Asynchronous event-based visual shape tracking for stable haptic feedback in microrobotics.

Zhenjiang Ni, Aude Bolopion, Joël Agnus, Ryad Benosman, Stéphane Régnier

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

Zhenjiang Ni[1], Aude Bolopion[2], Joël Agnus[2], Ryad Benosman[3] and Stéphane Régnier[1]

***Abstract*—Micromanipulation systems have recently received increasing interest. Teleoperated or automated micromanipulation are a challenging issue because of the need of high-frequency position or force feedback to guarantee stability. In addition, the integration of sensors in the micromanipulation platforms is complex. Vision is thus a commonly used solution for sensing; unfortunately the update rate of frame-based acquisition process of current available cameras can not ensure -at reasonable costs- stable automated or teleoperated control at the microscale level where low inertia induces 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 the biological retina enabling high update rates. The paper introduces an Event-based Iterative Closest Point algorithm able 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. Experiments on teleoperating spheres of $30-50$ μm of diameter using a piezoelectric gripper in a pick-and-place task are presented.**

I. Introduction

Versatile 3D manipulation systems that can operate in ambient conditions on micron-sized objects would greatly increase the potential applications of microtechnology [1]. However the development of such systems must face a major issue: the lack of position and force feedback [2]. Sensors have been developed [3], [4], but their integration into the dedicated tools induces an important increase in the complexity and the cost of their fabrication. In particular, even if some microgrippers offer sensing capabilities at the expense of complex designs [5], [6] most of them still lack force measurements [7], [8].

Vision is a promising solution to avoid the complexity of sensors' integration. Visual information can be converted to force measurement to monitor efforts applied on objects during manipulation [9], [10]. This is achieved by using deformable tools after a calibration step [11]. In particular, the stiffness of the tools must be determined to relate the measured deformations to the applied forces. However, the precise value of the force might not be necessary to control micromanipulation systems, in either teleoperated or automated mode. In these cases, position feedback obtained from vision sensors might as well be sufficient. In all cases, high 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 [12]. Thus they can not ensure a proper monitoring of high dynamic motion.

Conventional frame-based cameras' lack of dynamic information and their redundancies set an important limit to potential micromanipulations in automated or teleoperated modes. Event-based computer vision based on *Address Event Representation* (AER) provides a sound solution to high-speed vision problems [13]. This newly developed discipline is motivated by mimicking biological visual systems [14]. The Dynamic Vision Sensor (DVS) silicon retina used in this work reacts to changes of contrast that are converted in a stream of asynchronous time-stamped events [15]. 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 to benefit fully from the properties of this new framework [16]. The neural shape coding is a difficult issue as there is almost an infinite number of shapes' representation in the real world. A computationally efficient method is the HMax; it models the biological visual system to extract features of different forms for object recognition [17]. 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 to fulfill the high-frequency requirements needed in micromanipulation. The Iterative Closest Point (ICP) is an efficient algorithm dedicated to minimize the difference between a data point set and a model point set [18]. Many variants have been proposed to enhance both performance and adaptability [19]. However, the 2D image processing can only be achieved at several 10 Hz, depending on the number of points to be matched [20]. Although this is sufficient for many robotic applications, these frequencies are far from enough for automated or teleoperated micromanipulation that requires more than 1 kHz refreshing rates to ensure stability of the control loop because of high dynamics of physical phenomena [21].

This paper presents a modified Event-based Iterative Closest Point algorithm (EICP) directly applied on the silicon retina's output. It allows to track the manipulation tool at a frequency of 4 kHz. The positions of static objects are provided by a

[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

conventional frame-based camera. To validate this approach, a teleoperated pick-and-place task of 30 to 50 µm spheres using a piezoelectric gripper is proposed. 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. To ensure stability of such systems, frequencies higher than 1 kHz are commonly required [21]. Stable vision-based teleoperation is achieved in this work by the use of event-based retinas and the EICP algorithm.

