

Model-based Tracking of Miniaturized Grippers Using Particle Swarm Optimization

Stefano Scheggi, ChangKyu Yoon, David H. Gracias, and Sarthak Misra

Abstract—Micro-sized agents can benefit robotic minimally invasive surgery since they can be inserted into the human body and use natural pathways such as arteries and veins or the gastrointestinal tract, to reach their target for drug delivery or diagnosis. Recently, miniaturized agents with shape-changing and gripping capabilities have provided significant advantages in performing grasping, transportation, and manipulation tasks. In order to robustly perform such tasks, it is of utmost importance to properly estimate their overall configuration. This paper presents a novel solution to the problem of estimating and tracking the 3D position, orientation and configuration of the tips of miniaturized grippers from RGB marker-less visual observations obtained by a microscope. We consider this as an optimization problem, seeking for the gripper model parameters that minimize the discrepancy between hypothesized instances of the gripper model and actual observations of the miniaturized gripper. This optimization problem is solved using a variant of the Particle Swarm Optimization algorithm. The proposed approach has been evaluated on several image sequences showing the grippers moving, rotating, opening/closing and grasping biological material.

I. INTRODUCTION

Untethered miniaturized agents could significantly augment capabilities of minimally invasive surgery (MIS) by replacing tethered medical devices such as endoscopes and probes with smaller, maneuverable, reconfigurable structures. As the agent size goes down to the order of a single cell, previously inaccessible body sites would become available for high-resolution *in-situ* and *in-vivo* manipulations. Thanks to their potential in advancing MIS and micro-assembly, magnetically steered micro-agents have been gaining a growing interest in the research community.

Many traditional micro-agents however have no shape-changing capabilities which results in a limited dexterity that restricts the complexity of achievable tasks. In fact, besides the reduced size, miniaturized agents should have the capability to perform at least simple manipulation tasks in cluttered environments. Miniaturized grippers satisfy the aforementioned requirement which makes them suitable for a wide set of applications, like precise micro-assembly [1], minimally invasive surgery [2], genetics [3], and cell

manipulation [4]. They can be fabricated in sizes that are compatible with the major vascular conduits [5], and can be used to safely grasp, manipulate and transport biological material [6]. Self-folding miniaturized grippers use stimuli responsive mechanisms which provide an alternative to tethered actuation and offer the ability for smart behaviours such as autonomous responses in specific environments. Many of these technologies employ materials (such as hydrogels) which are able to swell and shrink significantly in response to a variety of stimuli like temperature, light, or chemical reactions [7]–[11]. Self-folding miniaturized agents can be used to grasp, manipulate and transport biological materials in MIS. In order to precisely move and control such miniaturized grippers, it is of utmost importance to properly detect and track them. For example, the knowledge of the configuration of the tips of their fingers would allow a user to determine if the object is firmly grasped. 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 [12]. To the best of our knowledge there is no work which tries to estimate the overall configuration of such agents, i.e., 3D pose, rotation, and tips' configuration. The majority of prior research has estimated the 2D pixel position of the miniaturized agents [13]–[17]. The main reason is related to significant technical difficulties in tracking the agents in 3D. Prior research on 3D tracking used multi-cameras [18], [19] and depth-from-focus techniques [20], [21] to estimate the 3D position of the miniaturized agents.

In this paper, we propose a model-based approach to address the problem of tracking the full pose and articulations of miniaturized grippers. We formulated 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 RGB images captured from a microscope. The optimization is performed with a variant of the Particle Swarm Optimization (PSO) algorithm, which exploits the stimuli responsive mechanisms of the miniaturized grippers. PSO has been successfully used to online track the human hands from RGB-D images [22], [23]; however, no one has demonstrated the possibility to use it in order to track miniaturized agents from RGB images. Experimental tests on challenging real data have been used to validate the efficacy of our method. We evaluated the proposed tracker using soft miniaturized grippers, and we believe that it can be successfully applied to different stimuli responsive grippers.

