A COMPARISON OF MODIFIED EVOLUTIONARY COMPUTATION ALGORITHMS WITH APPLICATIONS TO THREE-DIMENSIONAL ENDOSCOPIC CAMERA MOTION TRACKING

Ying Wan[†] *Xiangjian He*[†] *Xiongbiao Luo*[¶,†,*]

[†] School of Computing and Communications, University of Technology Sydney, Australia
[¶]Department of Computer Science, Fujian Key Laboratory of Sensing and Computing for Smart City, Xiamen University, Xiamen, China

ABSTRACT

Endoscope 3D motion tracking plays an irreplaceable role for computer-assisted endoscopy systems development. Without such tracking, it is impossible to synchronize pre- and intra-operative images in a reference coordinate frame. Currently available methods are comprised of video-based and electromagnetic tracking. These methods limit to either video image artifacts or inaccurate sensor measurements and dynamic errors. This paper proposes two modified evolutionary computation algorithms: (a) adaptive particle swarm optimization (APSO) and (b) observation-boosted differential evolution (OBDE), to augment current endoscopic camera motion tracking. The experimental results demonstrate that our modified algorithms, which combine endoscopic video images with sensor measurements to estimate endoscope movements, can improve tracking accuracy from 4.8 mm to 2.9 mm. OBDE outperforms APSO for endoscope tracking.

Index Terms— Endoscope tracking and navigation, evolutionary computation, particle swarm optimization, differential evolution, computer-assisted interventions

1. INTRODUCTION

Endoscope, which is commonly integrated with a video camera at its distal tip, is a widely used medical instrument during minimally invasive surgery. Physicians usually insert the endoscope into the body cavity (e.g., bronchus or colon) through natural orifices (e.g., mouth or nose) to directly observe and examine the interior of the body cavity. Under the endoscope with two-dimensional (2-D) video images, physicians also perform surgical procedures of interventions and treatment, e.g., a biopsy procedure that either takes a small sample of suspicious tumors for pathological analysis or removes samples of precancerous regions. However, during these procedures, it is difficult to determine the location (position and orientation) of the endoscope distal tip or its integrated video camera in the body cavity. Recently three-dimensional (3-D) endoscopic camera motion tracking has been introduced to solve the location problem of the endoscope tip. Such tracking seeks to locate the endoscopic camera at a spatial coordinate reference system, e.g., pre-operative image coordinate system of computed tomography (CT) scanners.

Current tracking methods use either endoscopic camera video images or an electromagnetic (EM) tracking device to estimate the 3-D endoscopic camera pose or location with position and orientation in the CT image coordinate system. Based on endoscopic video images, 2-D/3-D image registration is performed to align these video images to the 3-D CT images that are used to generate 2-D virtual rendering images by surface or volume rendering techniques [1]. The EM tracking device uses a miniature position sensor attached at the endoscope tip to measure the endoscopic camera motion. By calibrating the EM tracker and the 3-D CT volume, these motion measurements with six degrees of freedom (6DoF) position and orientation parameters can be transformed to the CT image coordinate system [2]. However, EM trackers have two main limitations: (1) inaccurate EM sensor measurements due to tissue deformation, particularly respiratory motion, and (2) dynamic errors because of magnetic field distortion. The motivation of this work is to address these limitations of EM trackers by using evolutionary computation.

Evolutionary computation, which is generally a family of artificial intelligence, uses biological evolution, e.g., mutation and natural selection, to solve dynamic optimization problems [3]. Its algorithms, particularly particle swarm optimization (PSO) and differential evolution (DE), with a meta-heuristic or stochastic optimization property, are increasingly introduced to various practical applications, such as image segmentation and object recognitions. This work employs the PSO and DE algorithms to tackle the limitations of EM trackers. Since the performance of PSO and DE depends heavily on their operators and evolutionary factors, we proposed adaptive particle swarm optimization (APSO) and observation-boosted differential evolution (OBDE). The main contributions of this works are two-fold: (1) PSO and DE algorithms were modified with improved performance, (2) the performance of APSO and OBDE was fully compared with applications to endoscopic camera motion tracking.

Asterisk indicates corresponding authors.

