

Predictive Camera Tracking for Bronchoscope Simulation with CONDensation

Fani Deligianni, Adrian Chung, and Guang-Zhong Yang

Department of Computing, Imperial College London
{fani.deligianni, adrian.chung}@ic.ac.uk,
g.z.yang@imperial.ac.uk
<http://vip.doc.ic.ac.uk>
<http://www.doc.ic.ac.uk/~ajchung/VISaVE>

Abstract. This paper exploits the use of temporal information to minimize the ambiguity of camera motion tracking in bronchoscope simulation. The condensation algorithm (Sequential Monte Carlo) has been used to propagate the probability distribution of the state space. For motion prediction, a second-order auto-regressive model has been used to characterize camera motion in a bounded lumen as encountered in bronchoscope examination. The method caters for multi-modal probability distributions, and experimental results from both phantom and patient data demonstrate a significant improvement in tracking accuracy especially in cases where there is airway deformation and image artefacts.

1 Introduction

In surgery, the value of minimally invasive procedures in terms of reduced patient trauma and recovery time has been established for many years. The limitation of the technique due to the complexity of instrument control and a loss of 3D vision and tactile feedback means effective training of visual spatial perception and hand-eye coordination is crucial to its safe practice. Flexible fiber-optic bronchoscopy, for example, is normally performed on patients who are fully awake or with light conscious sedation. The procedure can therefore entail considerable discomfort if it is not handled properly. Training according to the classical apprenticeship scheme is useful but can result in prolonged surgical procedures with increased patient discomfort and a potential risk for further complications. The use of computer simulation, particularly the reliance on patient specific data for building anatomical models both in terms of biomechanical fidelity and photorealism has attracted extensive interests in recent years [1],[2, 3]. Existing work has shown that by fusing real-bronchoscopy video with 3D tomographic data with the same patient, it is possible to generate photorealistic models that allow high fidelity, patient specific bronchoscope simulation [3]. The prerequisite of the technique, however, is accurate 2D/3D registration for recovering the pose of the camera in video bronchoscope sequences [2, 4]. Since bronchoscope images only provide localized views of the inner lumen, image-based technique cannot guarantee the convergence of the registration algorithm. To circumvent this problem, temporal constraints can be used to provide a predictive model by exploiting the temporal coherence of the camera movement.

For video bronchoscope navigation, Naga *et al* used Kalman filtering to increase the speed and accuracy of the registration algorithm [5]. Kalman filter, however, is generally restricted to situations where the probability distribution of the state variables is unimodal [6]. In bronchoscopy, tissue deformation, inter-reflection and view dependent specularly due to mucosa can limit the accuracy of image-based algorithms. The resultant probability density function of the state vector is typically multi-modal. Therefore, the observation probabilistic model cannot be approximated as a Gaussian distribution. The purpose of this paper is to develop a predictive tracking algorithm based on the Condensation algorithm [7]. The method is designed to cater for the general situation when several competing observations forming a non-Gaussian state-density. It uses a stochastic approach that has no restriction on the system/measurement models used and the distribution of error sources. An autoregressive algorithm is used as a predictive model [8], which is based on the fact that during bronchoscope navigation, the motion is restricted within a bounded area and a rapidly moving camera is more likely to slow down or change direction, rather than accelerate further. The proposed method provides a systematic learning procedure with modular training from the ground truth data such that information from different subjects are integrated for creating a dynamical model that accommodates the learnt behavior [9], [10].