The rest of the paper is organized as follows. Sect. II describes the main steps used to detect the miniaturized

S. Scheggi and S. Misra are affiliated with the Surgical Robotics Laboratory, Department of Biomechanical Engineering, MIRA - Institute for Biomedical Technology and Technical Medicine, University of Twente, 7522 NB, The Netherlands. S. Misra is also affiliated with the Department of Biomedical Engineering, University of Groningen and University Medical Center Groningen, 9713 GZ, The Netherlands.

C. Yoon and D.H. Gracias are with the Department of Materials Science and Engineering, The Johns Hopkins University, MD 21218, USA. D.H. Gracias is also affiliated with the Department of Chemical and Biomolecular Engineering, The Johns Hopkins University, MD 21218, USA.

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).

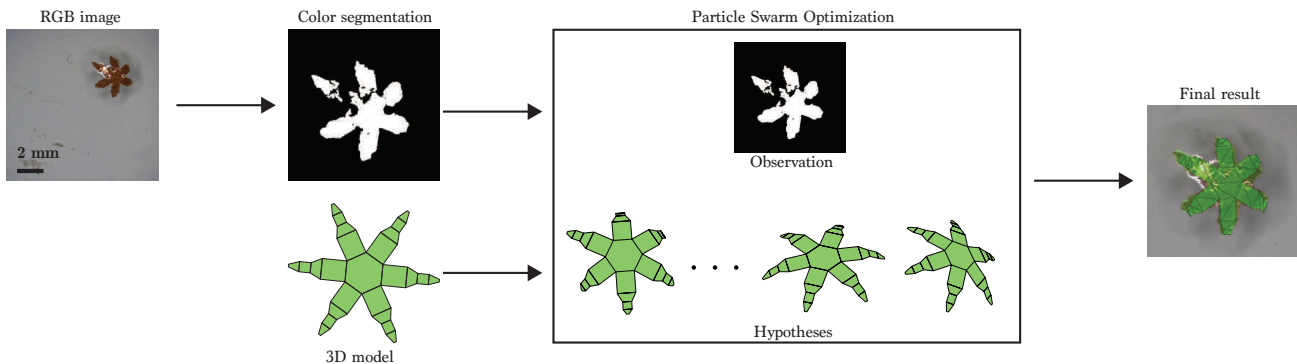


Fig. 1. Graphical illustration of the proposed model-based tracker. From left to right, an input image is taken from a microscope in Red-Green-Blue (RGB) colorspace. Color segmentation is performed in order to detect the miniaturized gripper. The employed gripper’s model and the segmented region of the image are used to recover the gripper configuration using the Particle Swarm Optimization algorithm. The algorithm generates several hypothetical configurations of the gripper. The hypothesis which best matches the observation is the solution to the tracking problem.

gripper and the model-based tracker used in this work. Sect. III introduces the Particle Swarm Optimization algorithm and shows how we used it in order to properly track the miniaturized agents. In Sect. IV the proposed model-based tracker is evaluated via comprehensive experiments on challenging real data and results are given and discussed. Finally, in Sect. V we summarize the main contributions of the paper, and we discuss possible avenues for future research.

II. MODEL-BASED TRACKING

A. Color segmentation

The input to the proposed method is a 1024×1024 RGB color image acquired using a microscope (see Fig. 1). Color segmentation is used to isolate the miniaturized agent. A major decision is the selection of the color space to be employed. Color spaces efficiently separating the chrominance from the luminance components of color are typically considered preferable. This is due to the fact that by employing chrominance-dependent components of color only, some degree of robustness to illumination changes can be achieved. In the proposed work, we selected the HSV color space. Then, the simplest approach for defining what constitutes miniaturized gripper color is to employ bounds on the coordinates of the selected space. We chose these bounds empirically, i.e. by examining the distribution of colors in a preselected set of images. Morphological filtering and opening have been used in order to properly segment the image. Finally, small blobs are removed and the resulting blobs which are close together are merged and kept for further consideration. The resulting image represents the observation \mathcal{O} related to our problem. In order to speed up the segmentation procedure, temporal continuity is exploited to track the grippers in a sequence of frames. Given the estimated position of the tracked gripper from the previous frame, the segmented pixels that are within a preset range from that estimation are kept, whereas the remaining color map is set to zero. The observation model \mathcal{O} that feeds the rest of the process consists of the 2D pixels of the segmented color image.