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

II. Setup

Several tools have been developed to manipulate micron-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 also have been demonstrated using two protruding tip cantilevers, but the complexity of the setup limits its applications [22]. A microgripper designed at the FEMTO-ST Institute[1] (Fig. 1) is used in this work to perform 3D manipulations that will enable a large range of applications, including microassembly. It is based on a pair of two-degrees-of-freedom piezoelectric beams called duo-bimorph as the actuation principle of the two fingers [23]. On each actuator, four electrodes referred to a central ground and two voltages are necessary to impose the displacements because of the deflections of the piezoelectric beam. This configuration offers numbers of capabilities: not only an open-and-close motion but also an up-and-down motion that allows, for instance, 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 DRIE (Deep Reactive-Ion Etching) process. They are designed to minimize the sticking effects between the end effector and the objects to facilitate the release of the objects. The end-effectors are fixed on the piezoelectric 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. The gripper is mounted on a 3-axis motorized micromanipulator[2] to allow an 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, and 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.

[1]http://www.femto-st.fr/

[2]Sutter Instrument, http://www.sutter.com/MP 285

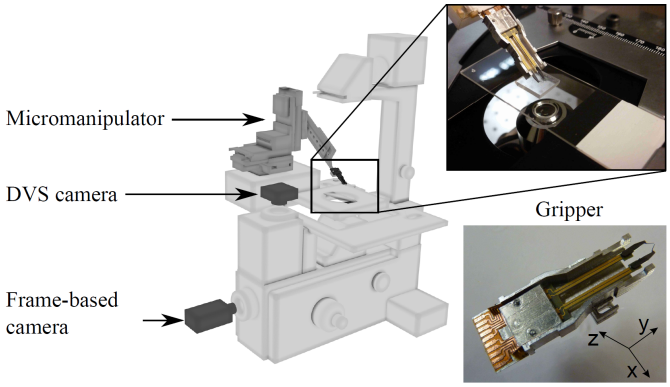

Fig. 1. Setup of the micromanipulation platform

As show in Fig. 1, the observed scene is monitored by two optical sensors, that 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).

The haptic feedback is provided to users by an Omega haptic device[3], with 3 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 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

Biological retinas, unlike frame-based cameras, transmit less-redundant information about a visual scene in an asynchronous manner. The various functionalities of the retina have been incorporated into neuromorphic vision sensors since the late 1980s in the pioneering work of Mahowald [24]. Since then, the most interesting achievements in neuromorphic retinas have been the development of activity-driven sensing. The event-based vision sensor output compressed digital data in the form of events, removing redundancy, reducing latency, and increasing dynamic range compared with conventional imagers. A complete review of the history and the existing sensors can be found in [14].

The Dynamic Vision Sensor (DVS) used in this work is an Address-Event Representation (AER) silicon retina with 128×128 pixels [15]. The DVS output consists of asynchronous address-events that signal scene reflectance changes at the times they occur. Each pixel is independent and detects changes in log intensity larger than a threshold since the last event it emitted (typically 15% contrast). As shown in Fig. 2(a), when the change in log intensity exceeds a set threshold, a $+1$ or -1 event is generated by the pixel depending on whether the log intensity increased or decreased.

[3]Force Dimension, http://www.forcedimension.com

Because the DVS is not clocked like conventional cameras, the timing of events can be conveyed with a very accurate temporal resolution of approximately 1 µs. Thus the "effective frame rate" is typically several kilohertz. The retina pixels also implement the local gain adaptation mechanism, which allows them to work over scene illuminations that range from 2 lux to more than 100×10^3 lux. When events are transmitted off-chip, they are time-stamped using off-chip digital components and then transmitted to a computer using a standard USB connection. The advantages of the sensor over conventional clocked cameras are that only moving objects produce data, thus reducing the load of post-processing.

The stream of events from the retina can be defined mathematically as follows: let $ev(\mathbf{p},t)$ be an event occurring at time t at the spatial location $\mathbf{p} = (x,y)^T$. The values $ev(\mathbf{p},t)$ are set to be -1 or $+1$ when a negative or a positive change of contrast is detected respectively. The absence of events when no change of contrast is detected implies that redundant visual information usually recorded in frames is not carried in the stream of events. Fig. 2(b) shows an example of the spatio-temporal visualization of a set of DVS events in response to the microgripper closing on a microsphere. An event $e(\mathbf{p},t)$ 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 a precise time t as:

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

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

B. 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). At each step, a rigid transformation between the known model and the data is estimated expressing 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 the positions of 2D model's points defining the shape of the gripper at time t. $M_{ev}(t)$ is the set of pixellic locations of active pixels of the silicon retina at time t defined as

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

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 current position and all elements of $G(t)$.

