



# A GPU-accelerated model-based tracker for untethered submillimeter grippers

Stefano Scheggi<sup>a</sup>, ChangKyu Yoon<sup>b</sup>, Arijit Ghosh<sup>c</sup>, David H. Gracias<sup>b,c</sup>, Sarthak Misra<sup>a,d,\*</sup>

<sup>a</sup> Department of Biomechanical Engineering, University of Twente, 7522 NB, The Netherlands

<sup>b</sup> Department of Materials Science and Engineering, The Johns Hopkins University, MD 21218, USA

<sup>c</sup> Department of Chemical and Biomolecular Engineering, The Johns Hopkins University, MD 21218, USA

<sup>d</sup> Department of Biomedical Engineering, University of Groningen and University Medical Center Groningen, 9713 GZ, The Netherlands

## ARTICLE INFO

### Article history:

Received 30 June 2017

Received in revised form 10 October 2017

Accepted 10 November 2017

Available online 5 December 2017

### Keywords:

Model-based tracker

Graphics Processing Unit

Particle Swarm Optimization

Miniaturized grippers

## ABSTRACT

Miniaturized grippers that possess an untethered structure are suitable for a wide range of tasks, ranging from micromanipulation and microassembly to minimally invasive surgical interventions. In order to robustly perform such tasks, it is critical to properly estimate their overall configuration. Previous studies on tracking and control of miniaturized agents estimated mainly their 2D pixel position, mostly using cameras and optical images as a feedback modality. This paper presents a novel solution to the problem of estimating and tracking the 3D position, orientation and configuration of the tips of submillimeter grippers from marker-less visual observations. We consider this as an optimization problem, which is solved using a variant of the Particle Swarm Optimization algorithm. The proposed approach has been implemented in a Graphics Processing Unit (GPU) which allows a user to track the submillimeter agents online. The proposed approach has been evaluated on several image sequences obtained from a camera and on B-mode ultrasound images obtained from an ultrasound probe. The sequences show the grippers moving, rotating, opening/closing and grasping biological material. Qualitative results obtained using both hydrogel (soft) and metallic (hard) grippers with different shapes and sizes ranging from 750 microns to 4 mm (tip to tip), demonstrate the capability of the proposed method to track the agent in all the video sequences. Quantitative results obtained by processing synthetic data reveal a tracking position error of  $25 \pm 7 \mu\text{m}$  and orientation error of  $1.7 \pm 1.3$  degrees. We believe that the proposed technique can be applied to different stimuli responsive miniaturized agents, allowing the user to estimate the full configuration of complex agents from visual marker-less observations.

© 2017 Elsevier B.V. All rights reserved.

## 1. Introduction

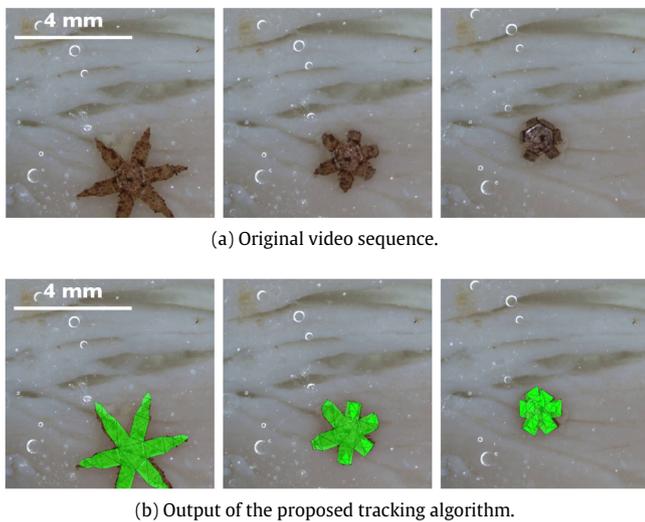
Untethered miniaturized agents offer the possibility to replace tethered medical devices such as endoscopes and probes with smaller, maneuverable, and reconfigurable structures. Due to their reduced size, magnetically-steered miniaturized agents have the potential to perform high-resolution *in-situ* and *in-vivo* manipulations in previously inaccessible body sites.

Many traditional miniaturized agents however have no shape-changing capabilities, which results in limited dexterity that restricts the complexity of achievable tasks [1–3]. In fact, besides the reduced size, miniaturized agents should have the capability to perform at least simple manipulation tasks in cluttered, confined and constrained environments. Miniaturized agents with grasping capabilities provide significant advantages in performing

complex tasks like precise micro-assembly, minimally invasive surgery, cell manipulation, and lab-on-a-chip applications [4–6]. They can be fabricated in sizes that are compatible with the major vascular conduits, and can be used to safely grasp, manipulate and transport biological material [7,8]. Self-folding miniaturized grippers can be composed of materials (e.g., hydrogels), which allow them to mimic soft biological systems as a consequence of changes in materials properties (e.g., swelling) without any complex feedback sensors and tethers. This feature offers the ability for smart behaviors such as autonomous responses in specific environments. For example, self-folding miniaturized grippers are able to swell and shrink significantly in response to a variety of stimuli like temperature, light, or chemical reactions [9–15]. Self-folding miniaturized agents can be used to access hard-to-reach areas within the human's body (e.g., gastrointestinal, urinogenital or cardiovascular conduits) and perform less invasive diagnostic, therapeutic and surgical interventions. In order to precisely move and control such agents, it is of utmost importance to properly detect and track their pose.

\* Corresponding author.

E-mail addresses: [s.scheggi@utwente.nl](mailto:s.scheggi@utwente.nl) (S. Scheggi), [s.misra@utwente.nl](mailto:s.misra@utwente.nl) (S. Misra).



**Fig. 1.** (a) A miniaturized hydrogel gripper is folding in order to grasp biological material (namely yolk from an egg) on porcine tissue (cropped regions from the original  $1024 \times 1024$  images, best viewed in color). (b) The configuration of the gripper (green) is estimated by the proposed tracking algorithm and superimposed on the original video sequence.

The majority of prior research estimated the 2D pixel position of the miniaturized agents. Image processing techniques and a standard Kalman filter were used to track the 2D pose of miniaturized hydrogel grippers in [16,17]. A particle filter was used to track the 2D positions of self-propelled micro-tubes in [18]. Random Sample Consensus (RANSAC) was used in combination with data filtering and optimal convergence process in order to detect and track micro-grippers and micro-objects in micro-assembly tasks [19]. Finally, Bollavaram et al. presented an accurate and robust method to detect the 2D positions and orientations of micro-scale objects in low contrast bright field microscopy [20]. The main reason for estimating the 2D pixel positions of miniaturized agents is related to significant technical difficulties in tracking the agents in 3D. Prior research on 3D tracking used multi-cameras [21,22] and depth-from-focus techniques to estimate the 3D position of the miniaturized agents [23,24]. Recently, a template-based hybrid visual tracking algorithm was presented to estimate the 3D posture of micro-objects in a scanning electron microscope [25]. The algorithm used luminance information to estimate the object displacement on the  $x$ - $y$  plane and its orientation around the  $z$ -axis. Defocus information was used to estimate the object depth. Digital holography was used to track the 3D position of a magnetic micro-robot (Neodymium–Iron–Boron (NdFeB) cylindrical magnet) [26]. Most of the aforementioned studies focused mainly on estimating the position of the agents. However, for miniaturized grippers it would be useful to estimate also the configuration of the tips of their fingers. This would allow a user to determine if the object is firmly grasped and possibly manipulate individual parts of the gripper [27]. Moreover, the knowledge of the overall configuration of the miniaturized grippers would allow a user to utilize grasp planners in order to better perform a grasping task [28].

In this paper, we propose a GPU-accelerated model-based approach to track online the full pose and articulations of submillimeter grippers (Fig. 1). We formulate the tracking problem as an optimization routine that minimizes the discrepancy between the hypothesized 3D gripper model instances and its actual visual observations. Observations come from images captured from a sensor (microscope, ultrasound (US) probe, etc.). The optimization is performed with a variant of the Particle Swarm Optimization

(PSO) algorithm, which exploits the structure and the stimuli responsive mechanisms of the submillimeter gripper. PSO has been successfully used to solve nonlinear, non-differentiable, multimodal optimization problems. Different from other population-based search approaches like genetic algorithms, PSO can compute a solution in a more computationally efficient way [29]. Several tracking problems used PSO [30–33]. However, the possibility to use it in order to track submillimeter agents has not yet been demonstrated. Tamadazte et al. used the *virtual visual servoing* approach to estimate the 3D pose of no shape-changing micro-objects from a single view of the scene [34]. Different from [34], the proposed approach is capable of estimating online the 3D pose and articulations of shape-changing agents, as they change their shape. Experimental tests on challenging real data are used to validate the efficacy of our method. We evaluate the proposed tracker using thermally responsive miniaturized grippers composed of both soft (hydrogel) and hard (metallic) materials, highlighting applicability for agents across material classes.

To the best of our knowledge, there is no work which tries to estimate online the overall configuration of miniaturized grippers, i.e., 3D pose and tips' configuration. Recently, we developed an offline model-based tracker for soft grippers using RGB camera images [35]. Compared to [35], the significant contributions of this study are: (1) An improved and more extended theory; (2) a new implementation on a GPU API which allows the tracker to run online ( $\sim 27$  Hz vs.  $\sim 0.3$  Hz without GPU); (3) a quantitative experimental validation performed on synthetic data; (4) an extended experimental validation performed on real video sequences acquired from both a camera and a US probe, including challenging dynamic backgrounds; (5) an extended experimental validation performed using different types of miniaturized grippers.