B. 3D Model and objective function

The adopted 3D model of the miniaturized gripper is made by a set of 46 appropriately assembled triangles. We empir-

ically set the triangles starting from the real measurements of the miniaturized gripper in order to better approximate its shape. We represented the gripper’s pose using a vector of 12 parameters (3 parameters for its position, 3 parameters for its orientation and 1 parameter for each *finger*). For the orientation, we considered the Euler representation. In the proposed work, we modeled each *finger* of the agent as a planar robot made by 4 joints and 4 links. We assumed that the joints which belong to the same *finger* bend in the same way, i.e., they have the same joint’s 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 12 gripper model parameters that minimize the discrepancy among the hypotheses and the actual observations. 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 Sect. II-A. 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} . In particular, given a gripper pose hypothesis \mathcal{H} and the camera calibration matrix \mathcal{C} , the projection of the miniaturized gripper on the camera image plane $\mathcal{P} = f(\mathcal{H}, \mathcal{C})$ is generated by means of computer graphic rendering.

A function $E(\mathcal{H}, \mathcal{O}, \mathcal{C})$, adapted from [22], 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,

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

being \mathcal{P} the projection of the hypothetical miniaturized gripper on the camera image plane, \wedge represents the logical AND operator, while \vee is the logical OR operator. $D(\cdot)$ measures the discrepancies between the pixels of the hypothetical model and the observation. The function $K(\cdot)$ adds a penalty

to possible collisions among the triangles of the model. In order to improve the tracking performance in presence of poor color segmentation and/or occlusions, in Eq. (1) we added the term $S(\cdot)$ which exploits the stimuli responsive mechanisms of the miniaturized grippers. In particular, since the grippers used in this work respond to temperature variations, in $S(\cdot)$ we used the information of the temperature of the environment to penalize unfeasible configurations of the miniaturized gripper. A similar approach can be used for grippers which respond to light, or chemical reactions.

III. 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 [24]. 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 one to effectively handle nonlinear optimization problems and it can be implemented with ease and scalability in order to achieve appropriate results in a faster and inexpensive 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. Since the dimensions of the multidimensional parameter space are bounded in some range, if during the configuration update (Eqs. (2),(3)) a velocity component forces the particle to move to a point outside such range, we constrained that particle to move to the point inside the bounds that minimizes the distance with its updated pose. In this specific work, PSO was used to optimize the objective function reported in Eq. (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. We exploited 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} .

Since PSO may suffer from premature convergence, we used a modified version of this algorithm in order to overcome this problem [25]. We applied a *partial randomization*

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 Eq. (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. A partial randomization of the population is performed every k_r generations.

```

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

```

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 from reaching 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 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 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.

IV. EXPERIMENTAL RESULTS

In this section, representative results of the proposed tracker are provided. The method has been tested and evaluated in a series of image sequences demonstrating challenging tracking scenarios. This is a common approach in the relevant literature because ground truth data for real-world image sequences is hard to obtain. Three video sequences have been considered which report the common motion of the hydrogel miniaturized grippers during manipulation and transportation tasks. In particular, two sequences report the motion of the miniaturized gripper in testbed scenarios, while in the last sequence the agent has to pick biological

plaque-like material on porcine tissue; *please refer to the accompanying video that shows the results of the tracking experiments*. The agents were magnetically controlled in a setup composed of a fluid reservoir and six orthogonally oriented electromagnets [15]. 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 and have been processed offline on an Intel Xeon laptop computer running MS Windows at 3.5 GHz. In the PSO algorithm we used $N = 40$ particles, $G = 30$ generations, and $C = 2$. We performed the partial randomization every $k_r = 5$ generations. Concerning the cost function, $\lambda_D = 10$, $\lambda_K = 5$, and $\lambda_S = 5$.