2. APPROACHES

Evolutionary computation usually generates a population $\mathcal{X}$ of potential or random solutions (or called individuals) $\mathcal{X} = \{\mathbf{x}_{i,g} \in \mathcal{R}^D\}_{i=1,2,\cdots,N,\ g=1,2,\cdots,G}$ (where N is the population size, G is the generation number, and D is the dimension of individual or vector $\mathbf{x}_{i,g}$) and propagates the population at each iteration (also called generation g) to approximate the optimal solution of global and stochastic optimization problems. Operation and selection are two primary forces to drive evolutionary computation. It is operation that evolves individuals by using different strategies, which lead to the difference between APSO and OBDE, as discussed as follows.

2.1. Adaptive Particle Swarm Optimizer

In PSO, the i-th individual or vector $\mathbf{x}_{i,g-1}$ at generation $g-1$ is evolved to $\mathbf{x}_{i,g}$ at generation g on the basis of speed $\mathbf{a}_{i,g}$ and inertia weight ω (to control previous speed $\mathbf{a}_{i,g-1}$) by

$$\mathbf{x}_{i,g} = \mathbf{x}_{i,g-1} + \mathbf{a}_{i,g}, \tag{1}$$

$$\mathbf{a}_{i,g} = \omega\mathbf{a}_{i,g-1} + \mu_1\eta_1(\mathbf{b}_i - \mathbf{x}_{i,g-1}) + \mu_2\eta_2(\mathbf{c} - \mathbf{x}_{i,g-1}), \tag{2}$$

where μ_1 and μ_2 are acceleration constants, η_1 and η_2 yield the uniform distribution with the interval [0, 1], $\mathbf{b}_i$ is the best solution obtained by the i-th individual so far, and $\mathbf{c}$ is the global best solution from population $\mathcal{B} = \{\mathbf{b}_i\}_{i=1}^N$. After iteration g, $\mathbf{b}_i$ and $\mathbf{c}$ are updated in terms of fitness $f(\cdot)$:

$$\mathbf{b}_i = \begin{cases} \mathbf{x}_{i,g} & \text{if} f(\mathbf{x}_{i,g}) > f(\mathbf{b}_i) \\ \mathbf{b}_i & \text{otherwise} \end{cases}, \tag{3}$$

$$\mathbf{c} = \arg\max_{\mathbf{b}_i \in \mathcal{B}} f(\mathbf{b}_i). \tag{4}$$

The performance of PSO depends on parameters ω, μ_1 and μ_2, speed $\mathbf{a}_{i,g-1}$, and fitness $f(\cdot)$. During the standard PSO, parameters ω, μ_1 and μ_2 are fixed constants and speed $\mathbf{a}_{i,g-1}$, and fitness $f(\cdot)$ do not involve the current observation information. Our idea of modifying PSO is to employ the current observation information to determine speed $\mathbf{a}_{i,g-1}$, and fitness $f(\cdot)$ and adaptively compute ω, μ_1 and μ_2. Suppose $\mathbf{o}_k$ be the current observation of one dynamic system at time k, we compute these variables with respect to $\mathbf{o}_k$:

$$f(\mathbf{x}_{i,g}) = p(\mathbf{o}_k|\mathbf{x}_{i,g}),\ \mathbf{a}_{i,g-1} = \Gamma(\mathbf{o}_k, \mathbf{o}_{k-1}), \tag{5}$$

$$\mu_1 = \frac{2p(\mathbf{o}_k|\mathbf{b}_i)}{p(\mathbf{o}_k|\mathbf{b}_i) + p(\mathbf{o}_k|\mathbf{c})},\ \mu_2 = \frac{2p(\mathbf{o}_k|\mathbf{c})}{p(\mathbf{o}_k|\mathbf{b}_i) + p(\mathbf{o}_k|\mathbf{c})}, \tag{6}$$

where function Γ computes the speed between $\mathbf{o}_k$ and $\mathbf{o}_{k-1}$.