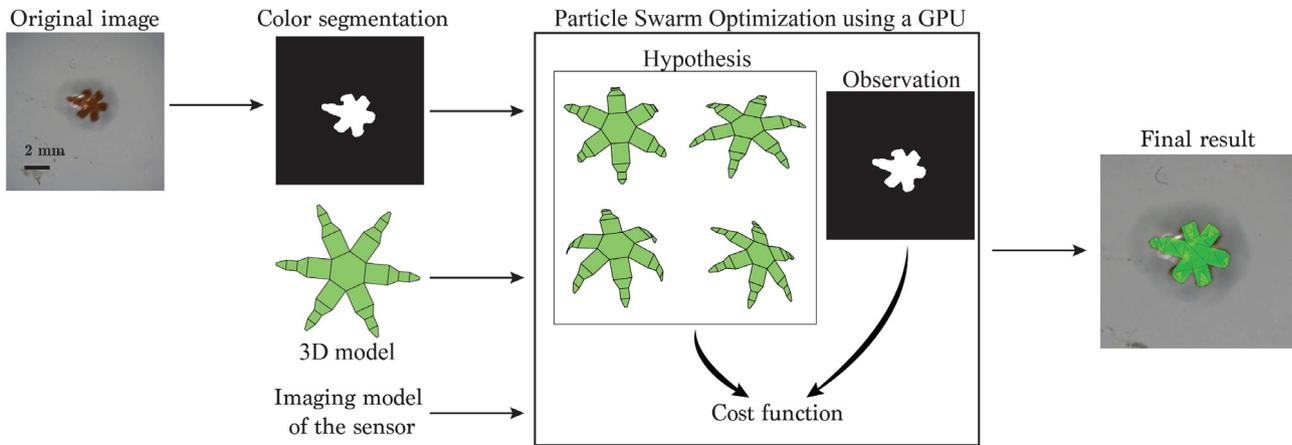
The rest of the paper is organized as follows. Section 2 describes the main steps of the model-based tracker. Section 3 introduces the Particle Swarm Optimization algorithm and shows how we tailored it in order to properly track the miniaturized agents. Section 4 gives the reader some insights on the implementation of the proposed tracker using the OpenGL Shading Language. In Section 5 the proposed tracker is evaluated on synthetic data and challenging real video sequences acquired from both a camera and an US probe. Finally, in Section 6, we summarize the main contributions of the paper, and we discuss possible avenues for future research.

## 2. Model-based tracking

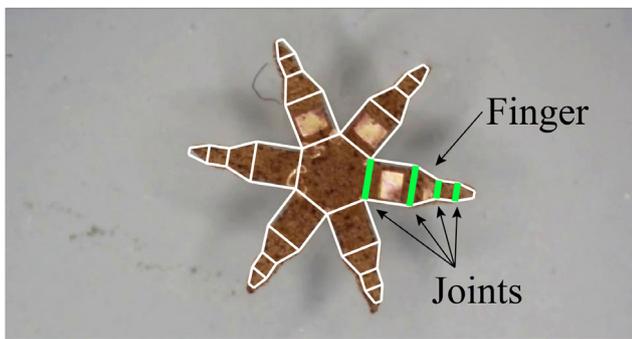
In this section, we introduce the proposed model-based tracker for miniaturized grippers (Fig. 2). In the initial phase, the algorithm detects the agent from color images. The output of this phase is a binary pixel mask where the pixel representing the gripper are set to one. Then, such mask is fed to a PSO algorithm along with a 3D model of the gripper. The algorithm uses the 3D model to generate several configurations of the miniaturized agent, and compares them with the input mask using a cost function. The configuration which minimizes the discrepancy with the input data is the solution to the proposed tracking problem. In what follows, we describe the steps used in the initial detection phase as well as the cost function that the PSO algorithm uses.

### 2.1. Color segmentation

Color segmentation is performed in order to isolate the miniaturized agent. We assume that the input to the proposed method is a color image. In the proposed work, we detect the miniaturized gripper by selecting a particular color space and employing bounds on the coordinates of the selected space. We choose these bounds empirically, i.e., by examining the distribution of



**Fig. 2.** Illustration of the proposed model-based tracker. From left to right, an input image (e.g., taken from a microscope in Red–Green–Blue (RGB) colorspace) is processed by the proposed algorithm. Color segmentation is performed in order to detect the miniaturized gripper. The 3D model of the gripper and the segmented region of the image are used to recover the gripper configuration using the particle swarm optimization algorithm. The hypothetical poses of the gripper are generated by means of computer graphic rendering on GPU. In order to speed up the process, several hypothetical configurations of the gripper are generated and evaluated at once. The hypothesis which minimizes the discrepancy with the observation is the solution to the tracking problem.



**Fig. 3.** The 3D model (white) of a miniaturized hydrogel gripper is composed of a set of 46 appropriately assembled triangles. The model consists of six *fingers*. Each *finger* is modeled as a planar kinematic chain made by 4 joints (one-dimensional hinges) and 4 links.

colors in a preselected set of images. Morphological filtering and morphological opening are used in order to properly segment the image. The resulting image represents the observation  $\mathcal{O}$  related to our problem. The observation  $\mathcal{O}$  is a binary pixel mask where the pixel representing the gripper are set to one. Further details about the color segmentation are reported in Section 5, where the proposed technique is evaluated on video sequences acquired by a camera and a US probe. In order to speed up the segmentation procedure and reduce erroneous estimations, temporal continuity is exploited to track the grippers in a sequence of frames. At each frame, the segmentation procedure uses the estimated position of the tracked gripper from the previous frame (cf. Section 3).

## 2.2. Cost function

The proposed method assumes the knowledge of the 3D model of the agent to be tracked. We empirically generate the 3D model starting from the real measurements of the miniaturized gripper in order to better approximate its shape. We represent the gripper's pose using 3 parameters for its position, 3 parameters for its orientation and 1 parameter for each *finger*. For example, the adopted 3D model for miniaturized six-fingers hydrogel grippers is represented using a vector of 12 parameters (Fig. 3). For the orientation, we consider the Euler representation. In the proposed work, we assume that the joints which belong to the same *finger*

bend in the same way, i.e., they have the same value. This is a reasonable assumption since its joints are usually made by the same swelling material and subject to the same stimulus.

The tracking problem consists of estimating the gripper model parameters that minimize the discrepancy among the hypotheses and the actual observations. For miniaturized six-finger hydrogel grippers, the model parameters consist in the 12 parameters that represent its pose. An appropriate objective function is thus formulated and a variant of PSO is employed to search for the optimal configuration. The result of this optimization process is the output of the method for the given frame. The observation model  $\mathcal{O}$  that feeds the rest of the process consists of the 2D pixels of the segmented color image obtained in Section 2.1. Having a parametric 3D model of a miniaturized gripper, the goal is to estimate the model parameters that are most compatible to the visual observation  $\mathcal{O}$ . Let  $\mathcal{H}$  be a pose hypothesis of the gripper, and let  $\mathcal{C}$  be the mathematical model of the sensor. The miniaturized agent  $\mathcal{P}$ , as imaged by sensor  $\mathcal{C}$ , is obtained by means of computer graphic rendering,  $\mathcal{P} = f(\mathcal{H}, \mathcal{C})$ . For video sequences acquired by a camera,  $\mathcal{C}$  contains the intrinsic parameters of the sensor and  $f(\mathcal{H}, \mathcal{C})$  is the projection of the 3D model of the gripper defined by  $\mathcal{H}$  on the image plane of the camera.

A function  $E(\mathcal{H}, \mathcal{O}, \mathcal{C})$  computes the distance measure between the hypothesis  $\mathcal{H}$  and the observation  $\mathcal{O}$ ,

$$E(\mathcal{H}, \mathcal{O}, \mathcal{C}) = \lambda_D D(\mathcal{O}, \mathcal{H}, \mathcal{C}) + \lambda_K K(\mathcal{H}) + \lambda_S S(\mathcal{H}), \quad (1)$$

where  $\lambda_D, \lambda_K, \lambda_S \in \mathbb{R}^+$  are gain factors and the function  $D(\cdot)$  is defined as [31],

$$D(\mathcal{O}, \mathcal{H}, \mathcal{C}) = \left(1 - \frac{2 \sum(\mathcal{O} \& \mathcal{P})}{\sum(\mathcal{O} \& \mathcal{P}) + \sum(\mathcal{O} | \mathcal{P})}\right),$$

being  $\&$  the logical AND operator, while  $|$  is the logical OR operator.  $D(\cdot)$  measures the discrepancies between the pixels of the hypothetical model and the observation. The sums are computed over the pixels of the observation  $\mathcal{O}$  and of the image  $\mathcal{P}$  generated by the hypothesis  $\mathcal{H}$ .

In (1),  $K(\cdot)$  adds a penalty to kinematically implausible gripper configurations and to possible collisions among the *fingers* of the model. Due to the simplified kinematic structure of the gripper, which considers the fingers as planar robots and assumes that the joints of the same finger have the same value, we estimate possible collisions by using the value of the joints of adjacent *fingers*. This procedure is faster than computing fast triangle–triangle intersections [36–38]. In order to improve the tracking

performance in presence of poor color segmentation and/or occlusions, in (1) we add the term  $S(\cdot)$  which exploits the stimulus responsive mechanisms of the miniaturized grippers. Since the grippers used in this work respond to temperature variations, in  $S(\cdot)$  we use the information of the temperature of the environment to penalize infeasible configurations of the miniaturized gripper. In particular, we consider two reference threshold values (and tolerances) for the joints of the model. One value is used to represent the configuration of the gripper when it is unfolded. The other value is used to represent the configuration of the gripper when it is folded. Moreover, we associate temperature values to such reference configurations. The joints values for the reference configurations depends on the type (material, shape) of the gripper. For miniaturized hydrogel grippers, possible values for the unfolded and folded configuration are about 0 and 60 degrees. Given the information of the temperature of the environment and a gripper pose hypothesis  $\mathcal{H}$ ,  $S(\cdot)$  detects the joint values of  $\mathcal{H}$  which represent infeasible configurations. Then,  $S(\cdot)$  computes the Mean Absolute Error (MAE) of the such values with respect to the reference ones. A similar approach can be used for grippers which respond to light, or chemical reactions.

