final keypoints and their respective orientations.

5. Non-rigid Reconstruction Algorithm

In this section we discuss the core elements of our approach. After formalizing the problem, we introduce the recursive model we use to represent the deformable coronary tree, as its particular representation will play a decisive role for being robust to high levels of occlusions and allowing to retrieve large deformations. We then describe the weak priors we use to build a probabilistic generative model of the tree, and finally we describe the iterative algorithm we propose that combines all the previous ingredients to solve for the shape and assign correspondences.

5.1. Problem formulation

Let $\mathcal{M}^{ref} = \{\mathbf{x}_1^{ref}, \ldots, \mathbf{x}_{n_m}^{ref}\}$ be the n_m model points we have extracted from the 3D reference scan and $\mathcal{U} = \{\mathbf{u}_1, \ldots, \mathbf{u}_{n_f}\}$ the n_f feature points extracted from the 2D X-ray image, corresponding to a projection of $\mathcal{M}$, a non-rigidly deformed version of $\mathcal{M}^{ref}$. Our goal is to retrieve both as many 3D-to-2D correspondences $\{\mathbf{x}^{ref}, \mathbf{u}\}$ as possible, and the 3D configuration of the deformed structure $\mathcal{M} = \{\mathbf{x}_1, \ldots, \mathbf{x}_{n_m}\}$. Note that since in practice the keypoints are obtained from uniformly sampling the segmented 3D scans and X-ray images, it may not exist a perfect one-to-one match between the sets $\mathcal{M}^{ref}$ and $\mathcal{U}$. However, in the synthetic results section we will show that our algorithm tolerates large amounts of noise and occlusions, which compensates for all these inaccuracies in the matches.

5.2. Recursive 3D Model Parameterization

Since we know the links between the 3D points of $\mathcal{M}$, we can represent the structure of the deformable model by the vector:

$$\mathbf{m} = [\mathbf{x}_1^\top, \boldsymbol{\rho}_2^\top, \ldots, \boldsymbol{\rho}_{n_m}^\top]^\top \qquad (1)$$

where $\mathbf{x}_1$ are the 3D coordinates of the root node, and $\boldsymbol{\rho}_j = [r_{ij}, \theta_{ij}, \phi_{ij}]^\top$ are the spherical coordinates of the vector joining the i-th and j-th nodes. Thus, the 3D position of a node $\mathbf{x}_k$ may be recursively written as:

$$\mathbf{x}_k = \mathbf{x}_1 + \sum_{i,j \in \mathcal{A}_k} \begin{bmatrix} r_{ij} \cos\theta_{ij} \sin\phi_{ij} \\ r_{ij} \sin\theta_{ij} \sin\phi_{ij} \\ r_{ij} \cos\phi_{ij} \end{bmatrix}$$

where $\mathcal{A}_k$ contains all the ancestors of the k-th node. Observe that using this formulation, when the 3D position of a node is updated, the location of all its ancestors is also updated. This is a remarkable novel contribution of our formulation, as it naturally introduces constraints that go beyond local neighborhoods. In addition, the propagation of the error using this recursive parameterization is very well suited to deal with structures like the coronary tree, in which the root node remains almost at constant position while the terminal nodes are usually highly deformed.
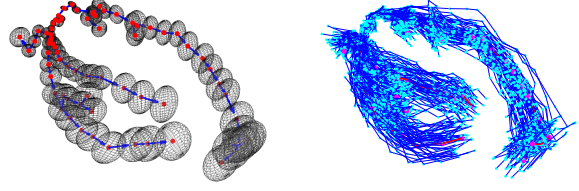

Figure 3. **Building weak deformation priors.** Since we do not explicitly use training data, we build a weak prior of the tree structure by assigning gaussian noise to its nodes (**left**). Note that since our parameterization is recursive, we assign larger noise values to the terminal nodes of the tree. Using this prior, we then produce multiple random shape samples (**right**) and fit a PPCA model.