We can then define the matching function as:

$$\begin{aligned} match : M_{ev}(t) &\rightarrow G(t) \\ \mathbf{p} \rightarrow \mathbf{p}_k,\ k \mid \min_{k \in \{1,\ldots,N_G\}} & d(\mathbf{p},\mathbf{p_k}), \end{aligned} \quad (3)$$

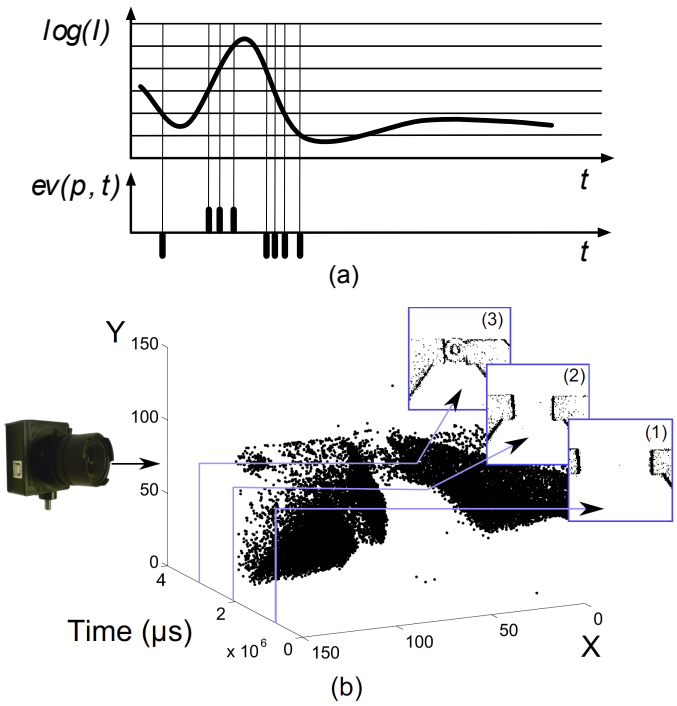

Fig. 2. (a) Principle of events' generation of DVS pixels, adapted from Lichtsteiner et al. [15]. Events with +1 or -1 polarity are emitted when the change in log intensity exceeds a predefined amount of change. (b) Events generated in the (X,Y,t) space when the gripper is closing on a sphere. Images (1)-(3) are shown at chosen temporal locations; they correspond to events' accumulation maps, namely, the projection of all events over a time interval on a single plane (X,Y) regardless of their timings.

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

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

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

Readers interested in the minimization details can refer to [18]. 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,**T**) according to equation (4).
5: Update the position of model points of $G(t)$ using (R,**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 14×10^3 events. The edge width of the gripper in the DVS focal plane is around 3 pixels. When the gripper finger passes, 10.6 events on average are generated on one pixel. Therefore, according to the timestamp, one pixel remains active during 2.46 ms. The decay function permits a pixel's activity to be considered

