Visual Servoed Micropositioning for Protein Manipulation Tasks

Youcef Mezouar, Peter K. Allen

Dept. of Computer Science and Northeast Structural Genomics Consortium
Columbia University, New York, USA
e-mail:{ymezouar,allen}@cs.columbia.edu

Abstract

In this paper, we present a framework for cell manipulation tasks with visual servoing micromanipulation strategies. A vision based micropositioner is designed in order to address the requirement of high precision needed to perform manipulation of objects under 100 μm in size. The system calibration (microscope-camera-micropositioner) and the model of the observed scene are not known. Experimental results for micropositoning tasks with respect to protein cells are presented and demonstrate the validity of the proposed approach.

1 Introduction

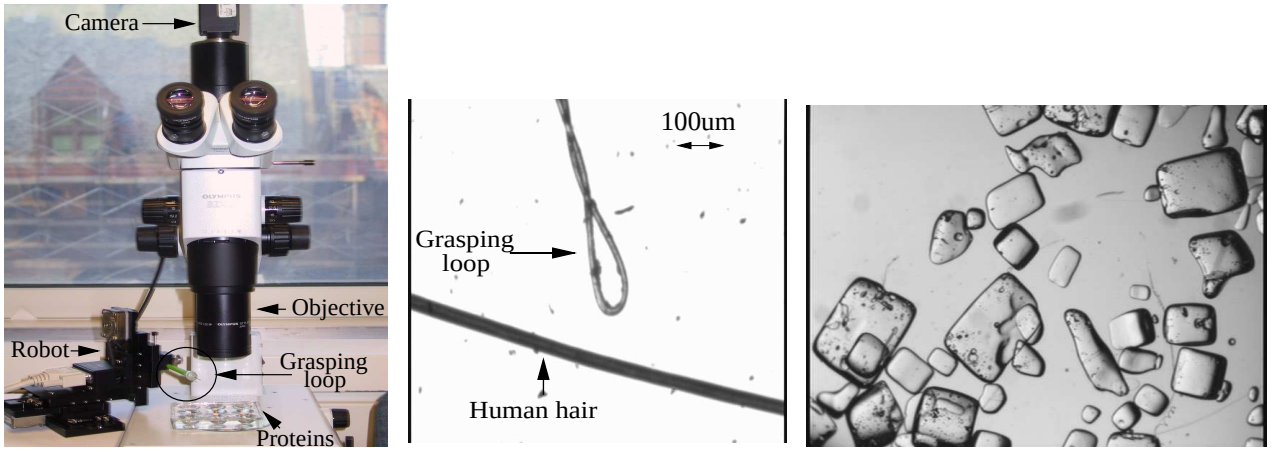
Crystallized proteins are an important source for research in the biological sciences. They can be used in light-scattering experiments, growth-rate measurements, calorimetry, and the evaluation of new crystallisation techniques and reagents, all of which contribute to a deeper understanding of molecular structure. This will impact a number of fields, including the emerging structural genomics field [12], structure-directed drug design [14], and the newly developed screening by X-ray crystallography [11], as well as small molecule applications. Protein crystals are utilized for downstream processing and purification in the industrial production of enzymes, they are involved in pharmaceutical formulation, and they are often useful in various other areas of biophysical experimentation. The development of strategies and devices for automated and fast manipulation of protein crystals can greatly improve the throughput in molecular research by reducing from hours to seconds the time for manipulation of crystals.

This work is aimed at using computer vision to provide the compliance and the robustness which precise protein manipulation requires without the need for extensive analysis of the physics of grasping or a detailed knowledge of the environment. One of the major advances in robotics over the last 20 years is the visual control of robotic manipulators [7]. The advent of fast and inexpensive digital imaging technology has allowed camera systems to be integrated as part of a closed loop feedback control system [1]. Visual servoing strategies had been successfully implemented at the microscale level for manipulation of known micro-electromechanical systems [10, 3] with calibrated devices (however in these papers the CAD models of the proteins crystals and the system calibration are not known). Vision can provide rich knowledge about the spatial arrangement of objects to be manipulate as well as knowledge about the means of manipulation, which in our case are the instruments needed to perform protein manipulation. Our goal is to visually monitor and control these instruments as they isolate and acquires proteins. For tasks such as cell or protein manipulation, this idea of visual feedback control becomes extremely important. Classical strategies of manipulation will not work at these scales (objects under 100 μm in size) due to the required precision (beyond the calibration range of conventional industrial precision devices) and additional problems relating to microscale phenomena. Currently, the mechanics of micromanipulation is poorly understood, and thus results of sensorless micromanipulation strategies are unpredictable. In this paper, we propose an integrated control system, consisting of a high resolution optical microscope, digital imaging system, image based servo-controllers and micromanipulator that will be able to precisely position a grasping loop with respect to a protein.