### 3. Particle Swarm Optimization

Particle Swarm Optimization (PSO) is a population-based, stochastic algorithm that optimizes an objective function through the evolution of the particles of a population [39]. The particles lie in the parameter space of the objective function to be optimized and represent potential solutions to the problem. Each particle searches for a better solution through the interaction with other particles. PSO allows to effectively handle nonlinear optimization problems and it has proven to be implemented with ease and scalability in order to achieve appropriate results in a faster and cheaper way.

In PSO, the particles evolve in runs, which are called generations, according to a policy which emulates social interaction. Let  $N \in \mathbb{N}^+$  be the dimension of the population and let  $G \in \mathbb{N}^+$  be the number of generations, every particle  $i \in \{1, \dots, N\}$  at each generation  $k \in \{1, \dots, G\}$  stores its current configuration in a vector  $\mathbf{x}_i^k$  and its current velocity in a vector  $\mathbf{v}_i^k$ . Vector  $\hat{\mathbf{x}}_i$  stores the configuration at which each particle achieved, up to the current generation  $k$ , the best value of the objective function. The swarm as a whole stores in vector  $\hat{\mathbf{g}}$  the best configuration encountered across all the particles of the swarm.  $\hat{\mathbf{g}}$  is broadcasted to the entire swarm such that every particle is aware of the global optimum. The basic velocity and configuration update rules are,

$$\mathbf{v}_i^{k+1} = w \mathbf{v}_i^k + c_1 r_1 (\hat{\mathbf{x}}_i - \mathbf{x}_i^k) + c_2 r_2 (\hat{\mathbf{g}} - \mathbf{x}_i^k), \quad (2)$$

$$\mathbf{x}_i^{k+1} = \mathbf{x}_i^k + \mathbf{v}_i^{k+1}, \quad (3)$$

where  $w$  is the inertia weight,  $c_1, c_2 \in \mathbb{R}^+$  are coefficients, and  $r_1, r_2$  are random samples in a uniform distribution in the range  $[0, \dots, 1]$ . The particles are usually initialized at random configurations and zero velocities. The dimensions of the multidimensional parameter space are bounded in some range. If during the configuration update ((2) and (3)) the particle moves to a point outside such range, we force that particle to move to the closest point inside the bounds. In this specific work, PSO was used to optimize the objective function reported in (1). In order to find the optimum, PSO operates in a 12-dimensional space. For every single frame, we run the PSO algorithm in order to find the configuration of the miniaturized gripper.

PSO, like other population-based optimization algorithms and evolutionary algorithms, relies on the initial population consisting of potential solutions. In this paper, we initialize the swarm by

**Algorithm 1** Particle Swarm Optimization: for each generation  $k \in \{1, \dots, G\}$ , the configuration  $\mathbf{x}_i^k$  of each particle  $i \in \{1, \dots, N\}$  is evaluated using the cost function reported in (1)  $\mathbf{x}_i^k$  represents the gripper pose hypothesis  $\mathcal{H}$ . If  $\mathbf{x}_i^k$  is better than the particle's best configuration  $\hat{\mathbf{x}}_i$ , then  $\mathbf{x}_i^k$  becomes the new particle's best configuration. The same evaluation is performed for the global best configuration  $\hat{\mathbf{g}}$ . Finally, the configuration  $\mathbf{x}_i^{k+1}$  and velocity  $\mathbf{v}_i^{k+1}$  of each particle are updated accordingly. When we initialize the particles, we exploit temporal continuity among frames, i.e., the solution obtained at frame  $f_t$  is used to initialize the population for the optimization problem at frame  $f_{t+1}$ . Finally, temperature information is used to perform a biased initialization of the joints values.

---

```

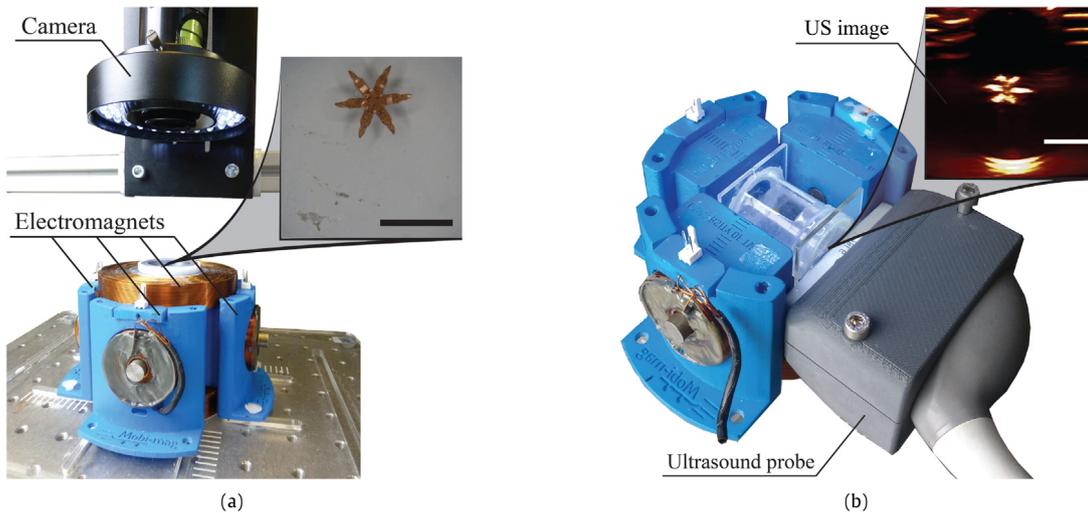
1: initialize_particles()
2:
3: for k ← 1, G do
4:   for i ← 1, N do
5:     if mod(k, k_r) == 0 then
6:       partial_randomization( $\mathbf{x}_i^k$ )
7:     end if
8:     if  $E(\mathbf{x}_i^k, \mathcal{O}, C) < E(\hat{\mathbf{x}}_i, \mathcal{O}, C)$  then
9:        $\hat{\mathbf{x}}_i \leftarrow \mathbf{x}_i^k$ 
10:    end if
11:    if  $E(\hat{\mathbf{x}}_i, \mathcal{O}, C) < E(\hat{\mathbf{g}}, \mathcal{O}, C)$  then
12:       $\hat{\mathbf{g}} \leftarrow \hat{\mathbf{x}}_i$ 
13:    end if
14:  end for
15:
16:  for i ← 1, N do
17:     $c_{1,i} \leftarrow \frac{C}{(N-1)^2} (\text{rank}_i - N)^2$ 
18:     $c_{2,i} \leftarrow C - c_{1,i}$ 
19:     $\mathbf{v}_i^{k+1} \leftarrow w \mathbf{v}_i^k + c_{1,i} r_1 (\hat{\mathbf{x}}_i - \mathbf{x}_i^k) + c_{2,i} r_2 (\hat{\mathbf{g}} - \mathbf{x}_i^k)$ 
20:     $\mathbf{x}_i^{k+1} \leftarrow \mathbf{x}_i^k + \mathbf{v}_i^{k+1}$ 
21:  end for
22: end for

```

---

exploit temporal continuity among frames. The solution obtained at frame  $f_t$  is used to generate the population for the optimization problem at frame  $f_{t+1}$ . In particular, among the  $N$  particles of the swarm, one particle is initialized with same parameters which define the best particle at the previous time frame. The remaining particles are generated by perturbing the previous solution with uniform pseudo-random numbers within a given range. For what concern the initialization of the joints values, we performed a biased initialization by using the temperature information and a simplified version of the hysteresis information presented in [16]. For example, if the temperature is increasing and it is in the range where the gripper will be more likely start folding, we initialize 75% of the swarm by increasing the joints values of a random number within a given range. The opposite is applied when the temperature decreases and it is in the range where the gripper will be more likely to unfold. In the remaining cases, 50% of the swarm is initialized with increasing joints values while the remaining 50% is initialized with decreasing joints values.

Since PSO may suffer from premature convergence, we used a modified version of this algorithm in order to overcome this problem [40]. We applied a *partial randomization* of the population by randomly perturbing one parameter of a subset of the population. This procedure is applied every  $k_r < G$  generations. Moreover, in order to avoid occasionally inordinate divergence which may prevent the algorithm to reach the best solution, we used a *coefficient adaptation technique*. The extension modifies the coefficients  $c_1, c_2$  depending on how far is the current particle from



**Fig. 4.** The microscopic camera datasets and the US datasets are generated using a magnetic system composed of iron-core electromagnets fixed in a frame around a reservoir with the liquids. In order to fold/unfold the miniaturized grippers, a Peltier element is placed below the reservoir and used to regulate the temperature of the water wherein the grippers are floating. The gripper is imaged using: (a) A microscopic camera positioned above the reservoir; and (b) a US probe. Magnetic forces are exerted on the gripper to control its pose. The scale bar is 4 mm. Please refer to the accompanying video that shows the results of the tracking experiments.

the global best solution  $\hat{\mathbf{g}}$ ,

$$c_{1,i} = \frac{C}{(N-1)^2} (\text{rank}_i - N)^2, \quad c_{2,i} = C - c_{1,i},$$

being  $C \in \mathbb{R}^+$  a constant, while  $\text{rank}_i$  is the rank of the  $i$ th particle's configuration. The adaptation technique allows particles in the lower rank to move toward the global optimum, while particles in a higher rank are close to the *best* solution and search around it. Algorithm 1 describes the PSO algorithm.

The estimated state  $\hat{\mathbf{g}}$  of the gripper at frame  $f_t$  is used by the color segmentation procedure (cf. Section 2.1) to determine the visual observation  $\mathcal{O}$  in the next video frame  $f_{t+1}$ . Among the detected pixels, the algorithm keeps the ones for which the probability of belonging to the gripper  $p(\mathcal{O}_{t+1} | \mathcal{P}_t)$  is above a certain value.  $\mathcal{P}$  is the generation of the miniaturized gripper  $\mathcal{P} = f(\hat{\mathbf{g}}, C)$ , being  $C$  the mathematical model of the sensor (cf. Section 2.2).