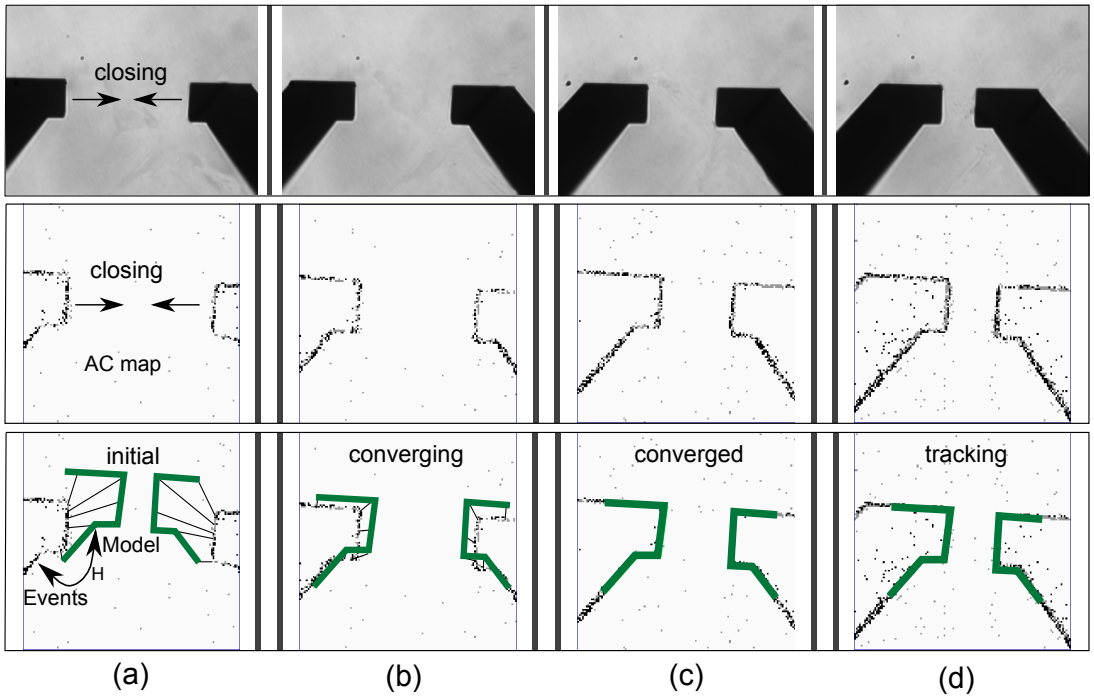

Fig. 3. The principle of event-based iterative closest point tracking (EICP). The first row is a sequence of conventional images showing the closing of the gripper. The middle row provides events' accumulation maps. The last row provides the convergence of the EICP model to the gripper edges. The four 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) for visualizing, 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.

during a certain period after the last event, which is tuned about $10-15$ ms. The EICP is even driven, and its update rate has a mean value of 4 kHz. The algorithm is implemented in Java under JAER open-source project [25].

Frame-based camera serves as a complement to DVS silicon retina solely for the 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 [26]. The homography is estimated off-line by extracting from both sensors' focal plane the coordinates of six corner points of the gripper fingers in both sensors 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 conventional camera output. Once detected, its location is converted into the event-based retina's focal plane coordinate systems. This operation provides both the gripper and the sphere locations in the same coordinate system. It then becomes possible to estimate the distance δ_x between the gripper's fingers and, 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 center of the sphere and the gripper δ_y (Fig. 4). These various distances will be used to compute the haptic feedback.

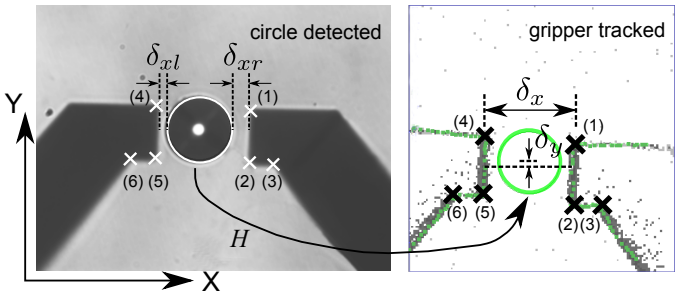

Fig. 4. The calibration between the classical image (left) and the DVS accumulation map (right). Six points (crosses (1)-(6)) have been chosen to calculate the homography transform H. The detected circle (left) is transferred by UDP socket so that the DVS part has both gripper and circle position available (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 center of the fingers and the center of the sphere.

IV. Haptic Coupling

A. Manipulation modes

The coupling between the haptic device and the microgripper is represented in Fig. 5. The operator applies a force $\mathbf{F_{op}}$ on 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 positioning

and the actuation of the gripper.

