A Direct Visual Servoing Scheme for Automatic Nanopositioning

Brahim Tamadazte, Nadine Le-Fort Piat, Eric Marchand

Abstract—This paper demonstrates an accurate nanopositioning scheme based on a direct visual servoing process. This technique uses only the pure image signal (photometric information) to design the visual servoing control law. With respect to traditional visual servoing approaches that use geometric visual features (points, lines ...), the visual features used in the control law is the pixel intensity. The proposed approach has been tested in term of accuracy and robustness in several experimental conditions. The obtained results have demonstrated a good behavior of the control law and very good positioning accuracy. The obtained accuracies are 89 nm, 14 nm, and 0.001 degrees in the x, y and θ axes of a positioning platform, respectively.

Index Terms—Microrobotics, Visual Servoing, Micromanipulation

I. OVERVIEW

THE heterogeneous integration of high-performance electronic devices, microelectromechanical structures (MEMS), and optoelectronic devices onto the same substrate is important for the development of low-cost, high performance, and compact microsystems [1]. To set up compact systems, it is necessary to be able to handle and assemble the various elements that constitute these microsystems. These operations (i.e. handling and assembling) must be achieved with high accuracy. In the last decade, considerable research has been performed, such as: development of robotic microassembly station [2], gripping systems [3], high accuracy actuators [4]–[6], micromanipulation and microassembly strategies [7]–[9]. Many works have dealt with the development of control approaches to automate the different micromanipulation and microassembly tasks such positioning, orientation, picking, placing and insertion of different micro-objects [9]–[11]. Most of these works use a vision system to control the behavior of the robotic structure of the microassembly station during the assembly process. The control of robots through real-time and continuous visual feedback, is generally, known as visual servoing [12] [13], and the continuous observation of objects of interest is referred to as visual tracking [14] [15]. Visual

Brahim Tamadazte, and Nadine Le-Fort Piat are with the FEMTO-ST Institute, UMR CNRS 6174-UFC/ENSMM/UTBM. Automatic Control and Micro-Mechatronic Systems Department (AS2M). 24 rue Alain Savary, 25000 Besançon, France. Corresponding e-mail : brahim.tamadazte@ens2m.fr. Phone: +33 (0)3 81 40 27 99, Fax: (33) 03 81 40 28 09, http://www.femto-st.fr. Eric Marchand is with Université de Rennes 1, IRISA, INRIA Rennes-Bretagne Atlantique, Rennes 35042, France

tracking of an object usually involves the detection of some known object features in the acquired images and, using these features, to estimate of the object position and orientation. A feature can be a distinctive part of the object and can exist naturally as a part of the geometry (i.e. a corner, or an edge), or as a deliberately fabricated marking on the surface (i.e. markers). This tracking process is one of the bottlenecks the development of visual servoing techniques.

That is why some works tend to alleviate this problem. An interesting way to avoid any tracking process is to use non geometric visual measurements as in [16] instead of geometric measurements as it is usually done. Of course, directly using non geometric visual features also avoids any tracking process. In that case, parameters of a 2D motion model have been used in [17]–[19]. Nevertheless, such approaches require a complex image processing step.

Recently, new visual servoing technique has been introduced to overcome these issues. In [20], it has been shown that this tracking process can be totally removed and show that no other information than the image intensity (the pure luminance signal) can be considered to control the robot motions. Indeed, to achieve this goal, it is possible to use as visual measurement and as visual feature the simplest that can be considered: the image intensity itself. This approach is referred as *photometric visual servoing* [20]. Other similar approaches use the image gradient [21] or the image entropy [22]. Considering that the control of robot is directly linked to the variation of the pixels intensity in the image, no geometric information has to be extracted and, consequently, tracked or matched over time. This avoid the development of complex visual tracking algorithms such as in [9] [14]. Let us also advocate that avoiding this feature extraction process allows to be naturally robust to error measurements increasing accuracy and repeatability of the positioning tasks. This accuracy is also substantially increased since highly redundant information (all the pixels of the image) is considered to design the control law.

In this paper, we will consider such a direct approach to control the microrobot motions with high accuracy. We consider a photometric visual servoing approach [20] since lighting conditions are constant. Despite the fact that no complex image processing algorithms are considered, we will show that these new techniques are robust to different perturbations such as the occlusion or addition of others micro-objects during the manipulation process.

It is then necessary to propose a control scheme in the joint space and develop a formulation of photometric visual servoing for eye-to-hand systems (corresponding to motionless camera mounted on an optical microscope observing a moving positioning platform). First, the developed meth-

ods are validated using a three Degrees of Freedom (DOF) microassembly workcell in nominal experimental conditions. Second, this development has also been validated in other conditions (i.e. with added external disturbances during the process). Finally, the proposed approach is validated in the case of planar positioning task with the control of optical microscope magnification simultaneously.

Section II presents some definitions about the principles of the eye-to-hand visual servoing. Section III describes the new direct visual servoing approach using only the pixels intensity of the image as visual features, and the design of the new control law. Section IV presents microassembly workcell which consists of a 3 DOF microassembly workcell equipped with a top view optical video microscope, and a semi-controlled environment. Section V discusses the experimental results using the developed method in terms of obtained accuracy quality during the different positioning and orientation tasks, and in different experimental conditions. The microscope magnification will also be controlled using the proposed approach.

II. TRADITIONAL VISUAL SERVOING APPROACHES

Visual servoing techniques consist of using the data provided by one or more cameras in order to control the motion of a robotic system [12] [13]. The camera can be mounted on the robot end-effectors (eye-in-hand configuration) or, alternatively, can be motionless and observes the robot motion (eye-to-hand configuration). Whatever the sensor configuration, the goal is to control the available DOF of the system through the minimization of an error in the image space. This error is defined as a distance between the current values of a set of visual features and their desired values. Usually, these visual features correspond to the positions of geometric features (such as points, lines...) computed from the images.

The desired value of the visual features can be computed or learned by moving the robot to its desired position. This latter approach particularly allows to be robust to camera and robot calibration errors [13] and improves repeatability. If the visual features are correctly chosen, then there is a unique position of the robot that allows to see these features at the desired position in the image.

More precisely, a set of visual features $\mathbf{s}$ has to be designed from the visual measurements $\mathbf{x}(\mathbf{q}(t))$ allowing control of the desired DOF. It has to be noted that visual measurements (and then visual features) are function of the robot joint positions $\mathbf{q}$ at time t. To simplify the notations, we denote $\mathbf{s}(\mathbf{q}) = \mathbf{s}(\mathbf{x}(\mathbf{q}(t)))$. $\mathbf{s}(\mathbf{q})$ is a vector that contains the chosen visual features observed by the camera at time t. The goal of visual servoing is then to design a control law so that these features $\mathbf{s}$ reach a desired value $\mathbf{s}^*$, defining a correct realization of the task.

Vision-based control schemes [12], [13] have been proposed is to regulate to zero the error $\mathbf{e}(\mathbf{q})$ defined by:

$$\mathbf{e}(\mathbf{q}) = \mathbf{s}(\mathbf{q}) - \mathbf{s}^* \tag{1}$$

Let us note that this can also be seen as an optimization problem where the we want to minimize the cost function

$C(\mathbf{q}) = \mathbf{e}(\mathbf{q})^\top \mathbf{e}(\mathbf{q})$ with respect to the robot articular position $\mathbf{q}$.