#### 4. GPU implementation

The most demanding part of the algorithm is the generation of the hypothesis  $\mathcal{H}$ , and consequently the computation of the cost function  $E(\cdot)$  (1). A possible way to speed up this process is to use a Graphics Processing Unit (GPU). GPU-accelerated computing consists in using a GPU together with a CPU to accelerate compute-intensive portions of the application.

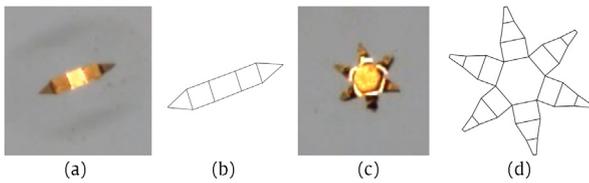
As reported in [31,41], we can take advantage of the GPU architecture by generating and processing multiple hypotheses simultaneously. In particular, four hypotheses are simultaneously generated using the OpenGL Shading Language. Different from [31,41], the hypotheses are generated with a reduced size with respect to the original image (usually their size is one fourth of the original dimension). A geometry shader is used to assemble the 3D model of the gripper from the values of its pose and joints. Such values represent the configuration of a particle  $i$  at a generation  $k$  and are stored in the vector  $\mathbf{x}_i^k$ , as described in Section 3. We generate the hypotheses using multiple Framebuffer Objects (FBO). Then, we use multiple Pixel Buffer Objects (PBO) to transfer the rendered hypotheses from an FBO to the CPU in an asynchronous way. The advantage of a PBO is that it allows to perform fast pixel data transfer to and from a graphics card through Direct Memory Access (DMA) without involving CPU cycles. Moreover PBO can perform asynchronous DMA transfer. The CPU is not involved in

transferring the pixel data to/from a PBO, which is managed by the GPU. Therefore, in order to properly take advantage of PBO, the program should actually have some computations to perform while it is waiting for the transfer to complete. In our case, in order to speed up the tracking process, we interleave the data transfer with: (1) PSO computations; (2) the evaluation of the cost function  $E(\cdot)$  on the CPU; (3) the generation of other hypotheses on the GPU.

The algorithm can be described as follows: During the initialization phase, the 3D model of the agent (number of *fingers*, number of links, length of the links, etc.) is transferred to the GPU and cached. Custom shaders (vertex, geometry and pixel) are designed to properly assemble the 3D model and generate  $\mathcal{P}$  based on the imaging model of the sensor (cf. Section 2.2). During the tracking process, at each generation of the PSO, particles are evaluated in groups of four. Each particle represents a gripper pose hypothesis  $\mathcal{H}_i$ ,  $i \in \{1, \dots, N\}$ . The four hypothesis are contemporary transferred to the GPU, and the related 3D models are properly assembled and rendered on a particular FBO. When the next four hypothesis are transferred to the GPU, the related 3D models are rendered on a different FBO, while the previous FBO is transferred to the CPU using a PBO. Since the pixel data transfer is a time consuming operation which does not involve CPU cycles, we interleave the pixel data transfer with CPU computations, among which the evaluation of  $K(\cdot)$  and  $S(\cdot)$  for the actual group of particles, and  $D(\cdot)$  for the previous particles (cf. Section 2.2). The use of the GPU allows the proposed algorithm to run online ( $\sim 27$  Hz). Without GPU, the algorithm runs at an average frame rate of  $\sim 0.3$  Hz [35].

#### 5. Experimental results

In this section, representative results of the proposed tracker are provided. We first evaluate the proposed technique on synthetic data with ground truth information. This is a common approach in the relevant literature because ground truth data for real-world image sequences are hard to obtain. Then, we test and evaluate the tracking method in a series of image sequences acquired by a microscopic camera, demonstrating challenging tracking scenarios (Fig. 4(a)). Finally, the proposed model-based tracker is evaluated in video sequences acquired by a US probe 18L6HD (Siemens ACUSON S2000, Siemens Healthcare, Mountain View, USA) (Fig. 4(b)). All the video sequences contain frames in which the miniaturized gripper is self-occluded. Also, the US sequences



**Fig. 5.** Two-finger (a) and six-finger (c) metallic grippers. We model each finger of the two-finger gripper as a planar kinematic chain made by 2 joints (one-dimensional hinges) and 2 links (b). For the six-finger gripper, each finger is represented as a planar kinematic chain made by 3 joints and 3 links (d).

present significantly more challenging dynamic backgrounds. The video sequences are processed off-line on an Intel Xeon laptop computer running MS Windows at 3.5 GHz, 16 GBs RAM and an Nvidia Quadro K2200 graphic card.

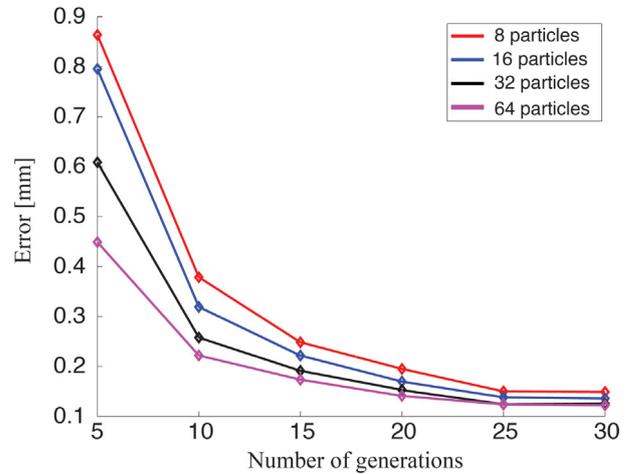
Three different miniaturized grippers are used: (1) Hydrogel grippers having 6 *fingers* and tip-to-tip length of 4 mm [15]; (2) metallic grippers with 6 *fingers* and tip-to-tip distance of 900  $\mu\text{m}$  [13]; (3) metallic grippers with 2 *fingers* and tip-to-tip length of 750  $\mu\text{m}$  [14]. For all the three different types of grippers, we generate 3D models that match the real shape and dimension of the agents (Figs. 3 and 5).

The miniaturized hydrogel grippers are composed of a stiff SU-8 and thermally responsive pNIPAM-AAc segmented bilayer. They open and close reversibly due to a lower critical solution temperature (LCST) phase transition and associated swelling or shrinkage in the pNIPAM-AAc layer in response to temperature changes. The range of temperatures used to fold/unfold such grippers is between 24  $^{\circ}\text{C}$  and 27  $^{\circ}\text{C}$  [16]. The submillimeter metallic grippers are composed of pre-stressed chromium/gold self-actuating hinges and a magnetic nickel layer to enable remote manipulation. The phalanges of the gripper consist of a magnetic layer of nickel (8.5  $\mu\text{m}$ ) sandwiched between two 0.5  $\mu\text{m}$  layers of gold. At room temperature, the trigger polymer of the gripper is stiff enough to prevent the phalanges from folding. As the temperature rises to or above the physiological temperature (37  $^{\circ}\text{C}$ ), the trigger softens and allows the folding of the submillimeter agent [13].

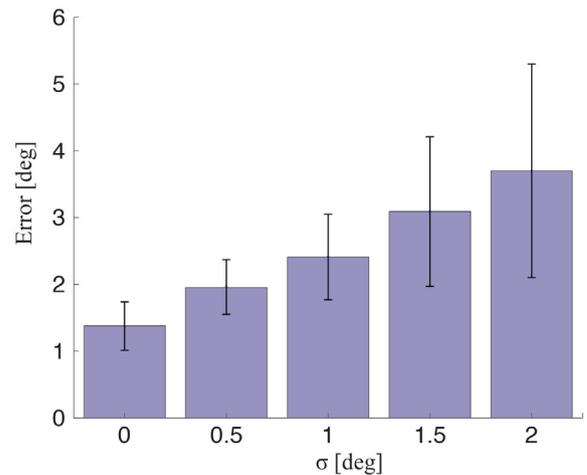
### 5.1. Synthetic dataset

Synthetic data is used for the quantitative evaluation of the proposed approach. Hydrogel grippers are simulated in the proposed evaluation. Several sequences are considered, showing the grippers moving, rotating about all their axis and folding/unfolding. The sequences comprise frames where the agent is self-occluded and/or it moves outside the field of view of the simulated sensor. Computer graphic rendering is used to generate the required observation  $\mathcal{O}$  for each simulated configuration of the gripper.

In the first experiment, we evaluate the tracking error depending on the amount of particles and generations used. In order to assess the tracking accuracy, we compute the tracking error by using the metric defined in [42]. Fig. 6(a) reports the average tracking error computed on 5614 frames. In particular, about 70% of the synthetic data presents occlusions which cover up to 80% of the shape of the miniaturized gripper. From Fig. 6(a), we can observe that employing more than 32 particles and more than 25 generations does not significantly improve the accuracy of the tracker. With such configuration, we obtain a position error of  $25 \pm 7 \mu\text{m}$ , orientation error of  $1.7 \pm 1.3$  degrees, and joints' estimation error of  $1.4 \pm 0.95$  degrees. Occlusions, motions outside the field of view, and big out-of-plane rotations decrease the accuracy of the tracker. This is due to the fact that the gripper is less visible. In particular, we experience a maximum position error of 72.3  $\mu\text{m}$ , maximum orientation error of 14.4, 18.7, and 10.2 degrees about the  $x$ -,  $y$ -,



(a)



(b)

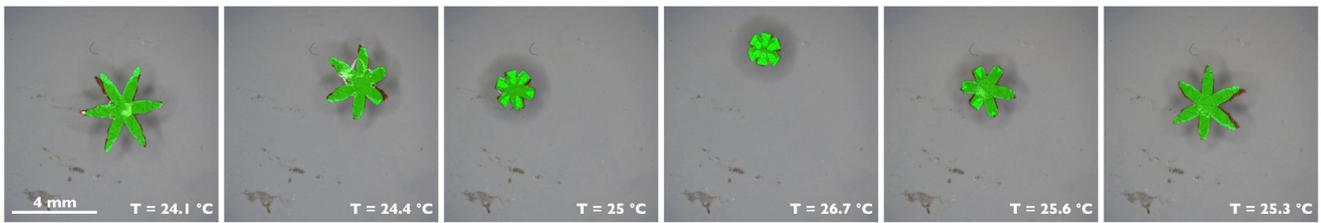
**Fig. 6.** Synthetic evaluation. (a) The proposed tracker is evaluated depending on the amount of particles and generations used. In order to evaluate the tracking accuracy, we measure the distance between the corresponding phalanx endpoints in the ground truth data and in the estimated configuration of the gripper. (b) We report the error of the tracker in estimating the joint values of the gripper. The ground truth joint values are corrupted with Gaussian noise having increasing standard deviation  $\sigma$ .