The remainder of this paper is organized as follows: In Section 1, we briefly describe our set-up and the optics model. The task to be achieved is presented in section 2. Section 3 presents methods to track the mobile loop observed by a static camera. Section 4 presents the visual servoing framework for positioning the loop with respect to protein crystals. In section 5, we present the experimental results.

2 Set-up and Optics model

In this section, after a short description of our workstation, we describe the optics model used to design our control scheme.



(a) (b) (c)

Figure 1: *(a) Workstation for protein crystals manipulation, (b) Grasping loop for protein crystals, (c) protein crystals*

2.1 Set-up

Our workstation (see Figure 1(a)) is centered around a Sutter MP-285 micropositioner platform with independent X, Y and Z motions. This positioner allows submicron resolution down to 40 nanometers. A loop (see Figure 1(b)) is mounted directly on the robot Z-axis. The loop is especially designed to grasp protein crystals (see Figure 1(c)). The loop and the protein crystals are observed through an Olympus SZX12 optical microscope. The microscope provides a total magnification from 8.4x to 108x. The unit has a CCD camera module adapter onto which a Sony XC-77 CCD camera is mounted.

2.2 Optics modeling

In our experiments, we are using a long working distance microscope SZX12 by Olympus with an infinity-focused objective. A simplified ray diagram for a such typical optical microscope is shown in Figure 2. The rays emanating from a given point are parallel between the objective and the tube lens. The parallel rays are focused onto the image plane by the tube lens. The total linear magnification α is given by :

$$\alpha = \frac{f_t}{f_o} \tag{1}$$

where f_t is the focal length of the tube lens and f_o the focal length of the objective. With our microscope, the focal length of the tube lens is adjustable so that the magnification varies from 8.4 to 108. A point M with coordinates $\mathbf{M} = [X\,Y\,Z]^T$ in the objective focal plane is projected in the image focal plane onto a point m of coordinates $\mathbf{m} = [x\,y]^T$ with :

$$\mathbf{M} = \alpha \mathbf{m} \tag{2}$$

To provide submicron accuracy, micro assembly

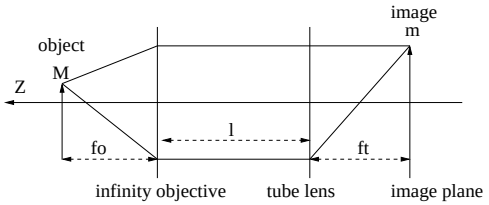


Figure 2: *Simplified ray diagram for an infinitly focused optical microscope*

tasks require high magnification of the observed objects. High magnification optical systems have a very small depth of field. This limited depth of field can be exploited to measure depth from the camera using techniques of depth-from-focus/defocus. For high numerical aperture systems the depth of field D is given by [8] :

$$D = \frac{\lambda n}{2A^2} + \frac{n}{7\alpha A}$$

where n is refractive index of the optics, A is the numerical aperture of the optics, λ the wavelength of incident light and α the total magnification given by (1). This equation means that a point with Z coordinate such that $f_t + f_o + l - D < Z < f_t + f_o + l + D$ appears as a blurred point in the image plane (l denoting the distance between the objective plane and the image plane). Since D is very small compared to $f = f_t + f_o + l$, an object in perfect focus is situated at a distance f of the image plane. When the observed object is moved out of focus such that the Z position of the object is $Z = f + dZ$ then a point belonging to the object is projected in the image plane into a disk of diameter db (blurred

image) with [8, 4] :

$$db = \underbrace{\frac{2\alpha A dZ}{n}}_{\text{geometric blur}} + \underbrace{\frac{1.22\lambda\alpha}{A^2}}_{\text{diffraction blur}} \quad (3)$$

If the blur is measured using a global focus measure in the image as the Tenengrad measure T_r, we assume that :

$$dT_r = \gamma db \quad (4)$$

where γ is a scaling factor and dT_r is the variation of the focus measure when the Z position of the object is moved from f to $f + dZ$. This assumption means that the blur measure varies as the diameter of the disk resulting from the projection in the image plane of the defocused points. This property will be exploited to control the Z axis of our manipulator.