The first video consists of 4148 frames. When the experiment starts, the agent is stationary and open. The tracker correctly finds its pose in the scene. Later, the miniaturized gripper starts rotating and moving at various directions and speeds. At some point in time, the miniaturized agent starts folding while it moves in the scene. Then it continues to move and rotate and the tracker can still track its pose. Finally, the experiment concludes with the miniaturized agent that unfolds while it continues to move (see Fig. 2).

The second video consists of 7960 frames. In order to evaluate the tracker in a more challenging scenario, we processed one frame out of every five frames and we randomly corrupted the color segmentation in order to simulate erroneous detection of the miniaturized agent. 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. When the experiment starts the agent is open and does not move. Then, it starts moving upward. Although the color segmentation allows us only to partially detect the gripper, the proposed tracker is still able to correctly track the agent (see Fig. 3). Finally, the gripper starts moving and rotating about the three axis (see Fig. 4). Also in this case, the tracker is able to correctly estimates the configuration of the agent.

The last video consists of 2558 frames. We processed one frame out of every two frames. A miniaturized gripper is used to grasp biological material namely yolk from an egg. Egg yolk was chosen due to its high content of cholesterol and fats (58% of its dry weight) in order to simulate plaque build up on porcine muscular tissue. 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 image plane, reaches the target and starts to fold (see Fig. 5). In this case the proposed tracker is able to detect and track the correct configuration of the gripper.

As can be verified from the snapshots, the tracking of the gripper is consistent throughout the whole sequences, which indicates that it is correctly tracked. Thus, the proposed tracker performs well in all the above cases, some of which are challenging. Also, as can be seen from the videos, errors in the tracking phase may be generated by

the soft and deformable nature of the miniaturized grippers which is not taken into account by the proposed tracker. 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 erroneously detected in several frames. Also, situations where the gripper is unfolded and rotated of 90 deg. about its x -, y -axis might lead to inaccurate results. With respect to the computational performance, the proposed prototypical version processes the video sequences presented previously at an average time of 3 s per frame (1024×1024 images). However, since GPU computations can benefit PSO algorithms in order to perform online model-based tracking [26], we believe that the computation time of the proposed algorithm can be reduced by exploiting parallel computations. Finally, although we evaluated the tracker with microscopic images, we think that the proposed approach may work also with clinical imaging modalities like ultrasound, assuming that a proper observation \mathcal{O} of the agent is computed.

V. CONCLUSIONS AND FUTURE WORK

In this paper, we presented a novel model-based method for efficient tracking of hydrogel miniaturized grippers from RGB images. The tracking problem consists in estimating the gripper model parameters that minimize the discrepancy between the gripper hypotheses and the actual observations. To this purpose, we used an appropriate objective function and we employed a variant of Particle Swarm Optimization (PSO) to search for the optimal configuration. The proposed method was tested on three different datasets of more than 14000 images. Our results show that the proposed method allows to properly track and detect the configuration of the miniaturized agent. With respect to existing methods, this is the first one which tries to estimate the full pose of miniaturized grippers from RGB images.

This work can be extended in many ways: first, we plan to exploit the inherent parallelism of PSO by performing most of the computation on a GPU. Second, we will conduct similar experiments using different types of grippers in order to further validate the proposed approach. Finally, we will consider using the ultrasound imaging modality which provides the position of the miniaturized gripper in situations where visual feedback cannot be provided via a microscope system.

REFERENCES

- [1] N. Dechev, W. L. Cleghorn, and J. K. Mills. Microassembly of 3-d microstructures using a compliant, passive microgripper. *J. of Microelectromechanical Systems*, 13(2):176–189, 2004.
- [2] T. Yoneyama, T. Watanabe, H. Kagawa, J. Hamada, Y. Hayashi, and M. Nakada. Force detecting gripper and flexible micro manipulator for neurosurgery. In *Proc. IEEE Int. Conf. on Engineering in Medicine and Biology Society*, pages 6695–6699, 2011.
- [3] A. Ichikawa, S. Sakuma, F. Arai, and 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*, pages 3795–3800, 2014.
- [4] M. M. S. Mousavi, G. De Pasquale, A. Somà, and 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*, pages 356–361, 2011.

