

Asynchronous Event-based Visual Shape Tracking for Stable Haptic Feedback in Microrobotics

Zhenjiang Ni¹, Aude Bolopion², Joël Agnus², Ryad Benosman³ and Stéphane Régnier¹

Abstract—Micromanipulation systems have recently been receiving increased attention. Tele-operated or automated micromanipulation is a challenging task due to the need for high-frequency position or force feedback to guarantee stability. In addition, the integration of sensors within micromanipulation platforms is complex. Vision is a commonly used solution for sensing; unfortunately the update rate of the frame-based acquisition process of current available cameras can not ensure—at reasonable costs—stable automated or tele-operated control at the microscale level, where low inertia produces highly unreachable dynamic phenomena. This paper presents a novel vision-based microrobotic system combining both an asynchronous Address Event Representation (AER) silicon retina and a conventional frame-based camera. Unlike frame-based cameras, recent artificial retinas transmit their outputs as a continuous stream of asynchronous temporal events, in a manner similar to the output cells of a biological retina, enabling high update rates. This paper introduces an Event-based Iterative Closest Point algorithm to track a microgripper’s position at a frequency of 4 kHz. The temporal precision of the asynchronous silicon retina is used to provide a haptic feedback to assist users during manipulation tasks, whereas the frame-based camera is used to retrieve the position of the object that must be manipulated. This paper presents the results of an experiment on tele-operating a sphere of diameter around 50 μm using a piezo-electric gripper in a pick-and-place task.

I. INTRODUCTION

Versatile 3D manipulation systems able to operate in ambient conditions on micrometer-sized objects would greatly increase the potential applications of microtechnology [1]. However, the development of such systems faces a major obstacle: the lack of position and force feedback [2]. Sensors have been developed [3], [4], but their integration into dedicated tools increases significantly the complexity and cost of the tool fabrication. In particular, even if some microgrippers offer sensing capabilities (at the expense of a complex design) [5], [6], [7], most still lack force measurement capabilities [8], [9].

Vision is a promising way to avoid the complexity of integrating sensors [10], [11]. Visual information can be converted to force measurements to monitor the efforts applied on the objects during manipulation [12], [13]. This is achieved

by using deformable tools after a calibration step [14]. In particular, the stiffness of the tools must be determined in order to relate the measured deformations to the applied forces. However, the precise value of the force might not be necessary for controlling micromanipulation systems, whether in the tele-operated or the automated mode. In these cases, the position feedback obtained from vision sensors might be sufficient. In all cases, highly dynamic phenomena due to the low inertia at this scale must be recorded. Most of the existing vision-based systems provide feedback at a couple of tens of Hertz, and their complexity depends on the size of the observed scenes [15]. Thus they can not ensure the proper monitoring of highly dynamic motion.

Conventional frame-based cameras’ lack of dynamic information and their redundancies set an important limit to potential micromanipulations in automated or tele-operated modes. Event-based computer vision based on *Address Event Representation* (AER) provides a sound solution to high-speed vision problems [16]. This newly developed discipline is motivated by mimicking biological visual systems [17]. The Dynamic Vision Sensor (DVS) silicon retina used in the research presented in this paper reacts to changes of contrast, which are then converted into a stream of asynchronous time-stamped events [18]. The reduction of redundant information makes this technique promising for high-speed tracking.

The use of event-based retinas requires the development of time-oriented event-based algorithms, in order to benefit fully from the properties of this new framework [19]. Neural shape coding is a difficult issue as there is almost an infinite number of representations of shapes in the real world. A computationally efficient method is HMax: it models a biological visual system to extract features of different forms for object recognition [20]. However, the shapes of micromanipulators are not very complex. As will be shown, simpler algorithms making full use of the high temporal resolution of the DVS (μs precision) allow fulfilling the high-frequency requirements of micromanipulation. The Iterative Closest Point (ICP) is an efficient algorithm, dedicated to minimizing the difference between a data point set and a model point set [21]. Many variants have been proposed to enhance its performance and adaptability [22]. However, 2D image processing can only be achieved at a frequency of several 10 Hz, depending on the number of points to be matched [23]. Although this is sufficient for many robotic applications, these frequencies are far from sufficient for automated or tele-operated micromanipulation, which requires refresh rates greater than 1 kHz to ensure the stability of the control loop, because of the highly dynamic physical phenomena involved [24].

¹ Institut des Systèmes Intelligents et de Robotique, Université Pierre et Marie Curie, CNRS UMR 7222, 4 Place Jussieu, 75005 Paris, France. {ni, regnier}@isir.upmc.fr

² Institut FEMTO-ST, UFC, ENSMM, UTBM, CNRS UMR 6174, 24 rue Alain Savary, 25000 Besançon, France. aude.bolopion@femto-st.fr, joel.agnus@ens2m.fr

³ Vision Institute, Université Pierre et Marie Curie, UMR S968 Inserm, UPMC, CNRS UMR 7210, CHNO des Quinze-Vingts, 17 rue Moreau, France. ryad.benosman@upmc.fr

This paper presents an Event-based Iterative Closest Point algorithm (EICP) directly applied to the silicon retina's output. It allows tracking the manipulation tool at a frequency of 4 kHz. The positions of static objects are provided by a conventional frame-based camera. This approach was validated by a tele-operated pick-and-place task using a piezo-electric gripper, involving a sphere with a diameter of about 50 μm . Haptic feedback directly estimated from the output of both the event-based retina and the conventional frame-based camera is provided to assist users during the manipulation. The first 3D pick-and-place manipulation with haptic feedback using a microgripper is successfully achieved. Stable vision-based tele-operation has been achieved in this project by the use of event-based retinas and the EICP algorithm.

This paper is based on several previous papers. In [19], DVS was used to develop an event-based Hough transformation to track specifically circles. Hough transformations rely on a voting scheme and maximum detection within the accumulation spaces to identify the location of a shape. That paper ends by showing that this can be used to detect the rapid dynamics of Brownian motion. In this paper, we develop a new method to track an arbitrary complex shape, which uses an incremental method that is able to compute, in real time at several kHz, the geometric transformation that maps a complex reference shape to the events acquired by the DVS. This method is related to the family of Iterative Closest Point techniques applied to asynchronous events. This paper also combines the use of a frame-based camera and an event-based camera. Apart from the new vision algorithm, the contribution of this paper includes the action of the DVS at very high frequency as a position sensor, providing valuable haptic feedback. A previous paper treated virtual haptic guides for pick-and-place operations at the microscale [25]. However, it used a homemade AFM composed of two independent cantilevers, which made the system very difficult to use. The cantilevers were equipped with force sensors (two optical levers) that provided haptic feedback, but this increased the complexity of the setup since the laser beams had to be aligned with the cantilevers and photodiodes. It was thus definitely not a system that could be used by non-expert users. For these reasons, a simple gripper was chosen for the project presented here. Compared to [25], where the two AFM cantilevers were equipped with force sensors, the gripper used in the present project is sensor-deprived. Vision is thus used to compute the haptic feedback. At the microscale, the integration of sensors inside the manipulation tools increases the complexity and cost of the fabrication. Here, we show that event-based cameras can provide this feedback. The integration of this sensor is then a sound solution for microscale applications. Its high frequency and low data load enable the monitoring of the rapid dynamics common at the micro scale.