$z$ -axis, respectively, and maximum joints' estimation error of 12.3 degrees. These results are obtained when the miniaturized gripper moves almost completely outside the field of view of the sensor.

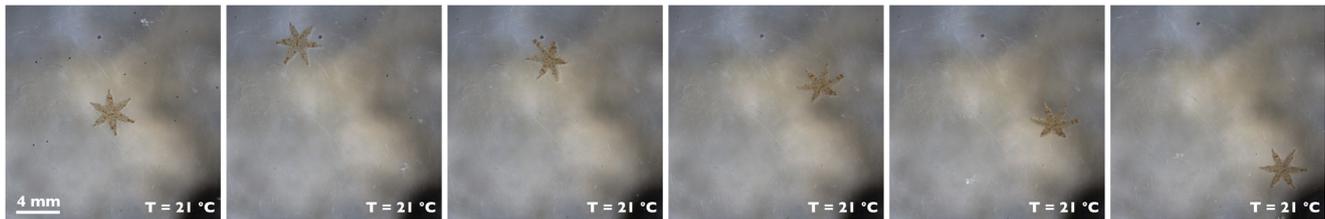
In the remainder of the experiments, we use  $N = 32$  particles,  $G = 25$  generations,  $w = 0.4$ , and  $C = 1$ , according to [43,44]. We performed the partial randomization every  $k_r = 8$  generations. Concerning the cost function,  $\lambda_D = 2$ ,  $\lambda_K = 0.15$ , and  $\lambda_S = 1$ . The tracker runs at an average frame rate of 27 Hz. In the proposed tracker, we model the miniaturized grippers assuming that the joints which belong to the same finger bend in the same way, i.e., they have the same value (cf. Section 2.2). In the second experiment, we evaluate the tracker in situations where the proposed assumption does not hold. In particular, we corrupt the joint values of the ground truth data with additive Gaussian noise, having zero mean and increasing standard deviation  $\sigma \in \{0, 0.5, 1, 1.5, 2\}$  degrees. Fig. 6(a) reports the error of the tracker in estimating the joint values of the gripper when they are corrupted with additive noise.

### 5.2. Microscopic camera dataset

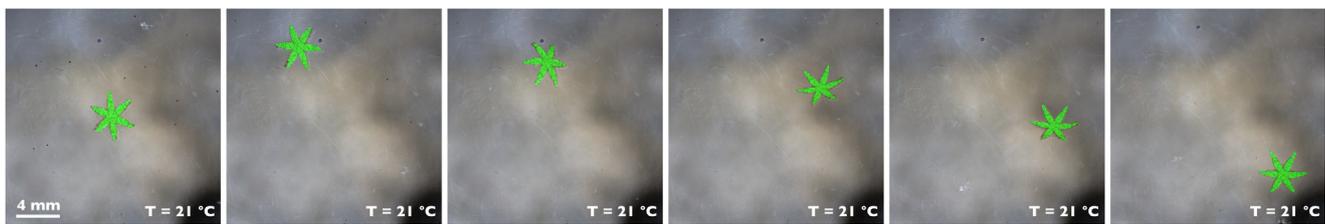
The microscopic camera dataset is composed of seven video sequences. Four of them report the motion of miniaturized hydrogel



**Fig. 7.** Characteristic snapshots from the tracking experiment on the first image sequence (best viewed in color). The miniaturized hydrogel gripper moves in the environment, folds performing a grasping task, and unfolds simulating a release task. For each snapshot, the estimated configuration of the gripper is displayed in green. Temperatures (T) are shown on the bottom-right corner of each snapshot. Please refer to the accompanying video that shows the results of the tracking experiments.



(a) Original video sequence.



(b) Outcome of the proposed tracking algorithm.

**Fig. 8.** Characteristic snapshots from the tracking experiment on the third image sequence (best viewed in color). The sequence shows a miniaturized hydrogel gripper rotating about its axes. In this sequence, the camera is placed further away from the agent. As a consequence, the dimension of the agent in the image plane is smaller and the background is more challenging. For each snapshot, the estimated configuration of the gripper is displayed in green. Temperatures (T) are shown on the bottom-right corner of each snapshot. Please refer to the accompanying video that shows the results of the tracking experiments.

grippers during manipulation and transportation tasks. Two video sequences depict the motion of submillimeter six-finger metallic grippers. The remaining video sequence shows the motion of a submillimeter two-finger metallic gripper; please refer to the accompanying video that shows the results of the tracking experiments.

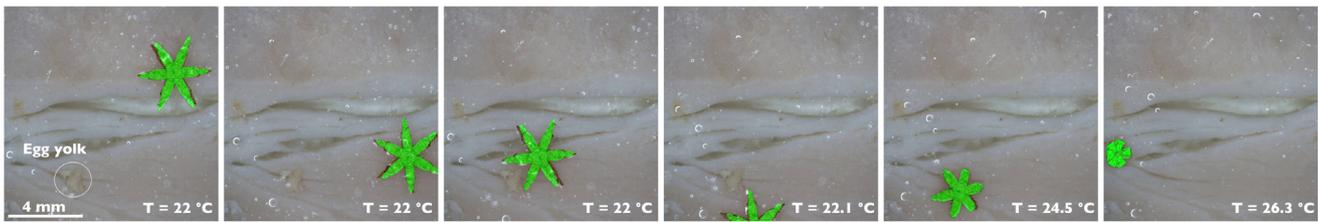
The video sequences are generated by magnetically controlling the agents in a setup composed of a fluid reservoir and six orthogonally oriented electromagnets (Fig. 4(a)) [18]. All of the videos were recorded using a Blackfly 1.4 MP Color GigE PoE (Point Grey Research Inc., Richmond, Canada) camera mounted on a Mitutoyo FS70 microscope unit (Mitutoyo, Kawasaki, Japan) using a Mitutoyo M Plan Apo 2 / 0.055 Objective. The video sequences consist of  $1024 \times 1024$  RGB color images. In order to detect the miniaturized agents, we perform the color segmentation steps described in Section 2.1. In particular, we convert the images to the Hue-Saturation-Value (HSV) color space and we employ bounds on the coordinates of the Saturation and Value spaces.

The first and second videos consist of 4148 and 3552 frames, respectively. The video sequences report the motion of a miniaturized hydrogel gripper in testbed scenarios, where the agent is clearly visible and easily detected against the background. During the experiment, the agent rotates and moves at various directions and speeds. At some point in time, we increase the temperature of the Petri dish using a Peltier element (Fig. 4(a)). The miniaturized gripper starts folding while it moves in the scene. Finally, the experiment concludes with the micro-sized agent that unfolds (as we cool down the temperature) while it continues to move (Fig. 7). The configurations of the gripper estimated by the tracker are consistent throughout the whole sequences.

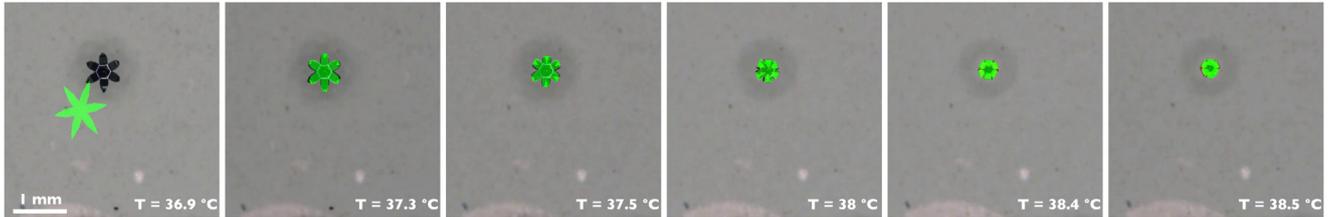
The third video consists of 7960 frames. With respect to the previous sequence, in this video the camera is placed further away from the agent. Thus the dimension of the agent in the image plane is smaller and the background is more challenging (Fig. 8(a)). In order to evaluate the tracker in a more challenging scenario, we process one frame out of every five frames. When the experiment starts, the agent is open and does not move. Then, it starts moving upward. Finally, the gripper starts moving and rotating about the three axis. Also in this case, the tracker is able to correctly estimate the configuration of the agent (Fig. 8(b)).

The fourth video consists of 2558 frames. A miniaturized gripper is used to grasp biological material namely yolk from an egg. We process one frame out of every two frames. When the experiment starts, the agent is stationary and open. Later, the miniaturized gripper starts rotating and moving toward the egg yolk. Then, the agent moves almost completely outside the field of view of the camera, reaches the target and starts to fold (Fig. 9). Throughout the whole sequence, the configuration of the gripper is correctly identified.