The equation that links the time derivative $\dot{\mathbf{s}} = \frac{\partial \mathbf{s}}{\partial t}$ of the visual feature $\mathbf{s}$ to the robot joint velocities $\dot{\mathbf{q}}$ is given by:

$$\dot{\mathbf{s}} = \mathbf{J_s} \dot{\mathbf{q}} \tag{2}$$

where $\mathbf{J_s}$ is the visual features Jacobian and $\dot{\mathbf{q}}$ the joint velocities. Using (1) and (2), we obtain the relationship that links the time derivative of the error to the robot joint velocity:

$$\dot{\mathbf{e}} = \mathbf{J_s} \dot{\mathbf{q}}. \tag{3}$$

If we want to try to ensure an exponential decoupled decrease of the error $\mathbf{e}$ (that is $\dot{\mathbf{e}} = -\lambda \mathbf{e}$), the control law is designed as follows:

$$\dot{\mathbf{q}} = -\lambda \mathbf{J_s}^+ \mathbf{e} \tag{4}$$

where λ is the proportional coefficient involved in the exponential convergence of the error, and $\mathbf{J_s}^+$ the pseudo-inverse of $\mathbf{J_s}$.

Dealing with the image Jacobian we have:

$$\dot{\mathbf{s}} = \mathbf{L_s} \mathbf{v} \tag{5}$$

where $\mathbf{L_s}$ represents the interaction matrix that links linearly $\dot{\mathbf{s}}$ and the relative camera instantaneous velocity $\mathbf{v}$ ($\mathbf{v} = (\boldsymbol{v}, \boldsymbol{\omega})$ with $\boldsymbol{v}$ the instantaneous linear velocity and $\boldsymbol{\omega}$ the instantaneous angular velocity) [13].

Since we want to control the robot using the joint velocities, and since we consider an eye-to-hand configuration. We have:

$$\mathbf{J_s} = -\mathbf{L_s}\, {}^c\mathbf{V}_{\mathcal{F}}\, {}^{\mathcal{F}}\mathbf{J}_n(\mathbf{q}) \tag{6}$$

where

- ${}^{\mathcal{F}}\mathbf{J}_n(\mathbf{q})$ is the robot Jacobian expressed in the robot reference frame $R_{\mathcal{F}}$. In our case, the robot Jacobian is computed for the 3 DOF positioning platform (see, Fig. 2) used to validate the proposed approach.
- ${}^c\mathbf{V}_{\mathcal{F}}$ is a spatial motion transform matrix that transforms velocity expressed in the camera frame R_c to the robot reference frame $R_{\mathcal{F}}$.

III. A DIRECT VISUAL SERVOING CONTROL LAW

In traditional visual servoing techniques, the visual features $\mathbf{s}$ are built from image measurements. $\mathbf{s}$ are mainly geometric features such as point or line coordinates, image moments, or even 3D features. These classical approaches require detection, tracking or matching steps and an estimation in each frame of the value of the visual features. This tracking or matching process has been, to date, considered as a necessary step but is also a non trivial issue.

In this section, we shall consider a direct visual control scheme. The proposed method uses the pure image signal to design the vision-based control law. Rather than extracting the traditional geometrical visual features $\mathbf{s}$ from the images measurements $\mathbf{x}(\mathbf{q}(t))$, we consider all the pixels of the image. In this case, the vector of visual features $\mathbf{s}$ is nothing but a vector $\mathbf{I}$ that contains the intensity of each pixel of the image.

Therefore, the vector $\mathbf{s}$ feature is nothing but the image itself. We can write:

$$\mathbf{s} = \mathbf{I} \tag{7}$$

where $\mathbf{I}$ is a vector representation of the image:

$$\mathbf{I} = (I(\mathbf{x}_{00}), \dots, I(\mathbf{x}_{ij}), \dots I(\mathbf{x}_{MN})) \tag{8}$$

where $I(\mathbf{x}_{ij})$ is the intensity of a pixel at line i (resp. column j) and M (resp. N) is the number of lines (resp. columns) in the image.

As already state, the visual servoing can be formulated as an optimization issue whose goal is to minimize the cost function $C(\mathbf{q}) = \mathbf{e}^\top \mathbf{e}$ with respect to the robot position ($\mathbf{q}$) . In our case this cost function is given by :

$$C(\mathbf{q}) = (\mathbf{I} - \mathbf{I}^*)^\top (\mathbf{I} - \mathbf{I}^*) \tag{9}$$

where $\mathbf{I}^*$ is the desired image. It is clear that, for such approach, the goal (that is $\mathbf{I}^*$) which is nothing but the way (we want to see the object of the scene) has to be learned in an off-line step. Once this goal is defined, for an unknown initial position of the robot, the controller has to regulate the cost function $C(\mathbf{q})$. When $C(\mathbf{q})$ is minimal, the relative position between the object and the robot corresponds to the desired one (see, Fig. 1) which shows the shape of the cost function in the subspace (x, y), given by (9).
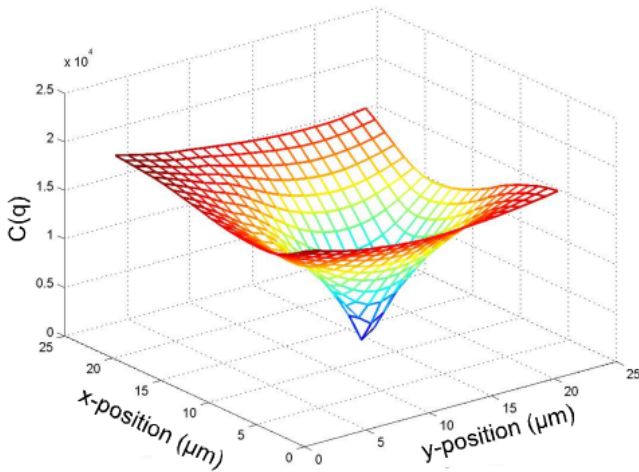

Fig. 1. Representation of the shape of the cost function $C(\mathbf{q})$ in the subspace (x, y).

To design the controller, we have to consider the interaction matrix that links the variation of the image intensity to the camera velocity [20], [23] as defined in (5). This interaction is deduced from the optical flow constraint equation (OFCE) hypothesis [24].

Let us assume that part of an object is at a pixel $\mathbf{x} = (x, y)$ in the image at a time t, and that at $t + dt$ it has moved through at $\mathbf{x} + \mathbf{dx}$. The OFCE states that the intensity of that part of the object $I(\mathbf{x}, t)$ remains the same during a short time interval dt. We have:

$$I(\mathbf{x}, t) = I(\mathbf{x} + \mathbf{dx}, t + dt). \tag{10}$$

A first order Taylor expansion of the equation (10) gives:

$$\frac{\partial I}{\partial x} dx + \frac{\partial I}{\partial y} dy + \frac{\partial I}{\partial t} dt = 0 \tag{11}$$

which can be written as follows:

$$\dot{I} = -\nabla I_x \dot{x} - \nabla I_y \dot{y} \tag{12}$$

with

$$\nabla I_x = \frac{\partial I}{\partial x} \tag{13}$$

and

$$\nabla I_y = \frac{\partial I}{\partial y} \tag{14}$$

Now, the temporal variations of the image must be linked to the camera displacements. For this, we introduce the interaction matrix of a point of the image which links the point velocity in the image to the camera velocity [13]. It is given by:

$$\dot{x} = \mathbf{L}_x \mathbf{v} \tag{15}$$

and

$$\dot{y} = \mathbf{L}_y \mathbf{v} \tag{16}$$

where $\mathbf{L}_x$ and $\mathbf{L}_y$ are the interaction related to the point:

$$\begin{aligned}\mathbf{L}_x &= \begin{pmatrix} -1/Z & 0 & x/Z & xy & -(1+x^2) & y \end{pmatrix} \\ \mathbf{L}_y &= \begin{pmatrix} 0 & -1/Z & y/Z & 1+y^2 & -xy & -x \end{pmatrix}\end{aligned}$$