To adaptively control ω, we first define spatial distribution factor γ_{g-1} at generation $g-1$ on the basis of average distance $d_{i,g-1}$ from $\mathbf{x}_{i,g-1}$ to all the other individuals:

$$d_{i,g-1} = \frac{1}{N-1} \sum_{i=1, i\neq j}^{N} \|\mathbf{x}_{i,g-1} - \mathbf{x}_{j,g-1}\|. \tag{7}$$

Based on maximal and minimal distances (d_{max}, d_{min}) from $\{d_{i,g-1}\}_{i=1}^N$ and average distance d_c between global best $\mathbf{c}$ and $\{\mathbf{x}_{i,g-1}\}_{i=1}^N$, distribution factor γ_{g-1} is computed by:

$$\gamma_{g-1} = (d_c - d_{min})/(d_{max} - d_{min}), \quad \gamma_{g-1} \in [0,\ 1]. \tag{8}$$

Since ω was suggested to range within the interval $[0.4, 0.9]$ to weighting the global and the local searching abilities [4], we use distribution factor γ_{g-1} and fitness $f(\mathbf{x}_{i,g-1})$ to adaptively calculate inertia weight ω for speed $\mathbf{a}_{i,g}$ by:

$$\omega = \frac{2}{2 + 3\exp(-1.28(p(\mathbf{o}_k|\mathbf{x}_{i,g-1}) + \gamma_{g-1}))}, \tag{9}$$

which shows a novel strategy to automatically control ω.

2.2. Observation-Boosted Differential Evolution

In DE, three operations are usually performed: (1) mutation, (2) crossover, and (2) selection. The DE performance depends on the mutation operator. Such an operator determines mutant vector $\mathbf{v}_{i,g}$ for individual or vector $\mathbf{x}_{i,g}$ at generation g by:

$$\mathbf{v}_{i,g} = \mathbf{x}_{i,g} + m_i \overbrace{(\mathbf{c} - \mathbf{x}_{i,g})}^{\tilde{\mathbf{v}}_{i,g}} + m_i \overbrace{(\mathbf{x}_{r_i^1,g} - \mathbf{x}_{r_i^2,g})}^{\hat{\mathbf{v}}_{i,g}}, \tag{10}$$

where m_i denotes the mutation factor and indexes r_i^1 r_i^2 are mutually exclusive integers chosen randomly from set $\{1, \cdots, i-1, i+1, \cdots, N\}$. The mutation operator in Eq. 10 brings good convergence performance during optimization since global best individual $\mathbf{c}$ was involved. However, best individual $\mathbf{c}$ might lose the population's diversity and lead to unstable convergence. To tackle this problem, we modify this operator by current observation $\mathbf{o}_k$ and two mutant factors:

$$\mathbf{v}_{i,g} = \mathbf{x}_{i,g} + \underbrace{\alpha_i\Gamma(\mathbf{o}_k, \mathbf{o}_{k-1})}_{observation} + \tilde{m}_i\tilde{\mathbf{v}}_{i,g} + \hat{m}_i\hat{\mathbf{v}}_{i,g}, \tag{11}$$

$$\tilde{m}_i = \frac{2p(\mathbf{o}_k|\mathbf{c})}{p(\mathbf{o}_k|\mathbf{x}_{i,g}) + p(\mathbf{o}_k|\mathbf{c})},\ \hat{m}_i = \frac{2p(\mathbf{o}_k|\mathbf{x}_{i,g})}{p(\mathbf{o}_k|\mathbf{x}_{i,g}) + p(\mathbf{o}_k|\mathbf{c})}, \tag{12}$$

where random number α_i controls the keep of the current observation and yields an uniformly distribution: $\alpha_i \in [0, 1]$.

After mutation, we perform a binomial crossover operation to determine trial vector $\mathbf{u}_{i,g} = \{u_{i,g}^1, \cdots, u_{i,g}^D\}$ using vectors $\mathbf{x}_{i,g} = \{x_{i,g}^1, \cdots, x_{i,g}^D\}$ and $\mathbf{v}_{i,g} = \{v_{i,g}^1, \cdots, v_{i,g}^D\}$:

$$u_{i,g}^j = \begin{cases} v_{i,g}^j & \text{if } (\delta \leq C_r) \textit{ or } (j = j_r) \\ x_{i,g}^j & \text{otherwise} \end{cases}, \tag{13}$$

where random number $\delta \in [0, 1]$, integer j_r is randomly selected from set $\{1, 2, \cdots, D\}$, and crossover rate C_r checks whether $u_{i,g}^j$ is copied from $v_{i,g}^j$ and is computed by:

$$C_r = \frac{p(\mathbf{o}_k|\mathbf{x}_{i,g}) + p(\mathbf{o}_k|\mathbf{v}_{i,g})}{2}. \tag{14}$$

