Constrained Evolutionary Diffusion Filter for Monocular Endoscope Tracking

Xiongbiao Luo*
Department of Computer Science and Technology, Xiamen University, Xiamen 361005, China
National Institute for Data Science in Health and Medicine, Xiamen University, Xiamen 361102, China
xiongbiao.luo@gmail.com

Abstract

Stochastic filtering is widely used to deal with nonlinear optimization problems such as 3-D and visual tracking in various computer vision and augmented reality applications. Many current methods suffer from an imbalance between exploration and exploitation due to their particle degeneracy and impoverishment, resulting in local optimums. To address this imbalance, this work proposes a new constrained evolutionary diffusion filter for nonlinear optimization. Specifically, this filter develops spatial state constraints and adaptive history-recall differential evolution embedded evolutionary stochastic diffusion instead of sequential resampling to resolve the degeneracy and impoverishment problem. With application to monocular endoscope 3-D tracking, the experimental results show that the proposed filtering significantly improves the balance between exploration and exploitation and certainly works better than recent 3-D tracking methods. Particularly, the surgical tracking error was reduced from 4.03 mm to 2.59 mm.

1. Introduction

Tracking a camera's 3-D motion is vital in various computer vision applications, e.g., augmented reality, 3-D reconstruction, computer assisted surgery, navigation and mapping, and robotics. Recent advances in 3-D tracking are widely discussed in the literature [9, 11, 13, 20, 21, 33]. Different from commonly used cameras in daily life, endoscopic cameras are typical hand-held devices (called endoscopes) used to inspect interior surfaces or inaccessible regions of tubular or hollow structures where the human visual system can hardly observe. While industrial endoscopes are powerful for examining unreachable areas of buildings or parts of machines, surgical endoscopes are useful to intuitively inspect cavities in the body. Monocular endoscopic 3-D tracking plays an essential role in precise industrial inspection, clinical diagnosis and treatment.

Unfortunately, surgical endoscopic cameras only provide 2-D video images without any depth information and cannot localize themselves and targets of interest like tumors in the surgical field. To this end, surgical 3-D tracking methods are widely developed to accurately localize surgical tools and targets and reduce inadvertent hurts in endoscopic or robotic surgery [16, 19, 26]. Such 3-D tracking is a nonlinear optimization problem as well as a multisensor or multisource information fusion procedure, which is commonly solved by stochastic optimization methods [4].

Stochastic filtering is widely used for 3-D tracking [23], and usually generates a population of particles (initial solutions) and propagates them to approximate the optimal solution. But it still limits itself to local optimums or premature convergence due to an imbalance between exploration and exploitation. Specifically, this imbalance results from the particle degeneracy and impoverishment after sequential resampling, leading to ineffective filtering. Theoretically, this work aims to solve the particle degeneracy-impoverishment problem to balance exploring and exploiting and create a new effective and powerful filtering strategy with robust optimization performance. Technically, this work also strives for addressing several challenges in current surgical 3-D tracking methods: (1) endoscopic image uncertainty or artifacts in vision-based 3-D tracking, (2) inaccurate and jitter measurements in sensor-based 3-D tracking, and (3) tissue deformation and patient movement in surgical procedures.

Technical contributions of this work are clarified as follows. First of all, two new spatial state constraints are introduced for nonlinear optimization problems, improving the optimization performance. More interestingly, a new strategy of evolutionary stochastic diffusion with adaptive history-recall differential evolution instead of sequential resampling can successfully resolve the particle degeneracy-impoverishment problem, effectively balancing between exploration and exploitation. We then propose constrained

*The author would like to give his special thanks to Professor Raymond Honfu Chan who is with Hong Kong Centre for Cerebro-cardiovascular Health Engineering and City University of Hong Kong. This work was supported in part by the National Nature Science Foundation of China under Grants 82272133 and 61971367, in part by the Fujian Provincial Technology Innovation Joint Funds under Grant 2019Y9091, and in part by the Fujian Provincial Natural Science Foundation under Grant 2020J01004.

evolutionary diffusion filtering (CEDF), which is a metaheuristic optimization algorithm and more ambidextrous than other filters. Additionally, a new hybrid bronchoscope 3-D tracking framework using the proposed filtering is developed to fuse multisource data including computed tomography (CT) or magnetic resonance (MR) images, surgical videos, and positional sensor measurements. Our framework can tackle these challenges discussed above.

2. Related Work

Camera 3-D tracking or egocentric motion estimation is widely discussed in computer vision. Salih et al. [23] compared stochastic filtering for 3-D tracking. Mur-Artal et al. [20] proposed a monocular simultaneous localization and mapping (SLAM) system using a very fast binary feature descriptor, while Barros et al. [2] used salient 2-D points to estimate 3-D head pose in real time. Forster et al. [9] discussed semidirect visual odometry for monocular and multicamera systems. Moreover, Wei et al. [31] employed a instant motion tracking method for smartphone camera-based augmented reality. Cavagna et al. [3] proposed a *SpaRTA* tracking method to address occlusions in multicamera systems, radars, and RGB-D systems with multiple targets, while Chang et al. [5] introduced a multisensor data fusion framework for 3-D tracking and forecasting with rich maps for autonomous driving. Event cameras are increasingly used since they provide more useful information than standard cameras for 3-D tracking. Kim et al. [13] used three decoupled probabilistic filters to predict 3-D motion of a single hand-held event camera without additional sensing, while Rebecq et al. [22] employed image-to-model alignment for 3-D event camera parallel tracking. More recently, Gallego et al. [11] proposed to use photometric depth maps to track the 3-D event camera.

Deep learning is increasingly used for 3-D tracking. Garon et al. [12] reported a temporal 3-D tracking method using deep learning to deal with occlusions while achieving real-time tracking. Tateno et al. [29] used deeply learned depth for real-time dense monocular SLAM. Pandey et al. [21] proposed an efficient egocentric 3-D tracking method for hand-held objects like mobile phones, while Frossard et al. [10] discussed a 3-D tracking-by-detection method on the basis of end-to-end learning multisensor data. More recently, Laidlow et al. [14] introduced a DeepFusion strategy of using single view depth and gradient estimation for monocular SLAM, while Zhou et al. [33] explored a DeepTAM framework for depth map estimation and dense keyframe-based tracking. Yang et al. [32] used a self-supervised learning method to estimate endoscopic depth and ego-motion, while Shao et al. [25] reconstructed stereo endoscopic images using a single-layer network.

Current surgical 3-D tracking methods are generally divided into three categories: (1) vision-based tracking , (2) sensor-based tracking , and (3) hybrid tracking (i.e., a combination of vision- and sensor-based tracking). Vision-based tracking is actually a 2-D/3-D registration procedure. Deep learning driven video-volume registration methods are increasingly introduced for depth and pose estimation [1, 15, 16, 19, 26]. Sensor-based tracking employs a positional sensor attached at the endoscope's distal end to track the endoscopic camera pose. Typically, electromagnetic (EM) tracking can estimate the endoscope 3-D motion by EM sensors fixed at the endoscope's distal end. But it still suffers from many bottlenecks, e.g., magnetic field distortion, tissue deformation, jitter errors, less smoothness, and inaccurate registration. EM tracking techniques are thoroughly discussed in a recent work [28]. Hybrid tracking is a combination of vision- and sensor-based tracking to address their disadvantages. A common way is to use the EM sensor tracked position and orientation as the initialization of 2-D/3-D registration for deterministic optimization [27]. Another way is to incorporate EM sensor outputs and endoscopic video images into stochastic optimization [18].