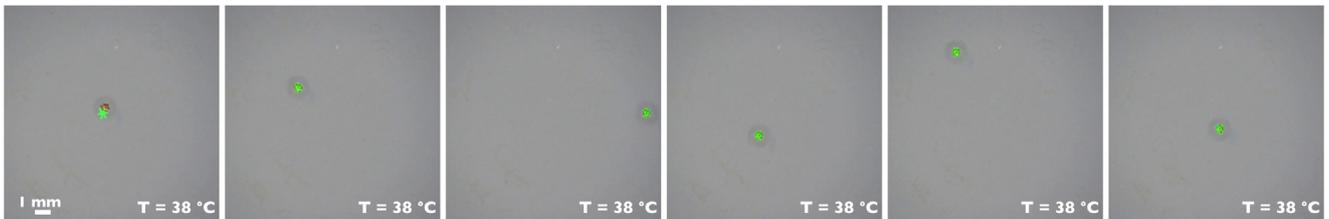
The next two video sequences show a submillimeter six-finger metallic gripper moving, rotating, and folding in order to simulate a grasping task. The video sequences consist of 1936 and 1320 frames, respectively. In the first one, the submillimeter gripper is steady. Then, it starts folding as the temperature is increased using a Peltier element (Fig. 10). In the second video sequence, the gripper is not completely folded. When the experiment starts, the agent starts moving and rotating at high speed in the environment. In this video the camera is placed further away from the agent.



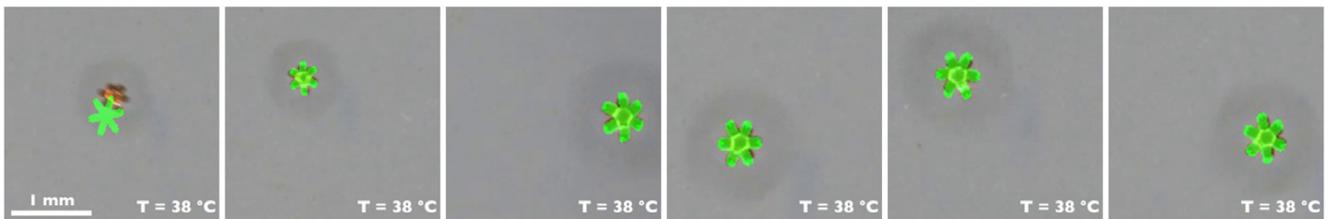
**Fig. 9.** Snapshots of the miniaturized hydrogel gripper moving on porcine muscle tissue and picking egg yolk (best viewed in color). Although the agent moves almost completely outside the field of view of the camera, the proposed algorithm is still able to track it. For each snapshot, the estimated configuration of the hydrogel gripper is displayed in green. Temperatures (T) are shown on the bottom-right corner of each snapshot. *Please refer to the accompanying video that shows the results of the tracking experiments.*



**Fig. 10.** Characteristic snapshots from the tracking experiment on the fifth image sequence (best viewed in color). The sequence shows the submillimeter six-finger metallic gripper folding. From left to right: Although the initial conditions used in the particle swarm optimization do not coincide with the actual configuration of the metallic gripper, the tracker is able to converge to the correct configuration of the gripper. For each snapshot, the estimated configuration of the gripper is displayed in green. Temperatures (T) are shown on the bottom-right corner of each snapshot. *Please refer to the accompanying video that shows the results of the tracking experiments.*

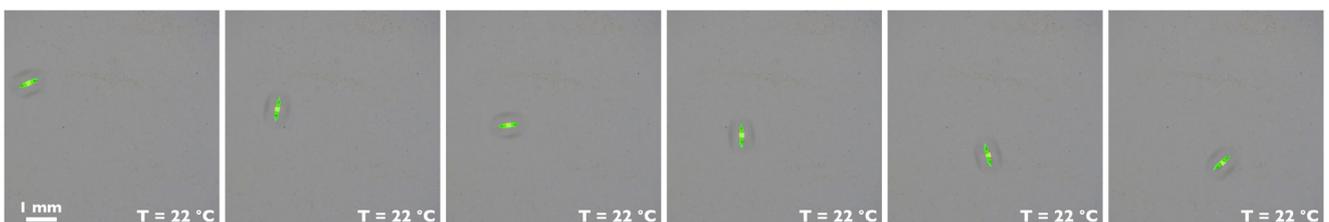


(a) Video sequence processed by the proposed algorithm.

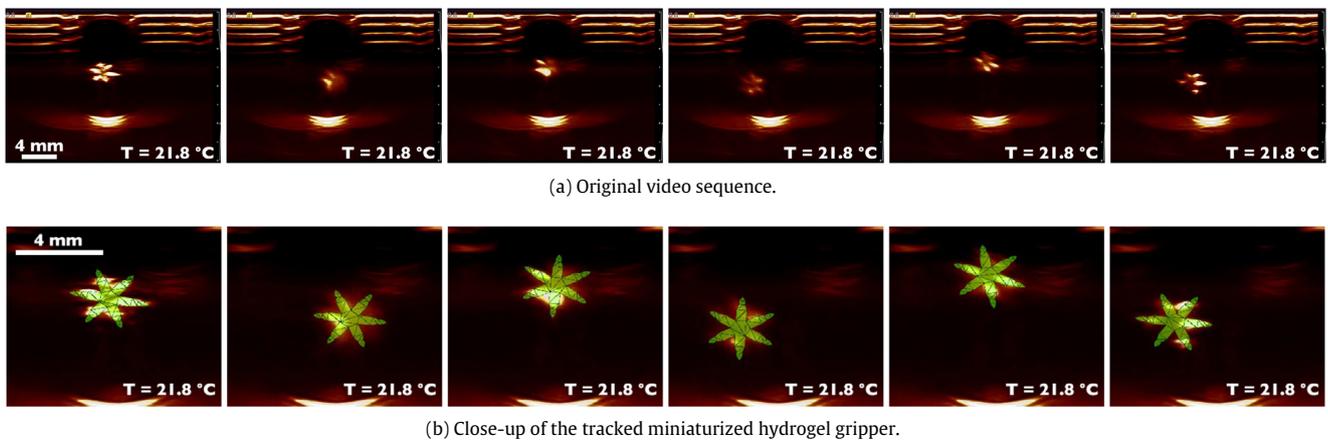


(b) Close-up of the tracked submillimeter metallic gripper.

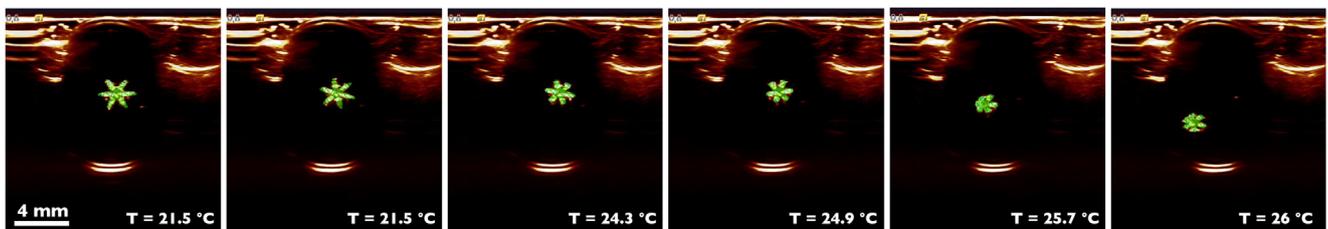
**Fig. 11.** Characteristic snapshots from the tracking experiment on the sixth image sequence (best viewed in color). The sequence shows the submillimeter six-finger metallic gripper moving and rotating in the environment at high speed. The gripper is not completely folded. In this sequence the camera is placed further away from the agent. Due to the small dimension of the agent, we consider this sequence challenging for the tracker. From left to right: Although the initial conditions used in the particle swarm optimization do not coincide with the actual configuration of the miniaturized gripper, the tracker is able to converge to the correct configuration of the gripper. For each snapshot, the estimated configuration of the gripper is displayed in green. Temperatures (T) are shown on the bottom-right corner of each snapshot. *Please refer to the accompanying video that shows the results of the tracking experiments.*



**Fig. 12.** Characteristic snapshots from the tracking experiment on the seventh image sequence (best viewed in color). The sequence shows the submillimeter two-finger metallic gripper moving and rotating in the environment. For each snapshot, the estimated configuration of the gripper is displayed in green. Temperatures (T) are shown on the bottom-right corner of each snapshot. *Please refer to the accompanying video that shows the results of the tracking experiments.*



**Fig. 13.** Characteristic snapshots from the tracking experiment on the first image sequence of the ultrasound dataset (best viewed in color). The agent rotates and moves in the workspace. The proposed algorithm is able to track the miniaturized hydrogel gripper even if the agent is not always completely visible. For each snapshot, the estimated configuration of the gripper is displayed in green. Temperatures (T) are shown on the bottom-right corner of each snapshot. Please refer to the accompanying video that shows the results of the tracking experiments.



**Fig. 14.** Characteristic snapshots from the tracking experiment on the second image sequence of the ultrasound dataset (best viewed in color). The miniaturized hydrogel gripper rotates, moves in the workspace and folds performing a grasping task. For each snapshot, the estimated configuration of the gripper is displayed in green. Temperatures (T) are shown on the bottom-right corner of each snapshot. Please refer to the accompanying video that shows the results of the tracking experiments.

Due to the small dimension of the agent, we consider this sequence challenging for the tracker Fig. 11(a). However, the proposed algorithm is able to correctly track the submillimeter agent Fig. 11(b). The last video sequence consists of 510 frames and reports the motion of a miniaturized two-finger gripper. Also in this case, the proposed tracker correctly identifies the configuration of the agent (Fig. 12).

### 5.3. Ultrasound dataset

Four video sequences have been considered in order to validate the proposed algorithm in more clinically relevant scenarios; please refer to the accompanying video that shows the results of the tracking experiments. In the video sequences, the miniaturized hydrogel gripper is magnetically controlled using the magnetic setup depicted in Fig. 4(b). The images are acquired using an ultrasound probe 18L6HD (Siemens ACUSON S2000, Siemens Healthcare, Mountain View, USA) operating with a frequency of 16 MHz with an in-plane resolution of approximately 0.09 mm per pixel. The video sequences consist of  $1024 \times 662$  RGB color images (Fig. 2). In order to detect the miniaturized agent, we apply a grayscale conversion to the images using the luminance information. Then, we employ bounds on the selected color space. Finally, we apply the steps described in Section 2.1.