Hence, introducing equations (15) and (16) in the equation (12), we obtain:

$$\dot{I} = -(\nabla I_x \mathbf{L}_x + \nabla I_y \mathbf{L}_y)\, \mathbf{v} \tag{17}$$

or

$$\dot{I} = \mathbf{L}_I\, \mathbf{v} \tag{18}$$

$\mathbf{L}_I$ is a 1×6 matrix that links the time derivative of a pixel intensity to the relative camera motion. Now, if we consider the entire image, is defined by the feature vector $\mathbf{I} = (I(\mathbf{x}_{00}), \dots, I(\mathbf{x}_{ij}), \dots I(\mathbf{x}_{MN}))$, and it is necessary to stack, as classical in visual servoing, the interaction matrices for each pixel, leading to:

$$\begin{aligned}\dot{\mathbf{I}} &= \begin{pmatrix} \mathbf{L}_{I_{00}} \\ \vdots \\ \mathbf{L}_{I_{MN}} \end{pmatrix} \mathbf{v} \tag{19} \\ &= \mathbf{L_I}\ \mathbf{v} \tag{20}\end{aligned}$$

where $\dot{\mathbf{I}}$ is a vector of size MN and $\mathbf{L_I}$ is a $MN \times 6$ matrix.

Knowing the interaction matrix $\mathbf{L_I}$, it is possible to design a control law. From (6) and (20) we compute the Jacobian matrix $\mathbf{J_I}$ which allows to build the control law adapted to our eye-to-hand configuration. Let us note that, although the 6 DOF can be controlled using this method, the configuration of our system allows us to control only 3 DOF (2 DOF for the platform translations and one rotating DOF). Therefore, the robot Jacobian ${}^\mathcal{F}\mathbf{J}_n(\mathbf{q})$ is a 6×3 matrix leading to an image Jacobian $\mathbf{J_I}$ of size $NM \times 3$.

The control law given in equation (4) can be considered but we use a control law inspired from the Levenberg-Maquardt optimization algorithm. This provides an efficient numerical

solution to the problem of minimizing the error function $\mathbf{I} - \mathbf{I}^*$ which is highly non-linear and allows to improve robustness and increase the convergence area [20]. More stable than a simple gradient descent, it converges faster that the Gauss-Newton scheme corresponding to equation (4). Therefore, the platform velocities $\dot{\mathbf{q}}$ is given by:

$$\dot{\mathbf{q}} = -\lambda \left(\mathbf{H} + \mu.diag(\mathbf{H}) \right)^{-1} \mathbf{J}_\mathbf{I}^\top \left(\mathbf{I} - \mathbf{I}^* \right) \qquad (21)$$

where $\mathbf{J}_\mathbf{I}$ represents the Jacobian matrix computed from interaction matrix (6) and (20) computed at the desired position. The parameters λ and μ are positive gains and $diag(\mathbf{H})$ is the matrix of diagonal terms of the combination matrix $\mathbf{H}$ which is given by: $\mathbf{H} = \mathbf{J}_\mathbf{I}^\top \mathbf{J}_\mathbf{I}$.

To improve the convergence rate, we implemented an adaptive gain λ (the gain increases when the error decreases). The parameter λ is a function of the difference between the current image $\mathbf{I}$ and the desired image $\mathbf{I}^*$. It is given by:

$$\lambda = \beta \parallel \mathbf{I} - \overline{\mathbf{I}^*} \parallel^2 \qquad (22)$$

where β is a positive gain which allows to adjust the initial value of λ (in the presented experiments, β have a value of $10^2/M$, M being the number of pixels in the image).

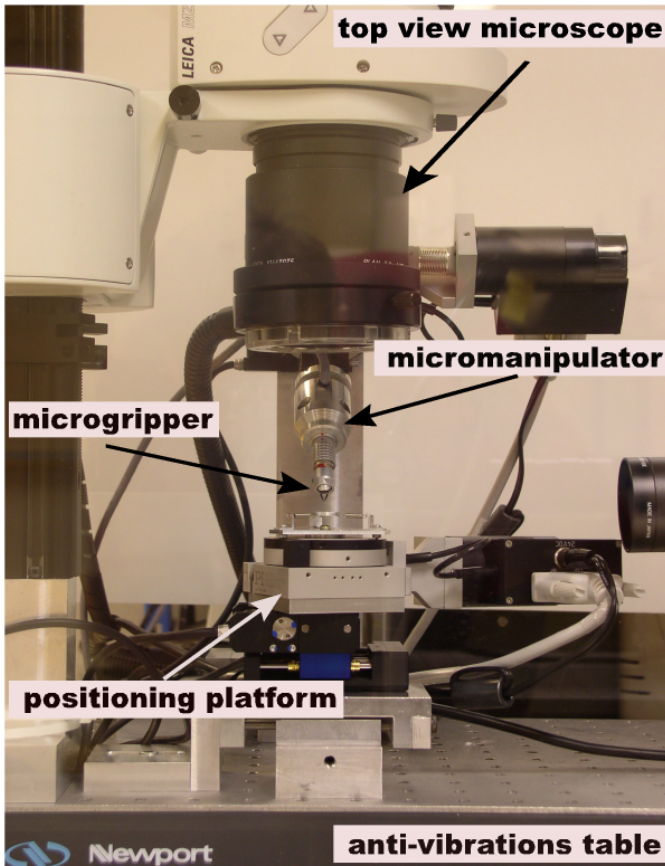## IV. EXPERIMENTAL RESULTS

A. Experimental Setup

Fig. 2.   Photography of the micromanipulation workcell.

The integration of the developed concept is done on the microassembly workcell illustrated in Figure 2. This MEMS microassembly station had been developed in our laboratory. It includes 3 DOF positioning platform with two linear stages i.e. xy and one rotating stage θ. The characteristics of these stages are summarized in Table I. The imaging system is a video stereo microscope of the type LEICA MZ 16 A. It delivers a top view of the work scene and it is equipped with a direct illumination source placed around the objective-lens. The zoom (and thus the magnification) and the focus are motorized and controlled by a PC. The Field of View varies from 700 μm $\times$ 900 μm (with a resolution of 1.4 μm/pixel) at the maximum magnification, to 20 mm $\times$ 25 mm (with a resolution of 21 μm/pixel) at the minimum magnification. The Depth of Field varies from 2.9 mm to 0.035 mm and the working distance is approximately 130 mm. This set-up is placed on an anti-vibrations table and inside a controlled environment.

TABLE I
LINEAR AND ANGULAR STAGES SPECIFICATIONS.