This paper is organized as follows. The experimental setup is presented in Section II. Section III gives details about the vision algorithms used to compute the haptic feedback, which in turn is described in Section IV. The proposed approach is validated by the experimental results presented in Section V. Finally, Section VI concludes the paper.

II. SETUP

Several tools have been developed to manipulate micrometer-sized objects in ambient conditions. The most common ones include cantilevers and grippers. Cantilevers can be used for 2D manipulations, such as pushing or pulling [2]. Pick and place operations have also been demonstrated using two protruding tip cantilevers, but the complexity of the setup limits its applicability [25]. A microgripper designed at the FEMTO-ST Institute¹ (Fig. 1) is used in this project to perform 3D manipulations that will enable a large range of applications, including microassembly. It is based on a pair of piezo-electric beams with two degrees of freedom, called a duo-bimorph, as the actuation principle of the two fingers [26]. On each actuator, four electrodes referred to a central ground and two voltages are necessary to impose the displacements, based on the deflections of the piezo-electric beam. This configuration offers a number of capabilities: not only an open-and-close motion, but also an up-and-down motion that allows, for example, a fine up-and-down approach or a fine alignment of the finger tips. Objects are grasped by the mean of two silicon end effectors, fabricated by a DRIE (Deep Reactive-Ion Etching) process. They are designed to minimize the sticking effects between the end effector and the objects, in order to facilitate the release of the objects. The end-effectors are fixed on the piezo-electric actuators with reversible thermal glue. The gripper is controlled with instructions sent from a PC to a high-voltage interface (four channels of ± 150 V) via an RS232 link. Such a microgripper presents a typical stroke of open/close motion and up/down motion of, respectively, 320 μm and 200 μm at the end of the finger tips for ± 100 V. The gripping force is on the order of a millinewton. The gripper is mounted on a 3-axis motorized micromanipulator² to allow accurate positioning with respect to the substrate (Fig. 1). The manipulator used relies on stepper motors with a step size of 0.040 μm . It is a cable-driven system with cross roller bearings, it has a sub micrometer resolution, and a travel range of 25 mm. The manipulator was originally controlled through a serial port. However, to increase the communication frequency, a joystick is emulated by programming the manipulator's parallel port using a PCI6259 National Instrument acquisition card.

As shown in Fig. 1, the observed scene is monitored by two optical sensors which record the same view. The light beam is divided into two optical paths, and redirected to an asynchronous event-based silicon retina and a conventional frame-based camera (GigE vision, Basler). The scene recorded by both sensors is magnified by a $10\times$ objective (Olympus).

Haptic feedback is provided to users by an Omega haptic device³, with three degrees of freedom for both displacement and force feedback. Forces higher than 5 N are saturated, to avoid any damage to the interface. Both the micromanipulator and the gripper are controlled through the use of this device.

A single PC (Intel Xeon core, 2.93 GHz) operating under

¹<http://www.femto-st.fr/>

²Sutter Instrument, http://www.sutter.com/MP_285

³Force Dimension, <http://www.forcedimension.com>

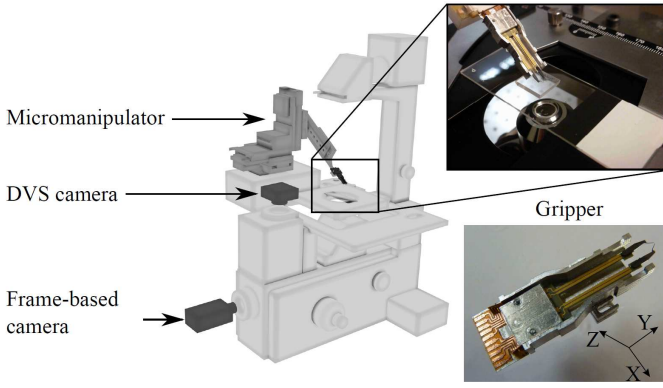


Fig. 1. Setup of the micromanipulation platform

Windows 7 runs the threads corresponding to the gripper, the micromanipulator, the vision detection, and the haptic feedback.

III. VISUAL TRACKING

A. Event-based artificial vision

Information about a visual scene is transmitted in an asynchronous manner by biological retinas, unlike frame-based cameras, and with less redundancy. The event-based vision sensor compresses the digital data in the form of events, removing redundancy, reducing latency, and increasing the dynamic range compared with conventional imagers. A complete review of the history and the existing sensors can be found in [17].

The Dynamic Vision Sensor (DVS) used in this work is an Address-Event Representation (AER) silicon retina with 128×128 pixels [18]. As shown in Fig. 2(a), when the change in scene reflectance in log units exceeds a set threshold, a +1 or -1 event is generated by the pixel, depending on whether the log intensity increased or decreased. Because the DVS is not clocked (as are conventional cameras), the timing of the events can be conveyed with a temporal resolution of approximately $1 \mu\text{s}$. Thus the ‘effective frame rate’ is typically several kilohertz. The absence of events when no change of contrast is detected implies that the redundant visual information usually recorded in frames is not carried in the stream of events. Events are transmitted to a computer using a standard USB connection.

Let us denote by $ev(\mathbf{p}, t)$ an event occurring at time t at the spatial location $\mathbf{p} = (x, y)^T$. Fig. 2(b) shows an example of the spatio-temporal visualization of a set of DVS events in response to the microgripper’s closing on a microsphere. An event $ev(\mathbf{p}, t)$ describes an activity in the spatio-temporal space. Similar to biological neurons, its influence lasts for a certain amount of time after it has been active. This temporal property of events can be introduced in the form of a decay function applied to model this phenomenon. We can then define $S(t)$, the spatio-temporal set of events active at time t , by

$$S(t) = \left\{ ev(\mathbf{p}, t_i) \mid e^{-\frac{t-t_i}{\tau}} > \delta_t \right\}, \quad (1)$$

with τ being the time constant parameter of the decay function and δ_t the predefined threshold.

