A Machine Learning Approach for Deformable Guide-Wire Tracking in Fluoroscopic Sequences

Olivier Pauly, Hauke Heibel, and Nassir Navab

Computed Assisted Medical Procedures,
Technische Universität München, Germany

Abstract. Deformable guide-wire tracking in fluoroscopic sequences is a challenging task due to the low signal to noise ratio of the images and the apparent complex motion of the object of interest. Common tracking methods are based on data terms that do not differentiate well between medical tools and anatomic background such as ribs and vertebrae. A data term learned directly from fluoroscopic sequences would be more adapted to the image characteristics and could help to improve tracking. In this work, our contribution is to learn the relationship between features extracted from the original image and the tracking error. By randomly deforming a guide-wire model around its ground truth position in one *single* reference frame, we explore the space spanned by these features. Therefore, a guide-wire motion distribution model is learned to reduce the intrisic dimensionality of this feature space. Random deformations and the corresponding features can be then automatically generated. In a regression approach, the function mapping this space to the tracking error is learned. The resulting data term is integrated into a tracking framework based on a second-order MAP-MRF formulation which is optimized by QPBO moves yielding high-quality tracking results. Experiments conducted on two fluoroscopic sequences show that our approach is a promising alternative for deformable tracking of guide-wires.

1 Introduction

During the last decade, the success of angiographic interventions relied on the ability of physicians to navigate in the patient's anatomy based only on their mental three-dimensional representation of the human body as well as on the haptic feedback from the instruments. Recent advances in computer aided planning and navigation techniques offer great potential of minimizing the risk of complications and improving the precision. In the case of angiographic applications, the most common imaging modality is X-ray fluoroscopy. Currently, in order to monitor guidance procedures, a roadmap, e.g. a digital subtracted angiography (DSA) showing vessel anatomy, is computed during the intervention. Unfortunately, such roadmaps cannot directly be fused with the intra-operative fluoroscopic sequence due to misalignment caused by respiratory motion. A fundamental step toward a successful integration of any navigation application into clinical routine is the estimation and compensation of such respiratory motion.

Determining this spatio-temporal information is a challenging task due to the fact that fluoroscopic X-ray images have a low signal to noise ratio, are subject to big changes in contrast and suffer from background clutter in the abdominal area. Moreover, the apparent motion of the guide-wire is a combination of multiple components. The major motion in the chest is caused by patient breathing. A second, deformable

T. Jiang et al. (Eds.): MICCAI 2010, Part III, LNCS 6363, pp. 343–350, 2010.
© Springer-Verlag Berlin Heidelberg 2010

component results from forces applied to the guide-wire by the physician and by surrounding organs which are subject to non-uniform motions during the breathing cycle. Furthermore, the guide-wire may sometimes partially vanish.

A recent approach dealing with the problem of guide-wire tracking in fluoroscopy is [5]. In this work, Heibel et al. proposed a scheme for deformable tracking based on a MAP-MRF formulation. However, their data term does not differentiate well between medical tools and anatomic background such as ribs and vertebrae. A learned data term being more robust and adapted to the image characteristics of fluoroscopic sequences could help to further improve the tracking. Since MRF formulations are derivative free optimization procedures, they ease the integration of such learning based energies for which analytical derivatives are hard to derive if possible at all. Learning permits to model complex relationships between the information contained in the images and the quality of alignment.

In the context of guide-wire tracking, we can distinguish two kinds of learning approaches: First, methods for the detection of the guide-wire in each frame and second, methods used for learning a data driven energy. A learning-based tracking approach by detection based on marginal space learning was presented by Barbu et al. in [1]. Later, Wang et al. proposed in [11] the combination of learning-based detectors and online appearance models. In the case of energy learning, Nguyen et al. [7] addressed the problem of modeling the error surface of parametric appearance models in order to minimize the number of local minima for image alignment and recently Pauly et al. suggested in [8] to learn the statistical relationship between two different imaging modalities to model a data term for multi-modal rigid registration.

