Endoscope 3-D Motion Tracking Using An Aggressive Particle Filtering for Boosting Electromagnetic Guidance Endoscopy

Xiongbiao Luo, Takayuki Kitasaka, and Kensaku Mori
1Information and Communications Headquarters, Nagoya University, Japan
2Faculty of Information Science, Aichi Institute of Technology, Japan
3Graduate School of Information Science, Nagoya University, Japan
xbluo@mori.m.is.nagoya-u.ac.jp, kitasaka@aitech.ac.jp, kensaku@is.nagoya-u.ac.jp

Abstract

The paper proposes a novel endoscope motion estimation method that bases an aggressive particle filter (APF) for enhancing electromagnetic tracking (EMT) during guided endoscopy. We explore an APF strategy to resolve two main limitations of EMT sensor measurements: (1) inaccuracy due to airway deformation and instability or jitter errors because of magnetic field distortion. During such a strategy, a swarm intelligent technique – particle swarm optimization that was modified to traceably determine evolutionary parameters in this paper was embedded into a generic particle filter (GPF) to address particle impoverishment that usually occurs in GPF. Experimental results from validating our method on a dynamic bronchial phantom demonstrate our method’s effectiveness and robustness during EMT-guided endoscopy. The tracking pose error was significantly reduced from \((4.3, 7.8)\) to \((2.7 \text{ mm}, 5.6^\circ)\).

1. Background

Electromagnetic guidance endoscopy (EGE) is now increasingly introduced to diagnosis, treatment, and staging of lung and bronchus cancer. During such an endoscopy, an EMT motion sensor with 6 degrees of freedom (DOF) measurements is usually fixed at the surface or the working channel of the endoscope distal tip and located by one magnetic field generator in an EMT system, e.g., 3D Guidance medSAFE Tracker from Ascension Technology Corporation. Unfortunately, EGE is somewhat limited to diagnostic yield or accuracy that ranges between 59% and 74%, as reported in [9]. We believe one main reason is that EMT location systems still suffer from drawbacks: (1) inaccurate location due to patient movements such as respiratory motion: since EMT sensor outputs provide position and orientation measurements relative to a fixed, world coordinate system, they do not correspond exactly to current camera pose; and (2) unstable measurements with stochastic jitter due to magnetic field distortion caused by ferrous metals or conductive material in the working volume.

Various research work to handle these drawbacks was published in the literature. Mori et al. [6] proposed a hybrid tracking method by integrating EMT and image registration. Their hybrid method was enhanced by Kalman filtering and a respiratory motion compensation method using a surrogate sensor, as discussed in [10]. Gergel et al. [3] tackled airway deformation using particle filtering and bronchial structure information on the basis of an assumption that an endoscope is always moving along the airway centerlines, which is broken easily in clinical applications. On the other hand, inaccurate EMT measurements are difficult to correct, unless in collaboration with optical tracking [2].

Distinguished from our previous work of utilizing a GPF method to boost electromagnetically guided endoscopy [5], this study proposed an APF strategy to tackle EMT sensor measurement inaccuracy and instability. We embed an improved particle swarm optimization (PSO) called traceable PSO to GPF for handling particle impoverishment that commonly happens and results in a local minima convergence in GPF. Several advantages of our proposed method are highlighted as follows. First, we proposed a traceable PSO (TPSO) algorithm that modify PSO to automatically control evolutionary parameters during its iterations. Furthermore, these parameters are estimated on the basis of spatial constraints and image intensity information, which makes TPSO more robust than PSO. Finally, TPSO provides an effective means to alleviate particle degeneracy and enhances particle exploration ability in APF.
2. Aggressive Particle Filtering

Particle filtering is commonly introduced to tackle measurement ambiguity during visual or object tracking. It belongs to sequential Monte Carlo techniques, which has its capacity to avoid local minima convergence and its applicability to uncertainty or non-Gaussian data. Unfortunately, one major problem of particle impoverishment happens in GPF and results in a loss of particle diversity. In our study, we propose a TPSO-based scheme to tackle such a problem and develop an aggressive particle filtering algorithm to deal with measurement uncertainties in EMT sensor outputs.