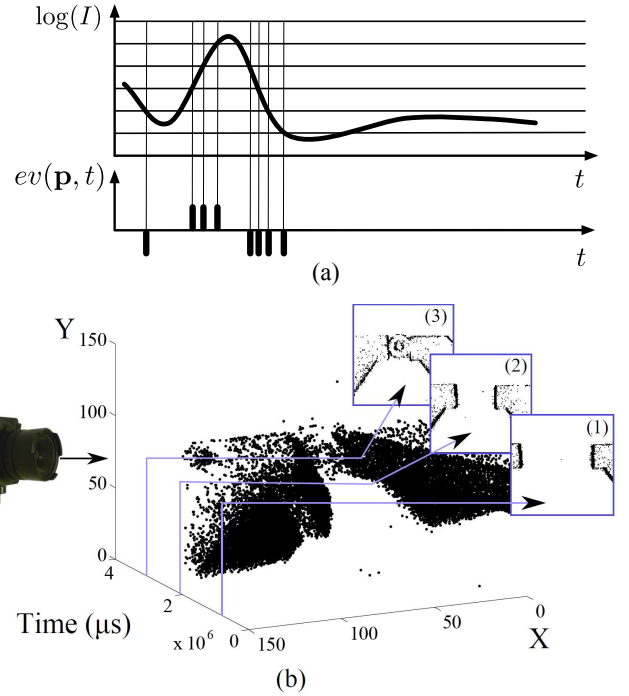


Fig. 2. (a) Principle of the generation of events of DVS pixels, adapted from Lichtsteiner et al. [18]. Events with +1 or -1 polarity are emitted when the change in log intensity exceeds a predefined threshold. (b) Events are generated in (X, Y, t) space when the gripper closes on a sphere. Images (1)–(3) are shown at chosen temporal locations; they correspond to the accumulation maps of the events, namely, the projection of all events over a time interval on a single plane (X, Y) regardless of their timings.

B. The Event-based Iterative Closest Point algorithm

The principle of ICP algorithms is to use iteratively a model point set delineating the desired object contour to match an acquired data point set (the matching step). Each step estimates a rigid transformation between the known model and the data, which expresses their geometric relationship (the tracking step). The ICP algorithm is particularly adapted to the task of tracking the gripper’s position, as most of its constituent shapes remain unchanged over time; more importantly, the scale of the observation remains unchanged during all the tracking. Let $G(t)$ be the set of positions of 2D model points defining the shape of the gripper at time t . $M_{ev}(t)$ is the set of pixelic locations of active pixels of the silicon retina at time t , defined by

$$M_{ev}(t) = \{ \mathbf{p} \in \mathbb{R}^2 \mid ev(\mathbf{p}, t) \in S(t) \}. \quad (2)$$

Following the ICP algorithm, a matching function is needed to pair the model points with the active pixels of the silicon retina. An active event is matched with an element of $G(t)$ by computing the minimal distance between the event’s position and all points of $G(t)$ that have not yet been paired.

We can then define the matching function by

$$\begin{aligned} match : M_e(t) &\rightarrow G(t) \\ \mathbf{p} &\mapsto \mathbf{p}_k, k \mid \arg \min_{k \in \{1, \dots, N_G\}} d(\mathbf{p}, \mathbf{p}_k), \end{aligned} \quad (3)$$

where $d(\mathbf{p}, \mathbf{p}_k)$ is the Euclidean distance between two points and N_G is the size of $G(t)$.

It then becomes possible to estimate the rigid body transformation (R, \mathbf{T}) between $M_{ev}(t)$ and $G(t)$ by minimizing a mean square cost function:

$$\min_{R \in SO(2), \mathbf{T} \in \mathbb{R}^2} \sum_{\mathbf{p} \in M_{ev}(t)} \|R\mathbf{p} - \mathbf{T} - match(\mathbf{p})\|^2, \quad (4)$$

where R is the rotation matrix belonging to the 2D special orthogonal group $SO(2)$ and \mathbf{T} is the translation vector. Readers interested in the details of this minimization can refer to [21]. Fig. 3 provides the principle of the event-based algorithm.

Algorithm 1 Event-based Iterative Closest Point Algorithm

Require: Event $ev(\mathbf{p}, t)$

- 1: **for** every incoming $ev(\mathbf{p}, t)$ **do**
 - 2: Update the content of $S(t)$ and $M_{ev}(t)$.
 - 3: Compute $match(\mathbf{p})$.
 - 4: Estimate (R, \mathbf{T}) according to equation (4).
 - 5: Update the position of model points of $G(t)$ using (R, \mathbf{T}) .
 - 6: **end for**
-

The gripper closes at a speed of 13 pixels per second (1 *pixel*=1.5 μm), producing a mean rate of 14k events per second. The edge width of the gripper in the DVS focal plane is around three pixels. When the gripper finger passes a pixel, 10.6 events on average are generated on it. Therefore, according to the timestamp, one pixel remains active for 2.46 ms. The decay function permits a pixel’s activity to be considered during a certain period after it has been spiked, which is tuned to about 10 ~ 15 ms. The EICP is event driven, and its update rate has a mean value of 4 kHz. The algorithm is implemented in Java under the JAER open-source project [27].

As far as we know, high speed tracking of relatively complex shaped micromanipulation tools by using software alone has not yet been reported. However, the general visual tracking performance of an ICP algorithm has been experimentally demonstrated. The closest performance can be found in a hardware implementation of a frame-based technique in [28], where it is shown that the processing speed reaches a rate of 200 fps for 512 data points and 2048 model points. That paper also demonstrates that a software implementation of the same algorithm runs at a rate of 39 Hz with extra optimization. The use to be presented here of an event-based camera using a classic PC with no particular software speeding process outperforms existing algorithms, since several kHz are reached.

The performance benefits of using DVS over frame-based cameras is significant. A dynamic vision sensor does not generate frames: its pixels are autonomous and react asynchronously to temporal contrast changes. The current shape registration algorithms are designed to work on image frames, usually acquired at a fixed rate. An acquired frame must then be processed, an edge detection must be applied, and in a second stage the registration algorithm is used to track the object. This operation is time consuming. The data load of a classic frame grabber at a frequency of 1 kHz would be so huge that the computation power needed would require specific

hardware or a computer grid. The benefits of a dynamic vision sensor is that it encodes during its acquisition process a compression and a natural edge detection. Thus, the amount of data provided by a DVS is really low: the seminal paper [18] shows that this can amount to 10% of an equivalent frame-camera of the same resolution. The second advantage of its computation is that it introduces an alternative approach to visual processing by introducing the timing of the changes as a main computational feature, and not grey level values.