In this work, our contribution is a learning approach for deformable tracking: we propose to learn a data term based on the relationship between features extracted from the original image and the tracking error. As features, we introduce local mean orthogonal intensity profiles representing information contained in the original image. Since deformable transformations have a high number of degrees of freedom, the intrisic dimensionality of the space spanned by these features is high. However, typical guide-wire deformations are lying on a subspace we propose to learn to reduce the complexity of our problem. A set of random deformations is then generated automatically and applied to the ground truth position of the guide-wire on a single reference image. A training set of data points from the corresponding local mean orthogonal profiles and their associated tracking error values is thereby created. Learning is then performed on this dataset with a support vector regression. The resulting data term is integrated into a tracking framework based on a MAP-MRF formulation which is solved with higher-order clique reduction techniques. Due to the higher-order nature of our problem and since we are dealing with non-submodular energy functions we chose a combination of the recently proposed reduction scheme of Ishikawa [6] and the QPBO [4] optimizer supporting improvements in order to deal with unlabeled nodes [9].

The remainder of the paper is organized as follows: Section 2 presents our regression approach to define an optimal data term for guide-wire tracking. Section 3 reports experiments performed on two fluoroscopic sequences. Results show that our approach presents a promising alternative for guide-wire tracking in fluoroscopic sequences. Section 4 concludes the paper and gives an outlook on future work.

2 Methods

2.1 Problem Statement

The goal of tracking is to identify the relative motion of an object in a series of consecutive frames. In most tracking algorithms, we can distinguish two phases: first, the detection of the object of interest in the initial frame followed by the actual tracking in each new frame given previous positions. In this paper, we focus on the problem of tracking a guide-wire through a fluoroscopic sequence knowing its initial position. Let us denote $\mathcal{C}$ our guide-wire model and $\{I_t\}_{t\in\{0,\ldots,T\}}$ the set of consecutive images in which we want to track the guide-wire. In fluoroscopic images, guide-wires appear as curvilinear structures which can be represented as B-spline curves. The advantage of such a representation is its low-dimensionality, its implicit smoothness and its local support of control points. Our guide-wire model $\mathcal{C}$ is defined as the following linear combination of control points:

$$\mathcal{C}(s) = \sum_{i=1}^{M} N_i(s) P_i \quad \text{where } s \in [0,1] \tag{1}$$

where N_i denote the basis functions and P_i the positions of M control points. By using this model, we want to estimate the optimal curve parameters, i.e. the best configuration of the control points, to match the visible structures in an image, and this, knowing its previous position. The tracking problem can be then formulated as a maximum a posteriori estimation:

$$\mathcal{C}_t^* = \underset{\mathcal{C}_t}{\operatorname{argmax}}\, P(I_t|\mathcal{C}_t)P(\mathcal{C}_t) \tag{2}$$

where $\mathcal{C}_t^*$ is the best curve estimate at instant t. $P(I_t|\mathcal{C}_t)$ is the likelihood of observing the data knowing the model and $P(\mathcal{C}_t)$ the prior or probability of the current curve configuration. Let us assume the likelihood to follow a Gaussian distribution and the prior a Gibbs' distribution, we can then reformulate Eq.2 as an energy minimization:

$$\mathcal{C}_t^* = \underset{\mathcal{C}_t}{\operatorname{argmin}}\, (E_{data}(I_t|\mathcal{C}_t) + E_{reg}(\mathcal{C}_t)) \tag{3}$$

$E_{reg}(\mathcal{C}_t)$ is a regularization term which constraints the space of possible model configurations. Assuming constant length of guide-wire segments in fluoroscopic sequences, we define the regularization term in order to penalize changes in length:

$$E_{reg}(\mathcal{C}_t) = \int_0^1 \left(1 - \frac{\|C_t'(s)\|}{\|C_0'(s)\|}\right)^2 ds \tag{4}$$

where $\mathcal{C}_t'$ and $\mathcal{C}_0'$ are the first derivatives at instant t and 0 respectively. Thank to the inherent smoothness of a B-spline representation, higher-order terms can be discarded. $E_{data}(I_t|\mathcal{C}_t)$ can be seen as a data term which drives the model according to the current image:

$$E_{data}(I_t|\mathcal{C}_t) = \int_0^1 \Phi(I_t(\mathcal{C}_t(s)))\, ds \tag{5}$$

A common choice for Φ is a function which enhances tubular structures similar to the ridgeness measure proposed by Frangi et al. [3]. Such measures can be tuned to emphasize only structures of the scale of the guide-wire and to remove outliers such as ribs or vertebrae. However, since the data term is only evaluated along the current position of the curve, the main drawback is a very low capture range and a lack of robustness in terms of outliers or partial occlusions.