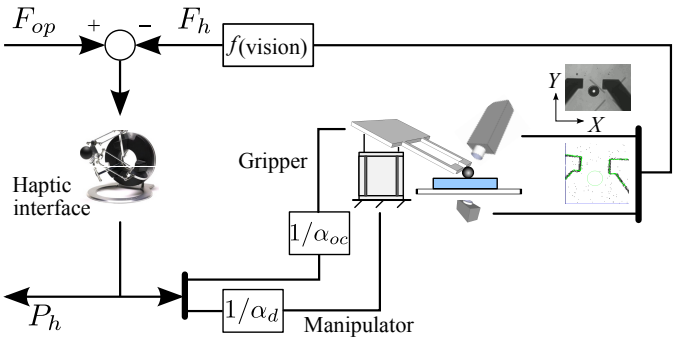

Fig. 5. Haptic coupling scheme. The user controls the position of the gripper and the actuation of the gripper's fingers using the haptic interface. He/she receives haptic feedback through the device. The haptic force is based on the distance between the gripper and the sphere, determined from vision algorithms. Scaling factors α_d and α_{oc} convert the position of the haptic handle to variables used to control the positioning 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: $\mathbf{F_h}^T = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$,
- 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: $\mathbf{F_h}^T = \begin{bmatrix} 0 & 0 & F_{hz} \end{bmatrix}$,
- 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 (detailed 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 brutal changes in force feedback during transitions between different modes, the haptic force is filtered by a second-order low-pass filter during 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 but controlled to avoid any damages to the object. The haptic feedback must assist the user for these two operations.

To help the user align the gripper with respect to 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:

$$F_{hy} = -k\delta_y \tag{5}$$

where δ_y is the distance between the center of the gripper and the center 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 interested here in computing the exact efforts applied on the object but only in deriving information to assist the user while performing a given task. The calibration process, which enables to relate the tool deformations to the applied force, is thus unnecessary. While closing the gripper, the user has to counteract a haptic force F_{hoc}:

$$F_{hoc} = \begin{cases} F_{max} e^{\frac{-\delta_x^{f\,2}}{\alpha}} & \text{if not in the contact zone} \\ F_{contact} & \text{if in the contact zone} \end{cases} \tag{6}$$

where δ_x^f is the free space between the two gripper's fingers. If the sphere is situated between the fingers $\delta_x^f = \delta_{xl} + \delta_{xr}$, 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 closed on the sphere but has not entered the contact zone yet. 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 $\delta_1 = 3$ µm (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 $\delta_2 = 7.5$ µm. This hysteresis avoids undesirable frequent transitions between contact and non-contact modes because of noises and tracking errors. The hysteresis values δ_1 and δ_2 are chosen according to our experiment to achieve a comfortable user sensation. The force step that is sent to the user when contact is detected is filtered to avoid brutal force changes. Even if the user does not receive the maximum force feedback at the instant of contact, he/she can distinctly feel the increase in the force, and infers that contact happened.

V. Experimental results

A. Influence of the sampling rate

To visualize the influence of the sampling rate on the stability of the haptic feedback, an experiment consisting in grasping the 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 native output of EICP result. For each frequency, the object is grasped and released three times (without being lifted). Results are given in Fig. 6. It can be seen that as the frequency decreases, oscillations appear. In addition, it is harder for the user to control the system since he/she has to

dissipate the excess of energy of the unstable system. For low frequencies the system's performances decrease, which makes it unsuitable for complex 3D manipulation.

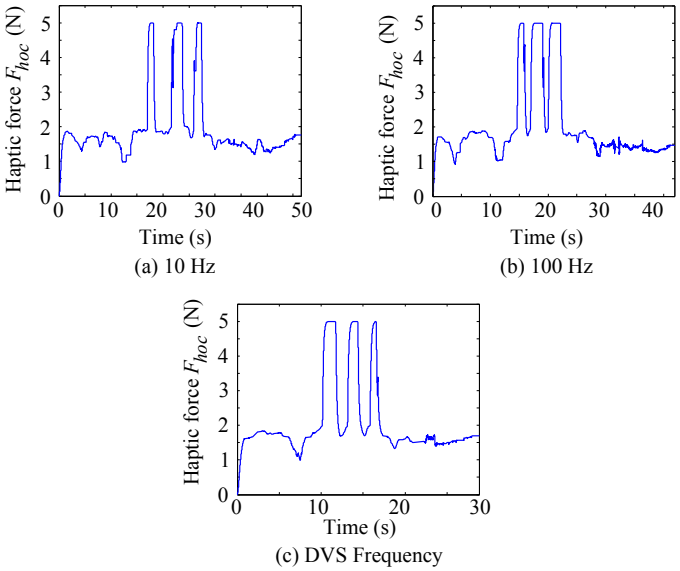

Fig. 6. Haptic forces during the grasping operation for different frequencies of the vision feedback. The haptic force F_{hoc} that helps monitor the grasping force is provided. Note that for each experiment, the user grasped the sphere 3 times, and released it. As the frequency decreases, oscillations appear.