The frame-based camera in our system serves as a complement to the DVS silicon retina solely for static object detection. The focal planes of both the DVS (128×128 pixels) and the frame camera (659×494 pixels) are related by a homography transform, as both observe the same 2D plane [29]. The homography is estimated off-line by extracting from both sensors’ focals plane the coordinates of six corner points of the gripper fingers and linking them to the actual metric of the gripper’s points in the scene (see Fig. 4). During the application, the circle corresponding to the sphere to be manipulated is detected using a frame-based Hough transform through the conventional camera output. Once detected, its location is converted into the coordinate system of the focal plane of the event-based retina. This operation provides the locations of both the gripper and the sphere in the same coordinate system. It then becomes possible to estimate the distance δ_x between the gripper’s fingers. If an object is detected between the two fingers, the relative finger-object distance on the left and the right sides δ_{xl} , δ_{xr} and the distance between the centre of the sphere and the gripper δ_y (Fig. 4) are estimated as well. These distances will be used to compute the haptic feedback.

IV. HAPTIC COUPLING

A. Manipulation modes

Fig. 5 represents the coupling between the haptic device and the microgripper. The operator applies a force \mathbf{F}_{op} to the haptic device and receives a force feedback \mathbf{F}_h based on the distance between the gripper and the sphere δ_x , δ_{xl} , δ_{xr} and δ_y determined from vision. Using the haptic interface, the user can control the displacement of the micromanipulator as well as the opening and the closing of the gripper. Scaling factors α_d and α_{oc} are introduced to convert the position of the haptic handle to the variables used to control the position and the actuation of the gripper.

To ensure ease of manipulation, different modes have been defined with adapted haptic feedback:

- Planar displacement: the operator controls the displacement of the gripper in a plane parallel to the substrate, the (x, y) plane. No force feedback is provided: $\mathbf{F}_h^T = [0 \ 0 \ 0]$,
- Vertical displacement: the operator controls the displacement of the gripper along the vertical direction. A constant repulsive force feedback is provided along the z direction so that the user has to make an effort to approach the gripper to the substrate, to avoid unexpected contact: $\mathbf{F}_h^T = [0 \ 0 \ F_{hz}]$,

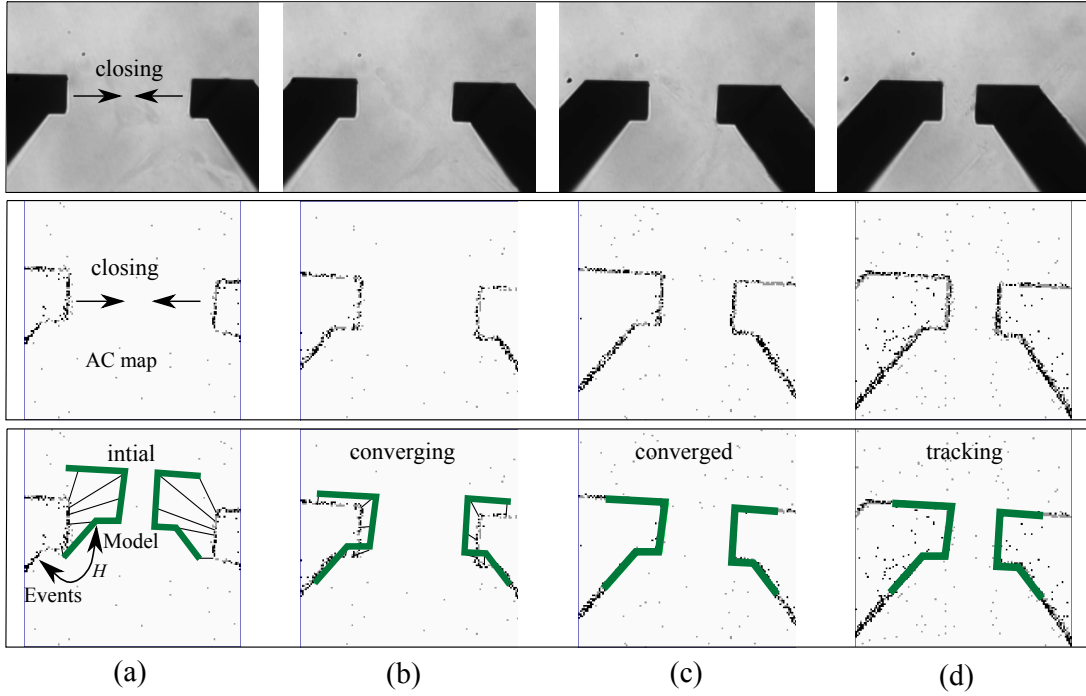


Fig. 3. The principle of event-based iterative closest point tracking (EICP). The first row is a sequence of conventional images showing the closing of the gripper. The middle row shows the events' accumulation maps. The last row is the convergence of the EICP model to the gripper edges. The four images in the first column (a) show the initial state when the gripper is fully opened. The model set (solid lines) is trying to match the corresponding closest events, which are represented on an accumulation map (AC map), and the rigid body transformation (R, T) is estimated to update the model position. In (b), the model is converging to the real gripper's position, until it converges to the gripper's location (c). Finally, in (d), while the gripper is closing, the model's position is updated simultaneously.

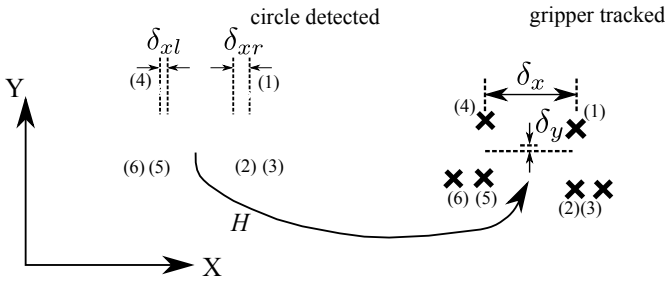


Fig. 4. The calibration between the classical image (left) and the DVS accumulation map (right). Six points (crosses (1)–(6)) have been chosen to calculate the homography transform H . The detected circle (left) is transferred by UDP socket so that the DVS part has both gripper and circle position available (right). δ_{xl} and δ_{xr} describe the distance between the gripper's fingers and the sphere on the left side and the right side, δ_x is the distance between the two fingers, and δ_y is the distance between the centre of the fingers and the centre of the sphere.

- Gripper control: the operator controls both the opening and closing of the gripper and its position along the y-axis to align the gripper with the middle line of the sphere. A 2D force feedback (explained in detail in the next paragraph) is provided to assist the user: $\mathbf{F}_h^T = \begin{bmatrix} F_{hoc} & 0 & F_{hy} \end{bmatrix}$.