3. Constrained Evolutionary Diffusion Filter

Stochastic filtering (also commonly called particle filtering) methods are basically to solve the Bayesian forward-backward recursion problem [6]. They generally generate and propagate a set of weighted particles to approximate the posterior density distribution and recursively search for the optimal estimate for each state of a dynamic system with noisy and incomplete or inaccurate observations.

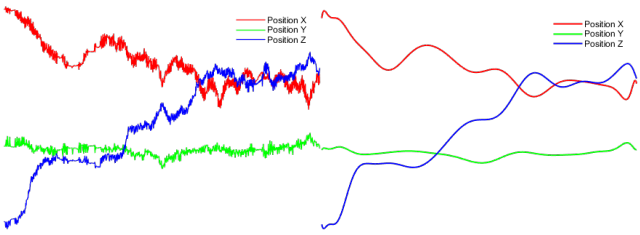
Suppose $\mathbf{x}_i$ and $\mathbf{y}_i$ be the current state and observation at time i of the dynamic system. The history observations are $\mathcal{Y}_i = \{\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_i\}$ ($i = 1, \cdots, N$, N is the number of observations). The set of particles is represented by $\mathcal{X}_i = \{(\mathbf{x}_i^j, \omega_i^j, \varphi_i^j), j = 1...M\}$ (M is the number of particles) with particle weight ω_i^j and accumulative weight φ_i^j. Stochastic filtering aims to approximate posterior probability distribution $p(\mathbf{x}_i|\mathcal{Y}_i)$ of current state $\mathbf{x}_i$. Based on the previous work [6], propagating the particle set $\mathcal{X}_i$ can approximate the posterior probability $p(\mathbf{x}_i|\mathcal{Y}_i)$ by

$$p(\mathbf{x}_i|\mathcal{Y}_i) \approx \sum_{j=1}^{M} \omega_i^j \delta(\mathbf{x}_i - \mathbf{x}_i^j), \tag{1}$$

where $\delta(\cdot)$ is the Dirac delta function and weight ω_i^j is

$$\omega_i^j \propto \omega_{i-1}^j \frac{p(\mathbf{y}_i|\mathbf{x}_i^j)p(\mathbf{x}_i^j|\mathbf{x}_{i-1}^j)}{\Pi(\mathbf{x}_i^j|\mathbf{x}_{i-1}^j, \mathbf{y}_i)}, \tag{2}$$

where the proposal $\Pi(\cdot)$ is an importance density function relative to the degree of the particle degeneracy. It is convenient to select $\Pi(\cdot)$ as prior $p(\mathbf{x}_i^j|\mathbf{x}_{i-1}^j)$: $\Pi(\mathbf{x}_i^j|\mathbf{x}_{i-1}^j, \mathbf{y}_i) = p(\mathbf{x}_i^j|\mathbf{x}_{i-1}^j)$ and then obtain $\omega_i^j \propto \omega_{i-1}^j p(\mathbf{y}_i|\mathbf{x}_i^j)$ [6].



(a) Jitter 3-D tracking (b) Smooth 3-D tracking

Figure 1. Trajectories of estimated 3-D positions of camera movements in jitter and smooth 3-D tracking, respectively

Stochastic filtering suffers from several limitations. The optimal importance distribution $\Pi(\cdot)$ is no guarantee of effective filtering. A particle degeneracy problem that most of these particles $\mathcal{X}_i$ tend to have very small or zero weight results in less effective particles to approximate the posterior probability distribution $p(\mathbf{x}_i|\mathcal{Y}_i)$. Consequently, it is necessary to resample the particles frequently to improve the particle effectiveness, e.g., sequential important resampling can intensively employ effective particles to simulate $p(\mathbf{x}_i|\mathcal{Y}_i)$. Unfortunately, excessive resampling leads to another big problem called particle impoverishment in which only certain particles dominate the posterior probability distribution, making $p(\mathbf{x}_i|\mathcal{Y}_i)$ losing its diversity [6].

The idea of this work contains two aspects to address the problems discussed above. From one point of view, particle (initial solution) $\mathbf{x}_i$ should be constrained as close as possible to the optimal solution, preventing particles from getting trapped in small or zero weights. From another point of view, these particles should be propagated to new states diversified as highly as possible to solve the exploration-exploitation dilemma without particle impoverishment for effective filtering. Based on the two points, this work develops two spatial state constraints and evolutionary diffusion filtering with adaptive history-recall differential evolution to establish a new nonlinear optimization algorithm.

3.1. Spatial State Constraints

Endoscopic cameras are routinely used to visually observe or examine diseases inside various tubular or cylindrical structures with bifurcations (e.g., the vessels, bronchi, colon, and urethra) in the body. This implies that the endoscopic camera should be physically located inside the tube or cylinder. On the other hand, hand-held camera 3-D movements are usually continuous and smooth, which implies that 3-D tracking estimates also should be smooth. Therefore, this work introduces two spatial state constraints for evolutionary diffusion filtering.

1) Smooth State Constraint. This condition is to obtain more smooth state $\mathbf{x}_i$ without jitter errors. Hand-held camera movements are spatially continuous. Unfortunately, the current observed camera movements are jitter and noise due

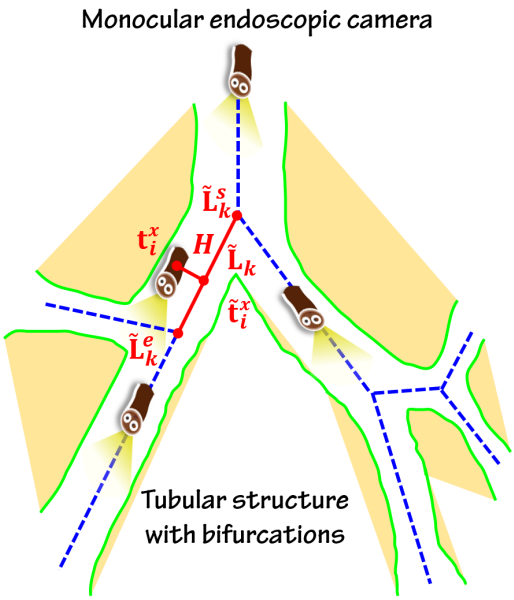


Figure 2. The centerline constraint projects the current state (camera position $\mathbf{t}_i^x$) on the centerline of the tubular structure

to observation distortion and uncertainty (Figure 1).