Instead of relying on the feature image intensities along the curve profile, we propose to extract features from the unprocessed image orthogonally to the curve, namely *local mean orthogonal intensity profiles*. We can then model a data term by learning a function Ψ relating the space $\mathcal{M}$ spanned by these features and the tracking error. By using a single fluoroscopic image and a set of local displacements around the ground truth position of our guide-wire, we can sample the space $\mathcal{M}$ by extracting the local mean orthogonal intensity profiles associated to each displaced curve. Each of these "points" of $\mathcal{M}$ is then associated to a tracking error derived from the corresponding curve parameters, hereby generating a set of data points. Finally Ψ is modeled by performing a regression on these points. The following section presents how to extract the mentioned features.

2.2 Local Mean Orthogonal Profiles

In a fluoroscopic image, a human being may recognize the guide-wire because of its curvilinear aspect and its darker intensities compared to its environment. For this reason, a common method would be to enhance this structure and to keep track of it along the sequence by using a data term based on the intensity profile along the curve. Unfortunately, in the case of larger displacements between two consecutive frames, it is hard to relocate the guide-wire in a heterogeneous region containing outliers without any information about the search direction. Indeed, such data terms suffer from an extremely narrow valley around the global extremum. To overcome this problem and benefit from an increased capture range, we propose features which describe the intensity profiles *orthogonally* to the curve. First, we subdivide our curve $\mathcal{C}_t$ into n segments $\left\{S_t^k\right\}_{k\in\{1,\dots,n\}}$. Each segment S_t^k is a spline we characterize by the following descriptor $\mathcal{J}_t^k$:

$$\mathcal{J}_t^k = \frac{1}{q}\sum_{j=1}^{q} \Lambda_t^{k,j}, \tag{6}$$

with q being the number of sample points along this segment. $\Lambda_t^{k,j}$ is an orthogonal intensity profile whose r^{th} element is defined as:

$$\Lambda_t^{k,j}(r) = I_t\left(S_t^k(u) + r\cdot\mathbf{n}(u)\right) \tag{7}$$

where $\mathbf{n}(u)$ is the normal vector at point $u = (j-1)/(q-1)$ and $r \in \{-R, ..., R\}$. The dimensionality of this vector is $2R+1$ which corresponds to the length of the profile centered on the segment. Note that since only the profile's shape is of interest, each profile $\Lambda_t^{k,j}$ is normalized between 0 and 1. Taking the mean over the segment provides a feature vector which is more robust to noise and outliers. Each curve $\mathcal{C}_t$ is then described by the following set $\left\{\mathcal{J}_t^k\right\}_{k\in\{1,\dots,n\}}$.

2.3 Data Points Generation by Motion Learning

The goal of our approach is to learn a function Ψ relating the local mean orthogonal profiles and the tracking error:

$$\Psi : \mathcal{M} \to \mathbb{R}, \tag{8}$$

with good characteristics for tracking purposes, namely convexity and smoothness. Therefore, the space $\mathcal{M}$ spanned by these features needs to be sampled thoroughly as a function of the relative displacement. Since the guide-wire is a deformable structure, the intrisic dimensionality of our features according to free deformations would be high and thus, hard to sample. However, in a real fluoroscopic sequence, a guide-wire is not subject to free deformations. Indeed, main displacements are due to breathing motions and additional small deformations. This means that in reality, our features do not describe the full space $\mathcal{M}$ but lie on a lower dimensional subspace. To reduce the complexity of our problem, we propose to learn the deformation probability distribution from a real sequence. Thus, random displacements can be automatically generated to build our training dataset.