The operator selects the appropriate manipulation mode on a graphical user interface developed in C++. To avoid any large and sudden changes in force feedback during transitions between different modes, the haptic force is filtered by a

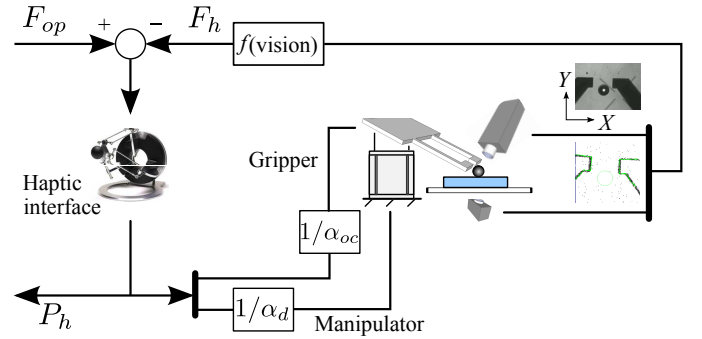


Fig. 5. Haptic coupling scheme. The user controls the position of the gripper and the actuation of the gripper's fingers using the haptic interface and receives haptic feedback through the device. The haptic force is based on the distance between the gripper and the sphere, determined from vision algorithms. Scaling factors α_d and α_{oc} convert the position of the haptic handle to the variables used to control the position and the actuation of the gripper.

second-order low-pass filter for the first couple of seconds after the selection of the desired mode. The filter is then deactivated to enable all the force variations to be sent to the user without smoothing. Note that even if the haptic feedback is delayed because of the time response of the filter, this is not an important issue as it occurs before the user actually starts to manipulate the gripper in the chosen mode.

B. Gripper control

To increase the success rate of the pick-and-place operation, two criteria should be met: the sphere should be grasped on its middle line, and the grasping force should be enough to lift the sphere yet controlled to avoid any damages to the object. The haptic feedback must assist the user in these two operations.

To help the user align the gripper with the middle line of the sphere, a haptic force corresponding to a spring of stiffness k between the position of the gripper and the sphere is provided,

$$F_{hy} = -k\delta_y, \quad (5)$$

where δ_y is the distance between the centre of the gripper and the centre of sphere along the y -axis (see Fig. 4).

A haptic feedback F_{hoc} is provided so that the user can monitor the grasping force. Contrary to what is commonly proposed in the literature, we are not here interested in computing the exact efforts applied to the object, but only in deriving information to assist the user while performing a given task. The calibration process which enables relating the tool deformations to the applied force is thus unnecessary. While closing the gripper, the user has to counteract a haptic force F_{hoc} ,

$$F_{hoc} = \begin{cases} F_{max} e^{-\frac{\delta_x^f}{\alpha}} & \text{if not in the contact zone} \\ F_{contact} & \text{if in the contact zone.} \end{cases} \quad (6)$$

Here, δ_x^f is the free space between the two fingers of the gripper. If the sphere is situated between the fingers, then $\delta_x^f = \delta_{xl} + \delta_{xr}$, which corresponds to the sum of the distances between each of the fingers and the sphere; otherwise, $\delta_x^f = \delta_x$. F_{max} is the maximum force that can be transmitted to the user when the gripper is close to the sphere but has not yet entered the contact zone. α is a constant chosen to tune the decrease of the haptic force as the distance between the two fingers increases. $F_{contact}$ is the force sent while the gripper is grasping the sphere. The step between F_{max} and $F_{contact}$ must be high enough to indicate clearly the contact between the sphere and the gripper. The contact zone is reached if δ_{xl} and δ_{xr} are less than a given distance $\delta_1 = 3 \mu\text{m}$ (which corresponds to 6% of the sphere diameter). The gripper will then reach the non-contact zone if δ_{xl} and δ_{xr} are greater than $\delta_2 = 7.5 \mu\text{m}$. This hysteresis avoids undesirable frequent transitions between contact and non-contact modes because of noise or tracking error. The hysteresis values δ_1 and δ_2 were chosen, based on our experience, for a comfortable user sensation. The force step that is sent to the user when contact is detected is filtered to avoid large and sudden force changes. Even if the user does not receive the maximum force feedback at the instant of contact, the increase in the force can be distinctly felt, and so it can be inferred that contact has happened.

V. EXPERIMENTAL RESULTS

A. The influence of the sampling rate on stability

To visualize the influence of the sampling rate on the stability of the haptic feedback, an experiment consisting of grasping a sphere is performed for different resampled frequencies from the vision algorithm output. The estimated

distances are transferred to the haptic thread with resampled dynamics manually set to 10 ms and 100 ms for comparison with the unfiltered output of the EICP. For each frequency, the object is grasped and released three times (without being lifted). The results are given in Fig. 6. It can be seen that as the frequency decreases, the effort is less smooth. This is very disturbing for the user, who has to counteract this variation. The user attenuates the oscillations by grasping firmly the haptic handle, which is equivalent to adding damping to the system. The plots show the attenuated oscillations. Fig. 6 does not represent an unstable response from a control point of view, since the user manages to limit the oscillations. However, the haptic feedback is highly uncomfortable and the user has to concentrate to counteract it. For low frequencies, the system's performance decreases, which makes it unsuitable for complex 3D manipulation.

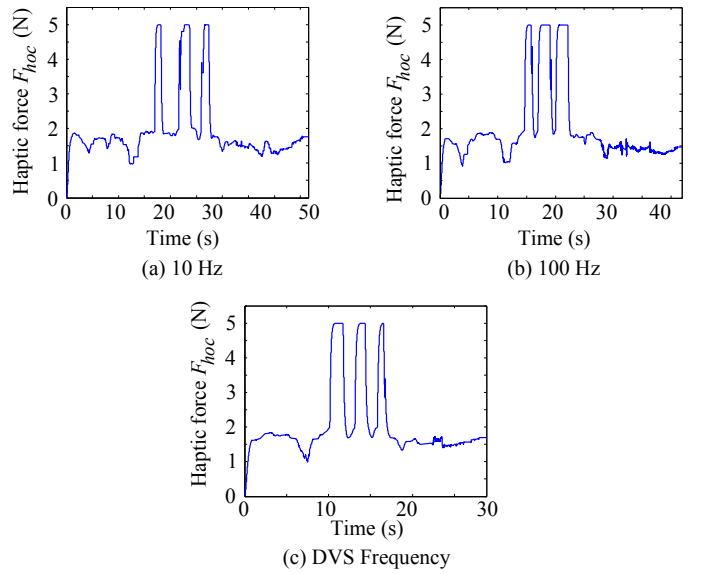


Fig. 6. Haptic forces during the grasping operation for different frequencies of the vision feedback. The haptic force F_{hoc} that helps monitor the grasping force is provided. Note that for each experiment, the user grasped the sphere three times, and released it. As the frequency decreases, oscillations appear.