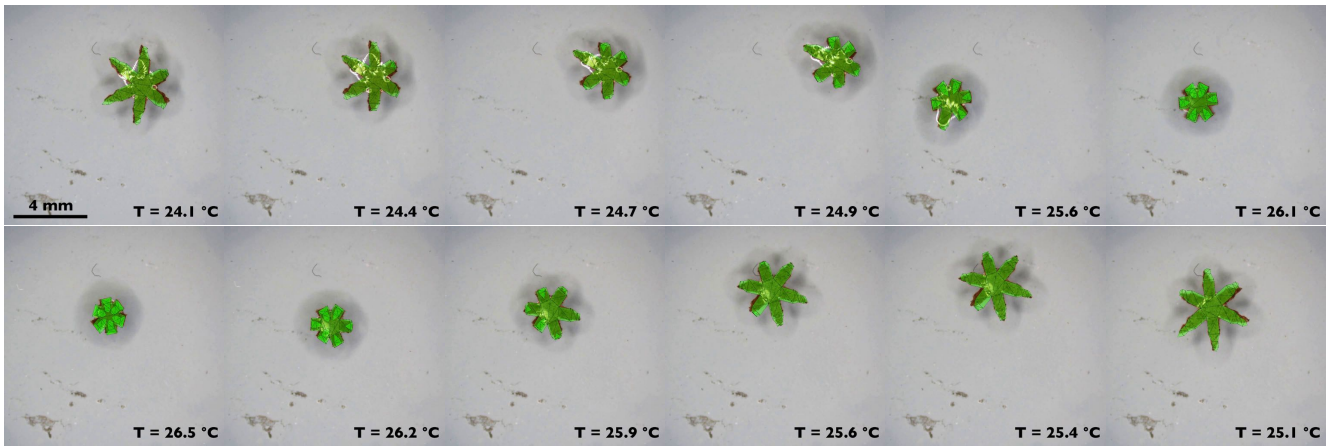


Fig. 2. Characteristic snapshots from the tracking experiment on the first image sequence (best view in colors). Top: the agent folds performing a grasping task. Bottom: the miniaturized gripper unfolds performing a release task. For each snapshot, the estimated configuration of the gripper is displayed 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.*

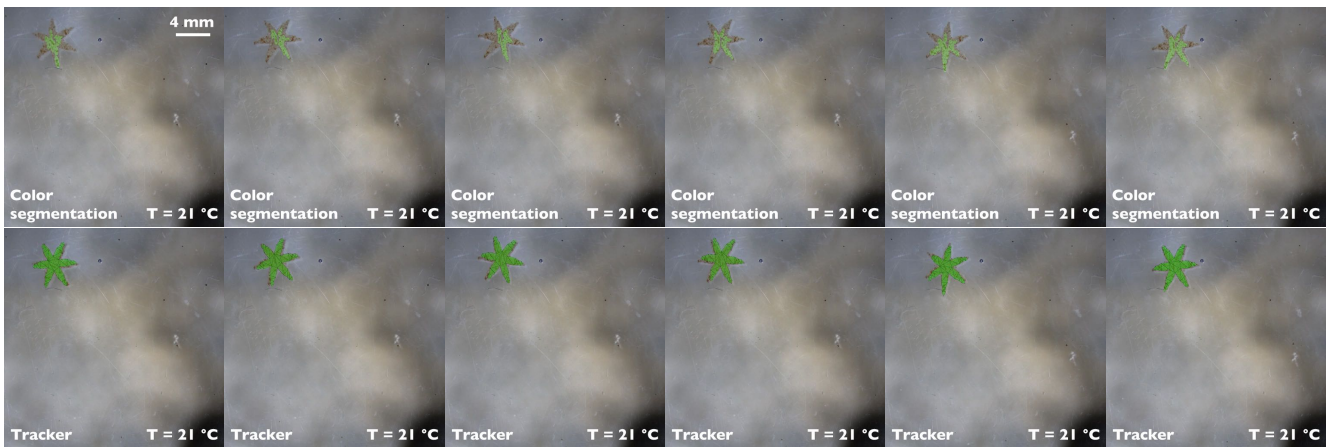


Fig. 3. Characteristic snapshots from the tracking experiment on the second image sequence (best view in colors). For each snapshot we reported the randomly corrupted color segmentation (top) and the estimated configuration of the gripper (bottom). Both the color segmentation and the estimated configuration of the gripper are 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.*