Learning guide-wire Motions: During a sequence, each segment S_t^k of our curve $\mathcal{C}_t$ is subject to a series of consecutive displacements we denote $\left\{D_t^k\right\}_{t\in\{0,\ldots,T-1\}}$. Each D_t^k is modeled by a vector containing the displacements of sample points of the segment between 2 consecutive frames. Its j^{th} element is defined as:

$$D_t^k(j) = S_{t+1}^k(u) - S_t^k(u) \tag{9}$$

These vectors are collected for all segments along the whole sequence and grouped in a training set $\mathcal{D} = \left\{D_t^k\right\}_{t\in\{0,\ldots,T-1\}}^{k\in\{1,\ldots,n\}}$. To learn the underlying probability distribution of these displacements, we propose to model it with a gaussian mixture model $\mathcal{G}$. The parameters of $\mathcal{G}$ can be estimated by using Expectation-Maximization. Once we have learned our gaussian mixture model, we can generate random segment displacements $\{D_i\}_{i\in\{1,\ldots,Q\}}$ from this probability distribution.

Data points generation: As shown on Fig.1, by using a reference fluoroscopic image, e.g. the first frame of the sequence, we can generate local mean orthogonal profiles $\{\mathcal{J}_i\}_{i\in\{1,\ldots,Q\}}$ by perturbating the segments of the ground truth curve with the

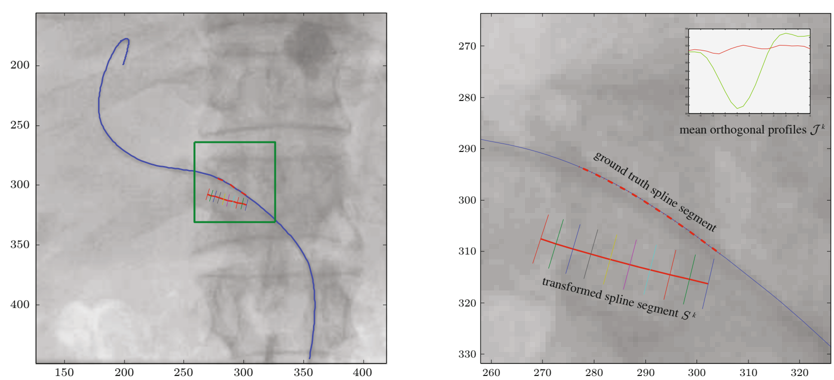

Fig. 1. Data term learning: by perturbating the ground-truth curve from a single frame with random displacements, we can build a training set of local mean orthogonal profiles with their associated tracking errors

randomly generated displacements $\{D_i\}_{i\in\{1,\dots,Q\}}$. The corresponding tracking error $\mathcal{E}_i$ associated to each $\mathcal{J}_i$ is computed as follows:

$$\mathcal{E}_i = \|D_i\|^2 \tag{10}$$

This procedure permits us to generate the set of pairs $\{(\mathcal{J}_i, \mathcal{E}_i)\}_{1,\dots,Q}$, on which the regression will be performed to learn our function Ψ.

2.4 Learning Data Term through Support Vector Regression

From previously generated data points, the function Ψ can be learned through non-parametric support vector regression. Let us consider the problem of fitting a function on the set of Q data points $\{(\mathcal{J}_i, \mathcal{E}_i)\}_{i\in\{1,\dots,Q\}}$. Ψ is modeled as the following function:

$$\Psi(\mathcal{J}) = \langle w\ ,\ \mathcal{J}\rangle + b, \tag{11}$$

where w is a weighting vector of dimensionality $\dim(\mathcal{M})$ and b a bias. This can be written as a convex optimization problem [10]:

$$\begin{aligned} &\text{minimize } \frac{1}{2}\|w\|^2 \\ &\text{subject to: } \begin{cases} \mathcal{E}_i - \langle w\ ,\ \mathcal{J}_i\rangle - b \leq \epsilon \\ \langle w\ ,\ \mathcal{J}_i\rangle + b - \mathcal{E}_i \leq \epsilon \end{cases} \end{aligned} \tag{12}$$