Suppose the current state $\mathbf{x}_i$ and observation $\mathbf{y}_i$ represents camera position $\mathbf{t}_i$ and direction (quaternion) $\mathbf{q}_i$:

$$\mathbf{x}_i = [\mathbf{t}_i^x, \mathbf{q}_i^x], \mathbf{y}_i = [\mathbf{t}_i^y, \mathbf{q}_i^y]. \tag{3}$$

To tackle the jitter or discontinuity, we introduce the Catmull–Rom spline interpolation and spherical linear interpolation to smooth the current state $[\mathbf{t}_i^x, \mathbf{q}_i^x]$ [7, 24]

$$\mathbf{t}_i^x = \Gamma \begin{bmatrix} 0 & 1 & 0 & 0 \\ -\eta & 0 & \eta & 0 \\ 2\eta & \eta-3 & 3-2\eta & -\eta \\ -\eta & 2-\eta & \eta-2 & \eta \end{bmatrix} \begin{bmatrix} \mathbf{t}_{d-1}^y \\ \mathbf{t}_d^y \\ \mathbf{t}_{d+1}^y \\ \mathbf{t}_{d+2}^y \end{bmatrix}, \tag{4}$$

$$\mathbf{q}_i^x = \begin{cases} \frac{\sin(1-\rho)\phi}{\sin\phi}\mathbf{q}_d^y + \frac{\sin\rho\phi}{\sin\phi}\mathbf{q}_{d+1}^y & \phi \geq 0 \\ \frac{\sin(1-\rho)\phi}{\sin\phi}\mathbf{q}_d^y - \frac{\sin\rho\phi}{\sin\phi}\mathbf{q}_{d+1}^y & \phi < 0 \end{cases}, \tag{5}$$

$$\phi = \arccos \frac{\langle \mathbf{q}_d^y, \mathbf{q}_{d+1}^y \rangle}{\|\mathbf{q}_d^y\| \, \|\mathbf{q}_{d+1}^y\|}, \tag{6}$$

where $\Gamma = [1 \;\; \rho \;\; \rho^2 \;\; \rho^3]$, the interpolation ratio $\rho = i/c - \lfloor i/c \rfloor$, $\mathbf{t}_{d-1}^y \cdots \mathbf{t}_{d+2}^y$ are positions of the continuous controlled points ($d = \lfloor i/c \rfloor$, the floor operator $\lfloor \cdot \rfloor$, the time spacing c), the tension parameter η impacts on the curvature at the control points (usually, $\eta = 0.5$), and symbol $\langle , \rangle$ denotes the dot-product operator.

2) Centerline State Constraint. This constraint aims to achieve more accurate particle states to approximate the optimal solution. Physically, endoscopic cameras are certainly

located inside the tubular structure. This implies that the current particle state should also be inside the tubular structure. Unfortunately, the current state $\mathbf{x}_i$ related to the current observation $\mathbf{y}_i$ is possibly outside the tubular structure due to inaccurate observations (e.g., noise and distortion).

To address the problem of observation uncertainties, we propose a centerline constraint to prevent all the observations from locating outside the tubular structure. The idea of this constraint is motivated by a fact that endoscopic cameras most commonly fly through the tubular structure along its centerline to obtain the largest field of view of the interior surface. This implies that the estimates of camera positions should be located around the centerlines of the tubular structure. Based on that, we project or constrain the current particle state $\mathbf{x}_i$ on the centerlines of the tubular structure, keeping all of the particle states interpolated by the current observation $\mathbf{y}_i$ inside the tubular structure.

We segment 3-D volumetric data to obtain a set of centerlines (curves) $\mathcal{C} = \{\mathcal{L}_k = (\mathcal{L}_k^s, \mathcal{L}_k^e)\}_{k=1}^K$ (here K is the number of the extracted centerlines, $\mathcal{L}_k^s$ and $\mathcal{L}_k^e$ are the centerline's start and end points, respectively), and actually $\mathcal{L}_k$ consists of a set of 3-D points $\mathcal{L}_k = \{\mathcal{L}_k^s, \mathcal{L}_k^{s+1} \cdots \mathcal{L}_k^e\}$.

We assign the closest centerline $\hat{\mathcal{L}}_k$ to the current interpolated state $\mathbf{t}_i^x$ by minimizing the Euclidean distance between $\mathbf{t}_i^x$ and the centerline set $\mathcal{C}$ (Figure 2)

$$\hat{\mathcal{L}}_k = \arg\min_{\mathcal{L}_k \in \mathcal{C}} \mathcal{H}(\mathbf{t}_i^x, \mathcal{L}_k), \tag{7}$$

where the Euclidean distance $\mathcal{H}(\mathbf{t}_i^x, \mathcal{L}_k)$ is calculated by

$$\mathcal{H}(\mathbf{t}_i^x, \mathcal{L}_k) = \begin{cases} \|\mathbf{t}_i^x - \mathcal{L}_k^s\| & \zeta < 0 \\ \|\mathbf{t}_i^x - \mathcal{L}_k^e\| & \zeta > \|\mathcal{L}_k\| \\ \sqrt{\|\mathbf{t}_i^x - \mathcal{L}_k^s\|^2 - \zeta^2} & \text{otherwise} \end{cases}, \tag{8}$$

where the centerline length $\|\mathcal{L}_k\| = \|\mathcal{L}_k^e - \mathcal{L}_k^s\|$ and ζ denotes the length of the vector $(\mathbf{t}_i^x - \mathcal{L}_k^s)$ that is projected on the centerline $\mathcal{L}_k$. The length ζ can be calculated by

$$\zeta = \langle \mathbf{t}_i^x - \mathcal{L}_k^s,\, \mathcal{L}_k^e - \mathcal{L}_k^s \rangle \,/ \|\mathcal{L}_k^e - \mathcal{L}_k^s\|, \tag{9}$$

where $\zeta < 0$ and $\zeta > \|\mathcal{L}_k\|$ indicate that projected point $\tilde{\mathbf{t}}_i^x$ is located on the previous and next centerlines of the current centerline $\mathcal{L}_k$, respectively; otherwise, it is located on $\mathcal{L}_k$.

We may obtain several closest centerlines $\{\hat{\mathcal{L}}_k\}_{k=1,2,\cdots}$ that have the same distance to $\mathbf{t}_i^x$ after the minimization procedure (Eq. (7)). We compute the angle between the orientation $\mathbf{q}_i^x$ in the $z-$direction $\mathbf{q}_i^{x,z}$ and centerline direction $\mathcal{Z}$ to determine the optimal centerline $\tilde{\mathcal{L}}_k$:

$$\tilde{\mathcal{L}}_k = \arg\min_{\hat{\mathcal{L}}_k} \arccos \left\langle \underbrace{\frac{\hat{\mathcal{L}}_k^e - \hat{\mathcal{L}}_k^s}{\left\|\hat{\mathcal{L}}_k^e - \hat{\mathcal{L}}_k^s\right\|}}_{\mathcal{Z}}, \frac{\mathbf{q}_i^{x,z}}{\|\mathbf{q}_i^{x,z}\|} \right\rangle. \tag{10}$$

After determining the optimal centerline $\tilde{\mathcal{L}}_k$, we project the current interpolated particle state $\mathbf{t}_i^x$ on the optimal centerline $\tilde{\mathcal{L}}_k$ and obtain the projected position $\tilde{\mathbf{t}}_i^x$ by

$$\tilde{\mathbf{t}}_i^x = \tilde{\mathcal{L}}_k^s + \frac{\left\langle \mathbf{t}_i^x - \tilde{\mathcal{L}}_k^s,\, \tilde{\mathcal{L}}_k^e - \tilde{\mathcal{L}}_k^s \right\rangle}{\left\|\tilde{\mathcal{L}}_k^e - \tilde{\mathcal{L}}_k^s\right\|} \frac{(\tilde{\mathcal{L}}_k^e - \tilde{\mathcal{L}}_k^s)}{\left\|\tilde{\mathcal{L}}_k^e - \tilde{\mathcal{L}}_k^s\right\|}. \tag{11}$$