B. Pick and place of microspheres

To validate the use of event-based vision to compute haptic feedback, experiments are performed. The microspheres are glass beads of 30 to 50 micrometers of diameter from Polysciences, Inc[4]. To avoid sticking issues while releasing the spheres, a Gel-Pack substrate has been selected. While taking off the sphere, the gripper enables to counteract adhesive forces. While releasing it, the Gel-Pack substrate provides enough adhesion to prevent the sphere from sticking to the gripper.

The experiment consists in positioning the gripper with respect to the sphere (in plane displacements), grasping it, taking it off, moving it, placing it down, and finally releasing it. The precision of gripper tracking during this process is depicted in Fig. 7, where the proportion of the mean ICP tracking error to the microsphere diameter is calculated. It can be seen that when the gripper closes, the tracking error converges to a constant value that corresponds to 7% of the sphere diameter. The "taking off" operation may cause a z-axis defocalization, so the error slightly increases. The "placing down" operation produces similar results. For the sake of clarity, it has been omitted in the figure.

To assist the user while aligning the gripper with respect to 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 release stage that is then omitted. Users control the position of the gripper along the y-axis. A haptic force F_{hy} that corresponds to a virtual stiffness

[4]http://www.polysciences.com/

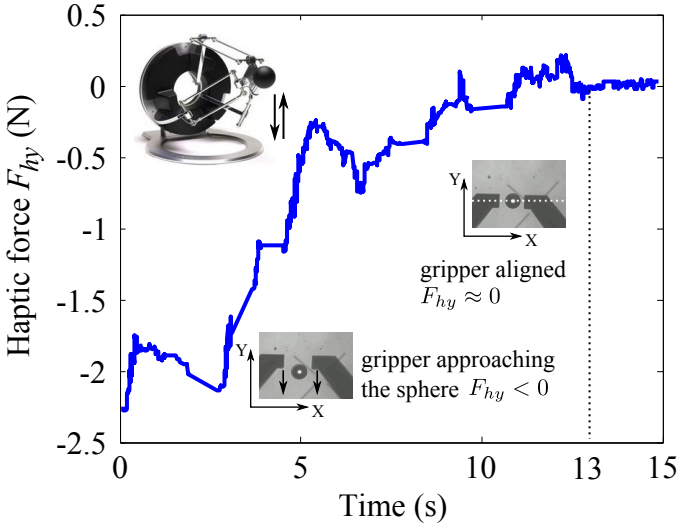

Fig. 8. The haptic force F_{hy} that assists the user to align the gripper with respect to the middle line of the sphere during grasping operation. Equation 5 is used with the following parameters: $k = 50000$ N·m^{-1} (a misalignment of 100 µm produces the maximum force admissible by the haptic interface,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

between the center of the gripper's fingers and the center of the sphere is transmitted to the operators. At the beginning of the experiment, the gripper is misaligned, and the user feels an attractive force that pulls him/her 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 users to monitor the grasping force, is given in Fig. 9 for both grasping and releasing stages. For time $t_1 < 11.8$ 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 in the contact zone, and the user feels a brutal increase on the haptic force. 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 $t_3 = 45.7$ s. As the gripper contacts the sphere he/she can feel a constant haptic force (equal to 5 N) that helps him open the gripper. At $t_4 = 47.6$ 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). During all the grasping and releasing operation, the user receives haptic feedbacks that help him perform the task.

During the lifting and the placing 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 at a 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.

[5]Different 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.