The selection operator updates population $\{\mathbf{x}_{i,g}\}_{i=1}^{N}$ to $\{\mathbf{x}_{i,g+1}\}_{i=1}^{N}$ at generation $g+1$ from $\{\mathbf{x}_{i,g}\}_{i=1}^{N} \cup \{\mathbf{u}_{i,g}\}_{i=1}^{N}$ in accordance with their fitness values:

$$\mathbf{x}_{i,g+1} = \begin{cases} \mathbf{u}_{i,g} & \text{if } f(\mathbf{u}_{i,g}) \geq f(\mathbf{x}_{i,g}) \\ \mathbf{x}_{i,g} & \text{otherwise} \end{cases}. \quad (15)$$

Finally, after the G-th iteration in optimization, global best solution $\mathbf{c}$ of using the DE algorithm is determined by:

$$\mathbf{c} = \arg \max_{\mathbf{x}_{i,G} \in \{\mathbf{x}_{i,G}\}_{i=1}^{N}} p(\mathbf{o}_k | \mathbf{x}_{i,G}). \quad (16)$$

2.3. Applications to 3-D Endoscopic Camera Tracking

Endoscope tracking is to continuously estimate the current camera pose with position and orientation in the CT image coordinate system. Suppose that $\mathbf{t}^k$ denotes the endoscopic camera position and quaternion $\mathbf{q}^k$ represents the endoscopic camera orientation. In evolutionary computation of APSO and OBDE, $\mathbf{x}_{i,g}$ becomes $\mathbf{x}_{i,g}^k$ that is a seven-dimensional vector ($D = 7$) including position $\mathbf{t}^k$ and quaternion $\mathbf{q}^k$:

$$\mathbf{x}_{i,g}^k = \left(\mathbf{t}^k \ \mathbf{q}^k\right) = (t_x^k, t_y^k, t_z^k, q_0^k, q_1^k, q_2^k, q_3^k), \quad (17)$$

where t_x^k, t_y^k, t_z^k are the position in the x-, y-, and z-axes of the CT coordinates and $(q_0^k)^2 + (q_1^k)^2 + (q_2^k)^2 + (q_3^k)^2 = 1$.

The inputs of endoscopic ca mare tracking are endoscopic camera video sequences, EM sensor measurements, and the CT images. Current observation $\mathbf{o}_k$ include endoscopic video image $\tilde{\mathbf{o}}_k$ and EM sensor measurement $\hat{\mathbf{o}}_k$ at time k. The CT images are used to generate 2-D virtual rendering image $\mathbf{I}(\mathbf{x}_{i,g}^k)$ corresponding to individual $\mathbf{x}_{i,g}^k$ using volume rendering techniques. During iterations, $\Gamma(\cdot)$ in Eqs. (5) and (11) is computed using EM sensor measurements $\hat{\mathbf{o}}_k$ and $\hat{\mathbf{o}}_{k-1}$:

$$\Gamma(\hat{\mathbf{o}}_k, \hat{\mathbf{o}}_{k-1}) = \hat{\mathbf{o}}_k - \hat{\mathbf{o}}_{k-1}, \quad (18)$$

Fitness value $f(\cdot)$ of each individual $\mathbf{x}_{i,g}^k$ is calculated by:

$$f(\mathbf{x}_{i,g}^k) = p(\tilde{\mathbf{o}}_k | \mathbf{x}_{i,g}^k) = s(\tilde{\mathbf{o}}_k, \mathbf{I}(\mathbf{x}_{i,g}^k)), \quad (19)$$

where intensity similarity $s(\tilde{\mathbf{o}}_k, \mathbf{I}(\mathbf{x}_{i,g}^k))$ between video image $\tilde{\mathbf{o}}_k$ and virtual rendering image $\mathbf{I}(\mathbf{x}_{i,g}^k)$ is defined as [5]:

$$s(\tilde{\mathbf{o}}_k, \mathbf{I}(\mathbf{x}_{i,g}^k)) = \frac{(2\xi_k\xi_i + C_1)(2\delta_{k,i} + C_2)}{(\xi_k^2 + \xi_i^2 + C_1)(\delta_k^2 + \delta_i^2 + C_2)}, \quad (20)$$

where $\delta_{k,i}$ is the correlation between images $\tilde{\mathbf{o}}_k$ and $\mathbf{I}(\mathbf{x}_{i,g}^k)$; ξ_k and ξ_i are the mean intensity values; δ_k and δ_i are the intensity variances, and C_1 and C_2 are two constants.