5.3. Introducing Synthetic Deformation Priors

Although we do not explicitly use training data besides the reference 3D scan, we synthetically define weak priors on the feasible deformations, which will be used within the optimization scheme to progressively fit the 3D model to the input data.

Given the 3D parameterization $\mathbf{m}^{ref}$ of the reference shape, we compute the weak priors by first generating multiple samples $\{\mathbf{s}_i\}$ from the normal distribution $\mathcal{N}(\mathbf{m}^{ref}, \boldsymbol{\Sigma}_m)$, where $\boldsymbol{\Sigma}_m$ is a $3n_m \times 3n_m$ diagonal covariance matrix generated by concatenating the covariances of $\mathbf{x}_1$ and the vectors $\{\boldsymbol{\rho}_j\}_{j=2,\ldots,n_m}$. As shown in Fig. 3 these covariances are set to relatively large values in order to deal with different types of deformations. Yet, in order to avoid completely random shapes with no physical meaning we slightly smooth the resulting sampled shapes. In addition, the variance associated to the inter-node lengths r_{ij} is set to a very small value, as we assume the vessels do not stretch. Note that although real vessels may stretch, we represent the tree by a sufficiently dense set of points that makes the inextensibility assumption locally correct.

We then learn a low-dimensional deformation model of the coronary tree, by applying Probabilistic Principal Component Analysis [25] over the set of deformed samples $\{\mathbf{s}_i\}$. By doing this, we can approximate the pose parameters of Eq. 1 as a weighted sum of a mean structure $\mathbf{m}_0$ and n_q deformation modes $\mathbf{Q} = [\mathbf{q}_1, \ldots \mathbf{q}_{n_q}]$:

$$\mathbf{m} = \mathbf{m}_0 + \sum_{i=1}^{n_q} \alpha_i \mathbf{q}_i = \mathbf{m}_0 + \mathbf{Q}\boldsymbol{\alpha} \qquad (2)$$

where $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_{n_q}]^\top$ are unknown *modal weights* that define the current structure, and whose covariance is defined by a $n_q \times n_q$ matrix $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}$.

5.4. Iterative Fitting and 3D-to-2D Matching

Representing the coronary tree by means of the modal weights of Eq. 2 allows rewriting the problem as that of estimating the parameters $\boldsymbol{\alpha}^*$ such that the following reprojection error is minimized
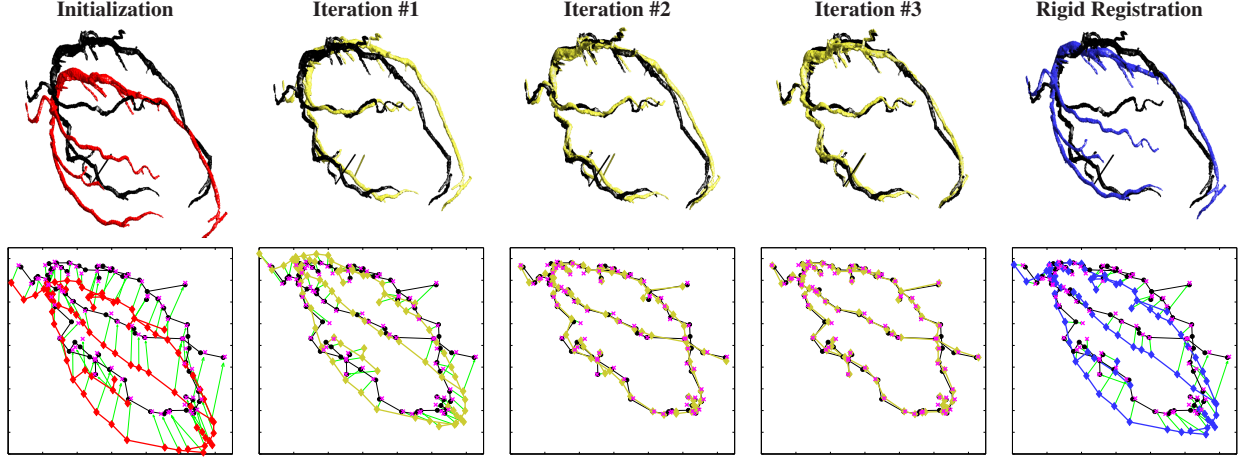
Figure 4. Iterative Fitting and 3D-to-2D Matching. We plot the 3D results (**top**) and 2D results (**bottom**) of the fitting process at different iterations and the matches we establish. The left-most figures show the initial prior (red) and the ground truth structure. The next 3 images show the fitting process of our approach (yellow) for different iterations. The figures on the right show what might be obtained using just a rigid registration, which is a standard solution to this kind of problems. Obviously the rigid approach yields large errors, specially on the terminal nodes of the coronary structure.

