Deep Learning Based Bayesian Filtering for Catheter Tip Tracking in X-ray Fluoroscopy

Author Name1
1 Address 1

Author Name2
2 Address 2

Author Name3
1

Editors: Under Review for MIDL 2020

Abstract

Medical instrument tracking is beneficial for motion compensation in image guidance during interventions. In this paper, we propose a deep learning based Bayesian filtering approach for accurate and robust tracking of catheter tip in X-ray fluoroscopy. The approach integrates the detection outcome of a deep neural network and the motion estimation between X-ray images using a particle filtering framework. The proposed method has been validated on 34 clinical X-ray sequences, achieving accurate catheter tip tracking in the experiment, showing the potential of being used in image guidance for interventions.

Keywords: Catheter tip tracking, deep learning, Bayesian filtering, particle filtering.

1. Introduction

Tracking medical instruments is relevant for motion compensation in image guidance during interventions. Particularly for coronary interventions, the respiratory motion of coronary arteries can be compensated via tracking the catheter tip in X-ray images (Baka et al., 2015). Many existing works rely on hand-crafted appearance and geometric features, or models and templates of specific tools that are hardly transferable for tracking of other instruments, e.g. (Ma et al., 2012, 2013). Modern deep learning based approaches provide more generic ways to detect medical instruments, but most of the time treat tracking as a detection or segmentation problem, where temporal or motion information between frames is not fully explored or explicitly modeled (Laina et al., 2017; Du et al., 2018; García-Peraza-Herrera et al., 2016). In this work, we propose a deep learning based Bayesian filtering method for detecting and tracking catheter tip in X-ray fluoroscopy. The method utilizes the information from the deep learning detection stage with the motion information between frames to predict the catheter tip location, leading to robust and accurate tracking.

2. Method

Let $x_k \in \mathbb{R}^2, k \in \mathbb{N}$ denote the true state of guiding catheter tip location in frame $k$, a 2D vector representing the coordinates of the tip in the image space. The transition of state is given by the system model $x_k = f_k(x_{k-1}, v_{k-1})$. The observation $z_k$, the $k$-th X-ray image...
of a sequence, is given by the observation model \( z_k = h_k(x_k, n_k) \). \( f_k \) and \( h_k \) are possibly nonlinear functions, \( v_{k-1} \) and \( n_k \) are i.i.d. noise sources. The system model \( f_k \) and the observation model \( h_k \), respectively, can also be equivalently represented using probabilistic forms, i.e. the state transition prior \( p(x_k|x_{k-1}) \) and the likelihood \( p(z_k|x_k) \) from which \( x_k \) and \( z_k \) can be obtained by sampling. With these definitions, Bayesian filtering estimates \( x_k \) based on the set of all available observations \( z_{0:k} = \{z_i, i = 0, ..., k\} \) via recursively computing the posterior \( p(x_k|z_{0:k}) \) \( (\text{Arunlampalam et al., 2002}) \) as Eq. (1):

\[
p(x_k|z_{0:k}) \propto p(z_k|x_k) \int p(x_k|x_{k-1})p(x_{k-1}|z_{0:k-1})dx_{k-1}.
\]

Based on Eq. (1), Bayesian filtering is computed in cycles of two steps: prediction and update. In the prediction step, the prior probability of \( x_k \) given previous observations, \( p(x_k|z_{0:k-1}) \), is estimated. In the update step, the prior is corrected with the current likelihood \( p(z_k|x_k) \) to obtain the posterior.

**A Deep Learning based Likelihood** We formulate the estimation of \( p(z_k|x_k) \) as a catheter tip detection problem. A deep neural network takes an X-ray image as input and outputs a 2D probability map of the catheter tip location. A 2D Gaussian probability density function (pdf) \( \mathcal{N}(x_k^t; x_k^t, \sigma^2I) \) centered at the ground truth tip location \( x_k^t \) in the image space is used as the label to train the network.

To build the network, we follow an encoder-decoder architecture with skip connections \( (\text{Ronneberger et al., 2015}) \). In the encoder and decoder, residual blocks \( (\text{He et al., 2016}) \) are adopted at each resolution level. In addition, we adopt the strategy by \( (\text{Laina et al., 2017}) \), using a catheter mask as an additional label to jointly train the network to output both the catheter segmentation and its tip probability map. The training loss \( L \) is a combination of the segmentation loss \( L_s \) (Dice loss) and the detection loss \( L_d \) (mean square error, MSE):

\[
L = L_s + \lambda L_d,
\]

where \( \lambda \) is a weight to balance the two terms.

**Posterior approximation** A particle filter can approximate the true posterior using discrete samples associated with weights \( \{x_k^i, w_k^i\}_{i=1}^{N_s} \) (a.k.a. particles). To define the state transition function \( f_k \), we propose to use the optical flow to estimate the motion field between two frames. A resampling step is applied to alleviate the weight degeneracy problem \( (\text{Arunlampalam et al., 2002}) \). After resampling, the weights of all samples are re-assigned as \( w_{k-1}^i = 1/N_s \), and \( w_k^i \) follows \( w_k^i \propto p(z_k|x_k^i) \). Assuming the initial density \( p(x_0|z_0) \) is known, \( p(x_k|z_{0:k}) \) can be computed recursively by associating the network output and the weighted samples for every frame. The final prediction of catheter tip location can then be computed as the expectation of \( x_k \), \( \hat{x}_k = \int x_k p(x_k|z_{0:k})dx_k \), the weighted sum of all samples.

3. Experiments And Results

In the experiments, the network was trained on 1086 images from 260 sequences of 25 patients. The final tracking results are reported on 1355 images from 34 sequences of another 18 patients. The baseline method uses the location of the pixel of the maximal intensity in the detection output of the deep neural network as the prediction.

Overall, the average tracking error on all test images of the baseline method is 5.62 (±15.91) mm and that of the proposed method is 1.29 (±1.76) mm. The longitudinal
Figure 1: The longitudinal view of tracking errors on two example test sequences. The x-axis denotes the time steps of a sequence, the y-axis is the tracking error (mm).

view in Figure 1 shows that the baseline method (Figure 1(a)) performs accurately on some frames, but it still can make very large tracking errors in other frames; whereas the proposed tracking approach (Figure 1(b)) avoids the large errors, resulting in more robust tracking. Figure 2 visualizes how the proposed method makes the prediction by combining temporal information with the detection outcome through steps.

4. Conclusion

In this paper, we have proposed a deep learning based Bayesian filtering method for tracking catheter tip in X-ray fluoroscopy. The method is based on the Bayesian filtering framework that recursively estimates the posterior of the true catheter tip location by combining the measurement likelihood computed by a deep neural network with dynamic weighted samples using particle filtering. The proposed method achieves an average tracking error of around 1.3 mm on 34 clinical X-ray sequences and outperforms the baseline approach that uses only the deep learning based detection, showing the potential of being used for motion compensation in further works.
References