2 Method

2.1 Temporal Tracking with the Condensation Algorithm

The main idea of the statistical framework is to maintain a time-evolving probability distribution $p(x_t | Z_t)$ of the tracker state x_t over time t . The *a posteriori* density $p(x_t | Z_t)$ represents knowledge about x_t deducible from the observation history $Z_t = \{z_1, \dots, z_t\}$. The Condensation algorithm, also known as Sequential Monte Carlo Tracking and particle filter, does not explicitly represent the density function. It uses instead Statistical Factor Sampling, which provides a way of approximating $p(x_t | Z_{t-1})$ by using a random number generator for sampling $p(x_t)$. The sampling technique is based on the Bayesian theory which gives [11]:

$$p(x | z) \propto p(z | x)p(x) \quad (1)$$

During statistical factored sampling, a sample set $\{s^{(1)}, \dots, s^{(N)}\}$ is generated from the prior density $p(x_t)$, where N is the number of sample sets. A weight is subsequently assigned to each particle according to the observation density of

$$\pi^{(n)} = \frac{p(Z | x = s^{(n)})}{\sum_{j=1}^N p(Z | x = s^{(j)})} \quad (2)$$

The weighted point set is then used as a representation of the *a posteriori* density, which is becoming more accurate as N increases. By evaluating given moments of the state density, it is possible to estimate the most probable state. During prediction, the method involves the evolution of the state vector with time. Sampling methods based on spatial

Markov processes are generally used for the Condensation Method to represent the dynamics of a moving object/camera. As an observation/measurement model, a pq -space based 2D/3D registration technique developed by Deligianni *et al* has been used [2]. The state of the system has been defined as the 6DoF pose of the camera.

2.2 Prediction Model

In order to construct a motion model for the endoscope camera that moves freely in the 3D tracheo-bronchial tree, an auto-regressive model is used [8]. This is in contrast to the ‘constant acceleration’ model used in [5], which effectively implies that the camera acceleration is expected to be constant during bronchoscope tracking. In this study, the auto-regressive model takes into account that during bronchoscope navigation, motion occurs within a bounded area, and a rapidly moving camera is expected to slow down or change in direction rather than accelerate further [12]. This ensures smooth navigation of the camera, and with modular training [9] multiple training sets can be used to obtain a more general model of the motion behavior. For a K^{th} -order auto-regressive model, the following equation can be used:

$$x_t = \sum_{k=1}^K A_k x_{t-k} + d + Bwt \tag{3}$$

where A_k represents the collection of damped harmonic oscillators associated with vibrational modes, d is a drift per unit time, and w the white noise with covariance coefficient B .

2.3 Training

In practice, it is possible to build a tracking model by approximating its general behavior to intuitive expectations of the observed motion. However, a hand-built model is a difficult in this study due to the high-dimensionality and complex motion involved. Mathematically, learning motion characteristics from a training sequence is to estimate the coefficients A_k , the mean value \bar{X} , and the random component B of an autoregressive process that best model the motion in a training sequence involving camera poses of x_1, \dots, x_M . The estimated pose x_t can be treated as the exact observation of the physical process, and by following the multi-variate algorithm of [8], the auto-correlation coefficients $R_{i,j}$ and $R'_{i,j}$ can be computed for $i, j = 0, 1, 2$ as

$$R_i = \sum_{t=3}^M x_{t-i}, \quad R_{i,j} = \sum_{t=3}^M x_{t-i} x_{t-j}^T, \quad R'_{i,j} = R_{i,j} - \frac{1}{M-2} R_i R_j^T \tag{4}$$

Subsequently, A_1, A_2 and D are given by:

$$\left\{ \begin{array}{l} A_2 = (R'_{02} - R'_{01} R'^{-1}_{11} R'_{12}) (R'_{22} - R'_{21} R'^{-1}_{11} R'_{12})^{-1} \\ A_1 = (R'_{01} - A_2 R'_{21}) R'^{-1}_{11} \\ D = \frac{1}{M-2} (R_0 - A_2 R_2 - A_1 R_1) \end{array} \right. \tag{5}$$

The mean of the AR process, on the other hand, is calculated as

$$\bar{X} = (\mathbf{I} - A_2 - A_1)^{-1} D \quad (6)$$

Finally, the covariance coefficient B is estimated as a matrix square root of C

$$C = \frac{1}{M-2} (R_{00} - A_2 R_{20} - A_1 R_{10} - D R_0^T) \quad (7)$$

For the purpose of bronchoscope simulation, it is more meaningful to collect several training sets from the same as well as different operators in order to construct a more representative dynamic model. To this end, the auto-correlation coefficients of each training set have been calculated individually and then combined in a linear fashion. Since each of these dynamic systems may have a different mean value, the use of pre-estimated mean value of the system can result in a prediction strongly biased. In this study, \bar{X} is estimated on-line as part of the state vector.