$$\boldsymbol{\alpha}^* = \arg\min_{\boldsymbol{\alpha}} \sum_{i=1}^{n_m} \|\mathsf{Proj}(\mathbf{x}_i; \boldsymbol{\alpha}) - \mathsf{Match}(\mathbf{x}_i, \mathcal{U})\| \quad (3)$$

where $\mathsf{Match}(\mathbf{x}_i, \mathcal{U})$ returns the match $\mathbf{u}_j \in \mathcal{U}$ for a 3D point $\mathbf{x}_i$, and $\mathsf{Proj}(\mathbf{x}_i; \boldsymbol{\alpha})$ is the perspective projection of the 3D point $\mathbf{x}_i$ given the modal weights $\boldsymbol{\alpha}$.

Note that we assume we know the projection matrix of the imaging system, including both the pose of the patient with respect to the CT and X-ray equipment, and the intrinsic parameters of the detector. In practice, though, there is always some translation in the position of the patient with respect to the global reference system. In addition, the heart deformation during the cardiac cycle and the global heart shift produced by the breathing introduce additional translation and small rotation effects. In order to address these issues, we have considered the translations and small rotations as part of the non-rigid model.

We next turn to the algorithmic steps to minimize Eq. 3. We achieve this by alternatively solving correspondences and fitting the 3D model, which we have initialized with the reference 3D model provided by the 3D scanner.

Establishing 3D-to-2D Correspondences. Let us assume that as input of an iteration we have $\boldsymbol{\alpha}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}$, the modal weights and their covariance matrix estimated at the previous iteration, the set $\mathcal{M}$ of 3D model points deformed according the weights $\boldsymbol{\alpha}$, and the set $\mathcal{U}$ of 2D features extracted from the X-ray image. We then compute 3D-to-2D matches with the following steps:

1. We project the 3D model points onto the image, considering the current configuration of modal weights. We denote these projections by $\mathcal{V} = \{\mathbf{v}_1, \ldots, \mathbf{v}_{n_m}\}$ where $\mathbf{v}_i = \mathsf{Proj}(\mathbf{x}_i; \boldsymbol{\alpha})$.

2. For each $\mathbf{v}_i$ we establish an uncertainty region surrounding it, by propagating the covariance $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}$ of the modal weights to the image plane. This region will be a Gaussian centered at each $\mathbf{v}_i$ and with covariance:

$$\boldsymbol{\Sigma}_{\mathbf{v}}^i = \mathbf{J}(\mathbf{x}_i)\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}\mathbf{J}(\mathbf{x}_i)^\top$$

where $\mathbf{J}(\mathbf{x}_i)$ is the $2 \times n_q$ Jacobian of the projection equation $\mathsf{Proj}(\mathbf{x}_i; \boldsymbol{\alpha})$ with respect to the modal weights $\boldsymbol{\alpha}$, evaluated at the 3D point $\mathbf{x}_i$.