Finally, the APSO- or OBDE-based tracking outputs the best estimate for current endoscopic camera pose $(\mathbf{t}_*^k \ \mathbf{q}_*^k)$:

$$\left(\mathbf{t}_*^k \ \mathbf{q}_*^k\right) = \mathbf{x}_*^k = \arg \max_{\{\mathbf{x}_{i,g}^k\}_{i=1}^{N}} p(\tilde{\mathbf{o}}_k | \mathbf{x}_{i,g}^k). \quad (21)$$

3. RESULTS AND DISCUSSION

We evaluate APSO- and OBDE-based tracking methods on six datasets of endoscopic videos and EM sensor measurements. We manually generated ground truth for these datasets and compute position and orientation errors between camera pose estimates and ground truth. We also define the visual quality as the similarity of video images and their corresponding 2-D virtual rendering images generated using camera pose estimates in the CT images. We compare four methods: (1) Mori et al. [6], directly combining EM sensor measurements and endoscopic video images to estimate camera motion, (2) Soper et al. [2], a hybrid method of using endoscopic video images and EM sensor measurements with Kalman filtering, and (3) APSO, and(4) OBDE, as discussed in Section 2.

Table 1 quantifies tracking position and orientation errors, processing time, and visual quality of using different methods. Fig. 1 plots examples of the tracking results. Fig. 2 visually compares virtual rendering images of using the four methods. All these experimental results demonstrate that two modified evolutionary computation algorithms of APSO and OBDE improve the tracking performance. We attribute such an improvement to the introduction of the current observation information of endoscopic video images and EM sensor measurements for evolutionary computation. The population with individuals was evolved on the basis of the current observation of the EM sensor that can positively guide individuals to approximate the best solution for the current endoscopic camera pose. Furthermore, evolutionary parameters, which were automatically on the basis of the fitness of each individual in APSO and OBDE, were beneficial to avoid the premature convergence of evolutionary computation. Additionally, the fitness of individuals was also computed on the basis of the current observation (video image) of the endoscopic camera.

In summary, we modified two evolutionary computation algorithms of APSO and OBDE that were boosted by observation information in stochastic optimization or dynamic systems. With application to 3-D endoscopic camera tracking, we demonstrated that observation-based evolutionary computation provides a more accurate and robust tracking method.

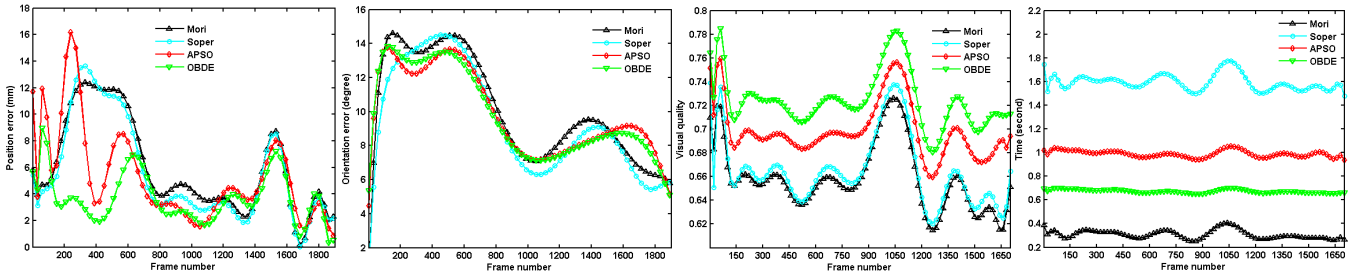
4. REFERENCES

[1] X. Luo, M. Feuerstein, D. Deguchi, T. Kitasaka, H. Takabatake, and K. Mori, "Development and comparison of new hybrid motion tracking for bronchoscopic navigation," *Medical Image Analysis*, vol. 16, no. 3, pp. 577–596, 2012.