We here introduce some notations used in our approach. Let \( Q_i \) with translation \( t_i \) and rotation \( R_i \) represents transformation between camera and CT coordinate systems at frame \( i \). Endoscope state \( p_i \) is also defined as homogeneous matrix: \( p_i \equiv Q_i \cdot \mathcal{Y}_i = \{y_1, ..., y_i\} \) denotes observation \( y_i \) with its history.

Before we perform iterations of our APF algorithm, we implement one particle randomness step, which seeks to enhance the diversity of the initialized particle set \( p_0 = \{p_0^j, w_0^j, c_0^j\} \) for \( P_i = \{p_i^j, w_i^j, c_i^j\} \) at time \( i \). The initialized particle set \( P_0 \) is basically randomized by a Gaussian transition model. The randomized particle set \( X_i = \{(x_i^j, w_i^j, c_i^j)\} \) is basically generated by:

\[
x_i^j \sim N(p_i^j, \Sigma),
\]

where \( \Sigma \) denotes a predetermined motion noise term. After that, we update all particle in \( P_i \) from \( X_i \) by: \( (p_i^j, w_i^j, c_i^j) = (x_i^j, w_i^j, c_i^j) \), if \( f(p_0^j) < f(x_i^j) \); otherwise, \( (p_i^j, w_i^j, c_i^j) = (p_0^j, w_0^j, c_0^j) \); \( f(\cdot) \) is a fitness value.

Note that such a particle randomness step is only performed once, i.e., it is only implemented at frame \( i = 1 \).

After obtaining two particle sets \( X_i \) and \( P_i = \{p_i^j, w_i^j, c_i^j, j = 1, 2, ..., M\} \) \( (M \) is the particle number and \( w_i^j \) and \( c_i^j \) respectively denote particle weight and accumulative weight) that is deterministically and stochastically propagated to approach posterior density of motion state \( p_{i+1} \) at frame \((i + 1)\), our new method with three steps of (1) diversification, (2) transition, and (3) likelihood is described in the following sections.

2.1. Diversification

We introduce a swarm intelligent technique – PSO to be modified to TPSO for diversifying particle swarm or cloud to avoid the particle impoverishment problem.

Basically, each particle \( x_i^j \) in \( X_i \) is moved to \( x_{i+1}^j \) from frames \( i \) to \((i + 1)\) with respect to an transition velocity vector \( v_i^j \) and an inertia weight \( \delta \) by:

\[
v_{i+1}^j = \delta v_i^j + \lambda_1 \eta_1 (p_i^j - x_i^j) + \lambda_2 \eta_2 (g_i^j - x_i^j),
\]

where \( \delta \) shows how much \( v_i^j \) to be preserved in \( v_{i+1}^j \), \( \lambda_1 \) and \( \lambda_2 \) are acceleration constants and \( \eta_1 \) and \( \eta_2 \) are from an uniform random generator with interval \([0.0, 1.0]\). \( p_i^j \) (for the local individual best) and \( g_i^j \) (for the global all best) are the best state found by particle \( i \) so far and the best state found by the whole swarm so far.

During such an iteration, three evolutionary factors \( \delta, \lambda_1, \) and \( \lambda_2 \) control the performance of PSO [4]. In TPSO, factors \( \lambda_1 \) and \( \lambda_2 \) are traceably computed by:

\[
\lambda_1 = 2f(p_i^j) / f(p_i^j) + f(g_i^j),
\]

\[
\lambda_2 = 2f(g_i^j) / f(p_i^j) + f(g_i^j),
\]

On the other hand, inertia weight \( \delta \) is calculated on spatial constraints and image information. We compute distance \( d^2 \) from one particle to all other particles and find \( d_{min} \) and \( d_{max} \) from \( \{d^2\}_{j=1}^M \). We then normalize distance \( d^2 \) between one particle and the current global best particle with \( d_{min} \) and \( d_{max} \) and obtain \( \gamma_i^j \) and assign it to each particle. Finally, since \( \delta \) was suggested within the interval \([0.4, 0.9]\) for weighting the global and the local searching abilities [7], it can be calculated by:

\[
\delta(f(x_i^j), \gamma_i^j) = 2 / (2 + 3 \exp(-1.28(f(x_i^j) + \gamma_i^j)) ),
\]

which shows a novel strategy to automatically control \( \delta \) by integrating spatial continuity constraint and image sequence information into our modified PSO algorithm.