3. Given the set $\{\mathbf{v}_i, \boldsymbol{\Sigma}_{\mathbf{v}}^i\}$ we solve the matching with the points $\mathbf{u}_j \in \mathcal{U}$ as a standard *Optimal Assignment Problem* using the Hungarian algorithm [12]. For this purpose, for each potential match $\{\mathbf{v}_i, \mathbf{u}_j\}$ we introduce a cost defined as a linear combination of the Mahalanobis distance between the two points and their similarity in the orientation computed in Section 4:

$$C_{ij} = \lambda_1 \mathsf{Mah}(\mathbf{v}_i, \mathbf{u}_j) + \lambda_2 \, \mathsf{Angle}(\mathbf{v}_i, \mathbf{u}_j)$$

where $\mathsf{Mah}(\mathbf{v}_i, \mathbf{u}_j) = (\mathbf{v}_i - \mathbf{u}_j)(\boldsymbol{\Sigma}_{\mathbf{v}}^i)^{-1}(\mathbf{v}_i - \mathbf{u}_j)^\top$, and $\mathsf{Angle}(\mathbf{v}_i, \mathbf{u}_j)$ is the difference in the orientations between the points of $\mathbf{u}_j$ and $\mathbf{x}_i$, the 3D model point projected on $\mathbf{v}_i$. Note that since the orientation vector of $\mathbf{x}_i$ is originally computed in the 3D space, we need to project it on the image plane to compare it with the orientation of $\mathbf{v}_i$. The terms λ_1 and λ_2 are constant scale factors used to give similar orders of magnitude to each of the components of the cost function. In practice, although the Hungarian algorithm is already an efficient technique for determining the optimal matching, we further reduce its complexity by only considering those costs C_{ij} which are below a certain threshold. This also prevents from fitting the shape to outlier correspondences or mismatches.
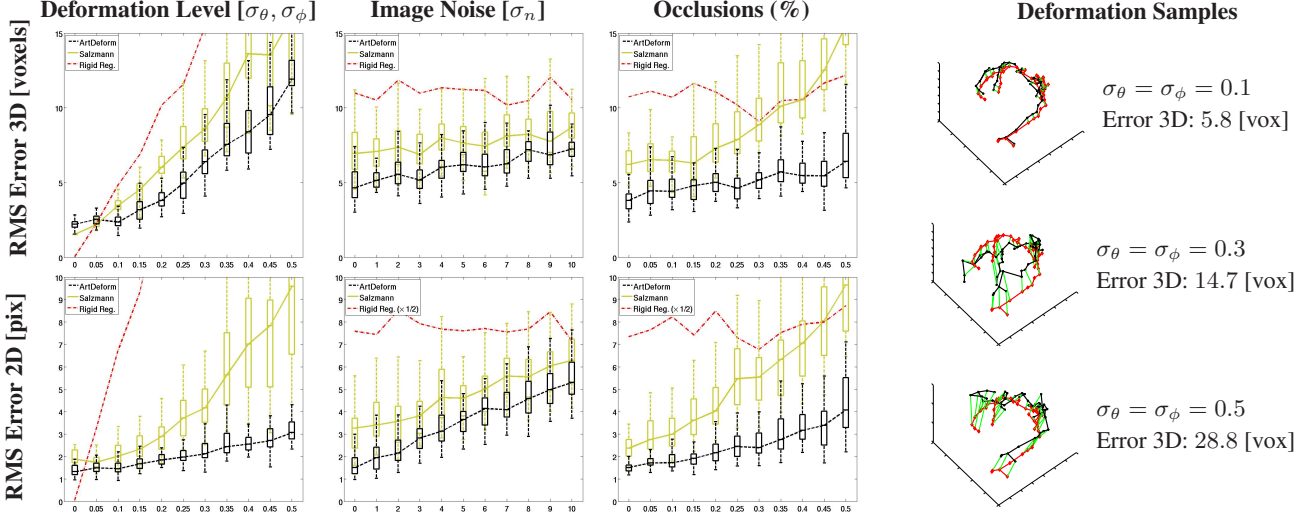
Figure 5. Synthetic experiments. RMS error distribution for 3D reconstruction (top) and 2D reprojection (bottom) for each experiment. The whiskers denote min. and max. errors, the box spans from first to third quartile and the inter-box lines show the mean RMS error.