3 Task description

In our system, the camera mounted on the microscope objective observes the protein crystals and the mobile loop mounted on the three degrees of freedom micropositioner (eye-to-hand configuration). Let $\mathcal{F}_c$ and $\mathcal{F}_r$ be respectively a frame attached to the fixed camera and a frame attached to the robot (see Figure 3). The Z-axis of the camera frame and of the robot control frame are parallel. Our goal is to align the center of gravity of the loop (more precisely of the loop hole) $\mathcal{M}$ with coordinates $\mathbf{M}_c = [X_c\, Y_c\, Z_c]^T$ in $\mathcal{F}_c$ and the center of an isolated protein $\mathcal{M}^*$ with coordinates $\mathbf{X}_c^* = [X_c^*\, Y_c^*\, Z_c^*]^T$ in $\mathcal{F}_c$. The protein is assumed motionless and in-focus. Three degrees of freedom must be controlled, we need thus to extract at least three independent measures from the image. Since the protein is assumed in focus, $Z_c = Z_c^*$ when the loop is also in-focus. As, we will see in the sequel, the Z-axis can thus be controlled using a focus measure T_r in the image.
Let $\mathbf{x} = [x\, y]^T$ (resp. $\mathbf{x}^* = [x^*\, y^*]^T$) be the coordinates in the image plane of the center of gravity of the loop (resp. of the protein). The robot motion along the X and Y axis will be controlled by minimizing the distance in the image space between $\mathbf{x}$ and $\mathbf{x}^*$. The positioning task can thus be achieved by first minimizing a focus measure and then by minimizing an error in the image space. To control the X and Y axis of the micropositioner we have to track the center of gravity of the loop in a complex scene. In the next section, we describe our tracking strategy.

4 Tracking the grasping loop

To track the center of gravity of the grasping loop, a motion detection is first realized. Since only the loop is in motion this step eliminates the complex

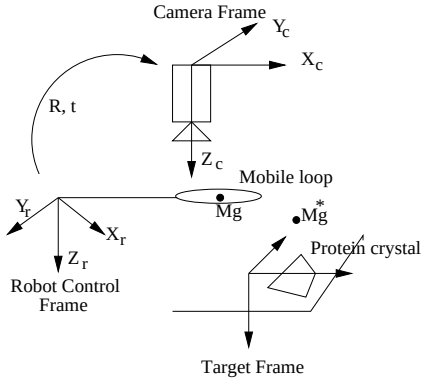


Figure 3: *Workstation frames*

background. Let I_t (resp. I_{t-dt}) be the luminance function at time t (resp. $t-dt$), and define the function :

$$F(x,y) = \frac{1}{N^2} \sum_{(u,v)\in V(x,y)} |I_{t-dt}(u,v) - I_t(u,v)| \quad (5)$$

where $V(x,y)$ is a $N \times N$ neighbor of the point (x,y). The binary image of temporal variations at time t is defined by :

$$I_t^b(x,y) = \begin{cases} 1 & \text{if } F(x,y) > \lambda \\ 0 & \text{else} \end{cases} \quad (6)$$

where λ is a chosen scalar defining the sensibility of the motion detector. We obtain thus a binary image I_t^b containing the loop at time $t-dt$ and t (see Figure 4(d)). In order to obtain a binary image with only the loop at time t, the binary image of temporal variations I_r^b (see Figure 4(c)) between a reference image (for example the image at the initial time) and the image at time t is computed and the logical operator AND is applied between I_t^b and I_r^b. The resulting image I_{rt}^b contains only the loop at time t (see Figure 4(e)). To eliminate potential small gaps in the binary image I_{rt}^b due to poor motion detection, we apply a closing morphological filter (see Figure 4(f)). Once again, since only the loop is moving, this step eliminates the background. The image is thus simplified in such a way that the center of gravity of the loop can easily be tracked by simply tracking the white points $\mathbf{x}_j$ of the hole loop with $\mathbf{x} = \frac{1}{n}\sum_{j=1}^{j=n} \mathbf{x}_j$, where n is the number of white points.