We also update the camera orientation in $z-$direction by

$$\tilde{\mathbf{q}}_i^{x,z} = \frac{(\tilde{\mathcal{L}}_k^e - \tilde{\mathcal{L}}_k^s)}{\left\|\tilde{\mathcal{L}}_k^e - \tilde{\mathcal{L}}_k^s\right\|}. \tag{12}$$

Finally, we obtain the current new state $\tilde{\mathbf{x}}_i$ after implementing the smooth and centerline constraints

$$\tilde{\mathbf{x}}_i = [\tilde{\mathbf{t}}_i^x, \tilde{\mathbf{q}}_i^x], \tag{13}$$

which is propagated by evolutionary diffusion filtering.

3.2. Evolutionary Diffusion Filtering

Our new evolutionary diffusion filtering performs the following three stages to approximate the optimal solution.

1) Constrained Randomization. Before the particle propagation or diffusion, we perform a constrained randomization procedure to initialize the particle set $\mathcal{X}_i$, which is also useful to enhance the diversity of the particle set$\mathcal{X}_i$.

Basically, the particle set $\mathcal{X}_i$ is randomly initialized by an uniform distribution model with respect to the current spatially constrained particle state $\tilde{\mathbf{x}}_i$ (obtained by Eqs. (4), (5), (11), (12), and (13). Actually, particle $\mathbf{x}_i^j$ is generated by randomly perturbing the constrained state $\tilde{\mathbf{x}}_i$

$$\mathbf{x}_i^j \sim \mathcal{U}(\tilde{\mathbf{x}}_i, \varepsilon_i^j), \tag{14}$$

where $\mathcal{U}(\varepsilon_i^j)$ denotes an uniform-density perturbation or noise term. After that, we also update ω_i^j and φ_i^j for each particle $\mathbf{x}_i^j$ before the particle diffusion step.

2) Evolutionary Stochastic Diffusion. To address the particle degeneracy and impoverishment problems, we skip sequential important resampling on the particle set $\mathcal{X}_i$ and replace it by an evolutionary computing strategy called adaptive history-recall differential evolution.

Different from stochastic diffusion in conventional particle filtering, our new evolutionary stochastic diffusion to propagate $\mathbf{x}_i^j$ to the new state $\tilde{\mathbf{x}}_i^j$ is formulated as

$$\tilde{\mathbf{x}}_i^j = \Psi(\hat{\mathbf{x}}_i^j, \hat{\varepsilon}_i^j),\ \hat{\mathbf{x}}_i^j = \mathbf{E}(\mathbf{x}_{i-1}^j, \mathbf{x}_i^j, \Delta\mathbf{x}_i), \tag{15}$$

where $\Psi(\cdot)$ is a propagation function, $\hat{\varepsilon}_i^j$ is Gaussian noise. We define a function or operator $\mathbf{E}$ as adaptive history-recall differential evolution that is a new stochastic propagation using the current constrained observation $\Delta\mathbf{x}_i$ and historical particle $\mathbf{x}_{i-1}^j$ to obtain the new particle $\mathbf{E}(\mathbf{x}_{i-1}^j, \mathbf{x}_i^j, \Delta\mathbf{x}_i)$.

Differential evolution (DE) is a powerful tool for various stochastic optimization problems [8]. Compared to other methods such as particle swarm optimization and artificial bee colony, DE generally provides much better repeatability and the quality of obtained solutions [17]. DE consists of three main steps of mutation, crossover, and selection. Both mutation and crossover play an essential role in balancing convergence and computational efficiency. The selection should be appropriate to precisely evaluate the quality or fitness of the particle. Improper mutation and crossover potentially result in premature convergence and local minima [8]. This work proposes adaptive history-recall differential evolution that integrates the historical estimate and current constrained observation to propagate and diversify set $\mathcal{X}_i$. Such an evolution runs the following three steps.

Adaptive Mutation. Let $\mathbf{x}_{i-1}^*$ and $\mathbf{x}_i^*$ be the particles with the best or maximal weight in populations $\mathcal{X}_{i-1}$ and $\mathcal{X}_i$. For a target particle $\mathbf{x}_i^j$, its mutant vector $\mathbf{v}_i^j$ is defined by a new mutation strategy with the current observation $\Delta\mathbf{x}_i$.

$$\mathbf{v}_i^j = \mathbf{x}_{i-1}^* + \lambda\Delta\mathbf{x}_i + \mu_i^*(\mathbf{x}_i^* - \mathbf{x}_i^j) + \mu_i^r(\mathbf{x}_i^{r_1} - \mathbf{x}_i^{r_2}), \quad (16)$$

where λ is an inertia weight with uniform distribution $[0, 1]$, μ_i is a mutation factor, vector $(\mathbf{x}_i^* - \mathbf{x}_i^j)$ disturbs the base state $\mathbf{x}_i^j$, $(\mathbf{x}_i^{r_1} - \mathbf{x}_i^{r_2})$ denotes the difference vector, r^1 and r^2 are mutually exclusive integers chosen randomly from set $\{1, \cdots, j-1, j+1, \cdots, M\}$. The new mutation strategy absorbs the current observation, improving the diversity of the population $\mathcal{X}_i$. We adaptively calculate the coefficients:

$$\mu_i^* = \frac{2\mathcal{W}(\mathbf{x}_i^*)}{\mathcal{W}(\mathbf{x}_i^*) + \mathcal{W}(\mathbf{x}_i^j)}, \ \mu_i^r = \frac{2\mathcal{W}(\mathbf{x}_i^j)}{\mathcal{W}(\mathbf{x}_i^*) + \mathcal{W}(\mathbf{x}_i^j)}, \quad (17)$$

where $\mathcal{W}(\mathbf{x}_i^*)$ and $\mathcal{W}(\mathbf{x}_i^j)$ are the fitness of $\mathbf{x}_i^*$ and $\mathbf{x}_i^j$.