Updating Modal Weights. Given the set $\{\mathbf{x}_i, \mathbf{u}_i\}_{i=1,\dots,n_c}$ of estimated correspondences we then use Kalman filter equations to update the modal weights $\boldsymbol{\alpha}$ and its covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}$. Since we simultaneously use all of the estimated matches, we define an extended $3n_c$ vector of 3D points $\hat{\mathbf{x}} = [\mathbf{x}_1^\top, \dots, \mathbf{x}_{n_c}^\top]^\top$, an extended $2n_c$ vector of 2D locations $\hat{\mathbf{u}} = [\mathbf{u}_1^\top, \dots, \mathbf{u}_{n_c}^\top]^\top$, and an extended $2n_c \times n_q$ Jacobian matrix $\hat{\mathbf{J}} = [\mathbf{J}(\mathbf{x}_1)^\top, \dots, \mathbf{J}(\mathbf{x}_{n_c})^\top]^\top$. We then update $\boldsymbol{\alpha}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}$ as

$$\boldsymbol{\alpha}^+ = \boldsymbol{\alpha} + \mathbf{K}\left(\hat{\mathbf{u}} - \mathsf{Proj}(\hat{\mathbf{x}}; \boldsymbol{\alpha})\right)$$
$$\boldsymbol{\Sigma}_{\boldsymbol{\alpha}^+} = (\mathbf{I} - \mathbf{K}\hat{\mathbf{J}})\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}$$

where $\mathbf{K}$ is the Kalman gain and $\mathbf{I}$ the identity matrix.

Iterating and Increasing Flexibility. The matching and modal weight updating processes are iteratively repeated until the convergence of Eq. 3.

In order to adapt the reference model to highly deformed structures, such as those shown in Fig. 1, the number of modes we use is increased at each iteration. This allows to progressively fit the structure, starting from the most rigid parts up to the more deformed ones. In addition, using more rigid structures at early stages yields robustness to mismatches, preventing to adapt the model towards outlier 2D features. In practice, for a coronary tree with n_m points, and thus with a maximum number of $3n_m$ modes, at iteration number n_{iter} we used $n_q = 3n_m/(10 - n_{iter})$ modes. We found the number of iterations to converge to be always ≤ 5. For instance, Fig. 4 shows an example in which convergence was achieved after 3 iterations.

6. Results

We now present the results on both synthetic and real data. In the synthetic results we compare our approach (de-noted *ArtDeform*) to [18][1], which is a representative example of the state-of-the-art in articulated pose recovery that has been shown successful in recovering human and hand pose. We also compare it to the solution obtained from a 3D-to-3D rigid transformation between the reference 3D shape and the ground truth shape associated to the input image, which would represent the best solution that a rigid 2D-to-3D registration algorithm could obtain. For this purpose, we use a standard technique for absolute orientation estimation [10]. Note that both in [18] and [10], the correspondences are assumed to be known, while in our approach we simultaneously estimate them with the shape.

6.1. Synthetic Experiments

In this section we extensively evaluate the performance of each algorithm against noise in the correspondences, different levels of deformation and partial occlusions. We generated random shapes of 50 nodes within a volume of $300 \times 300 \times 300$ voxels, such as the tree-like structure shown in Fig. 1, and simulated the deformations undergone in the coronary tree by applying increasing levels of noise $\{\sigma_\theta, \sigma_\phi\}$ to the joint angles. We then projected each 3D shape on a 512×512 image and added gaussian noise of standard deviation σ_n to the 2D correspondences. In addition, a percentage p_o of the projected points was randomly removed in order to simulate partial occlusions. Given the original reference shape and the set of projected points of the deformed shape, we then performed the reconstruction with each of the algorithms.

Three different types of experiments were performed. We initially evaluated the amount of deformation each algorithm was able to recover by sweeping the variances of

---

[1]We thank Dr. Mathieu Salzmann for kindly testing the data of our experiments on the algorithm proposed in [18].