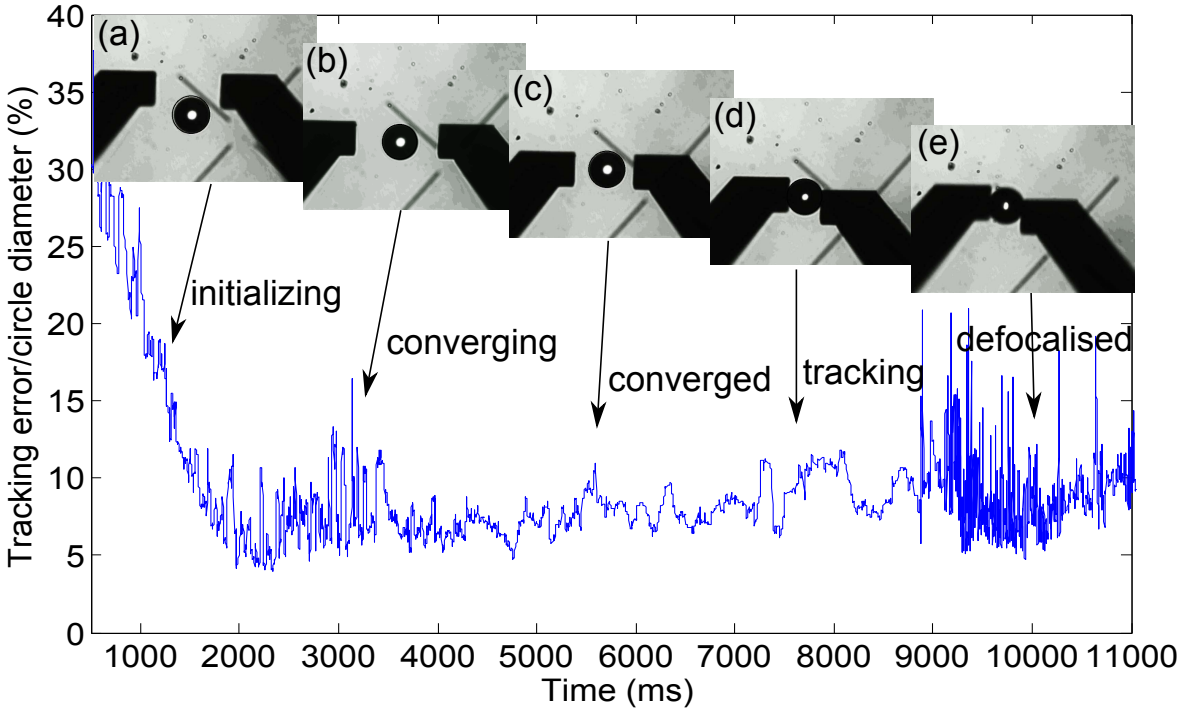

Fig. 7. Gripper position-tracking error. Images represent different steps. a-b): the gripper is converging to the gripper contour from its initial position; c-d): the gripper closes, and its position is tracked; e): the sphere is lifted and the gripper starts to defocalize. This induces a smoothing of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

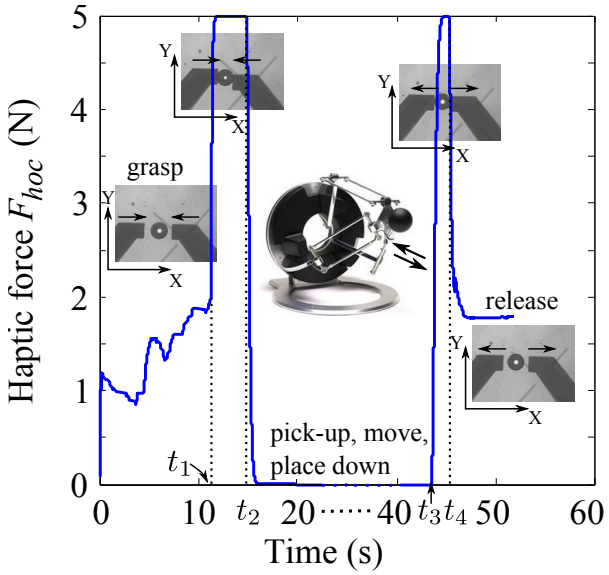

Fig. 9. Haptic force F_{hoc} used to control the grasping force. It is computed using Eq. 5 with the following coefficients: $F_{contact} = 5$ N (set to the maximum admissible force of the haptic interface), $F_{max} = 2$ N, $\alpha = 1.44 \cdot 10^{-8}$ (a distance of 100 µm produces a haptic force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