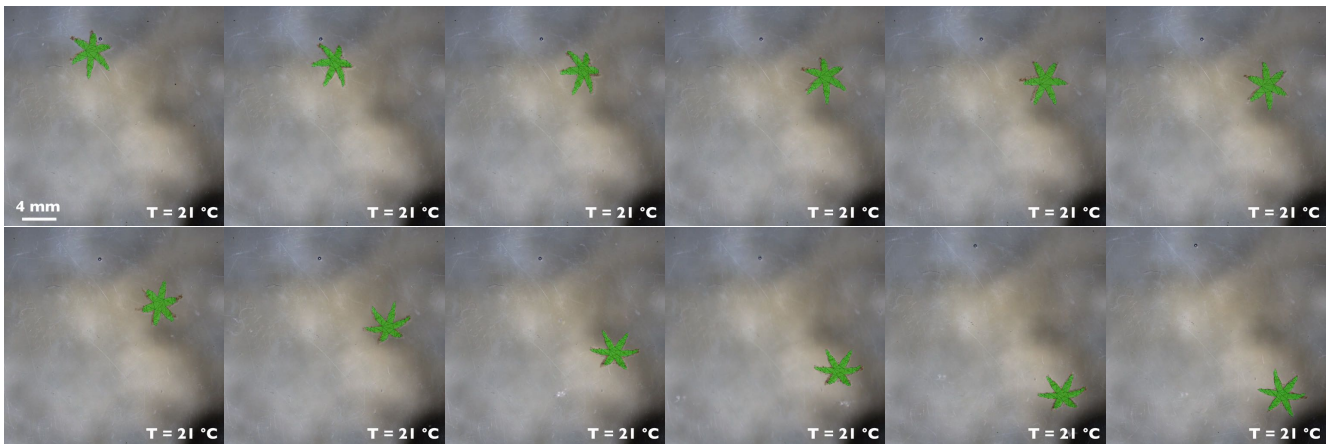


Fig. 4. Characteristic snapshots from the tracking experiment on the second image sequence (best view in colors). The sequence shows the miniaturized gripper rotating about its axes. For each snapshot, the estimated configuration of the gripper is displayed in green. Temperatures (T) are shown on the bottom-right corner of each snapshot. *Please refer to the accompanying video that shows the results of the tracking experiments.*

[5] F. Hansen, P. Mangell, B. Sonesson, and T. Lanne. Diameter and compliance in the human common carotid artery: variations with age and sex. *Ultrasound in medicine & biology*, 21(1):1–9, 1995.
 [6] F. Beyeler, A. Neild, S. Oberti, D. J. Bell, Y. Sun, J. Dual, and 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):7–15, 2007.