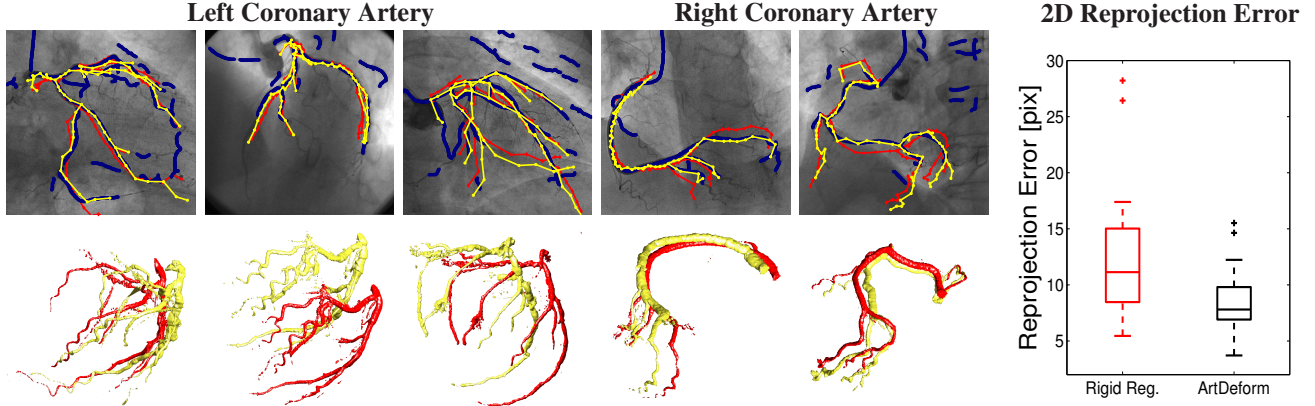**Left Coronary Artery** | **Right Coronary Artery** | **2D Reprojection Error**

Figure 6. **Reconstruction results on real vessel structures. Left:** Several samples of the LCA and RCA reconstructions. Upper row: initial model (red) and the final non-rigid tree (yellow) overlaid on the input X-ray images. We also plot the results of the 2D segmentation (blue), which, as can be seen, contains certain amounts of false positives. Bottom row: Initial and retrieved 3D shapes. Note that despite the segmentation suffers from occlusions and false positive regions our approach yields an accurate solution. **Right:** 2D reprojection error, which is used to quantitatively assess the quality of the real results, because the 3D ground truth is not available.

the joint angles within the range $[\sigma_\theta, \sigma_\phi] \in [0 - 0.5]$ rad, and randomly setting $\sigma_n \in [1 - 3]$ and $p_o \in [5 - 20]\%$. To give significance to the levels of deformation and reconstruction errors, Fig. 5-right depicts different deformations of the model corresponding to specific values of $\{\sigma_\theta, \sigma_\phi\}$. In a second experiment we analyzed the robustness to image noise by synthetically introducing random noise of $\sigma_n \in [0 - 10]$, and setting $\{\sigma_\theta, \sigma_\phi\} \in [0.15 - 0.3]$ and $p_o \in [5 - 20]\%$. Finally, we evaluated the effect of occlusions in an interval $p_o \in [0 - 50]\%$, with $\{\sigma_\theta, \sigma_\phi\} \in [0.15 - 0.3]$ and $\sigma_n \in [1 - 3]$.

For each set of parameters we performed 50 trials. The graphs on Fig. 5 depict the mean 2D reprojection error, expressed in pixels, and the mean 3D reconstruction error, expressed in voxels. Observe that our algorithm consistently outperforms [18] in all experiments. This difference is specially remarkable when dealing with occlusions, for which we obtain reconstruction errors below 5 voxels even when a 50% of the model is occluded. Observe in the top right graph of Fig. 5 that these amounts of error correspond to very good approximations. It is fair to mention, though, that [18] is a general algorithm easily adaptable to different domains, from articulated structures to deformable surfaces, while our algorithm is specifically designed to handle tree-like and articulated structures. In Fig. 5 we also plot the results that would be recovered using a rigid registration and, as expected, the errors are significantly larger. Note that the error values for this case are scaled by a factor $1/2$ for displaying purposes.