Linear motion stages	
resolution	0.007 μm
increment	0.05 μm
max velocity	1.5 mm/s
stroke	25 mm
Angular motion stages	
resolution	26 μrad
increment	26 μrad
max velocity	0.78 rad/s

B. Validation in Nominal Conditions

The proposed method has been validated on our microassembly workcell. The task studied in this paper concerns the automatic positioning of silicon micro-objects. Object's dimensions are range from 300 to 600 micrometers. In the first experiment, only the 3 DOF of the positioning platform are controlled (i.e. planar positioning). In the second experiment, the controller of the optical microscope magnification (zoom) has been included in the visual servoing scheme. Various experimental situations were tested considering additional external disturbances (i.e. light changing, adding other micro-object) during the positioning process.

The first experiment concerns the positioning of a micro-object with dimensions of 400 μm $\times$ 400 μm $\times$ 100 μm. Figure 3 illustrates a sequence of images of the scene acquired during the positioning task. Figure 3(a) and Figure 3(b) show the initial and desired position of the micro-object, respectively. Figures 3(c) to 3(e) represent the error between the acquired image (current image $\mathbf{I}$) and the desired position (reference image $\mathbf{I}^*$). This is nothing but a graphical representation of the error vector $\mathbf{I} - \mathbf{I}^*$ (the error is null when all the pixels are gray). Figure 3(f) shows the error at the end of the positioning task.

Figure 4 shows the joint velocities (on the 3 DOF $xy\theta$) of the positioning platform during the positioning process. It can be seen the good convergence behavior of the proposed control law. In this Figure, abscissa unit is iteration. Considering
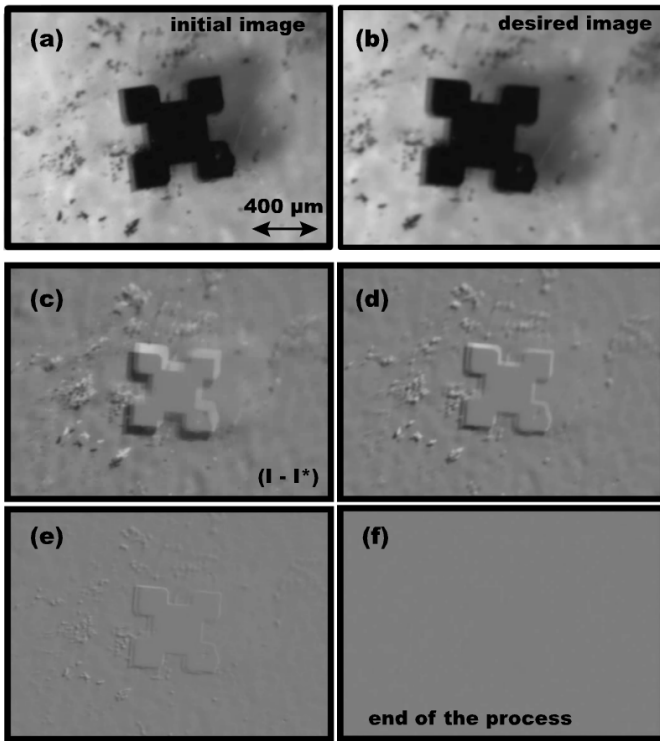
Fig. 3. Sequence of images captured during the achieved positioning task in the first experiment. (a) represents the initial position of the micro-object, and (b) show the desired position to attain. (c) to (e) show the error ($\mathbf{I}$ - $\mathbf{I}^*$) between the current image $\mathbf{I}$ and the desired image $\mathbf{I}^*$, and the image (f) shows the error at the end of the positioning task.

that we have a close-loop system, an iteration correspond to one step of the control scheme (it includes image processing and the computation of the control law). Considering that the image acquisition rate is 25 frames per second, each iteration lasts 0.040 second. The complete positioning has been achieved in 5.2 seconds.
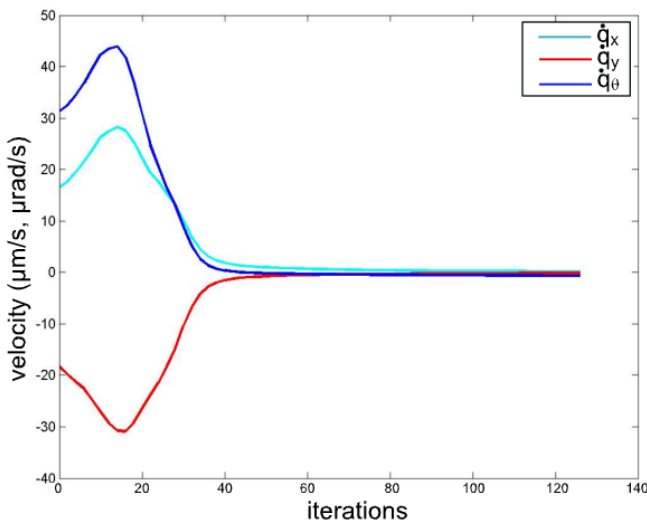
Fig. 4. Joint velocities components (in μm/s and μrad/s) at each iteration of control scheme ($\dot{\mathbf{q}}_\mathbf{x}$, $\dot{\mathbf{q}}_\mathbf{y}$ and $\dot{\mathbf{q}}_\theta$) for the first experiment shown on Figure 3

The obtained accuracies in the presented experiment are

estimated to 89 nm, 14 nm, and 0.001 degrees in the x, y and θ axes, respectively. These accuracies are measured using a double planes mirror SIOS SP 120 interferometer characterized by a high resolution of 0.1 nm (see, Fig. 5). The accuracy measurements are obtained as follows: the first interferometer beam is placed opposite to the x positioning platform motion stage and the second interferometer beam is placed opposite to the y positioning platform motion stage.
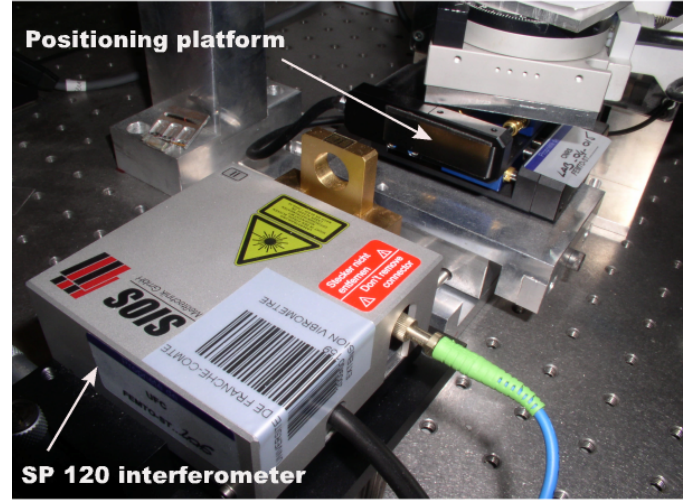

Fig. 5. Photography of the double plane mirror SIOS SP 120 interferometer used to validate the obtained positioning accuracies.

C. Image preprocessing

To reach these accuracies, the image are filtered using a 3×3 Median filter in order to reduce the acquisition noise of the CCD sensor. This filter is given by:

$$M_I(p) = med\big(\mathbf{I}(q)|q \subset \mathbf{W}(p)\big) \qquad (23)$$

where M_I is the image filtered such that for every pixel p, gray level $M_I(p)$ is the median of gray levels $\mathbf{I}(q)$ of q pixels in the window $\mathbf{W}(p)$.