Eventually, we update \( g_{i+1}^j \) and each particle in \( P_i \) to \( P_{i+1} \) in terms of \( X_{i+1} \) by the following equations:

\[
p_i^{j+1} = \begin{cases} x_{i+1}^j & \text{if } f(x_{i+1}^j) > f(p_i^j) \\ p_i^j & \text{otherwise} \end{cases},
\]

\[
g_i^{j+1} = \arg\max_{p_{i+1}^j} f(p_{i+1}^j). \]

2.2. Transition

We first select particles to be propagated. The standard condensation algorithm selects particle \( p_{i+1}^j \) at frame \((i + 1)\) by identifying the smallest \( c_{i+1}^j \) that satisfies: \( c_{i+1}^j > r \), \( r \) is a random number that is produced by an uniform generator. We then obtain swarm \( \bar{P}_{i+1} = \{ (\bar{p}_{i+1}^j, \bar{w}_{i+1}^j, \bar{c}_{i+1}^j) \} \) selected from \( P_{i+1} \).

Now, these selected particles are deterministically drifted and stochastically diffused to new states \( \bar{P}_{i+1} \):

\[
\bar{p}_{i+1}^j = \mathcal{F}(A\bar{p}_{i+1}^j, B(n_i^j)),
\]

where \( \mathcal{F} \) is the transmission function and the stochastic diffusion part \( B(n_i^j) \) is a noise term with an independent stochastic variable \( n_i^j \) that was assumed to yield a
normal distribution in our implementation. The deterministic drift part \( \mathbf{A} \) is calculated on the basis of two consecutive EMT sensor outputs \( \hat{\mathbf{Q}}_i \) and \( \hat{\mathbf{Q}}_{i+1} \):

\[
\mathbf{A} = \hat{\mathbf{Q}}_{i+1}(\hat{\mathbf{Q}}_i)^{-1}.
\]

### 2.3. Likelihood

After particle propagation and obtain \( \{\hat{\mathbf{p}}^j_{i+1}\} \), observation probability

\[
Pr(y_{i+1} | \hat{\mathbf{p}}^j_{i+1}) = \tilde{w}^j_{i+1} = \exp \left(-S^j_{i+1}/\sigma^2_r\right),
\]

where \( \sigma_r \) is the mean intensity of \( I^R_{i+1} \) and \( S^j_{i+1} \) is the similarity between \( I^R_{i+1} \) and \( I_V \) and is calculated by [1]:

\[
S^j_{i+1} = \text{MoMSE}(\hat{\mathbf{p}}^j_{i+1} - \mathbf{I}_V). \tag{12}
\]

We update cumulative weight \( \tilde{c}^j_{i+1} \) and normalize \( \tilde{c}^j_{i+1} \):

\[
\tilde{c}^j_{i+1} = \tilde{c}^j_{i+1} + \tilde{w}^j_{i+1}, \quad \tilde{c}^j_{i+1} = \tilde{c}^j_{i+1}(\sum_{j=1}^M \tilde{c}^j_{i+1})^{-1}. \tag{13}
\]

Finally, estimate \( \hat{\mathbf{Q}}^*_{i+1} \) of current endoscope motion pose is to choose particle \( \hat{\mathbf{p}}^*_{i+1} \) with the maximum weight from \( \{(\hat{\mathbf{p}}^j_{i+1}, \tilde{w}^j_{i+1}, \tilde{c}^j_{i+1})\}: \hat{\mathbf{Q}}^*_{i+1} \equiv \hat{\mathbf{p}}^*_{i+1} \).

### 3. Results and Discussion

Due to a shortage of patient data, we evaluate our proposed method on a dynamic bronchial phantom that imitates respiratory motion by the respiratory rate with the range of 0~10 breaths per minute, which corresponds to maximum breathing motion of 0~24-mm. We compare our proposed method to other three previous approaches, as reported in the work of [8, 6, 5].

Table 1 quantifies all tracking results with average tracking accuracy from different methods by comparing to ground truth datasets. The average position and orientation errors of our proposed approach were 2.7 mm and 5.6°, which are much better than those of other previous methods that had average errors of at least 4.3 mm and 7.8°. We also visually inspected all tracking results by manually checking whether the real images resembled the virtual images. Fig. 1 displays examples of real images and the corresponding virtual images generated from camera motion posed estimated by different method. This visual investigation of the successfully processed frames further demonstrates the effectiveness of our proposed method. Additionally, the current runtime of our method is about 1.5 seconds per frame.