6.2. Real Data

We also evaluated our approach on real CT data and X-ray data collected during ordinary pre-operative diagnosis and percutaneous intervention. We collected CT data of 7

patients, using a Philips Brilliance iCT, at the 75% of the heart cycle, with slice thickness 0.67 or 0.8mm, and pixel resolution between 0.38×0.38 and 0.45×0.45 mm. We collected a total of 17 X-ray sequences, 10 of Left Coronary Artery (LCA) and 7 of Right Coronary Artery (RCA), using a single plane Philips INTEGRIS Allura Flat Detector. Image and camera calibration was performed using the catheter width and the geometrical information on the C-Arm position. For each sequence, one image in which the contrast liquid was sufficiently visible was selected.

For each of pair CT scan/X-ray image, we then extracted 3D and 2D features as described in Sect. 4. In all experiments, we represented the segmented CT volume as a tree with 75-nodes, and extracted 500 feature points from each X-ray image. Starting with the initial tree of the CT scan, we then iteratively fit the model and established 3D-to-2D matches. In all experiments we achieved convergence in less than 5 iterations, taking about 8 seconds per iteration. Note that this represents in fact a significantly faster algorithm compared to competing methods. For instance, [8] reports computation times of about 7 minutes per image.

The 2D registration and 3D reconstruction results of a few sample experiments are depicted in Fig. 6. Observe that even when the segmentation contains false positive regions, or does not detect some branches of the coronary tree our approach is able to provide an accurate solution.

Since the 3D ground truth does not exist for the deformed artery tree, we quantitatively evaluated the performance of our algorithm based on the 2D reprojection error with respect to ground truth centerlines manually annotated by an expert physician. This error is shown in the bar plot of Fig. 6-right, which summarizes the results for all the 17 experiments. Observe that our non-rigid approach clearly outperforms a method that rigidly registers the original CT

scan. In fact, considering an average calibrated pixel resolution of 0.22 mm, the median error of our method is of about 1.9 mm. This compares very well with the 1 mm error reported in [11], especially considering that they restrict their evaluation to X-ray images acquired at systole and diastole time instants, where the coronary tree deformation is minimal compared to other cardiac cycle instants we contemplate.

7. Conclusion

We have presented a novel approach to estimate the 3D structure of the coronary tree from single X-ray images. In order to handle the large amounts of deformation, noise and occlusions present in this kind of images we have introduced a generative model based on a recursive parameterization that progressively increases its complexity. We have integrated this model within a Kalman filter framework which, making use of very weak priors on the structure, iteratively guides the matching process while recovering coarse-to-fine levels of deformation.

The formulation we propose is fairly general, and allows integrating additional features. As part of future work, we consider exploiting motion coherence for tracking heart beat sequences in real time. Moreover, we believe that the inextensibility constraints between neighboring nodes may be relaxed, thus allowing to handle stretchable structures.

Acknowledgments

References

[1] A. Agarwal and B. Triggs. 3d human pose from silhouettes by relevance vector regression. In *CVPR*, 2004.

[2] D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, and J. Davis. Scape: shape completion and animation of people. *ACM SIGGRAPH*, 24:408–416, 2005.

[3] C.Bauer, H.Bischof. Edge Based Tube Detection for Coronary Artery Centerline Extraction. *MICCAI Workshop*, 2008.

[4] F. Benmansour, L. Cohen, M. Law, and A. Chung. Tubular anisotropy for 2d vessels segmentation. In *CVPR*, 2009.

[5] A. Ecker, A. D. Jepson, and K. N. Kutulakos. Semidefinite programming heuristics for surface reconstruction ambiguities. In *ECCV*, 2008.