The acquisition noise (random image noise) can be shown in Figure 6. This noise image is reconstructed as follows:

$$\begin{array}{l} \mathbf{I}_t(x,y) - \mathbf{I}_{t+1}(x,y) = 0, \; the \; pixel \; is \; white, \\ \mathbf{I}_t(x,y) - \mathbf{I}_{t+1}(x,y) \neq 0, \; the \; pixel \; is \; black, \; else \end{array} \qquad (24)$$

This means that the images $\mathbf{I}_t$ and $\mathbf{I}_{t+1}$ represent the same image at the times t and $t+1$ without any displacements, and in stable conditions of illuminations. In an ideal case, Figure 6 should be uniformly white. However, as it can be noticed it is not the case. This is a source of imprecision in the positioning process, and the Median filter is a solution to this issue.

To test the efficiency of the proposed median filter to reduce the CCD noise acquisition, a series of images was taken in stable conditions. The variations of the pixel intensities in a line of the image are observed. Figure 7(a) shows the variations of an image line l during the capture of two successive images. It can be seen the oscillation of gray values of pixels without any change of position or condition of acquisition of the image. This causes a problem to attain the maximum accuracy of the
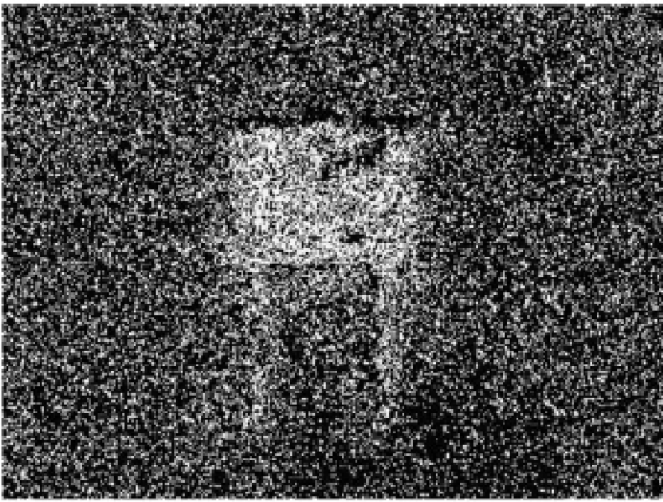
Fig. 6. Illustration of the acquisition noise of the CCD sensor used in the experimental validations.

proposed visual servoing control law. However, Figure 7(b) illustrates the same line of pixels after the application of the filtering method. It can be noticed that the fluctuations of the pixels gray levels are largely attenuated. For example the standards deviations of the pixels fluctuations before and after the filtering method are σ_1 = 1.3769 and σ_2 = 0.3964, respectively. Thus, we can see a reduction of the fluctuations by a factor 3.47.
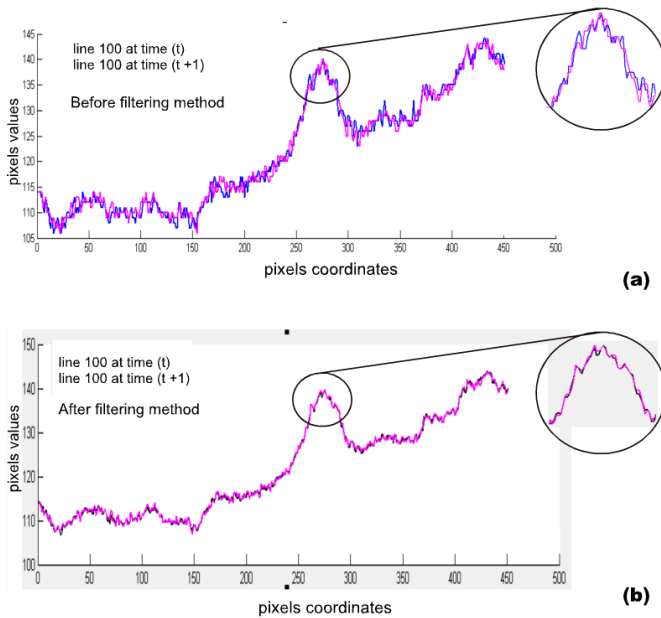

Fig. 7. Illustration of the efficiency of the proposed filtering method. (a) represents the variation of a line of an image at the times (t) and $(t + 1)$ before the filtering method and (b) shows the variation of the same line at times (t) and $(t + 1)$ after the application of the filtering method.

D. Robustness to Occlusions and Lighting Variations

The proposed vision-based control law has been tested in order to measure its robustness in terms of accuracy and

convergence in non-perfect conditions. The first test consists in the addition of others micro-objects during the visual servoing achievement (positioning task). These micro-objects are not present in the desired image but in all images acquired during the positioning task. We also provoked lighting reflection on the silicon micro-objects during the positioning task in order to test the robustness of the proposed control law to the lighting change. Figure 8 shows a sequence of images representing this test. Figure 8(a) illustrates the image (i.e. desired image) of a single silicon micro-object, and Figures 8(b) to 8(f) show the error $\mathbf{I}$ - $\mathbf{I}^*$ between the current image and the desired image. It can be noticed that in the current image, there is a presence of additional micro-objects not present in the desired image. Despite the modification of the external conditions, it can be seen that the quality of the convergence and the accuracy remains good. Figure 9 shows the robot joint velocities during this experiment.
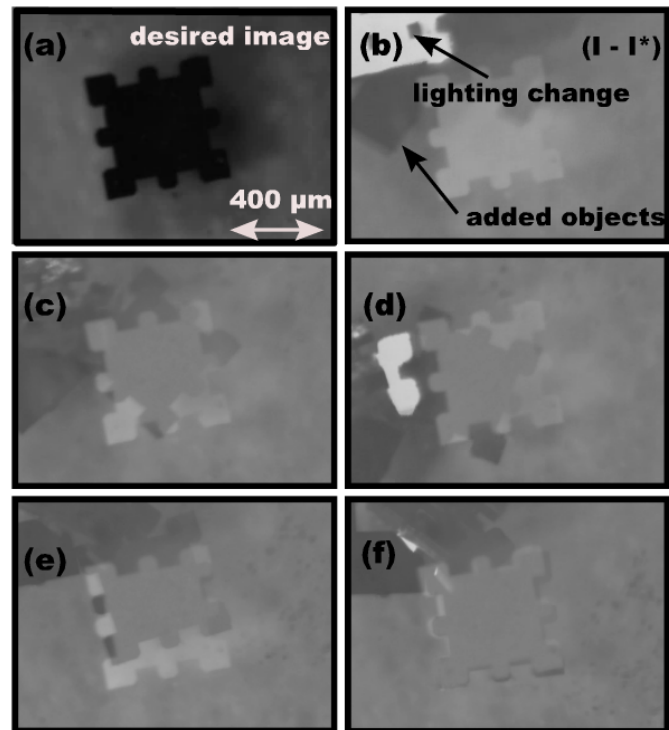

Fig. 8. Sequence of images captured during the visual servoing process in the case of the insertion of additional micro-objects in the scene. Image (a) represents the initial position of the micro-object and the images (b) to (f) show the error $(\mathbf{I} - \mathbf{I}^*)$ between the current image $\mathbf{I}$ and the desired image $\mathbf{I}^*$.