The first video consists of 544 frames. When the experiment starts, the agent is stationary and open. Later, the miniaturized gripper starts rotating and moving at various directions and speeds. Although the agent is partially visible during its motion (Fig. 13(a)), the tracker can still track its pose using the temperature information (Fig. 13(b)).

The second video consists of 2033 frames. With respect to the previous sequence, in this sequence the miniaturized gripper folds

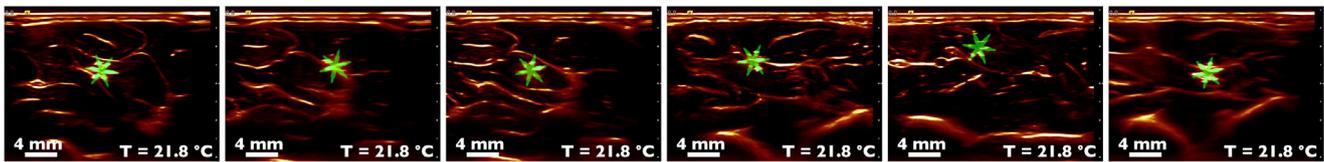
and moves simulating a transportation task. When the experiment starts, the agent is open and does not move. Then, it starts folding. Finally, the miniaturized gripper starts moving downward (Fig. 14). Also in this case, the tracker is able to correctly estimate the configuration of the agent.

The third and fourth video sequences are generated by combining the first two video sequences with two dynamic backgrounds (See Figs. 15 and 16). The dynamic backgrounds are obtained by scanning the abdomen of a human volunteer. With respect to the first and second video sequences, the tracker is less accurate. This is due to the presence of the dynamic background and occlusions. However, the tracker is able to successfully detect the agent through the whole sequences.

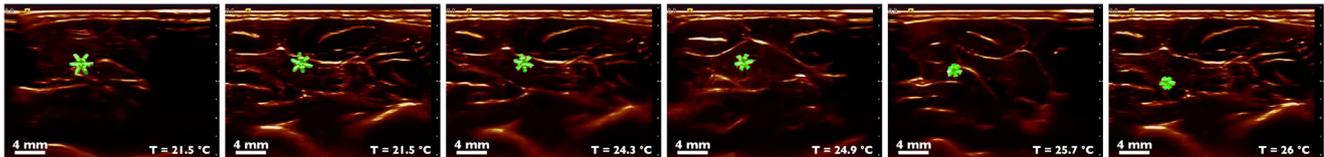
The proposed method mainly relies on the observation  $\mathcal{O}$  and on the term  $S(\cdot)$  to properly estimate the configuration of the agent. Thus, possible limitations might arise when the gripper is severely occluded or erroneously detected in several frames. In crowded manipulation workspaces where multiple agents and object can be tracked, these limitations can be leveraged by using a tracker which takes into account all agents simultaneously, instead of considering each agent independently [45].

## 6. Conclusions and future work

In this paper, we present a novel GPU-accelerated model-based method for efficient tracking of submillimeter grippers. The tracking problem consists of estimating the gripper model parameters that minimize the discrepancy between the gripper hypotheses and the actual observations. We use an appropriate objective function and we employ a variant of Particle Swarm Optimization (PSO) which exploits the stimuli responsive mechanism of the miniaturized grippers. GPU-accelerated computing is used to properly



**Fig. 15.** Characteristic snapshots from the tracking experiment on the third image sequence of the ultrasound dataset (best viewed in color). The sequence is obtained by combining the first sequence of the ultrasound dataset with a dynamic background. The sequence shows a submillimeter hydrogel gripper moving and rotating in the environment. The gripper is unfolded. For each snapshot, the estimated configuration of the gripper is displayed in green. Temperatures (T) are shown on the bottom-right corner of each snapshot. Please refer to the accompanying video that shows the results of the tracking experiments.



**Fig. 16.** Characteristic snapshots from the tracking experiment on the fourth image sequence of the ultrasound dataset (best viewed in color). The sequence is obtained by combining the second sequence of the ultrasound dataset with a dynamic background. A submillimeter hydrogel gripper rotates, moves in the environment and folds performing a grasping task. For each snapshot, the estimated configuration of the gripper is displayed in green. Temperatures (T) are shown on the bottom-right corner of each snapshot. Please refer to the accompanying video that shows the results of the tracking experiments.

track and detect the configuration of the micro-sized agent online ( $\sim 27$  Hz).

Extensive experimental results demonstrate the validity of the proposed method. The method is applicable to both hydrogel (soft) and metallic (hard) grippers with different shapes and tip-to-tip sizes ranging from 4 mm to 750 microns. A quantitative evaluation of the proposed tracker using synthetic data reveals a position error of  $25 \pm 7 \mu\text{m}$ , orientation error of  $1.7 \pm 1.3$  degrees, and estimation error of the joints' values of  $1.4 \pm 0.95$  degrees. Several datasets, acquired using a microscopic camera and an ultrasound probe, demonstrate the capability of the proposed method to track submillimeter grippers in challenging scenarios, including dynamic backgrounds. More than 20000 frames are used to validate the proposed method. With respect to existing methods, the proposed algorithm is the first one which tries to estimate the full 3D pose and the articulations of untethered submillimeter grippers from marker-less visual information. This will allow a user to fully control the motion of the agent by exploiting the information of its 3D pose, i.e., more tasks can be achieved. Moreover, manipulation tasks can be performed in a more accurate way by using the knowledge of the pose and articulations of the gripper. We believe that the proposed technique can be applied to different stimuli responsive miniaturized agents, allowing the user to estimate the full configuration of complex agents from visual marker-less observations.

In future work, we will consider the possibility to extend the proposed approach to multiple gripping agents. Moreover, we will evaluate the proposed tracker in more clinically relevant scenarios, e.g., in fluidic micro-channels with time-varying flow rates to enable eventual applicability *in-vivo*.

### Acknowledgments

This project (ROBOTAR) has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation programme (Grant Agreement #638428). The research was also supported by the National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health (NIH) under Award Number R01EB017742. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

### Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.robot.2017.11.003>.

### References

- [1] B. Jurado-Sánchez, S. Sattayasamitsathit, W. Gao, L. Santos, Y. Fedorak, V.V. Singh, J. Orozco, M. Galarnyk, J. Wang, Self-propelled activated carbon janus micromotors for efficient water purification, *Small* 11 (4) (2015) 499–506.
- [2] S.K. Srivastava, M. Medina-Sánchez, B. Koch, O.G. Schmidt, Medibots: Dual-action biogenic microdaggers for single-cell surgery and drug release, *Adv. Mater.* 28 (5) (2016) 832–837.
- [3] C. Pacchierotti, S. Scheggi, D. Prattichizzo, S. Misra, Haptic feedback for micro-robotics applications: A review, *Frontiers in Robotics and AI* 3 (2016) 53.
- [4] T. Yoneyama, T. Watanabe, H. Kagawa, J. Hamada, Y. Hayashi, M. Nakada, Force detecting gripper and flexible micro manipulator for neurosurgery, in: *Proc. IEEE Int. Conf. on Engineering in Medicine and Biology Society*, 2011, pp. 6695–6699.
- [5] A. Ichikawa, S. Sakuma, F. Arai, S. Akagi, Untethered micro-robot with gripping mechanism for on-chip cell surgery utilizing outer magnetic force, in: *Proc. IEEE Int. Conf. on Robotics and Automation*, 2014, pp. 3795–3800.
- [6] M.M.S. Mousavi, G. De Pasquale, A. Somà, E. Brusa, A novel su-8 microgripper with external actuator for biological cells manipulation, in: *Proc. Symposium on Design, Test, Integration & Packaging of MEMS and MOEMS*, 2011, pp. 356–361.
- [7] F. Hansen, P. Mangell, B. Sonesson, T. Lanne, Diameter and compliance in the human common carotid artery variations with age and sex, *Ultrasound in medicine & biology* 21 (1) (1995) 1–9.
- [8] F. Beyeler, A. Neild, S. Oberti, D.J. Bell, Y. Sun, J. Dual, B.J. Nelson, Monolithically fabricated microgripper with integrated force sensor for manipulating microobjects and biological cells aligned in an ultrasonic field, *J. of Microelectromechanical Systems* 16 (1) (2007) 7–15.
- [9] David H. Gracias, Stimuli responsive self-folding using thin polymer films, *Current Opinion in Chemical Engineering* 2 (1) (2013) 112–119.
- [10] S. Fusco, M.S. Sakar, S. Kennedy, C. Peters, R. Bottani, F. Starsich, A. Mao, G.A. Sotiriou, S. Pané, S.E. Pratsinis, D. Mooney, B.J. Nelson, An integrated microbotic platform for on-demand, targeted therapeutic interventions, *Adv. Mater.* 26 (6) (2014) 952–957.
- [11] H. Li, G. Go, S.Y. Ko, J.-O. Park, S. Park, Magnetic actuated ph-responsive hydrogel-based soft micro-robot for targeted drug delivery, *Smart Mater. Struct.* 25 (2) (2016).
- [12] N. Bassik, B.T. Abebe, K.E. Laffin, D.H. Gracias, Photolithographically patterned smart hydrogel based bilayer actuators, *Polymer* 51 (26) (2010) 6093–6098.
- [13] E. Gultepe, J.S. Randhawa, S. Kadam, S. Yamanakaand, F.M. Selaru, E.J. Shin, A.N. Kallou, D.H. Gracias, Biopsy with thermally-responsive untethered microtools, *Adv. Mater.* 25 (4) (2013) 514–519.
- [14] S. Pandey, E. Gultepe, D.H. Gracias, Origami inspired self-assembly of patterned and reconfigurable particles, *Journal of Visualized Experiments* (72) (2013).
- [15] J.C. Breger, C. Yoon, R. Xiao, H.R. Kwag, M.O. Wang, J.P. Fisher, T.D. Nguyen, D.H. Gracias, Self-Folding thermo-magnetically responsive soft microgrippers, *ACS Applied Materials & Interfaces* 7 (5) (2015) 3398–3405.
- [16] F. Ongaro, C. Yoon, F. van den Brink, M. Abayazid, S.H. Oh, D.H. Gracias, S. Misra, Control of soft untethered grippers for pick-and-place tasks, in: *Proc. IEEE Int. Conf. on Biomedical Robotics and Biomechatronics*, 2016, pp. 299–304.
- [17] F. Ongaro, C. Pacchierotti, C. Yoon, D. Prattichizzo, D.H. Gracias, S. Misra, Evaluation of electromagnetic system with haptic feedback for control of