Binomial Crossover. This step is further to diversify the population to avoid getting trapped into local minima. It explores a m-dimensional trial vector by exchanging the target and mutant vectors. This work uses the binomial crossover to generate trial vector $\mathbf{u}_i^j = \{u_i^{j,1} \cdots u_i^{j,m}\}$ with respect to $\mathbf{x}_i^j = \{x_i^{j,1} \cdots x_i^{j,m}\}$ and $\mathbf{v}_i^j = \{v_i^{j,1} \cdots v_i^{j,m}\}$:

$$u_i^{j,m} = \begin{cases} v_i^{j,m} & \text{if } (r_a \le C_r) \text{ or } (m = m_r) \\ x_i^{j,m} & \text{otherwise} \end{cases}, \quad (18)$$

where random number r_a yields uniform distribution, m_r is randomly selected from $\{1, 2, \cdots m\}$, and C_r is the crossover rate or probability that determines if $u_i^{j,m}$ duplicates $v_i^{j,m}$. Das et al. [8] suggested that the crossover probability C_r ranges within an internal $[0, 1]$ for balancing the global and local searching abilities. Instead of predefining C_r , we still adaptively calculate it for better crossover

$$C_r = \frac{\mathcal{W}(\mathbf{x}_i^j) + \mathcal{W}(\mathbf{v}_i^j)}{2}. \quad (19)$$

History-Recall Selection. We obtain the trial population $\mathcal{U}_i$ after the crossover and select the particle based on its fitness. Differential evolution does not recall the historical population $\mathcal{X}_{i-1}$ that are the best solutions at frame $i-1$. In this work, the selection step recalls the population $\mathcal{X}_{i-1}$ to determine the new particle state $\hat{\mathbf{x}}_i^j$ by the fitness $\mathcal{W}(\cdot)$

$$\hat{\mathbf{x}}_i^j = \mathbf{E}(\mathbf{x}_{i-1}^j, \mathbf{x}_i^j, \Delta\mathbf{x}_i) = \arg \max_{\mathbf{x} \in \{\mathbf{x}_{i-1}^j, \mathbf{x}_i^j, \mathbf{u}_i^j\}} \mathcal{W}(\mathbf{x}), \quad (20)$$

which shows that $\hat{\mathbf{x}}_i^j$ has the biggest fitness after selecting.

3) Marginal Likelihood Probability. After the evolutionary stochastic diffusion, we obtain a set of new particle $\tilde{\mathbf{x}}_i^j$. Since stochastic filtering employs the history and current observations to approximate the probability density function of the current state $\mathbf{x}_i$, it must determine the observation likelihood $p(\mathbf{y}_i|\mathbf{x}_i = \tilde{\mathbf{x}}_i^j)$ that is the conditional probability of the current observation $\mathbf{y}_i$ given the current state $\mathbf{x}_i$. Basically, the conditional probability $p(\mathbf{y}_i|\mathbf{x}_i = \tilde{\mathbf{x}}_i^j)$ is approximated by the weight $\tilde{\omega}_i^j$ of new state particle $\tilde{\mathbf{x}}_i^j$

$$p(\mathbf{y}_i|\mathbf{x}_i = \tilde{\mathbf{x}}_i^j) \propto \frac{\tilde{\omega}_i^j}{\sum_{j=1}^M \tilde{\omega}_i^j}. \quad (21)$$

We also update the accumulative weight $\tilde{\varphi}_i^j$

$$\tilde{\varphi}_i^j = \varphi_i^j + \tilde{\omega}_i^j. \quad (22)$$

After stages *2)* and *3)*, we can obtain a population of particles with new states and weights $\tilde{\mathcal{X}}_i = \{(\tilde{\mathbf{x}}_i^j, \tilde{\omega}_i^j, \tilde{\varphi}_i^j)\}$.

3.3. Application to Bronchoscope 3-D Tracking

Monocular endoscope tracking is to estimate the position and orientation (direction) of surgical endoscopic cameras in a reference space. It is actually a multisource or multisensor data fusion procedure. In this work, multisource data include intraoperative endoscopic video sequences (observations), intraoperative EM sensor measurements (observations), and preoperative CT images. To fuse these data, we develop a surgical hybrid 3-D tracking method to robustly predict the surgical camera pose in the 3-D CT image space.

Our hybrid 3-D tracking is established exactly in accordance with constrained evolutionary diffusion filtering. In this dynamic system, we have two observations of EM sensor measurement $\mathbf{y}_i$ and endoscopic image $\mathbf{I}_i$ at time i. EM sensor measured pose $\mathbf{y}_i$ is use to calculate spatial state constraints and determine the current constrained observation $\Delta\mathbf{x}_i$ in Eq. (15). According to Eq. (3), $\mathbf{x}_i$ represent the camera pose with position $\mathbf{t}_i^x$ and quaternion $\mathbf{q}_i^x$. This implies $\mathbf{x}_i$ is a 7-dimensional vector and $m = 7$ in Eq. (18).

The fitness $\mathcal{W}(\mathbf{x}_i^j)$ and weight $\tilde{\omega}_i^j$ must be determined in evolutionary diffusion filtering. This work defines them as the selective structural similarity $\mathcal{S}(,)$ between another observation (endoscopic image) $\mathbf{I}_i$ and 2-D virtual images

$$\mathcal{W}(\mathbf{x}_i^j) = \mathcal{S}(\mathbf{I}_i, \hat{\mathbf{I}}(\mathbf{x}_i^j)), \ \tilde{\omega}_i^j = \tilde{\omega}(\tilde{\mathbf{x}}_i^j) = \mathcal{S}(\mathbf{I}_i, \hat{\mathbf{I}}(\tilde{\mathbf{x}}_i^j)), \quad (23)$$

Algorithm 1: Hybrid bronchoscope 3-D tracking

Input: Bronchoscopic video sequences, EM sensor measurements, and volumetric CT data
Output: 3-D camera pose (position and direction)

for $i = 1$ **to** N (frames or EM measurements) **do**
- ❶ Spatial constrains on $\mathbf{x}_i$ *Eqs. (4)∼(13)*;
- ❷ Constrained randomization *Eq. (14)* to generate M initialized particles $\mathcal{X}_i$;
- ❸ Evolutionary stochastic diffusion on $\mathcal{X}_i$: **for** $j = 1$ **to** M (particle number) **do**
  - ❹ Perform adaptive history-recall differential evolution with steps 1), 2), 3)
  - 1) Adaptive mutation to calculate the mutation vector $\mathbf{v}_i^j$ *Eqs. (16)∼(17)*;
  - 2) Binomial crossover to calculate the trial vector $\mathbf{u}_i^j$ *Eqs. (18)∼(19)*;
  - 3) History-recall selection *Eq. (20)* to get $\hat{\mathbf{x}}_i^j$;
  - ❺ Propagate $\hat{\mathbf{x}}_i^j$ to new state $\tilde{\mathbf{x}}_i^j$ *Eq. (15)*;
  - ❻ Update $\tilde{\omega}_i^j$ and $\tilde{\varphi}_i^j$ *Eqs. (21)∼(23)*;
  - $j = j + 1$;
- **end**
- ❼ Propagated particles $\tilde{\mathcal{X}}_i = \{(\tilde{\mathbf{x}}_i^j, \tilde{\omega}_i^j, \tilde{\varphi}_i^j)\}$;
- ❼ Determine the optimal estimate $\mathbf{x}_i$ *Eq. (24)*;
- ❾ Go to the next frame $i = i + 1$;

end

where virtual images $\hat{\mathbf{I}}(\mathbf{x}_i^j)$ and $\hat{\mathbf{I}}(\tilde{\mathbf{x}}_i^j)$ are generated by volume rendering on the basis of poses $\mathbf{x}_i^j$, $\tilde{\mathbf{x}}_i^j$, and CT images.