B. Pick-and-place of microspheres

Some experiments were performed to validate the use of event-based vision in computing haptic feedback. The microspheres are glass beads of about $50\mu\text{m}$ in diameter from Polysciences, Inc.⁴. To avoid issues of sticking while releasing the spheres, a Gel-Pack substrate was selected. When the gripper lets go of the sphere, the Gel-Pack substrate provides enough adhesion to prevent the sphere from sticking to the gripper.

The experiment consists of positioning the gripper with respect to the sphere (in plane displacements), grasping it, picking it up, moving it, putting it down, and, finally, releasing it. The precision of gripper tracking during this process is depicted in Fig. 7, where the ratio of the mean ICP tracking error to the microsphere diameter is calculated. It can be seen

⁴<http://www.polysciences.com/>

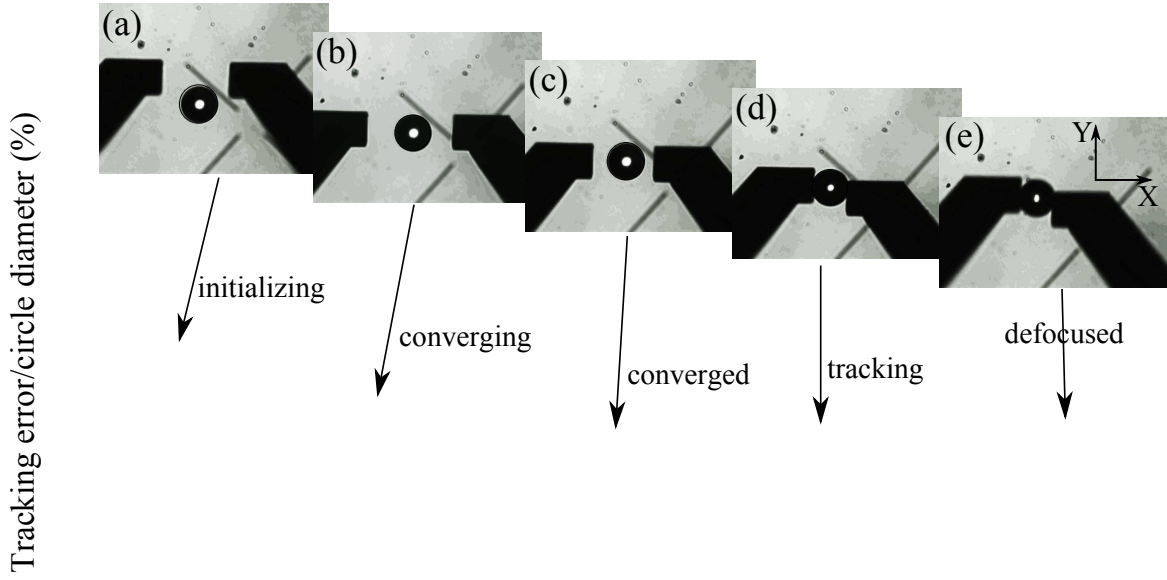


Fig. 7. Gripper position-tracking error. Images represent different steps. a–b): the gripper converges to the gripper contour from its initial position; c–d): the gripper closes, and its position is tracked; e): the sphere is lifted and the gripper starts to defocus. This induces a smoothing of contours that then lowers the tracking precision.

that when the gripper closes, the tracking error converges to a constant value, which corresponds to 7% of the diameter of the sphere. The ‘picking up’ operation may cause a z-axis defocusing, so the error slightly increases. The ‘putting down’ operation produces similar results. For the sake of clarity, it has been omitted from the figure.

To assist the user in aligning the gripper with the sphere, haptic feedback is provided for both the grasping and the releasing tasks. The results are given in Fig. 8 for the grasping. They are similar for the releasing stage, which is hence omitted. The user controls the position of the gripper along the y-axis. A haptic force F_{hy} that corresponds to a virtual stiffness between the centre of the gripper’s fingers and the centre of the sphere is transmitted to the operator. At the beginning of the experiment, the gripper is misaligned, and the user feels an attractive force that pulls the user to the correct position. After 13 s, the gripper is correctly aligned, and the haptic feedback drops to zero.

The evolution of F_{hoc} , the haptic force that helps the user monitor the grasping force, is given in Fig. 9 for both the grasping and releasing stages. When $t_1 < 11.8$ s, the user closes the gripper on the sphere. As the free space between the gripper’s fingers and the sphere decreases, the operator has to counteract an increasing haptic force F_{hoc} . At $t_1 = 11.8$ s, the gripper enters the contact zone, and the user feels a large and sudden increase of the haptic force. The sphere is grasped. Due to an initial misalignment of the finger tips (along the vertical z-axis), this grasping causes a rotation of the sphere between the two fingers and amplifies the misalignment. Thus, the right finger goes down, and left finger goes up. Since the

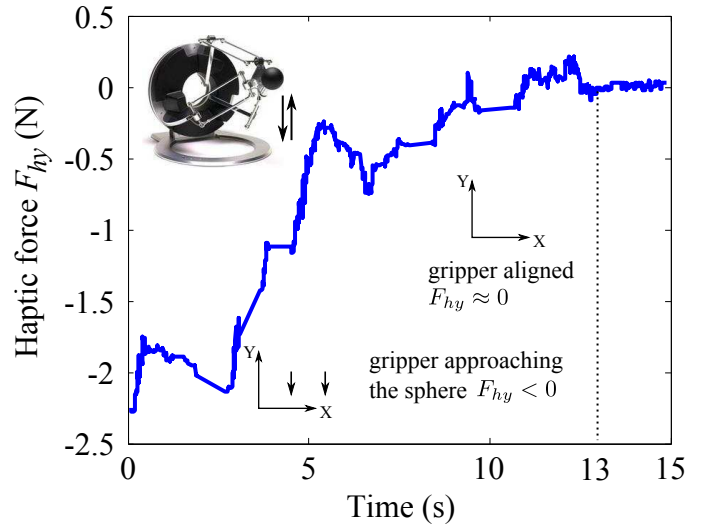


Fig. 8. The haptic force F_{hy} that assists the user to align the gripper with the middle line of the sphere during the grasping operation. Eq. (5) is used with the following parameters: $k = 50000 \text{ N} \cdot \text{m}^{-1}$ (a misalignment of $100 \mu\text{m}$ produces the maximum force admissible by the haptic interface, 5 N). The displacement scaling factor along the y-axis is set to $\alpha_d^y = 2.5 \times 10^3$.

gripper is mounted with an angle of about 45° with respect to the horizontal plane (see Fig. 1), a shift between the finger tips along the y-axis can be observed by the projection on the top view in the inset of Fig. 9 (or more clearly in the insets of Figs. 7d) or 7e)). At $t_2 = 15.6$ s, the operator begins the pick-and-place operation. To avoid any disturbance during this operation, F_{hoc} is set to zero. The user starts to release

the sphere at time $t_3 = 43.5$ s. As the gripper contacts the sphere, a constant haptic force is felt (equal to 5 N), which helps the user open the gripper. At $t_4 = 45.3$ s, the gripper is opened, and the fingers are out of the contact zone. The haptic force drops suddenly. It can be noted that it does not reach 0, as the force is still assisting the user to open the gripper (and avoid unexpected closing). For the entirety of each of