E. Validation using a Mask as a Desired Image

The third experiment is based on the use of an image mask $\mathbf{I}_m$ the as desired image. Only the image of the object is considered and the background will not be considered in the desired image. The goal of this experiment is to demonstrate that the scene background (corresponding to the platform itself) can be filtered and that the object does not have to be located at a specific position on the platform. The desired
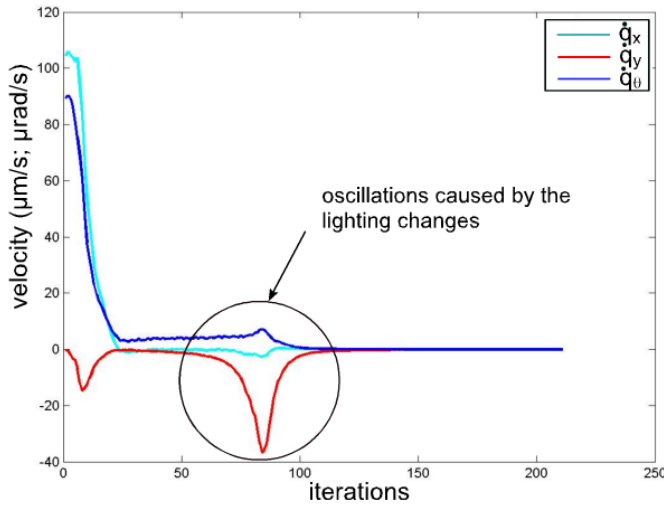
Fig. 9.   Joint velocities components (in μm/s and μrad/s) at each iteration of control scheme ($\dot{\mathbf{q}}_{\mathbf{x}}$, $\dot{\mathbf{q}}_{\mathbf{y}}$ and $\dot{\mathbf{q}}_{\theta}$) for the experiment shown on Figure 8

image is obtained as:

$$\begin{aligned} \mathbf{I}^*{}_m(x,y) &= I^*(x,y), \; if \; object, \\ \mathbf{I}^*{}_m(x,y) &= 255, \; else \end{aligned} \quad (25)$$

Figure 10 shows a sequence of images captured during the nanopositioning task. Figure 10(a), represents the desired image which consists of the image of the micro-object inserted in a white image. Figures 10(b) to 10(f) represent the image error ($\mathbf{I}$ - $\mathbf{I}^*{}_m$) during the visual servoing task.
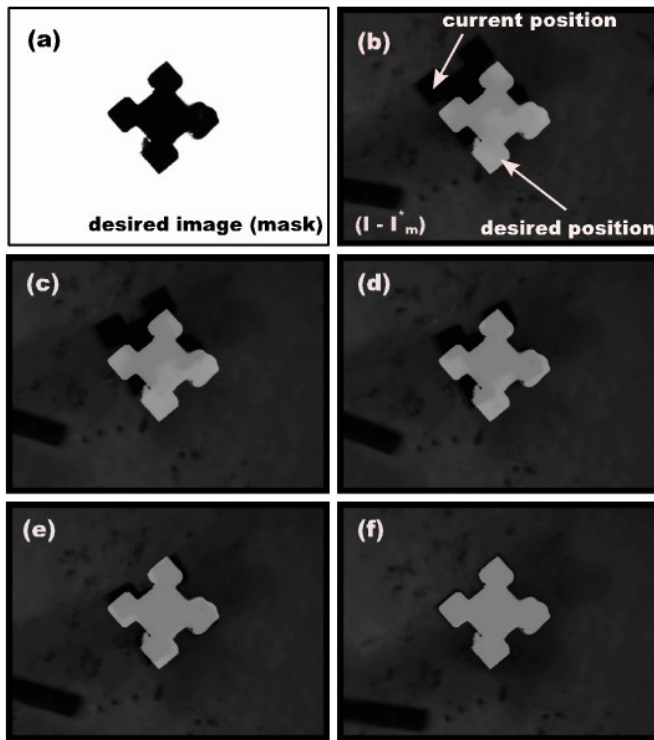

Fig. 10.   Sequence of images captured during the positioning process using a mask as a desired image. (a) shows the mask used as a desired image. (b) to (e) represent the error ($\mathbf{I}$ - $\mathbf{I}^*$) and (f) shows the end of the visual servoing process.

Fig. 11.   Joint velocities components (in μm/s and μrad/s) at each iteration of control scheme ($\dot{\mathbf{q}}_{\mathbf{x}}$, $\dot{\mathbf{q}}_{\mathbf{y}}$ and $\dot{\mathbf{q}}_{\theta}$) for experiment shown in figure 10.

Despite the presence of more than 50% of white pixels (virtual pixels) in the desired image $\mathbf{I}^*{}_m$ do not have corresponding pixels in the current image, the proposed control law remains robust and converges to zero as shown in Figure 11. By analyzing the behavior of the control law, we remark that there is a presence of small oscillations at the beginning because of the large number of pixels in the current image that have no correspondences in the desired image.

F. Validation using low Magnification Optical Microscope

The proposed approach has also been tested using a low magnification optical microscope ($\times$0.65). Figure 12 illustrates images captured during the positioning task. As can be seen, the desired image contains additional objects not present in the current images. These objects have been added in order to test the robustness of the proposed method in terms of disturbance. It can be noticed that the proposed control law converges to the desired position as shown in Figure 12(f). Considering a sequence of positioning tasks with increasing magnification will allows to increase the convergence area of the global task. Here a 2 mm motion is considered, while for greater magnification (experiment reported on Figure 3) a 500 μm motion is considered.

G. Joint Control of the Positioning Platform and Optical Microscope Magnification

The last experiment presented in this paper propose to control both the positioning platform and the optical microscope magnification with the same photometric visual servoing approach. We then propose a decoupled control law. The motion of the platform will be controlled using the control law proposed in previous section. The control of the magnification consists of controlling the displacement of the optical microscope objective-lens along the optical axis. This latter is coaxial with the vertical axis z which carried the rotation stage θ of the positioning platform.
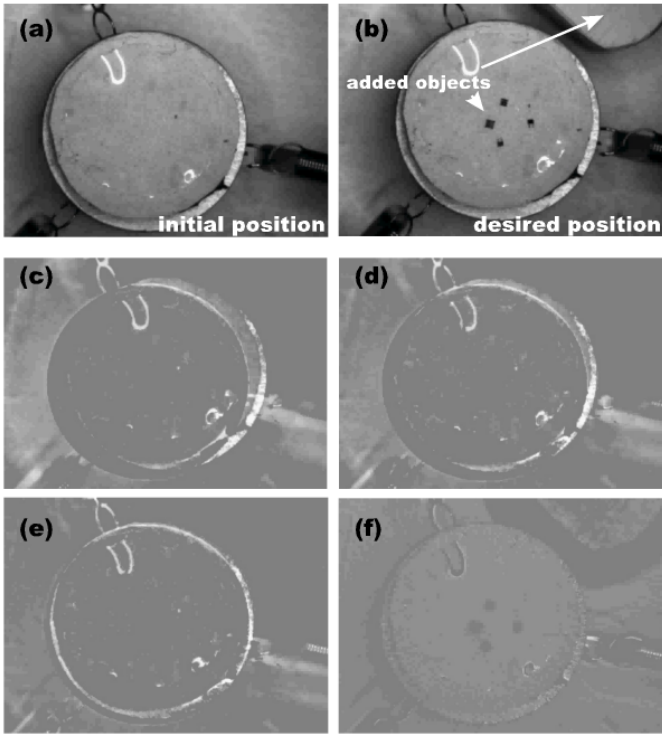
Fig. 12. Positioning task using a low magnification optical microscope. (a) represents the initial position (i.e. the initial image) of the micro-object, and (b) its desired position (i.e. the desired image $\mathbf{I}^*$). (c) to (e) show the error $(\mathbf{I} - \mathbf{I}^*)$ between the current image $\mathbf{I}$ and the desired image $\mathbf{I}^*$, and the (f) shows the end of the visual servoing process where $\mathbf{I} - \mathbf{I}^* = 0$.