After evolutionary diffusion filtering, we obtain a set of particles with new states $\tilde{\mathcal{X}}_i = \{(\tilde{\mathbf{x}}_i^j, \tilde{\omega}_i^j, \tilde{\varphi}_i^j)\}$. Each new particle $\tilde{\mathbf{x}}_i^j \in \tilde{\mathcal{X}}_i$ denotes a potential solution or optimal estimate for the current camera pose $\mathbf{x}_i$. Eventually, we find and determine the optimal camera pose $\mathbf{x}_i$ at frame i by

$$\mathbf{x}_i = \arg\max_{\tilde{\mathbf{x}}_i^j \in \tilde{\mathcal{X}}_i} \mathcal{S}(\mathbf{I}_i, \hat{\mathbf{I}}(\tilde{\mathbf{x}}_i^j)), \tag{24}$$

which means that optimal estimate $\mathbf{x}_i$ has the best weight.

Algorithm 1 summarizes a new hybrid bronchoscope 3-D tracking method that uses CEDF for multisoure information fusion. As well known, the performance of stochastic nonlinear optimization algorithms depends on the exploration-exploitation balance. While exploration aims to explore the search space more thoroughly and find more diverse solutions, exploitation employs local information in the search process to generate better solutions at close vicinity of the current ones. CEDF uses adaptive history-recall differential evolution to diversify solutions and evaluates the particle with large weight and fitness on the basis of the current observation, resulting in strong exploration (global optimums and mature convergence) and strong exploitation (large fitness) that will be demonstrated in experiments.

Table 1. Quantitative comparison of average position and direction errors (e_p, e_d), smoothness (τ, ψ), and visual quality ξ of using the four methods (the error and smoothness units are (mm, degree))

Methods	Metrics				
	e_p	e_d	τ	ψ	ξ
Luo et al. [18]	4.03	10.7	3.57	5.91	0.74
Shen et al. [26]	6.03	16.5	4.76	10.9	0.72
Banach et al. [1]	8.90	21.6	7.37	15.4	0.67
CEDF	**2.59**	**8.42**	**1.73**	**2.86**	**0.79**

4. Validation

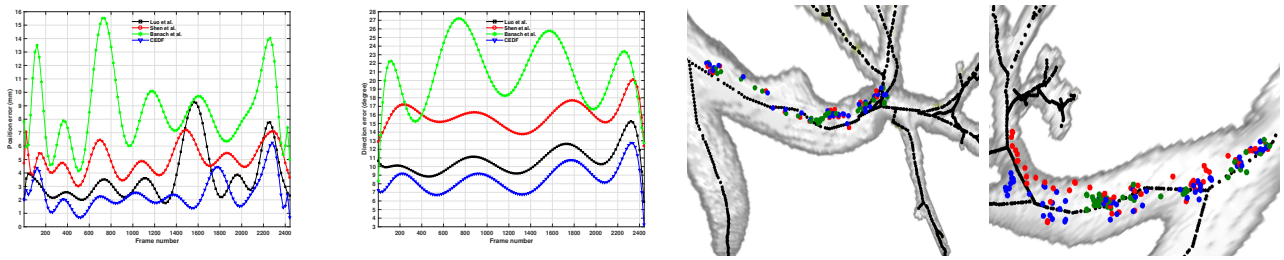
We acquired 32,167 video frames from 17 patient cases (17 volumetric CT datasets) from bronchoscopic procedures under a protocol approved by the research ethics board of the university. Three experts manually generated ground-truth data with 15,657 frames corresponding to 15,657 EM sensor measurements from 6 cases using our developed surgical planing software. Several metrics are used to evaluate hybrid 3-D tracking. We first compute the tracked position and direction errors (e_p, e_d) between the estimated and ground-truth poses. Since the trajectory of an endoscope is continuous and smooth, it is necessary to evaluate the smoothness of hybrid 3-D tracking results. The position smoothness and direction smoothness (τ, ψ) are defined as the average Euclidean distance of estimated positions and direction among continuous endoscopic images.

All the tracking results must be intuitively visualized in augmented reality assisted surgical procedures. We use volume rendering to generate virtual 2-D images that correspond to estimated 3-D camera poses. Then, subjective assessment of the tracked results is performed by three surgeons who manually compare if the current video real and virtual images resemble each other. On the other hand, objective (quantitative) assessment is to calculate the visual quality (similarity) ξ between the real and virtual images on the basis of the structural similarity index measure [30].

We also compare several surgical 3-D tracking approaches: (1) Luo et al. [18]: hybrid 3-D tracking using evolutionary computation, (2)Shen et al. [26]: deep learning based depth estimation and 2-D/3-D image registration, (3) Banach et al. [1]: deep learning based depth estimation and iterative closest point based 3-D/3-D image registration, and (4) the CEDF method discussed in Section 3.

5. Results and Discussion

1) Results. Table 1 shows the results of the four methods evaluated on 6 ground-truth cases. While the position and rotation errors were reduced to (2.59 mm, 8.42°), the smoothness and visual quality of CEDF were much better



(a) Position error (b) Direction error (c) Compared trajectories (Case 2) (d) Compared trajectories (Case 3)

Figure 3. (a) and (b) plot the position and direction errors of using the different endoscope 3-D tracking methods, while (c) and (d) illustrate the trajectories of ground truth (*red dots*), Shen et al. [26] (*green dots*) and CEDF (*blue dots*).

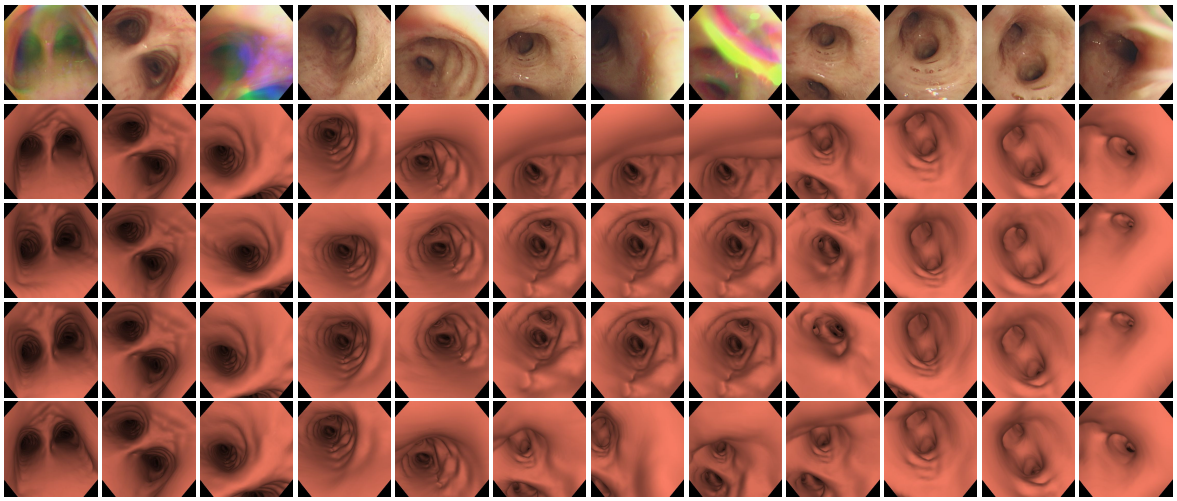


Figure 4. Visual comparison of the tracking results of using the different tracking methods. The first row shows the input endoscopic images and Rows 2∼5 the generated virtual images corresponding to Luo et al. [18], Shen et al. [26], Banach [1], and CEDF, respectively.