This aims at minimizing the norm of w to penalize the model complexity and the regression errors on the data points with a regression tolerance denoted by ϵ. Equation (12) corresponds to minimizing the following functional:

$$\begin{aligned} &\text{minimize } \frac{1}{2}\|w\|^2 + \mathrm{C}\sum_{i=1}^{Q}\left(\xi_i^+ + \xi_i^-\right) \\ &\text{subject to: } \begin{cases} \mathcal{E}_i - \langle w\ ,\ \mathcal{J}_i\rangle - b \leq \epsilon + \xi_i^+ \\ \langle w\ ,\ \mathcal{J}_i\rangle + b - \mathcal{E}_i \leq \epsilon + \xi_i^- \\ \xi_i^+, \xi_i^- \geq 0 \end{cases} \end{aligned} \tag{13}$$

where C weights the impact of the errors and thus the flexibility of the model. According to the Representer theorem, a solution w_{opt} of this minimization is always a linear combination of the training vectors in $\mathcal{M}$ with weights $\{\alpha_i\}_{i\in\{1,\dots,Q\}}$:

$$w_{opt} = \sum_{i=1}^{Q} \alpha_i\ \mathcal{J}_i \tag{14}$$

which leads to the following model:

$$\Psi(\mathcal{J}) = \sum_{i=1}^{Q} \alpha_i\ \langle \mathcal{J}_i\ ,\ \mathcal{J}\rangle + b \tag{15}$$

Finally, the global data term computed on all segments can be written as:

$$E_{data}^{learn}(\mathcal{C}_t) = \frac{1}{n}\sum_{k=1}^{n} \Psi(\mathcal{J}_t^k) \tag{16}$$

3 Experiments and Results

In the following experiments, we show the successful application of our machine learning approach for the tracking of guide-wires in fluoroscopic images. The two sequences we used for our experiments were accquired during liver chemoembolizations. In this procedure, a guide-wire is inserted into the femoral artery and threaded into the aorta. The catheter is then advanced into the hepatic artery. Once the branches that feed the liver cancer are reached, the chemotherapy is infused. In both sequences, the catheter is already inserted in the artery and we aim at recovering from breathing motions.

Motion learning: A set of inter-frame segment displacements is computed from a reference sequence where the guide-wire positions were manually annotated. A gaussian mixture model is then fitted to this dataset by using EM algorithm. The analysis of Bayes' Information Criterion leads to the choice of two gaussian components.

Data term learning: A quadratic B-spline is fit to each hand-labeled point set by minimizing discontinuities in the second derivative [2]. Given the previously learned gaussian mixture model, a set of $Q = 3000$ random segment displacements is automatically generated. By perturbating the ground-truth curve from a single frame with these random displacements, we can build a training set of 3000 local mean orthogonal profiles with their associated tracking errors. Note that the choice of Q is a compromise between complexity and accurate modeling of the data term. During the experiments profiles with different radii are evaluated. Finally, the data term is learned by performing a support vector regression.

Tracking experiments: Experiments are conducted on two clinical sequences of 142 and 228 frames with a resolution of 512×512 pixels and respective pixel spacings of 0.432×0.432 mm and 0.308×0.308 mm. In order to evaluate the tracking results, guide-wires are manually annotated in each frame. The following distance measure has been used throughout all experiments to assess the quantitative tracking quality:

$$\chi = \frac{1}{2}\left(\frac{1}{|\mathcal{C}_t|}\sum_{x_i \in \mathcal{C}_t} \min_{y \in \mathcal{C}_{\mathrm{GT}}} d(x_i, y)^2 + \frac{1}{|\mathcal{C}_{\mathrm{GT}}|}\sum_{y_j \in \mathcal{C}_{\mathrm{GT}}} \min_{x \in \mathcal{C}_t} d(x, y_j)^2\right). \quad (17)$$

Here $\mathcal{C}_{\mathrm{GT}}$ is the manually annotated curve and $\mathcal{C}_t$ the tracking result of an individual frame.