2.4 Validation

In order to assess the accuracy of the proposed algorithm, an airway phantom made of silicone rubber and painted with acrylics was constructed. The phantom has a cross sectional diameter of 12cm at the opening and narrows down to 5cm at the far end. The inside face was created such to give the surface a specular finish that looks similar to the surface of the lumen. A real-time, six degrees-of-freedom Electro-Magnetic (EM) motion tracker (FASTRAK, Polhemus) was used to validate the 3D camera position and orientation, as illustrated in Fig. 2(c). The EM-tracker has an accuracy of 0.762mm RMS. The tomographic model of the phantom was scanned with a Siemens Somatom Volume Zoom four-channel multi-detector CT scanner with a slice thickness of 3mm and in-plane resolution of 1mm. A CMOS camera and NTSC standard with frame rate of 30fps was used.

For *in vivo* validation, bronchoscopy examination was performed in five patients according to a standard clinical protocol. During the bronchoscope procedure two similar type videoscopes (Olympus BF Type; with field of view 120°) were used. Video images from the bronchoscope examination were transferred to digital videotapes in PAL format at 25fps. Since the original endoscopic video frames contain both the endoscopic image and redundant black background, only the endoscopic view was digitized and cropped to images of 454×487 pixels. All images were converted to grayscale before the pq -space analysis. Similar to the phantom study, the CT images were acquired from the Siemens Somatom Volume Zoom four-channel multi-detector CT scanner with a slice width of 3mm and collimation of 1mm, and the acquisition volume covered from the aortic arch to the dome of hemidiaphragm. Pre-processing of the video images was necessary in order to alleviate the effects of interlacing, lens distortion and unnecessary texture information. To remove noise and image artifacts, anisotropic filtering was applied to each image.

3 Results

Figure 1 illustrates the effectiveness of the training process involved in this study. The ground truth data of the camera pose from four different patients have been used to

train the auto-regressive model. Subsequently, the performance of the trained model was evaluated on the fifth patient data. The Euclidean distance between the first and subsequent camera positions predicted from the condensation algorithm was used for error analysis. Similar analysis for the error in orientation was also performed. In Figure 1(a), the system mean has been predefined according to the training sequences, whereas for (b), it was included in the state vector of the predictive model such that it was updated in real-time. It is evident that without continuous updating of \bar{X} , the derived tracking value is heavily biased.

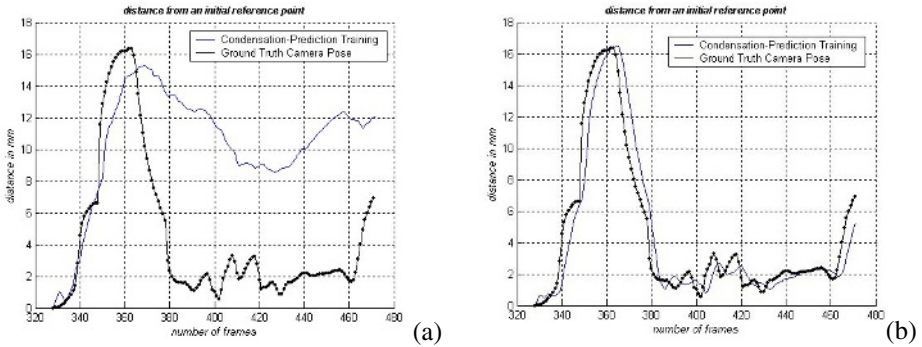


Fig. 1. Assessment of the accuracy of the training model and the effect of excluding (a), and including (b), mean value estimation of \bar{X} as part of the state vector