[2] T. D. Soper, D. R. Haynor, R. W. Glenny, and E. J. Seibel, "In vivo validation of a hybrid tracking system for navigation of an ultrathin bronchoscope within peripheral airways," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 3, pp. 736–745, 2010.

Table 1: Position and orientation errors, processing time per frame, and visual quality of using different methods.

Approaches	Position (mm)	Orientation (°)	Time (second)	Visual quality
Mori et al. [6]	4.84±4.73	9.82±4.74	0.31	0.658
Soper et al. [2]	4.38±3.80	9.24±5.48	1.59	0.664
APSO	3.33±2.41	9.63±4.24	0.99	0.698
OBDE	2.92±2.62	9.37±4.12	0.67	0.724



(a) Position error of Dataset 1 (b) Orientation error of Dataset 1 (c) Visual quality of of Dataset 3 (d) Processing time of of Dataset 3

Fig. 1: Examples of plotted tracking errors, processing time per frame, and visual quality of using different methods.

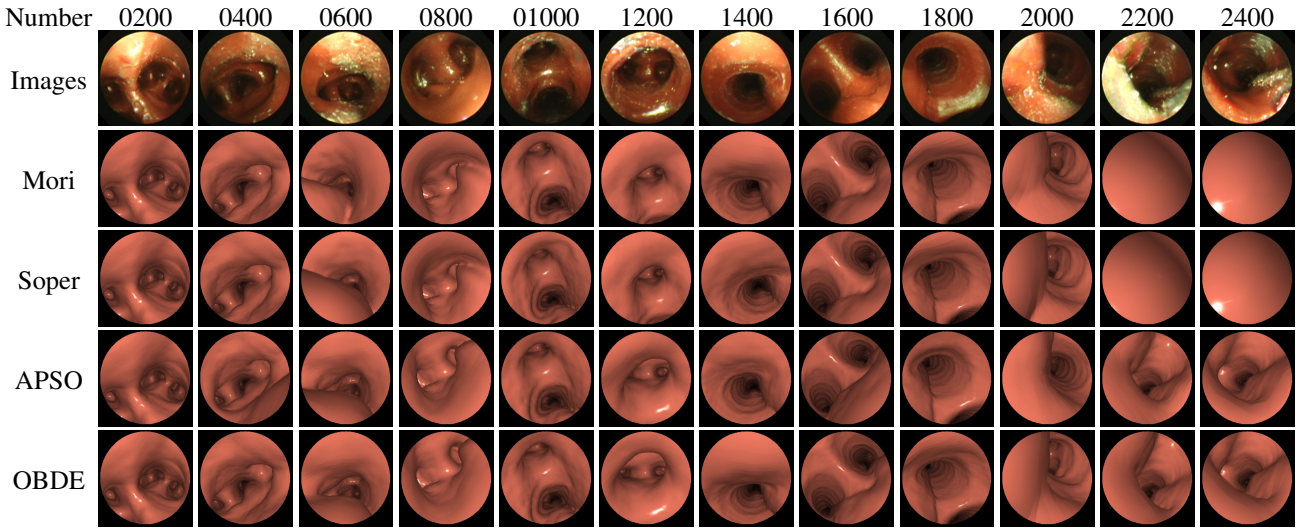


Fig. 2: Examples of visual comparison of camera pose estimates on Dataset 6. Top row shows frame numbers and second row their corresponding video images. Other rows display virtual rendering images generated from estimates of different methods. A method outperforms another if video images resemble better to virtual rendering ones. OBDE shows better performance.

[3] Agoston E. Eiben and J. E. Smith, *Introduction to Evolutionary Computing*, Springer, 2007.

[4] D. Parrott and X. Li, "Locating and tracking multiple dynamic optima by a particle swarm model using speciation," *IEEE Transactions on Evolutionary Computation*, vol. 10, no. 4, pp. 440–458, 2006.

[5] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[6] K. Mori, D. Deguchi, K. Akiyama, T. Kitasaka, C. R. Maurer Jr., Y. Suenaga, H. Takabatake, M. Mori, and H. Natori, "Hybrid bronchoscope tracking using a magnetic tracking sensor and image registration," in *Proc. MICCAI 2005*, 2005, vol. LNCS 3750, pp. 543–550.