- untethered, soft grippers affected by disturbances, in: Proc. IEEE Int. Conf. on Biomedical Robotics and Biomechatronics, 2016, pp. 908–913.
- [18] C. Pachierotti, V. Magdanz, M. Medina-Sánchez, O.G. Schmidt, D. Prattichizzo, S. Misra, Intuitive control of self-propelled microjets with haptic feedback, *J. of Micro-Bio Robotics* 10 (1–4) (2015) 37–53.
- [19] J. Zhang, Z. Zhang, D. Xu, W. Zhang, Aligning micro-gripper to ring object in high precision with microscope vision, in: Proc. IEEE Int. Conf. on Computer Science and Automation Engineering, 3, 2011, pp. 123–127.
- [20] M. Bollavaram, P. Sane, S. Chowdhury, S.K. Gupta, A.G. Banerjee, Automated detection of live cells and microspheres in low contrast bright field microscopy, in: Proc. IEEE Int. Conf. on Manipulation, Automation and Robotics at Small Scales, 2016, pp. 1–6.
- [21] T. Xu, G. Hwang, N. Andreff, S. Regnier, Planar path following of 3-d steering scaled-up helical microswimmers, *IEEE Trans. on Robotics* 31 (1) (2015) 117–127.
- [22] S.J. Lee, K. Kim, D. Kim, J. Park, G.T. Park, Recognizing and tracking of 3d-shaped micro parts using multiple visions for micromanipulation, in: Proc. IEEE Int. Symposium on Micromechatronics and Human Science, 2001, pp. 203–210.
- [23] C. Bergeles, K. Shamaei, J.J. Abbott, B.J. Nelson, Single-camera focus-based localization of intraocular devices, *IEEE Trans. on Biomedical Engineering* 57 (8) (2010) 2064–2074.
- [24] Z. Zhang, C.H. Menq, Three-dimensional particle tracking with subnanometer resolution using off-focus images, *Applied optics* 47 (13) (2008) 2361–2370.
- [25] L. Cui, E. Marchand, S. Haliyo, S. Regnier, Three-dimensional visual tracking and pose estimation in scanning electron microscopes, in: Proc. IEEE/RSJ Int. Conf. on Intelligent Robotic Systems, 2016, pp. 5210–5215.
- [26] A. Hong, B. Zeydan, S. Charreyron, O. Ergeneman, S. Pané, M.F. Toy, A.J. Petruska, B.J. Nelson, Real-Time holographic tracking and control of micro-robots, *IEEE Robotics and Automation Letters* 2 (1) (2017) 143–148.
- [27] D. Prattichizzo, J.C. Trinkle, Grasping, in: B. Siciliano, O. Khatib (Eds.), *Springer handbook of robotics*, Springer Science & Business Media, 2008, pp. 671–700. Chapter 28.
- [28] A.T. Miller, S. Knoop, H.I. Christensen, P.K. Allen, Automatic grasp planning using shape primitives, in: Proc. IEEE Int. Conf. on Robotics and Automation, 2, 2003, pp. 1824–1829.
- [29] R. Hassan, B. Cohanin, O. De Weck, G. Venter, A comparison of particle swarm optimization and the genetic algorithm, in: Proc. of the 1st AIAA multidisciplinary design optimization specialist conference, 2005, pp. 18–21.
- [30] X. Zhang, W. Hu, S. Maybank, X. Li, M. Zhu, Sequential particle swarm optimization for visual tracking, in: Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition, 2008, pp. 1–8.
- [31] I. Oikonomidis, N. Kyriazis, A. Argyros, Efficient model-based 3d tracking of hand articulations using kinect, in: Proc. British Machine Vision Conference, 2011, pp. 101.1–101.11.
- [32] I. Oikonomidis, N. Kyriazis, A. Argyros, Tracking the articulated motion of two strongly interacting hands, in: Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition, 2012, pp. 1862–1869.
- [33] C. Qian, X. Sun, Y. Wei, X. Tang, J. Sun, Realtime and robust hand tracking from depth, in: Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition, 2014.
- [34] B. Tamadazte, E. Marchand, S. Dembele, N. Le Fort-Piat, Cad model-based tracking and 3d visual-based control for mems microassembly, *Int. J. of Robotic Research* 29 (11) (2010) 1416–1434.
- [35] S. Scheggi, C. Yoon, D.H. Gracias, S. Misra, Model-based tracking of miniaturized grippers using particle swarm optimization, in: Proc. IEEE/RSJ Int. Conf. on Intelligent Robotic Systems, 2016, pp. 454–459.
- [36] T. Moller, A fast triangle-triangle intersection test, *Journal of graphics tools* 2 (2) (1997) 25–30.
- [37] O. Tropp, A. Tal, I. Shimshoni, A fast triangle to triangle intersection test for collision detection, *Computer Animation and Virtual Worlds* 17 (5) (2006) 527–535.
- [38] O. Devillers, P. Guigue, *Faster Triangle-Triangle Intersection Tests*, PhD. Thesis, INRIA, 2002.
- [39] J. Kennedy, R. Eberhart, Particle swarm optimization, in: Proc. IEEE Int. Conf. on Neural Networks, vol. 4, 1995, pp. 1942–1948.
- [40] T. Yasuda, K. Ohkura, Y. Matsumura, Extended pso with partial randomization for large scale multimodal problems, in: *World Automation Congress*, 2010, pp. 1–6.
- [41] N. Kyriazis, I. Oikonomidis, A. Argyros, A gpu-powered computational framework for efficient 3d model-based vision, *Technical Report TR420* (2011).
- [42] H. Hamer, K. Schindler, E. Koller-Meier, L. Van Gool, Tracking a hand manipulating an object, in: Proc. IEEE Int. Conf. on Computer Vision, 2009, pp. 1475–1482.
- [43] M. Clerc, J. Kennedy, The particle swarm-explosion, stability, and convergence in a multidimensional complex space, *IEEE Trans. on Evolutionary Computation* 6 (1) (2002) 58–73.
- [44] V. Kadiramanathan, K. Selvarajah, P.J. Fleming, Stability analysis of the particle dynamics in particle swarm optimizer, *IEEE Trans. on Evolutionary Computation* 10 (3) (2006) 245–255.

- [45] N. Kyriazis, A.A. Argyros, Scalable 3d tracking of multiple interacting objects, in: Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition, 2014, pp. 3430–3437.



**Stefano Scheggi** received the M.Sc. and Ph.D. degrees in Computer Engineering from the University of Siena, Italy, in 2007 and 2012, respectively. In 2011, he was a Visiting Ph.D. Student at the Department of Computer Science, George Mason University, Fairfax, USA, under the supervision of Prof. Jana Kosecka. Since 2012, he held a postdoctoral position at the University of Siena, Italy. Since 2015, he holds a postdoctoral position at the University of Twente, The Netherlands. His research interests include computer vision, medical robotics, haptics, and augmented/virtual reality.



**Changkyu Yoon** received his Ph. D. from Johns Hopkins University and currently is doing post-doctoral work at Max Planck Institute for Intelligent Systems. His current research focuses on theoretical and technical challenges to study stimuli responsive biomedical soft robotics.



**Arijit Ghosh** has completed his Ph.D from Indian Institute of Science, Bangalore, in 2015. His research interests included artificial magnetic nanorobots, nanoscale swimming and active microrheology. Currently he is a Postdoctoral Fellow at the Johns Hopkins University, Baltimore, where he is investigating the therapeutic applications of self folding microrobots.



**David H Gracias** received the Ph.D. degree from the University of California at Berkeley, Berkeley, CA, USA, and did post-doctoral work at Harvard University, Cambridge, MA, USA. He is a Professor and has been a Principal Investigator an interdisciplinary laboratory at The Johns Hopkins University, Baltimore, MD, USA, since 2003. His current research interests include self-folding, interfacial/thin film/smart materials science, miniature robots, and 3-D/hybrid fabrication with approximately 150 technical publications and 28 issued patents in these areas.



**Sarthak Misra** received the M.E. degree in mechanical engineering from McGill University, Montreal, QC, Canada, and the Ph.D. degree from the Department of Mechanical Engineering, The Johns Hopkins University, Baltimore, MD, USA. He was a Dynamics and Controls Analyst with the International Space Station Program, MacDonald, Dettwiler and Associates, Richmond, BC, Canada, for three years. He joined the University of Twente, Enschede, The Netherlands, in 2009. He is currently a Professor in the Department of Biomechanical Engineering, Faculty of Engineering Technology, University of Twente, where he also directs the Surgical Robotics Laboratory, MIRA-Institute for Biomedical Technology and Technical Medicine. He also has a part-time appointment with the Department of Biomedical Engineering, University of Groningen, Groningen, The Netherlands, and the University Medical Center Groningen, Groningen. His broad research interests are primarily in the area of applied mechanics at both macro and micro scales. He is interested in the modeling and control of electro-mechanical systems with applications to medical robotics. Dr. Misra was a recipient of the European Research Council Starting Grant, Netherlands Organization for Scientific Research (NWO) VENI and VIDI Awards, a Link Foundation Fellowship, a McGill Major Fellowship, and a NASA Space Flight Awareness Award. He is the Co-Chair of the IEEE Robotics and Automation Society Technical Committee on Surgical Robotics and the Area Co-Chair of the IFAC Technical Committee on Biological and Medical Systems.