Table 1. Quantitative assessment of the pq -space based registration with and without the condensation algorithm (pq - CD), respectively

Case	Frames	Absolute Error							
		Position (mm)				Angle (rad)			
		Pq		CD		Pq		CD	
		Mean	\pm Std	Mean	\pm Std	Mean	\pm Std	Mean	\pm Std
Phantom	300	36.5	22.7	5.89	5.5	0.14	0.1	0.12	0.07
Pat1	100	7.0	2.0	3.3	2.2	0.8	0.2	0.03	0.02
Pat2	120	13.3	5.8	2.6	1.7	0.35	0.14	0.11	0.06
Pat3	243	26.98	13.0	1.9	1.52	1.02	0.6	0.11	0.06
Pat4	144	5.87	5.5	3.15	2.1	0.06	0.1	0.05	0.04
Pat5	100	14.4	10.12	2.26	1.65	0.7	0.4	0.19	0.15

The detailed assessment results for the phantom and patient studies are summarized in Table 1. It is evident that in both cases the 2D/3D registration accuracy has been increased significantly by the use of the proposed predictive tracking algorithm. More notably, the method permits more stable tracking results in patients where image artifact (e.g. partial occlusion of the images due to mucosa or bleeding) and sudden airway deformation due to coughing can introduce large propagational errors to the original pq -space registration technique. In Figure 2, we demonstrate the extent of this

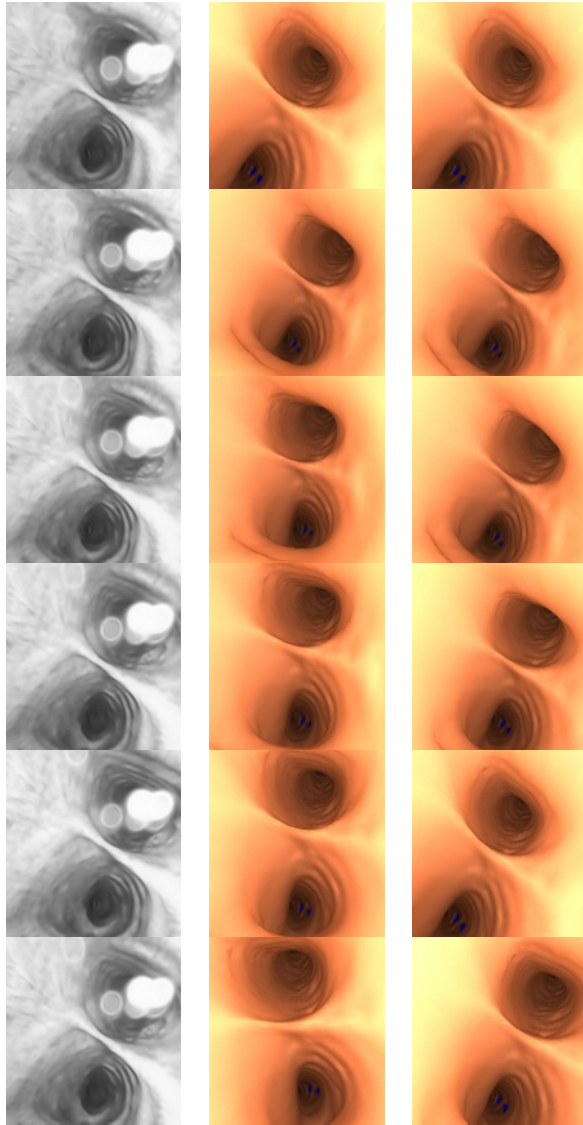


Fig. 2. The effect of airway deformation and partial occlusion of the image due to mucosa and blood on the accuracy of the 2D/3D registration technique without (mid-column) and with (right-column) predictive camera tracking

effect on the actual bronchoscope views. The left column shows the original frames from the bronchoscope video, whereas the middle and right columns are the virtual views of the 3D model by using pq -space registration without and with predictive camera pose tracking. It is worth noting that pre-processing, including radial distortion correction, de-interlacing and anisotropic filtering, has been applied to the real bronchoscope images before the registration step.