Results: Tab.1 shows mean errors on whole sequences where the data term is trained on the first frame of one sequence, and tested in tracking in both sequences. Submillimeter yet subpixel tracking accuracy can be achieved with our learned data-term and this, for a frame rate of 1.5 frame/s on a 3 Ghz duo core. Moreover, cross-validation illustrates the robustness of our approach even if it has been trained on another sequence showing different contrasts, motions and background. Note that since the Seq.1 presents motions of higher amplitude, its mean error is slightly bigger than for the other sequence. The great advantage is the ability to model the convexity and smoothness of this term. Indeed, its convexity properties can be designed by replacing the tracking error function 10. The choice of hyper-parameter C from equation (13) influences the flexibility of the regression and thus the smoothness of the resulting function.

Table 1. Tracking errors in real fluoroscopic sequences

Tracking Results								
Trained on	Seq.1				Seq. 2			
Tested on	Seq.1		Seq.2		Seq.1		Seq.2	
Profile Radius	5 pixels	10 pixels	5 pixels	10 pixels	5 pixels	10 pixels	5 pixels	10 pixels
χ mean (mm^2)	0.7115	0.5249	0.1636	0.1622	0.6632	0.5815	0.1796	0.1700
χ std dev (mm^2)	0.4289	0.2715	0.1633	0.1185	0.6184	0.3366	0.1771	0.1645

4 Discussion and Conclusion

In this work, our contribution was to learn the relationship between features extracted from an unprocessed image and the tracking error in order to model a data term. Experiments conducted on two fluoroscopic sequences show that our approach is a promising alternative for deformable guide-wire tracking. Indeed, our method is robust to changes in contrast, background clutter and partial occlusions of the guide-wire during the sequence, and this, even if training was performed on another dataset. Since the feature space under free deformations is high-dimensional, we proposed to model the distribution of the reduced space of typical guide-wire motions with a gaussian mixture model. In turn, this permitted us to automatically generate random guide-wire deformations from this distribution for the sake of regression. Going further, the space of relative motions between consecutive frames could be constrained during tracking to expected guide-wire motions. In future work, we will explore the possibility of deriving an adapted regularization term from this motion distribution model.

References

1. Barbu, A., Athitsos, V., Georgescu, B., Boehm, S., Durlak, P., Comaniciu, D.: Hierarchical learning of curves application to guidewire localization in fluoroscopy. In: CVPR (2007)
2. Dierckx, P.: Curve and surface fitting with splines. Oxford University Press, Inc., New York (1993)
3. Frangi, A.F., Niessen, W.J., Vincken, K.L., Viergever, M.A.: Multiscale vessel enhancement filtering. In: Wells, W.M., Colchester, A.C.F., Delp, S.L. (eds.) MICCAI 1998. LNCS, vol. 1496, p. 130. Springer, Heidelberg (1998)
4. Hammer, P.L., Hansen, P., Simeone, B.: Roof duality, complementation and persistency in quadratic 01 optimization. Mathematical Programming 28, 121–155 (1984)
5. Heibel, T.H., Glocker, B., Groher, M., Paragios, N., Komodakis, N., Navab, N.: Discrete tracking of parametrized curves. In: CVPR (2009)
6. Ishikawa, H.: Higher-order clique reduction in binary graph cut (2009)
7. Nguyen, M.H., de la Torre, F.: Metric learning for image alignment. In: IJCV (2009)
8. Pauly, O., Padoy, N., Poppert, H., Esposito, L., Navab, N.: Towards application-specific multimodal similarity measures: a regression approach. In: MICCAI Workshop, PMMIA (2009)
9. Rother, C., Kolmogorov, V., Lempitsky, V., Szummer, M.: Optimizing binary mrfs via extended roof duality (2007)
10. Smola, A.J., Schölkopf, B.: A tutorial on support vector regression. Statistics and Computing (2004)
11. Wang, P., Chen, T., Zhu, Y., Zhang, W., Zhou, S.K., Comaniciu, D.: Robust guidewire tracking in fluoroscopy. In: CVPR (2009)