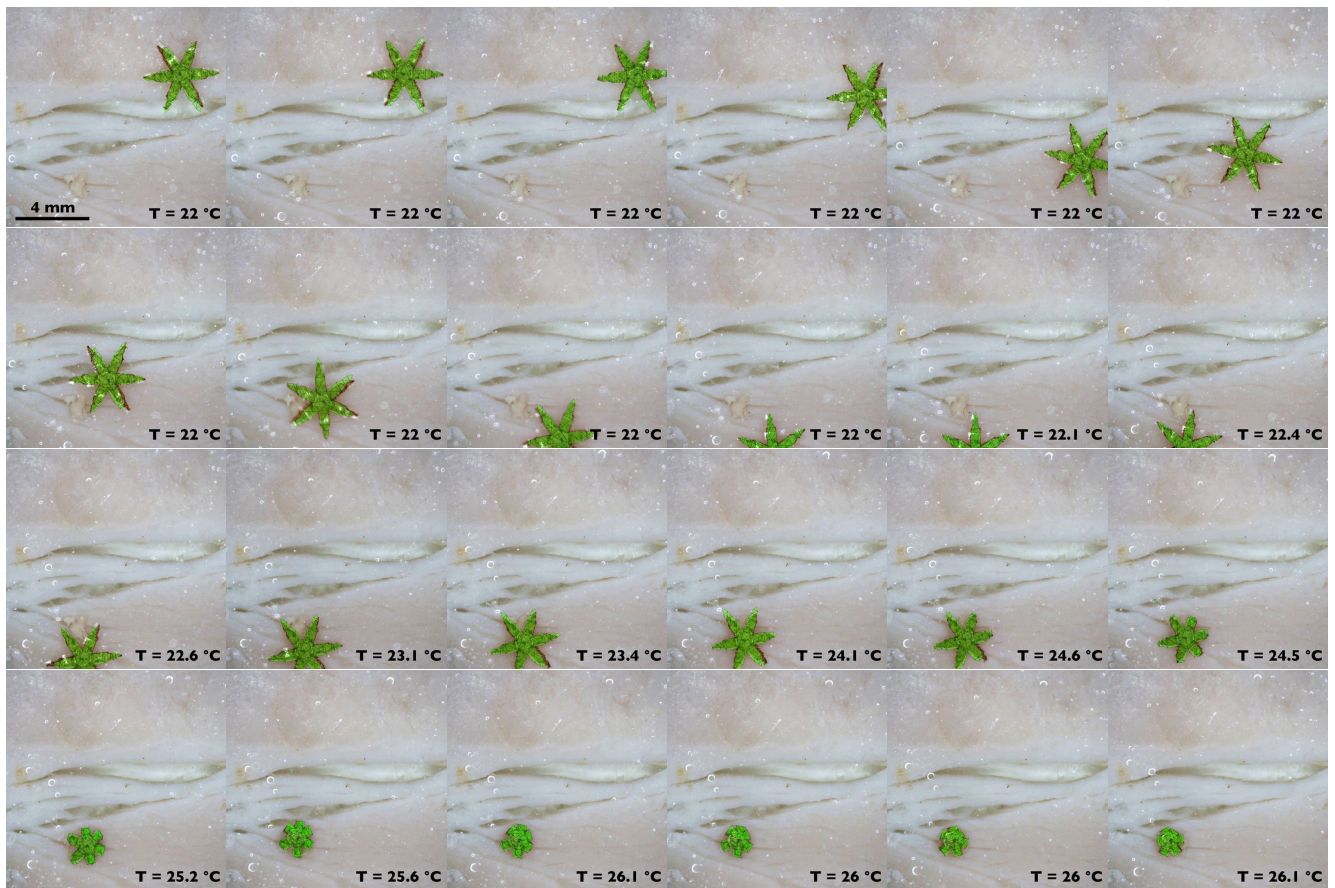


Fig. 5. Snapshots of the miniaturized gripper moving on porcine muscle tissue and picking egg yolk (best view in colors). For each snapshot, the estimated configuration of the gripper is displayed in green. Temperatures (T) are shown on the bottom-right corner of each snapshot. Please refer to the accompanying video that shows the results of the tracking experiments.