4 Discussion and Conclusions

In this paper, we have described the use of predictive camera tracking for increasing the general accuracy and robustness of 2D/3D registration involved in virtual bronchoscope modeling. A stochastic filter is used to resolve the inherent global ambiguity in tracking by exploiting the temporal coherence of the camera tip. The use of the condensation algorithm permits the use of multi-modal probability distributions, and our results from both phantom and patient data demonstrate a significant improvement in tracking accuracy especially in cases where there is airway deformation and image artifacts. The method effectively avoids the registration algorithm being trapped in local minima. The use of auto-regressive model based on the principles of the maximum likelihood learning and extension to modular learning has facilitated the incorporation of multiple sequences from different patients. The proposed method can be further extended to multi-class motion description such that the dynamic behavior of camera navigation in different parts of the tracheo-bronchial tree can be incorporated.

References

1. Indelicato, D., "Virtual Reality in Surgical Training", *The Dartmouth Undergraduate Journal of Science*, vol. I, no. 1, 1999.
2. Deligianni, F., Chung, A., and Yang, G.-Z., "pq-Space 2D/3D Registration for Endoscope Tracking", in Proceedings of Medical Image Computing & Computer Assisted Intervention (MICCAI03), Montréal, Québec, Canada, 2003, pp. 311-318.
3. Chung, A. J., Deligianni, F., Shah, P., Wells, A., and Yang, G.-Z., "Enhancement of Visual Realism with BRDF for Patient Specific Bronchoscopy Simulation", in Proceedings of MICCAI, Rennes, France, 2004, pp. 486-493.
4. Mori, K., Deguchi, D., Sugiyama, J., Suenaga, Y., Toriwaki, J.-i., Jr., C. R. M., Takabatake, H., and Natori, H., "Tracking of a bronchoscope using epipolar geometry analysis and intensity-based image registration of real and virtual endoscopic images", *Medical Image Analysis*, vol. 6, no. 3, 2002, pp. 181-336.
5. Naga, J., Mori, K., Enjouji, T., Deguchi, D., Kitasaka, T., Suenaga, Y., Hasegawa, J.-i., Toriwaki, J.-i., Takabatake, H., and Natori, H., "Fast and Accurate Bronchoscope Tracking Using Image Registration and Motion Prediction", in Proceedings of Medical Image Computing and Computer-Assisted Intervention (MICCAI04), Saint-Malo, France, 2004, pp. 551-558.
6. Welch, G. and Bishop, G., "An Introduction to the Kalman Filter", in Proceedings of ACM SIGGRAPH, 2001.
7. Isard, M. and Blake, A., "CONDENSATION-Conditional Density Propagation for Visual Tracking", *International Journal of Computer Vision*, vol. 29, no. 1, 1998, pp. 5-28.
8. Blake, A. and Isard, M., *Active Contours*, London, Springer, 1998.
9. Wildenberg, A. P., *Learning and Initialisation for Visual Tracking*, University of Oxford, Oxford, UK, Modular Training for Visual Tracking Thesis, 1998.
10. North, B., Blake, A., Isard, M., and Rittscher, J., "Learning and classification of complex dynamics", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 9, 2000, pp. 1016-1034.
11. Arulampalam, M. S., Maskell, S., Gordon, N., and Clapp, T., "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking", *IEEE Transactions on Signal Processing*, vol. 50, no. 2, 2002, pp. 174-188.
12. Davison, A. J., "Simultaneous localization and map-building using active vision", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, 2002, pp. 865-880.