5 Eye-to-hand Visual servoing with a microscope

Visual servoing is classified into two main approaches [15, 6, 7]. The first one [16, 9], based on the computation of a 3-D Cartesian error, requires a perfect

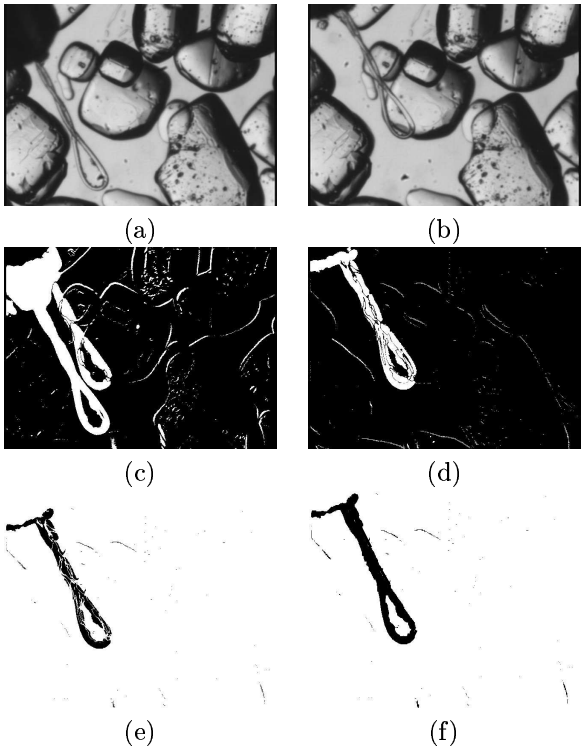


Figure 4: *(a) Reference image, (b) image in which the grasping loop has to be detected, (c) image I_r^b, (d) image I_t^b, (e) image I_{rt}^b, (f) final image*

CAD-model of the object and a calibrated camera to obtain unbiased pose estimation. In the second approach, the pose estimation is omitted and the control loop is directly closed in the image space. That ensures local convergence and stability in presence of modeling errors and noise perturbations [2]. In our case, the models of the observed targets are unknown and the system is not calibrated, we thus use the image-based approach. Central to this approach is the image jacobian $\mathbf{L}$ (also called interaction matrix). It relates the differential motion $\mathbf{ds}$ of some image features $\mathbf{s}$ to the differential motion in the camera coordinates $\mathbf{dP}$: $\mathbf{ds} = \mathbf{LdP}$. In our application, the camera observes the robot to be controlled; such configuration is known as eye-to-hand systems [5] by opposition to the more classical eye-in-hand configuration where the camera is mounted on the robot effector and observes the object to be manipulated. In the eye-to-hand case, the image jacobian has to take into account the mapping from the camera frame $\mathcal{F}_c$ onto the robot control frame $\mathcal{F}_r$. If we note $[\mathbf{R}, \mathbf{t}]$ this mapping ($\mathbf{R}$ being the rotational matrix and $\mathbf{t}$ the translation vector), the eye-to-hand jacobian $\mathbf{L}_t$ is related to the eye-in-hand one $\mathbf{L}$ by (for the general case of a 6-DOF motion):

$$\mathbf{L}_t = -\mathbf{L} \begin{bmatrix} \mathbf{R} & -\mathbf{R}[-\mathbf{R}^T\mathbf{t}]_x \\ \mathbf{0} & \mathbf{R} \end{bmatrix} \tag{7}$$

where $[-\mathbf{R}^T\mathbf{t}]_x$ is the skew symmetric matrix associated with vector $-\mathbf{R}^T\mathbf{t}$.

The first step of our control strategy is to bring the loop in focus. To realize the control along the Z-axis, the small depth of field of the microscope can be used. Since the target (i.e the proteins crystals) is assumed in focus, the positioning of the micromanipulator along Z can be done by moving along the Z axis until a measure of focus in the image is maximized. We chose as focus measure the Tenengrad measure T_r. Tenengrad is a measure of thresholded gradient magnitude [13]:

$$\begin{cases} T_r = \sum_x \sum_y \|\vec{\delta}I\|^2 & \\ \quad = \sum_x \sum_y (\delta I_x^2 + \delta I_y^2) & \text{and } \|\vec{\delta}I\|^2 > \epsilon \end{cases} \tag{8}$$