- [7] D. H. Gracias. Stimuli responsive self-folding using thin polymer films. *Current Opinion in Chemical Engineering*, 2(1):112–119, 2013.
- [8] 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, and B. J. Nelson. An integrated microrobotic platform for on-demand, targeted therapeutic interventions. *Advanced Materials*, 26(6):952–957, 2014.
- [9] H. Li, G. Go, S. Y. Ko, J.-O. Park, and S. Park. Magnetic actuated ph-responsive hydrogel-based soft micro-robot for targeted drug delivery. *Smart Materials and Structures*, 25(2), 2016.
- [10] N. Bassik, B. T. Abebe, K. E. Laffin, and D. H. Gracias. Photolithographically patterned smart hydrogel based bilayer actuators. *Polymer*, 51(26):6093–6098, 2010.
- [11] E. Gultepe, J. S. Randhawa, S. Kadam, S. Yamanaka and F. M. Selaru, E. J. Shin, A. N. Kalloo, and D. H. Gracias. Biopsy with thermally-responsive untethered microtools. *Advanced materials*, 25(4):514–519, 2013.
- [12] A. T. Miller, S. Knoop, H. I. Christensen, and P. K. Allen. Automatic grasp planning using shape primitives. In *Proc. IEEE Int. Conf. on Robotics and Automation*, volume 2, pages 1824–1829, 2003.
- [13] J. Zhang, Z. Zhang, D. Xu, and 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*, volume 3, pages 123–127, 2011.
- [14] A. Burkle and S. Fatikow. Computer vision based control system of a piezoelectric microrobot. In *Proc. Int. Conf. on Computational Intelligence for Modelling, Control and Automation*, 1999.
- [15] C. Pacchierotti, V. Magdanz, M. Medina-Sánchez, O. G. Schmidt, D. Prattichizzo, and S. Misra. Intuitive control of self-propelled microjets with haptic feedback. *J. of Micro-Bio Robotics*, 10(1-4):37–53, 2015.
- [16] F. Ongaro, C. Yoon, F. van den Brink, M. Abayazid, S. H. Oh, D. H. Gracias, and S. Misra. Control of soft untethered grippers for pick-and-place tasks. In *Proc. IEEE RAS/EMBS Int. Conf. on Biomedical Robotics and Biomechanics*, pages 299–304, 2016.
- [17] F. Ongaro, C. Pacchierotti, C. Yoon, D. Prattichizzo, D. H. Gracias, and S. Misra. Evaluation of electromagnetic system with haptic feedback for control of untethered, soft grippers affected by disturbances. In *Proc. IEEE RAS/EMBS Int. Conf. on Biomedical Robotics and Biomechanics*, pages 908–913, 2016.
- [18] S. J. Lee, K. Kim, D. Kim, J. Park, and G. T. Park. Recognizing and tracking of 3d-shaped micro parts using multiple visions for micro-manipulation. In *Proc. IEEE Int. Symposium on Micromechatronics and Human Science*, pages 203–210, 2001.
- [19] I. S. M. Khalil, V. Magdanz, S. Sanchez, O. G. Schmidt, and S. Misra. Three-dimensional closed-loop control of self-propelled microjets. *Applied physics letters*, 103(17):172404, 2013.
- [20] C. Bergeles, K. Shamaei, J. J. Abbott, and B. J. Nelson. Single-camera focus-based localization of intraocular devices. *IEEE Trans. on Biomedical Engineering*, 57(8):2064–2074, 2010.
- [21] Z. Zhang and C. H. Menq. Three-dimensional particle tracking with subnanometer resolution using off-focus images. *Applied optics*, 47(13):2361–2370, 2008.
- [22] I. Oikonomidis, N. Kyriazis, and A. Argyros. Efficient model-based 3d tracking of hand articulations using kinect. In *Proc. British Machine Vision Conference (BMVC)*, pages 101.1–101.11, 2011.
- [23] C. Qian, X. Sun, Y. Wei, X. Tang, and J. Sun. Realtime and robust hand tracking from depth. In *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition*, 2014.
- [24] J. Kennedy and R. Eberhart. Particle swarm optimization. In *Proc. IEEE Int. Conf. on Neural Networks*, volume 4, pages 1942–1948 vol.4, 1995.
- [25] T. Yasuda, K. Ohkura, and Y. Matsumura. Extended pso with partial randomization for large scale multimodal problems. In *World Automation Congress*, pages 1–6, 2010.
- [26] N. Kyriazis, I. Oikonomidis, and A. Argyros. A gpu-powered computational framework for efficient 3d model-based vision. *Technical Report TR420*, 2011.