In that case the interaction matrix $\mathbf{K_I}$ that links zoom velocity to the time derivative of a pixel intensity to the relative camera motion is given by:

$$\dot{I} = -\left(\nabla\mathbf{I}_x x + \nabla\mathbf{I}_y y\right)\dot{f}$$

or

$$\dot{I} = \mathbf{K}_I \dot{f}$$

When considering all the image pixels, we have $\dot{\mathbf{I}} = \mathbf{K_I}\dot{f}$.

Therefore, to control the magnification in the camera frame, we have:

$$\mathbf{v}_f = -\beta\mathbf{K_I^+}\left(\mathbf{I} - \mathbf{I}^*\right) \tag{26}$$

where β a fixed positive gain and $\mathbf{K}_I^+$ the pseudo-inverse of the interaction matrix $\mathbf{K_I}$.

To do this, the desired image is captured with a high magnification, i.e. $\times 9$ (high resolution/low Field of View) (Fig. 13(b)), and the visual servoing starts with a low magnification, i.e. $\times 2$ (low resolution/high Field of View) (see Fig. 13(a)). After that, the magnification increases during the realization of the task until the desired magnification in which the desired image is acquired.

Figure 14 illustrates the achievement planar positioning task with the optical microscope magnification control. From Figure 14(a) to Figure 14(d), it can be seen the error $\mathbf{I} - \mathbf{I}^*$ between the desired image and the current image as the function of the magnification factor during the positioning task (see, also Fig. 15).
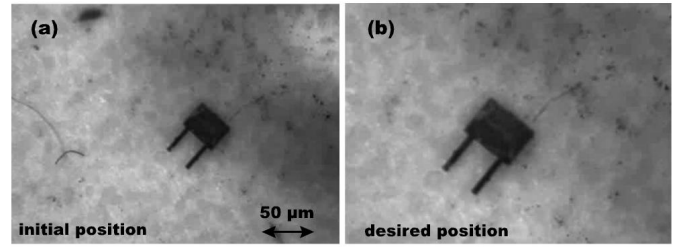

Fig. 13. (a) and (b) represent the initial position captured with a low magnification and the desired position acquired with a high magnification, respectively.
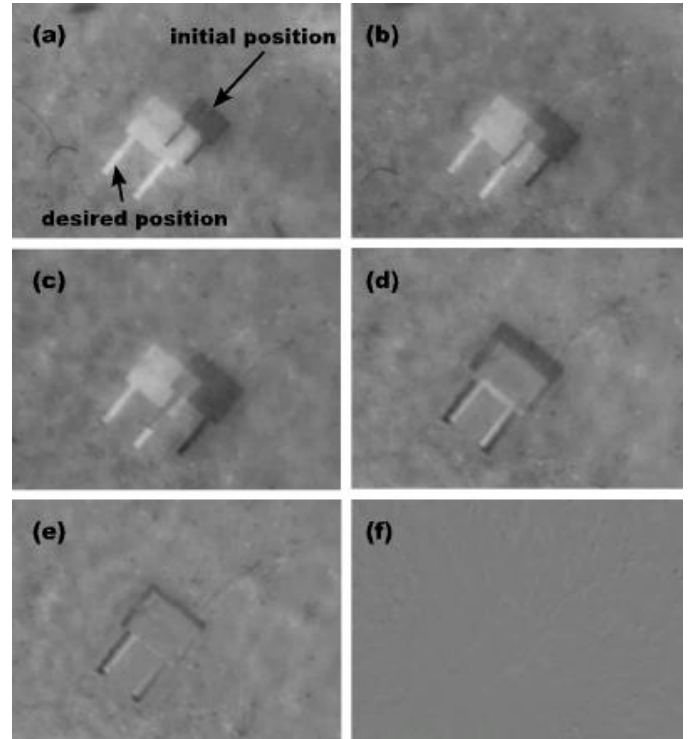

Fig. 14. Sequence of images captured during the positioning task with the control of the magnification of the optical microscope simultaneously. (a) shows the first step of the process. (b) to (e) represent of the error $(\mathbf{I} - \mathbf{I}^*)$ during the positioning task and the magnification increasing in the same time. (f) illustrates the end of the positioning task.

V. CONCLUSION

The problem of the automation of micromanipulation and microassembly of MEMS using a direct vision-based control approach was studied in this paper. Classical visual servoing schemes use geometrical visual features (points, lines, edges ...) extracted and tracked in the image sequence. This kind of approaches requires the use of image processing algorithms to extract these features from the images. It is very time consuming and complex to set up and often not robust to external conditions modifications. Therefore, we have proposed a new control law based on the use of the pure image signal (pixels intensity). This photometric-based approach has been tested on our microassembly workcell. From the different validation tests of the proposed control law, it appears that it is robust to global light variations, to important occlusions and to different perturbations. The proposed approach allows
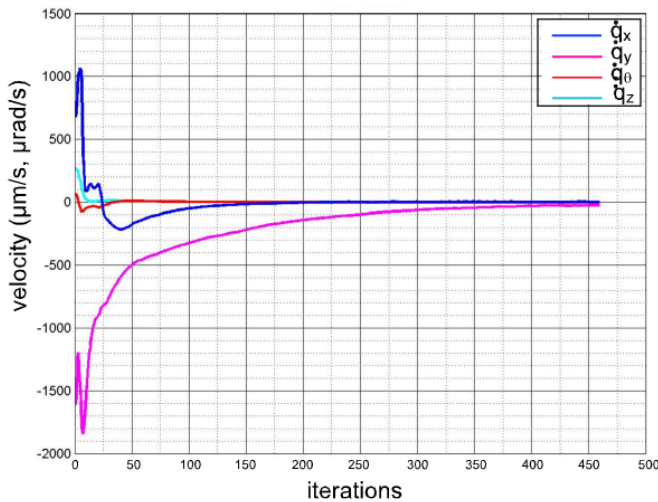
Fig. 15. Joint velocities components (in μm/s and μrad/s) at each iteration of control scheme ($\dot{q}_x$, $\dot{q}_y$, $\dot{q}_z$, and $\dot{q}_\theta$) for experiment shown on Figure 14 along with the motion of the camera zoom velocity.

highly accurate positioning task (below 100 nm in translation and 0.001 degrees in rotation along the platform axis). This process has also been used to perform a positioning task with the control of the optical microscope magnification in the same time. So, the first experimental results are promising in terms of accuracy and of the designed control law behavior. Future work will attempt to apply the developed method for use with a scanning electron microscope (SEM) in order to control a serial 3 DOF micromanipulator (MM3 Kleindiek). This microrobot will be associated to a high accurate 3 DOF positioning platform. The whole system will be placed inside the SEM. The objective of this work is to perform a full-automatic handling on carbon nanotubes of few nanometers of diameter with a subnanometer accuracy.

REFERENCES

[1] S. Shet, V. R. Mehta, A. F. Fiory, M. P. Lepselter, and N. M. Ravindra, "The Magnetic Field-Assisted Assembly of Nanoscale Semiconductor Devices: A New Technique," *Journal of materials (JOM)*, vol. 56(10), pp. 32–34, October 2004.