[6] A. Frangi, W. Niessen, K. Vincken, and M. Viergever. Multiscale vessel enhancement filtering. In *MICCAI*, 1998.

[7] G.Gonzalez, E.Turetken, F.Fleuret, P.Fua. Delineating trees in noisy 2d images and 3d image-stacks. In *CVPR*, 2010.

[8] M. Groher, D. Zikic, and N. Navab. Deformable 2d-3d registration of vascular structures in a one view scenario. *IEEE Trans. on Medical Imaging*, 2009.

[9] J. Hipwell, G. Penney, R. McLaughlin, K. Rhode, P. Summers, T. Cox, J. Byrne, J. Noble, and D. Hawkes. Intensity based 2d-3d registration of cerebral angiograms. *IEEE Trans. on Medical Imaging*, 2003.

[10] B. Horn, H. Hilden, and S. Negahdaripour. Closed-form solution of absolute orientation using orthonormal matrices. *Journal Optical Society of America*, 1988.

[11] C. Metz et al. Patient specific 4d coronary models from ecg-gated cta data for intra-operative dynamic alignment of cta with x-ray images. In *MICCAI*, 2009.

[12] J. Munkres. Algorithms for the assignment and transportation problems. *J. Soc. Ind. and Applied Mathematics*, 1957.

[13] A. Myronenko and X. Song. Point set registration: Coherent point drift. *IEEE Trans. PAMI*, 2010.

[14] M. Perriollat, R. Hartley, and A. Bartoli. Monocular template-based reconstruction of inextensible surfaces. In *BMVC*, 2008.

[15] R. Prim. Shortest connection networks and some generalizations. *Bell System Technical Journal*, 1957.

[16] D. Ruijters, B. M. ter Haar Romeny, and P. Suetens. Vesselness-based 2d3d registration of the coronary arteries. *Int. J. Comp. Ass. Radiology and Surgery*, 2009.

[17] M. Salzmann, F. Moreno-Noguer, V. Lepetit, and P. Fua. Closed-form solution to non-rigid 3d surface registration. In *ECCV*, 2008.

[18] M. Salzmann and R. Urtasun. Combining discriminative and generative methods for 3d deformable surface and articulated pose reconstruction. In *CVPR*, 2010.

[19] J. Sanchez, J. Ostlund, P. Fua, and F. Moreno-Noguer. Simultaneous pose, correspondence and non-rigid shape. In *CVPR*, 2010.

[20] M. Schaap et al. Standardized evaluation methodology and reference database for evaluating coronary artery centerline extraction algorithms. *Medical Image Analysis*, 2009.

[21] G. Shakhnarovich, P. Viola, and T. Darrell. Fast pose estimation with parameter-sensitive hashing. In *ICCV*, 2003.

[22] L. Sigal, A. O. Balan, and M. J. Black. Combined discriminative and generative articulated pose and non-rigid shape estimation. In *NIPS*, 2007.

[23] L. Sigal and M. J. Black. Humaneva: Synchronized video and motion capture dataset for evaluation of articulated human motion. In *Technical Report*, Brown University, 2006.

[24] C. Sminchisescu, A. Kanaujia, and D. Metaxas. Learning joint top-down and bottom-up processes for 3d visual inference. In *CVPR*, 2006.

[25] M. Tipping and C. Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society*, 1999.

[26] G.A. Turgeon et al. 2d-3d registration of coronary angiograms for cardiac procedure planning and guidance. *Medical Physics*, 2005.

[27] R. Urtasun and T. Darrell. Local probabilistic regression for activity-independent human pose inference. In *CVPR*, 2008.

[28] C. Wang and O. Smedby Coronary artery segmentation and skeletonization based on competing fuzzy connectedness tree. In *MICCAI*, 2007.

[29] J. Yao and R. Taylor. Assessing accuracy factors in deformable 2d/3d medical image registration using a statistical pelvis model. In *ICCV*, 2003.