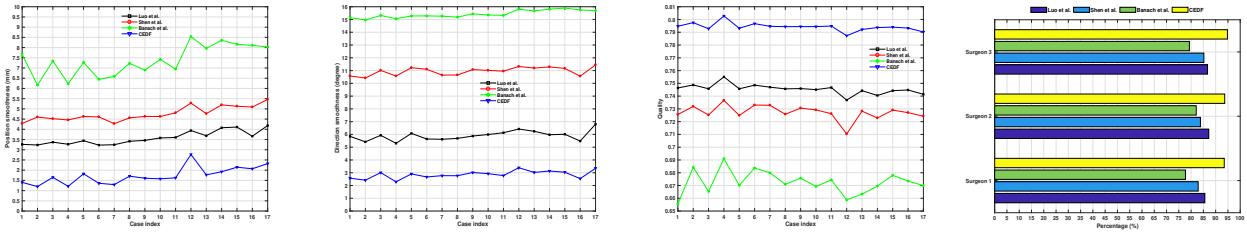
than the others. While Fig. 3(a)(b) shows a case of plotted position and rotation errors of the four methods, Fig. 3(c)(d) illustrate the trajectories of using Shen et al. [26] and ours. The estimated path of CEDF follows ground truth better than Banach et al. [1]. Fig. 4 visually compares the tracking results and investigates if the current real endoscopic image resembles to the virtual image generated by volume rendering on the basis of the estimated pose and CT images. The virtual 2-D rendering images corresponding to the 3-D camera poses tracked by our method resemble much more similar to the real images, demonstrating that CEDF can provide surgeons with better tracking results. Fig. 5(a)(b)(c) display the position and rotation smoothness and the visual quality of the four methods on 17 cases, while Fig. 5(d) shows the subjective assessment results by three surgeons who manually inspect if 15,657 frames of virtual images resemble to the real video images. The average percentage of the resembled virtual images was 86.5%, 84.0%, 79.7%, and 94.0% of using Luo et al. [18], Shen et al. [26], Banach et al. [1], and ours, respectively. We also conduct an ablation study

Table 2. Quantitative results of the ablation study: S and C denote the smooth and centerline constraints, respectively

Methods	Metrics				
	e_p	e_d	τ	ψ	ξ
Baseline [18]	4.03	10.7	3.57	5.91	0.74
Baseline+S	3.83	10.5	3.42	5.68	0.74
Baseline+C	3.52	10.1	2.99	4.77	0.76
Baseline+S+C	3.14	9.26	2.29	3.46	0.77
CEDF	**2.59**	**8.42**	**1.73**	**2.86**	**0.79**

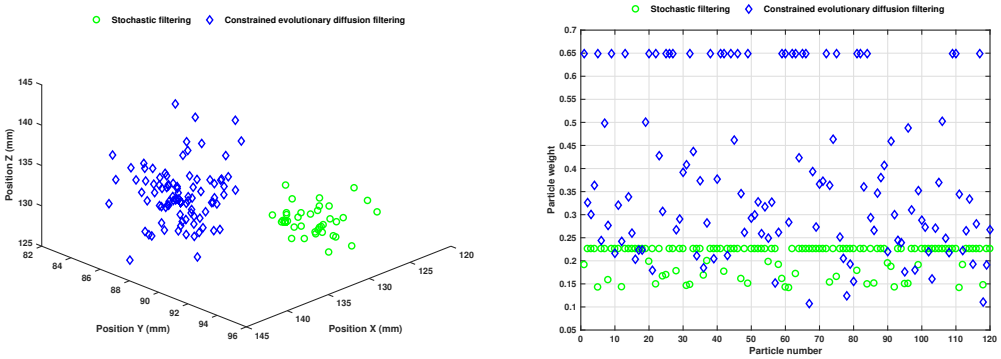
and use [18] as a baseline. Table 2 lists the results of the ablation study, which shows the effectiveness of each part. Additionally, CEDF requires 0.39 seconds to process one frame, which works better than [26] and [1].

2) Effectiveness. Dynamic balancing of exploration and exploitation is a key for nonlinear stochastic optimization methods, especially for population-based optimization algorithms that depend on local search, global search, and randomization. A good balance should engage in sufficient



(a) Position smoothness (b) Direction smoothness (c) Visual quality (d) Surgeon assessment

Figure 5. Smoothness, visual quality, and subjective assessment of using the four 3-D tracking approaches



(a) Particle position diversity (b) Particle weight distribution

Figure 6. Population diversity and weight distribution of stochastic filtering and CEDF: *Blue* and *green* regions are totally different since *blue* successfully tracked the endoscope while *green* failed to. Obviously, CEDF significantly improves the particle diversity and weight.

exploitation to guarantee and continuously enhance the current viability of a population of particles while equally providing sufficient additional focus on the exploration to ensure future viability. As one of nonlinear stochastic optimizers, stochastic filtering usually gets trapped in the exploration-exploitation dilemma due to the particle degeneracy and impoverishment. This work develops a new constrained evolutionary diffusion filtering framework.

The effectiveness of the new framework lies in two aspects. We first introduce the smooth and centerline constrains to spatially keep the current state close to the optimal solution vicinity, which prevents particles from obtaining small or zero weights and successfully addresses the particle degeneracy. More interestingly, without using sequential resampling, CEDF uses evolutionary stochastic diffusion in accordance with adaptive history-recall differential evolution to propagate the particles, certainly resolving the particle impoverishment problem to improve the particle diversity. Fig. 6 demonstrates our proposed method can generate much more diverse particles with larger weights compared to conventional stochastic filtering, achieving a good balance between exploitation and exploration to improve the filtering performance. Therefore, the new filtering method is generally ambidextrous and metaheuristic.

3) Limitations. Accurate calculation fitness and weight to correctly evaluate a swarm of particles is important to population-based stochastic optimization methods. This work defines the fitness-weight evaluation model as an image similarity function that characterizes the difference between the current real video image and virtual rendering images. Unfortunately, surgical cameras collect egocentric videos with limited lighting and field of view, while surgical videos also contain completely uninformative images in complex tubular environments, leading to incorrect fitness and weight computations. Another potential limitation is that the centerline constraint in stochastic optimization approaches requires to accurately segment or extract the centerline of tubular structures. Inaccurate centerline extraction possibly fails to constrain the current observation to the correct centerline, improperly estimating the camera pose.

In summary, this work develops a new constrained evolutionary diffusion filtering strategy for nonlinear stochastic optimization. Specifically, we introduces spatial state constraints and evolutionary diffusion filtering to resolve the degeneracy and impoverishment of stochastic filtering. With application to surgical hybrid 3-D tracking, the experimental results demonstrate that our strategy works much more effective and robust than other 3-D tracking methods.