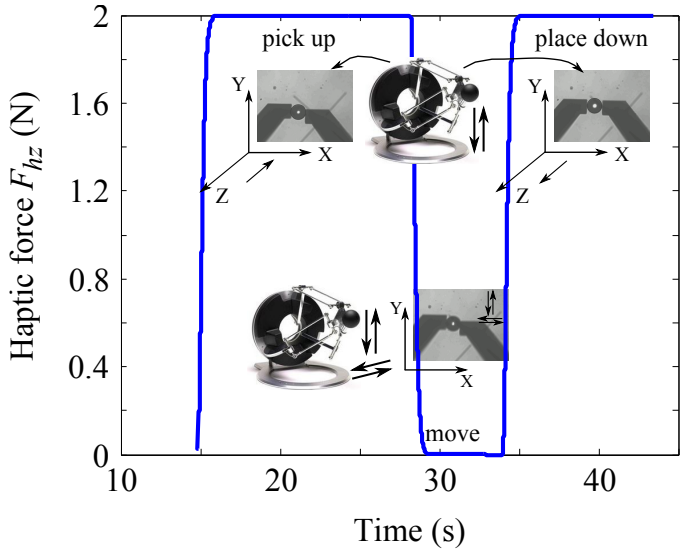

Fig. 10. The vertical haptic force F_{hz} sent to users during lifting, moving and placing down operations. To perform the vertical displacements a scaling factor $\alpha_d^z = 12.5 \times 10^3$ is used.

The vision sensors and detection algorithms provide a high-frequency feedback that enables the derivation of a stable haptic system. Users successfully performed a 3D teleoperated manipulation on micron-sized objects. This work will surely benefit teleoperated or automated microassembly and opens new perspectives for complex micromanipulation.

VI. Conclusion

To overcome the lack of sensing capabilities at microscale, a vision-based system is proposed. To enable a wide range of applications, in particular for automated or teleoperated micromanipulations, the frequency of the vision feedback must be higher than 1 kHz as low inertia at this scale induces high 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) is 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 is tested on a pick-and-place experiment of glass spheres with a diameter between 30 and 50 µm using a piezoelectric gripper. The task is realized by teleoperation 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 are shown. A successful pick-and-place task of micron-sized objects with 3D haptic feedback based on vision tracking is performed with this system.

This work 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 a calibration step of the tool. Fully automated manipulation also would benefit from the high frequency of the feedback to guarantee the system's stability. Future works include the use of a model of the gripper to avoid the tracking drift because of the loss of focus while doing out-of-plane movements.

VII. Acknowledgment

This work was supported by the French National Agency of Research, through the NANOROL and the NANOROBUST projects.

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] 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.

[8] 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.

[9] 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.

[10] 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.

[11] D. Cappelleri, G. Piazza, and V. Kumar, "Two-dimensional, vision-based µN force sensor for microrobotics," in *IEEE International Conference on Robotics and Automation*, Kobe, May 2009, pp. 1016–1021.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] Z. Ni, C. Pacoret, R. Benosman, S. Ieng, , and S. Régnier, "Asynchronous event based high speed vision for micro-particles tracking," *Journal of microscopy*, in press, 2011.

[17] M. Riesenhuber and T. Poggio, "Models of object recognition," *Nature Neuroscience*, vol. 3, pp. 1199–1204, 2000.

[18] 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.

[19] S. Rusinkiewicz and M. Levoy, "Efficient variants of the ICP algorithm," in *Third International Conference on 3D Digital Imaging and Modeling (3DIM)*, June 2001.

[20] 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.

[21] 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.

[22] A. Bolopion, H. Xie, D. S. Haliyo, and S. Régnier, "Haptic teleoperation for 3-d microassembly of spherical objects," *IEEE/ASME Transactions on Mechatronics*, 2010, available on line.

[23] 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.

[24] M. Mahowald, "Vlsi analogs of neuronal visual processing: a synthesis of form and function," Ph.D. dissertation, California Institute of Technology, Pasadena, CA, 1992.

[25] [Online]. Available: http://sourceforge.net/apps/trac/jaer/wiki

[26] R. Hartley and A. Zisserman, *Multiple View Geometry in computer vision*. Cambridge University Press, 2003.