It can be proved that the best focused image is obtained at the global maximum of the focus measure T_r if the threshold is zero [13]. We thus use the Tenengrad measure with $\epsilon = 0$. Combining the Equations (3) and (4), and differentiating the resulting equation lead to :

$$^cV_z = \Gamma \dot{T}_r$$

where $\Gamma = \frac{n}{2\gamma\alpha A}$, and ^cV_z is the Z control expressed in the camera frame. Noting that the Z axis of the camera frame $\mathcal{F}_c$ and of the robot control frame $\mathcal{F}_r$ have been defined parallel and in the same direction, we have also :

$$^rV_z = \Gamma \dot{T}_r \tag{9}$$

where ^rV_z is the Z control expressed in the robot control frame. We want to maximize T_r, by moving the robot in the Z direction. One simple way to realize the maximisation of T_r is to control the Z motion using (9). The norm of Γ can be fixed as a gain of a proportional control law and the sign of Γ ensuring the maximisation of T_r can be obtained off-line by observing the effect of the robot Z motion on T_r. After that the loop is focused, the X and Y motions have to be controlled in order to align the center of gravity of the loop and the center of a protein crystal.

Since we are only interested in a 3-DOF motion, the interaction matrix of a single point is a 2×3 matrix. Differentiating equation (2), the eye-in-hand image jacobian relationship for a single image feature point is :

$$\dot{\mathbf{x}} = \begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix}^T = \begin{bmatrix} \alpha & 0 & 0 \\ 0 & \alpha & 0 \end{bmatrix} \begin{bmatrix} ^cV_x \\ ^cVy \\ ^cVz \end{bmatrix} = \mathbf{L}_{xyz}\, ^c\mathbf{V}_{xyz} \tag{10}$$

where $\mathbf{L}_{xyz}$ denotes the eye in hand interaction matrix and $^c\mathbf{V}_{xyz}$ the control vector expressed in the

camera frame. The previous equation can obviously be simplified as follow :

$$\dot{\mathbf{x}} = \begin{bmatrix} \alpha & 0 \\ 0 & \alpha \end{bmatrix} \begin{bmatrix} {}^cV_x \\ {}^cVy \end{bmatrix} = \mathbf{L}_{xy}\, {}^c\mathbf{V}_{xy} \qquad (11)$$

where $\mathbf{L}_{xy}$ and ${}^c\mathbf{V}_{xy}$ are respectively the reduced eye-in-hand interaction matrix and the reduced control vector expressed in the camera frame. Noticing once again that the Z-axis of the camera frame and of the control frame are parallel in the same direction, and combining the equation (7) and (11), the eye-to-hand jacobian relationship for a single feature point is :

$$\dot{\mathbf{x}} = \mathbf{L}_{xy}\mathbf{R}_{xy} \begin{bmatrix} {}^rV_x \\ {}^rVy \end{bmatrix} = \mathbf{L}_{\mathbf{t}xy}\, {}^r\mathbf{V}_{xy} \qquad (12)$$

where $\mathbf{L}_{\mathbf{t}xy} = \alpha\mathbf{R}_{xy}$ is the eye to hand interaction matrix and $\mathbf{R}_{xy}$ is a 2×2 rotation matrix that maps the reduced control vector expressed in the camera frame ${}^c\mathbf{V}_{xy}$ and the reduced control vector expressed in the robot control frame ${}^r\mathbf{V}_{xy}$. A suitable control law to make the error vector $\mathbf{e} = \mathbf{x} - \mathbf{x}^*$ decrease exponentially (i.e, $\dot{\mathbf{e}} = -\beta\mathbf{e}$) to $\mathbf{0}$ is given by :

$${}^r\mathbf{V}_{xy} = -\beta\widehat{\mathbf{L}}^+_{\mathbf{t}xy}(\mathbf{x} - \mathbf{x}^*) \qquad (13)$$