References

[1] Artur Banach, Franklin King, Fumitaro Masaki, Hisashi Tsukada, and Nobuhiko Hata. Visually navigated bronchoscopy using three cycle-consistent generative adversarial network for depth estimation. *Medical Image Analysis*, page 102164, 2021. 2, 6, 7

[2] J. Maria Diaz Barros, F. Garcia, B. Mirbach, and D. Stricker. Real-time monocular 6-dof head pose estimation from salient 2d points. In *IEEE International Conference on Image Processing (ICIP)*, pages 121–125, 2017. 2

[3] Andrea Cavagna, Stefania Melillo, Leonardo Parisi, and Federico Ricci-Tersenghi. Sparta tracking across occlusions via partitioning of 3d clouds of points. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019. 2

[4] M. Cavazzuti. *Optimization Methods: From Theory to Design*. Springer-Verlag Berlin Heidelberg, 2013. 1

[5] Ming-Fang Chang, John Lambert, Patsorn Sangkloy, Jagjeet Singh, Slawomir Bak, Andrew Hartnett, De Wang, Peter Carr, Simon Lucey, Deva Ramanan, and James Hays. Argoverse: 3d tracking and forecasting with rich maps. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8748–8757, 2019. 2

[6] Dan Crisan and Boris Rozovskii, editors. *The Oxford Handbook of Nonlinear Filtering*. Oxford University Press, 2011. 2, 3

[7] Thomas L. Curtright, David B. Fairlie, and Cosmas K. Zachos. A compact formula for rotations as spin matrix polynomials. *Symmetry, Integrability and Geometry: Methods and Applications*, 10:084–1–15, 2014. 3

[8] Swagatam Das, Sankha Subhra Mullick, and P. N. Suganthan. Recent advances in differential evolution – an updated survey. *Swarm and Evolutionary Computation*, 27:1–30, 2016. 5

[9] C. Forster, Z. Zhang, M. Gassner, M. Werlberger, and D. Scaramuzza. Svo: Semidirect visual odometry for monocular and multicamera systems. *IEEE Transactions on Robotics*, 33(2):249–265, 2017. 1, 2

[10] Davi Frossard and Raquel Urtasun. End-to-end learning of multi-sensor 3d tracking by detection. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 635–642, 2018. 2

[11] Guillermo Gallego, Jon E.A. Lund, Elias Mueggler, Henri Rebecq, Tobi Delbruck, and Davide Scaramuzza. Event-based, 6-dof camera tracking from photometric depth maps. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(10):2402–2412, 2019. 1, 2

[12] Mathieu Garon and Jean-Francois Lalonde. Deep 6-dof tracking. *IEEE Transactions on Visualization and Computer Graphics*, 23(11):2410–2418, 2017. 2

[13] Hanme Kim, Stefan Leutenegger, and Andrew J. Davison. Real-time 3d reconstruction and 6-dof tracking with an event camera. In *European Conference on Computer Vision (ECCV)*, pages 349–364, 2016. 1, 2

[14] Tristan Laidlow, Jan Czarnowski, and Stefan Leutenegger. Deepfusion: Real-time dense 3d reconstruction for monocular slam using single-view depth and gradient predictions. In *International Conference on Robotics and Automation (ICRA)*, pages 4068–4074, 2019. 2

[15] Shan Lin, Albert J. Miao, Jingpei Lu, and et al. Semantic-super: A semantic-aware surgical perception framework for endoscopic tissue classification, reconstruction, and tracking. *arXiv:2210.16674*, 2022. 2

[16] X. Liu, A. Sinha, M. Ishii, G. D. Hager, Austin Reiter, Russell H. Taylor, and Mathias Unberath. Dense depth estimation in monocular endoscopy with self-supervised learning methods. *IEEE Transactions on Medical Imaging*, 2019. 1, 2

[17] Ying Liu, Aixin Sun, Han Tong, and et al. *Advances of Computational Intelligence in Industrial Systems*. Springer-Verlag Berlin Heidelberg, 2008. 5

[18] X Luo, Y Wan, X He, and K Mori. Observation-driven adaptive differential evolution and its application to accurate and smooth bronchoscope three-dimensional motion tracking. *Medical Image Analysis*, 24(1):282–296, 2015. 2, 6, 7

[19] Nader Mahmoud, Toby Collins, Alexandre Hostettler, Luc Soler, Christophe Doignon, and Jose Maria Martinez Montiel. Live tracking and dense reconstruction for handheld monocular endoscopy. *IEEE Transactions on Medical Imaging*, 38(1):79–89, 2019. 1, 2

[20] Raul Mur-Artal, J. M. M. Montiel, Juan D. Tardos, and et al. Orb-slam: A versatile and accurate monocular slam system. *IEEE T. Robotics*, 31:1147–1163, 2015. 1, 2

[21] Rohit Pandey, Pavel Pidlypenskyi, Shuoran Yang, and Christine Kaeser-Chen. Efficient 6-dof tracking of handheld objects from an egocentric viewpoint. In *European Conference on Computer Vision (ECCV)*, pages 426–441, 2018. 1, 2

[22] Henri Rebecq, Timo Horstschaefer, Guillermo Gallego, and Davide Scaramuzza. Evo: A geometric approach to event-based 6-dof parallel tracking and mapping in real time. *IEEE Robotics and Automation Letters*, 2(2):593–600, 2017. 2

[23] Yasir Salih and Aamir Saeed Malik. Comparison of stochastic filtering methods for 3d tracking. *Pattern Recognition*, 44(10–11):2711–2737, 2011. 1, 2

[24] David Salomon. *The computer graphics manual*. Springer, 2011. 3

[25] Shuwei Shao, Zhongcai Pei, Weihai Chen, and et al. Self-supervised monocular depth and ego-motion estimation in endoscopy: Appearance flow to the rescue. *Medical Image Analysis*, 77:102338, 2022. 2

[26] M. Shen, Y. Gu, N. Liu, and G.-Z. Yang. Context-aware depth and pose estimation for bronchoscopic navigation. *IEEE Robotics and Automation Letters*, 4(2):732–739, 2019. 1, 2, 6, 7

[27] 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*, 57(3):736–745, 2010. 2

[28] Angela Sorriento, Maria Bianca Porfido, Stefano Mazzoleni, Giuseppe Calvosa, Miria Tenucci, Gastone Ciuti, and Paolo Dario. Optical and electromagnetic tracking systems for biomedical applications: A critical review on potentialities

and limitations. *IEEE Reviews in Biomedical Engineering*, 13:212–232, 2020. 2

[29] Keisuke Tateno, Federico Tombari, Iro Laina, and Nassir Navab. Cnn-slam: Real-time dense monocular slam with learned depth prediction. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6565–6574, 2017. 2

[30] 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*, 13(4):600–612, 2004. 6

[31] Jianing Wei, Genzhi Ye, Tyler Mullen, Matthias Grundmann, Adel Ahmadyan, and Tingbo Hou. Instant motion tracking and its applications to augmented reality. *arXiv:1907.06796*, 2019. 2

[32] Bo Yang, Siyuan Xu, Hongrong Chen, and et al. Reconstruct dynamic soft-tissue with stereo endoscope based on a single-layer network. *IEEE Transactions on Image Processing*, 31:5828–5840, 2022. 2

[33] Huizhong Zhou, Benjamin Ummenhofer, and Thomas Brox. Deeptam: Deep tracking and mapping with convolutional neural networks. *International Journal of Computer Vision*, 128:756–769, 2020. 1, 2