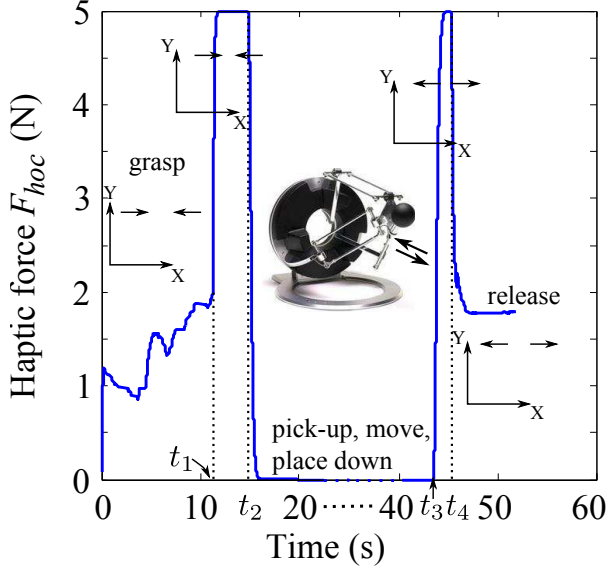


Fig. 9. Haptic force F_{hoc} used to control the grasping force. It is computed using Eq. (5) with the following coefficients: $F_{contact} = 5$ N (set to the maximum admissible force of the haptic interface), $F_{max} = 2$ N, $\alpha = 1.44 \cdot 10^{-8}$ (a distance of 100 μm produces a haptic force of 1 N). The opening/closing scaling factor is set to $\alpha_{oc} = 1.8 \times 10^{-3} \text{ m} \cdot \text{V}^{-1}$.

the grasping and releasing operations, the user receives haptic feedback which helps in the performance of the task.

During the lifting and the putting operations, a constant repulsive haptic force field, set to 2 N, is provided to avoid any involuntary contact with the substrate (Fig. 10). When the sphere has been lifted above the substrate to the desired height, the user can move it freely in the (x, y) plane parallel to the substrate ($\alpha_d^x = 4.0 \times 10^3$, $\alpha_d^y = 2.5 \times 10^3$)⁵. For this operation, the haptic feedback is turned off.

A high frequency capability for the sensor is of the utmost importance for real-time applications at the microscale, since the low inertia of the objects induces rapid dynamics. In addition, due to the scale difference, position and/or force scaling factors are necessary to decrease the movements performed by the user so that they can be used as an input for the micromanipulation system and to enhance the force so that it can be used as a haptic feedback (if the interaction force between the gripper and the sphere is measured). To ensure stability, the ratio of these scaling factors multiplied by the square of the sampling rate of the system and the stiffness of the contact must be less than the inertia of the haptic device (see [30]). Ensuring stability is possible either by adapting the scaling factors or by increasing the sampling

⁵Different factors are used along the three axes of the micromanipulator to achieve easy positioning. They are set according to the user's comfort of manipulation.

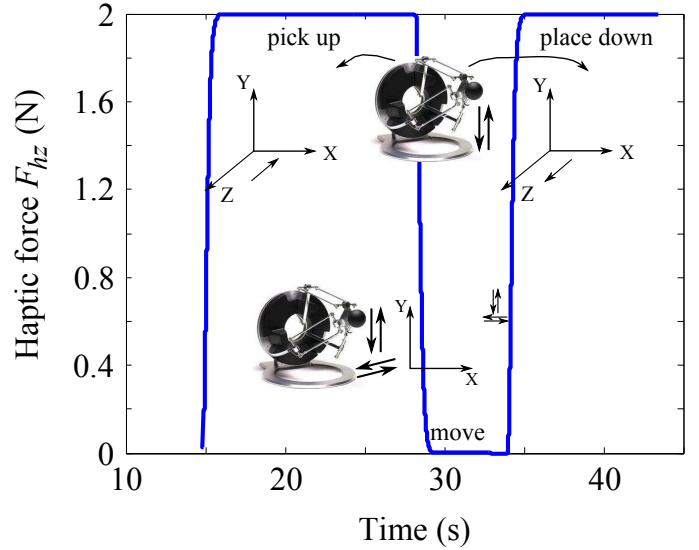


Fig. 10. The vertical haptic force F_{hz} sent to users during lifting, moving, and putting down operations. To perform the vertical displacements, a scaling factor $\alpha_d^z = 12.5 \times 10^3$ is used.

rate. Adapting the scaling factors leads to time consuming experiments or very weak haptic feedback. In practice, this renders the system uncomfortable and difficult to use. A sensing capability higher than 1 kHz is commonly recommended [24]. Getting high frequency sensing is the only solution for providing usable haptic feedback systems. In this paper, the vision sensors and detection algorithms provided a high-frequency feedback that enabled users to successfully perform a 3D tele-operated manipulation on micrometer-sized objects. The results in this paper will surely benefit tele-operated or automated microassembly and open new perspectives for complex micromanipulation.

VI. CONCLUSION

To overcome the lack of sensing capabilities at the microscale, a vision-based system has been proposed. To enable a wide range of applications, in particular for automated or tele-operated micromanipulations, the frequency of the vision feedback must be higher than 1 kHz, as the low inertia present at this scale induces highly dynamic phenomena. This is ensured by the output of the DVS sensor, which conveys temporal contrast in the scene in the form of time-stamped events. An Event-based Iterative Closest Point algorithm (EICP) has been proposed, to track the tool at more than 4 kHz. This feedback is combined with the output of a classical frame-based camera, used to derive information about static parts of the scene, and in particular the position of the object that must be manipulated. This approach was tested with a pick-and-place experiment of a glass sphere with a diameter about 50 μm using a piezo-electric gripper. The task was realized by tele-operation with haptic feedback. This application is especially challenging as a frequency of more than 1 kHz is required for the system's stability. The influence of the frequency rate on the system's stability has been experimentally highlighted, and the benefits of the DVS sensor over conventional frame-based

cameras with lower frequencies was shown. A successful pick-and-place task of micrometer-sized objects with 3D haptic feedback based on vision tracking was performed with this system.

These results can be easily extended to other applications, involving different objects or tools. In particular, vision-based force measurement could be performed with the DVS sensor after the calibration of the tool. Fully automated manipulation also would benefit from the high frequency of the feedback to guarantee the system's stability. Future research projects include the use of a model of the gripper to avoid tracking drift from the loss of focus while performing out-of-plane movements.