[2] N. Dechev, L. Ren, W. Liu, W. Cleghorn, and J. Mills, "Development of a 6 Degree of Freedom Robotic Micromanipulation for use in 3D MEMS Microassembly," in *IEEE International Conference on Robotics and Automation, ICRA*, Orlando, Fl., May 2006, pp. 281–288.

[3] J. Agnus, D. Hériban, M. Gauthier, and V. Pétrini, "Silicon End-Effectors for Microgripping Tasks," *Precision Engineering*, vol. 33, pp. 542–548, February 2009.

[4] J.-Y. Gauthier, A. Hubert, J. Abadie, N. Chaillet, and C. Lexcellent, "Nonlinear Hamiltonian Modelling of Magnetic Shape Memory Alloy-based Actuators," *Sensors and Actuators A Physical*, vol. 141, pp. 536–547, February 2008.

[5] M. Rakotondrabe, Y. Haddab, and P. Lutz, "Development, Modeling, and Control of a Micro-/Nanopositioning 2 DOF Stick-Slip Device," *IEEE/ASME Transactions on Mechatronics*, vol. 14, pp. 733–745, December 2009.

[6] L. Yangmin and X. Qingsong, "Development and Assessment of a Novel Decoupled xy Parallel Micropositioning Platform," *IEEE/ASME Transactions on Mechatronics*, vol. 15, pp. 125–135, February 2010.

[7] M. Probst, C. Hrzeler, R. Borer, and B. J. Nelson, "A Microassembly System for the Flexible Assembly of Hybrid Robotic MEMS Devices," *International Journal of Optomechatronics*, vol. 3, no. 2, pp. 69–90, April 2009.

[8] T. W. Seo, and D. S. Kang, and H. S. Kim, and J. Kim, "Dual Servo Control of a High-Tilt 3 DOF Microparallel Positioning Platform", *IEEE/ASME Transactions on Mechatronics*, vol. 15, no. 5, pp. 616–625, October 2009.

[9] B. Tamadazte, E. Marchand, S. Dembélé, and N. Le Fort-Piat, "CAD Model-based Tracking and 3D Visual-based Control for MEMS Microassembly," *The International Journal of Robotics Research, IJRR*, vol. 29, pp. 1416–1434, September 2010.

[10] M. Probst, R. Borer, and B. J. Nelson, "A Microassembly System for Manufacturing Hybrid MEMS," in *IFTOMM World Congress*, Besançon, France, June 2007, pp. 37-42.

[11] M. Savia and H. N. Koivo, "Contact Micromanipulation: Survey of Strategies," *IEEE/ASME Transactions on Mechatronics*, vol. 14, pp. 504–514, August 2009.

[12] S. Hutchinson, G. Hager, and P. Corke, "A Tutorial on Visual Servo Control," *IEEE Transactions on Robotics and Automation*, vol. 12, no. 5, pp. 651–670, October 1996.

[13] F. Chaumette and S. Hutchinson, "Visual Servo Control, part 1: Basic Approaches," *IEEE Robotics and Automation Magazine*, vol. 13, no. 4, pp. 82–90, December 2006.

[14] K. B. Yesin and B. J. Nelson, "A CAD Model-based Tracking System for Visually Guided Microassembly," *Robotica*, vol. 23, pp. 409–418, July 2005.

[15] B. Tamadazte, N. Le Fort-Piat, and S. Dembélé, "Robotic Micromanipulation and Microassembly using Monoview and Multiscale Visual Servoing," *IEEE/ASME Transactions on Mechatronics*, accepted, to appear, 2011.

[16] S. Benhimane and E. Malis, "Homography-based 2D Visual Tracking and Servoing," *International Journal of Computer Vision, IJCV*, vol. 26, no. 7, pp. 661–676, July 2007, special IJCV/IJRR issue on vision for robots.

[17] V. Sundareswaran, P. Bouthemy, and F. Chaumette, "Exploiting Image Motion for Active Vision in a Visual Servoing Framework," *International Journal of Robotics Research, IJRR*, vol. 15, no. 6, pp. 629–645, June 1996.

[18] J. Santos-Victor and G. Sandini, "Visual Behaviors for Docking," *Computer Vision and Image Understanding*, vol. 67, no. 3, pp. 223–238, September 1997.

[19] A. Crétual and F. Chaumette, "Visual Servoing based on Image Motion," *International Journal of Robotics Research, IJRR*, vol. 20, no. 11, pp. 857–877, November 2001.

[20] C. Collewet and E. Marchand, "A Photometric Visual Servoing," *IEEE Transactions on Robotics and Automation*, accepted, to appear, 2011 .

[21] E. Marchand and C. Collewet, "Using Image Gradient as a Visual Feature for Visual Servoing," in *IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS*, Taipei, Taiwan, October 2010, pp. 5687–5692.

[22] A. Dame and E. Marchand, "Entropy-based Visual Servoing," in *IEEE International Conference on Robotics and Automation, ICRA*, Kobe, Japan, May 2009, pp. 707–713.

[23] E. Marchand, "Control Camera and Light Source Positions using Image Gradient Information," in *IEEE International Conference on Robotics and Automation, ICRA*, Roma, Italia, April 2007, pp. 417–422.

[24] B. Horn and B. Schunck, "Determining Optical Flow," *Artificial Intelligence*, vol. 17, no. 1-3, pp. 185–203, August 1981.

Brahim Tamadazte holds a Ph.D. in Automation and Computer Science from the "Université de Franche-Comté" in 2009 and a MS degree in Robotics and Intelligent Systems from the " Université Pierre et Marie Curie (Paris VI)" in 2005, in France. Before, Brahim has also trained as an engineer in Automation Systems at the "Université Mouloud Mammeri de Tizi-Ouzou" from Algeria. During his Ph.D thesis, Brahim has developed some vision algorithms and vision-based control laws (image-based visual servoing and pose-based visual servoing) for the automation of MEMS (Micro-Electro-Mechanical-Systems) handling and assembly. He also has an interest to the development of new calibration methods dedicated to modeling imaging systems equipped with high magnification lens as the optical microscopes. Actually, Brahim is an associate professor at the ENSMM (National Engineering School of Mechanics and Microtechniques) and makes his research works at the AS2M (Automatic Control and Micro-Mechatronic Systems) department of FEMTO-ST Institute.

Nadine Le-Fort Piat received her Ph.D. in Automation and Signal Processing at the University of Technology of Compiègne (UTC) in 1984. She became associate professor at UTC Compiègne in 1984 and professor at the ENSMM (National Engineering School of Mechanics and Microtechniques - Besançon) in 1998. She is a member of the AS2M department of the Institute FEMTO-ST, UMR CNRS 6174, in Besançon. Her research areas concern perception and advanced control strategies based on visual servoing and reinforcement learning for automation of distributed microrobotic systems.

Eric Marchand is professor at Université de Rennes 1 in France and a member of the INRIA Lagadic team. He received the Ph.D degree and "Habilitation á Diriger des Recherches" in Computer Science from the Université de Rennes 1 in 1996 and 2004, respectively. He spent one year as a Postdoctoral Associate in the AI laboratory of the Department of Computer Science at Yale University. He has been an INRIA research scientist ("Chargé de recherche") at IRISA-INRIA Rennes from 1997 to 2009. His research interests include robotics, perception strategies, visual servoing, real-time object tracking and augmented reality. Since 2010, he is an associate editor for IEEE Transactions on Robotics.