The objective of this work is to precisely and stably determine endoscope pose during electromagnetic guidance endoscopy. In general, our APF framework provides a more accurate and robust means to estimate endoscope motion than previous approaches. Two reasons behind such a beneficial performance of our APF strategy. First, TPSO enhance the performance of PSO by traceably or adaptively determining evolutionary parameters based on spatial constraints and image information to make PSO iterations meaningful. On the other hand, we introduce a TPSO-based scheme to tackle particle degeneracy phenomena that often happen in generic particle filtering. We believe that TPSO can diversify particle cloud very well and distribute particles reasonably to approximate the posterior probabilistic density of camera motion poses in a solution space.

Additionally, we must illuminate one potential drawback of our proposed method. We assume particle fitness values could be correctly weighted, which depends on likelihood function that models image intensity information. Unfortunately, image artifacts that appear in endoscopic video may collapse likelihood model to successfully calculate particle fitness values. To handle such a limitation, a more robust likelihood model or function involved image intensity information, which should be slightly insensitive to illumination changes or other image artifacts, must be explored in future work.

### 4. Conclusions

This paper proposes a new endoscope motion tracking approach on the basis of aggressive particle filtering to boost electromagnetic guidance endoscopy. Since a modified particle swarm optimization was introduced in generic particle filtering, particle impoverishment was address well, which makes our aggressive particle filtering much effective and robust in 3-D motion tracking, as demonstrated by experimental results with tracking accuracy that was improved from (4.3, 7.8) to (2.7, 5.6°).

### References

Table 1. Quantitative results of average tracking accuracy of compared methods in accordance with position and orientation errors between all estimated results and ground truth datasets.

<table>
<thead>
<tr>
<th>Experiments</th>
<th>Data 1</th>
<th>Data 2</th>
<th>Data 3</th>
<th>Data 4</th>
<th>Data 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schwarz et al. [8]</td>
<td>4.2±2.6 mm</td>
<td>5.3±3.5 mm</td>
<td>5.6±2.8 mm</td>
<td>6.0±2.6 mm</td>
<td>7.2±3.5 mm</td>
</tr>
<tr>
<td></td>
<td>6.7±5.2°</td>
<td>8.8±6.2°</td>
<td>7.9±5.3°</td>
<td>9.6±6.0°</td>
<td>13.5±11.1°</td>
</tr>
<tr>
<td>Mori et al. [6]</td>
<td>3.8±3.2 mm</td>
<td>4.9±4.2 mm</td>
<td>5.4±3.2 mm</td>
<td>5.8±3.6 mm</td>
<td>6.8±4.4 mm</td>
</tr>
<tr>
<td></td>
<td>6.1±4.1°</td>
<td>7.6±5.5°</td>
<td>6.8±6.2°</td>
<td>8.8±5.6°</td>
<td>12.9±13.4°</td>
</tr>
<tr>
<td>Luo et al. [5]</td>
<td>3.1±2.2 mm</td>
<td>3.9±2.1 mm</td>
<td>4.1±2.5 mm</td>
<td>4.6±3.2 mm</td>
<td>5.6±4.3 mm</td>
</tr>
<tr>
<td></td>
<td>3.6±3.1°</td>
<td>4.3±2.3°</td>
<td>4.6±2.8°</td>
<td>7.8±4.5°</td>
<td>10.2±10.5°</td>
</tr>
<tr>
<td>Our method</td>
<td>1.9±2.1 mm</td>
<td>2.2±1.6 mm</td>
<td>2.6±2.2 mm</td>
<td>3.1±2.5 mm</td>
<td>3.6±2.8 mm</td>
</tr>
<tr>
<td></td>
<td>2.9±2.1°</td>
<td>3.4±2.5°</td>
<td>4.5±3.3°</td>
<td>8.0±4.2°</td>
<td>9.3±9.6°</td>
</tr>
</tbody>
</table>

Frame number 0272 0383 0506 0551 0646 0731 0803 0953 1016 1145 1333 1445

RB images

Schwarz et al. [8]

Mori et al. [6]

Luo et al. [5]

Our method

Figure 1. Visual comparison of tracking results of Data 5. Top row shows selected frame numbers, and second row shows their corresponding real images. Other rows display virtual images generated from tracking results estimated from all different approaches discussed above.