VII. ACKNOWLEDGMENT

This work was supported by the French Agence Nationale de la Recherche, through the NANOROBUST project (contract ANR 2011 NANO 006).

REFERENCES

- [1] V. Sariola, M. Jääskeläinen, and Q. Zhou, "Hybrid microassembly combining robotics and water droplet self-alignment," *IEEE Transactions on Robotics*, vol. 26, no. 6, pp. 965–977, 2010.
- [2] N. Chaillet and S. Régnier, *Microrobotics for Micromanipulation*. J. Wiley & Sons, 2010.
- [3] S. Fahlbusch and S. Fatikow, "Force sensing in microrobotic systems—an overview," in *International Conference on Electronics, Circuits and Systems*, vol. 3, 1998, pp. 259–262 vol.3.
- [4] F. Beyeler, S. Muntwyler, and B. J. Nelson, "A six-axis MEMS force-torque sensor with micro-Newton and nano-Newtonmeter resolution," *Journal of Microelectromechanical Systems*, vol. 18, no. 2, pp. 433–441, Apr. 2009.
- [5] T. C. Duc, G. Lau, J. F. Creemer, and P. M. Sarro, "Electrothermal microgripper with large jaw displacement and integrated force sensors," *Journal of Microelectromechanical Systems*, vol. 17, no. 6, pp. 1546–1555, Dec. 2008.
- [6] K. Kim, X. Liu, Y. Zhang, and Y. Sun, "Micronewton force-controlled manipulation of biomaterials using a monolithic MEMS microgripper with two-axis force feedback," in *IEEE International Conference on Robotics and Automation*, May 2008, pp. 3100–3105.
- [7] [Online]. Available: <http://www.femtotools.com>
- [8] B. Lopez-Walle, M. Gauthier, and N. Chaillet, "Principle of a submerged freeze gripper for microassembly," *IEEE Transactions on Robotics*, vol. 24, no. 4, pp. 897–902, 2008.
- [9] K. N. Andersen, D. H. Petersen, K. Carlson, K. Molhave, O. Sardan, A. Horsewell, V. Eichhorn, S. Fatikow, and P. Boggild, "Multimodal electrothermal silicon microgrippers for nanotube manipulation," *IEEE Transactions on Nanotechnology*, vol. 8, no. 1, pp. 76–85, Jan. 2009.
- [10] Y. L. Zhang, M. L. Han, J. Vidyakshmi, C. Y. Shee, and W. T. Ang, "Automatic control of mechanical forces acting on cell biomembranes using a vision-guided microrobotic system in computer microscopy," *Journal of Microscopy*, vol. 236, no. 1, pp. 70–78, 2009.
- [11] Y. Zhang, B. Chen, X. Liu, and Y. Sun, "Autonomous robotic pick-and-place of micro objects," *IEEE Trans. Robotics*, vol. 26, no. 1, pp. 200–207, 2010.
- [12] V. Chawda and M. K. O'Malley, "Vision-Based force sensing for nanomanipulation," *IEEE/ASME Transactions on Mechatronics*, vol. 16, no. 6, pp. 1177 – 1183, 2011.
- [13] A. N. Reddy, N. Maheshwari, D. K. Sahu, and G. K. Ananthasuresh, "Miniature compliant grippers with Vision-Based force sensing," *IEEE Transactions on Robotics*, vol. 26, no. 5, pp. 867–877, Oct. 2010.
- [14] D. Cappelleri, G. Piazza, and V. Kumar, "Two-dimensional, vision-based μ N force sensor for microrobotics," in *IEEE International Conference on Robotics and Automation*, Kobe, May 2009, pp. 1016–1021.
- [15] M. Greminger and B. Nelson, "Vision-based force measurement," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 3, pp. 290–298, Mar. 2004.
- [16] T. Delbruck, "Frame-free dynamic digital vision," in *Proceedings of Intl. Symp. on Secure-Life Electronics, Advanced Electronics for Quality Life and Society*, 2008, pp. 21–26.
- [17] T. Delbruck, B. Linares-Barranco, E. Culurciello, and C. Posch, "Activity-driven, event-based vision sensors," in *IEEE International Symposium on Circuits and Systems*, 2010, pp. 2426 –2429.
- [18] P. Lichtsteiner, C. Posch, and T. Delbruck, "A 128*128 120 db 15us latency asynchronous temporal contrast vision sensor," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 2, pp. 566–576, 2008.
- [19] Z. Ni, C. Pacoret, R. Benosman, S. Ieng, , and S. Régnier, "Asynchronous event based high speed vision for micro-particles tracking," *Journal of microscopy*, vol. 245, pp. 236–244, 2012.
- [20] M. Riesenhuber and T. Poggio, "Models of object recognition," *Nature Neuroscience*, vol. 3, pp. 1199–1204, 2000.
- [21] P. Besl and H. McKay, "A method for registration of 3-d shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239 –256, Feb 1992.
- [22] S. Rusinkiewicz and M. Levoy, "Efficient variants of the ICP algorithm," in *Third International Conference on 3D Digital Imaging and Modeling (3DIM)*, June 2001.
- [23] J. L. Martínez, J. González, J. Morales, A. Mandow, and A. J. García-Cerezo, "Mobile robot motion estimation by 2d scan matching with genetic and iterative closest point algorithms," *Journal of Field Robotics*, vol. 23, pp. 21–34, 2006.
- [24] J. Colgate and G. Schenkel, "Passivity of a class of sampled-data systems: Application to haptic interfaces," *Journal of Robotic Systems*, vol. 14, no. 1, p. 3747, 1997.
- [25] A. Bolopion, H. Xie, S. Haliyo, and S. Régnier, "Haptic teleoperation for 3D microassembly of spherical objects," *IEEE/ASME Transaction on Mechatronics*, vol. 17, no. 1, pp. 116–127, 2012.
- [26] J. Agnus, D. Hériban, M. Gauthier, and V. Pétrini, "Silicon end-effectors for microgripping tasks," *Precision Engineering*, vol. 33, no. 4, pp. 542–548, Oct. 2009.
- [27] [Online]. Available: <http://sourceforge.net/apps/trac/jaer/wiki>
- [28] M. Belshaw and M. Greenspan, "A high speed iterative closest point tracker on an fpga platform," in *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on*, 27 2009-Oct. 4 2009, pp. 1449 –1456.
- [29] R. Hartley and A. Zisserman, *Multiple View Geometry in computer vision*. Cambridge University Press, 2003.
- [30] S. R. A. Bolopion, B. Cagneau and S. Régnier, "Variable gain haptic coupling for molecular simulation," in *IEEE WorldHaptics*, 2011, pp. 469–474.