Where $\mathbf{x}^*$ is the desired position in the image space of the center of gravity of the loop, $\widehat{\mathbf{L}}_{\mathbf{t}xy}$ is the estimated image jacobian and β is a proportional gain. In order to chose a suitable estimation of the image jacobian, we can study the stability domain of the control law (13). It is well known that the control law (13) is asymptotically stable if $\mathbf{L}_{xy}\widehat{\mathbf{L}}^+_{xy}$ is positive. We have thus to study the positiveness of the eigenvalues of $\mathbf{L}_{\mathbf{t}xy}\widehat{\mathbf{L}}^+_{\mathbf{t}xy}$ ($= \mathbf{L}_{\mathbf{t}xy}\widehat{\mathbf{L}}^{-1}_{\mathbf{t}xy}$ in this case):

$$\mathbf{L}_{\mathbf{t}xy}\widehat{\mathbf{L}}^{-1}_{\mathbf{t}xy} = \frac{\alpha}{\widehat{\alpha}}\mathbf{R}_{xy}\widehat{\mathbf{R}}^{-1}_{xy} \qquad (14)$$

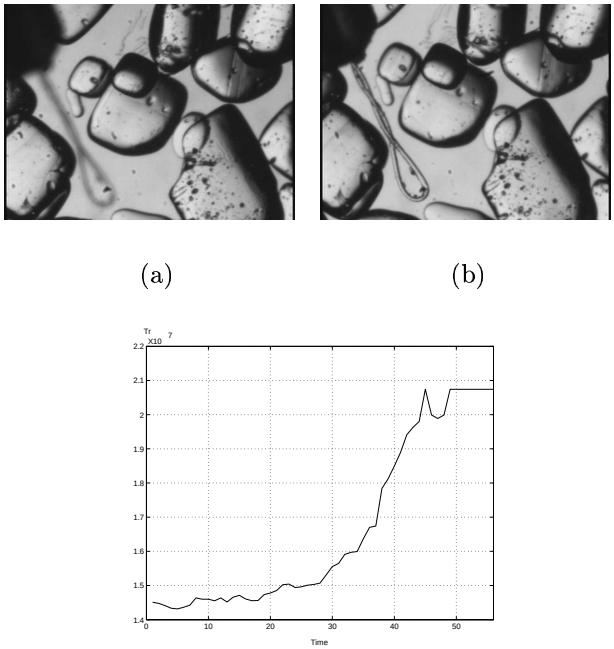
where $\widehat{\alpha}$ and $\widehat{\mathbf{R}}_{xy}$ are estimated values of g and $\mathbf{R}_{xy}$. According to (14), the eigenvalues of $\mathbf{L}_{xy}\widehat{\mathbf{L}}^{-1}_{xy}$ are the real part of the eigenvalues of $\frac{\alpha}{\widehat{\alpha}}\mathbf{R}_{xy}\widehat{\mathbf{R}}^{-1}_{xy}$ which are $(\frac{\alpha}{\widehat{\alpha}}e^{i\theta}, \frac{\alpha}{\widehat{\alpha}}e^{-i\theta})$ where θ is the angle of the rotation $\mathbf{R}_{xy}\widehat{\mathbf{R}}^{-1}_{xy}$. This means that if $\widehat{\alpha}$ is chosen as a positive scalar, the system is locally asymptotically stable, for any rotation axis, if :

$$-\frac{\pi}{2} < \theta < \frac{\pi}{2}$$

Judging from this stability study, we can easily fix $\widehat{\mathbf{L}}_{xy} = \widehat{\alpha}\widehat{\mathbf{R}}$ to a constant value while ensuring the convergence of control. For our application, $\widehat{\alpha}$ is fixed to the value given by the manufacturer. The estimated rotation matrix $\widehat{\mathbf{R}}_{xy}$ is fixed as the closest rotation matrix to $\frac{1}{\widehat{\alpha}}\widehat{\mathbf{L}}^0_{xy}$ where $\widehat{\mathbf{L}}^0_{xy}$ is the interaction matrix numerically computed off-line by observing the repercussion of the robot motion in the image space.

6 Experimental results

The proposed methods have been tested on our esperimental platform (see Figure 1). The specified visual task consists in a positioning task of the grasping loop with respect to a crystal protein. First, the focus of the grasping loop is performed. The initial defocused image is given in Figure 5(a) and the in focus image is illustrated by the Figure 5(b). The tenengrad measure is given by the Figure 5(c). After that the focusing task was achieved the grasping loop is visually guided to its desired position. The images corresponding to the desired and final configurations are given in Figures 6(a) and 6(b) respectively (the initial and desired position of the center of gravity of the grasping loop are represented by white disks). The figure 6(c) gives the error on the coordinates of the center of gravity of the grasping loop between its current and its desired location in the image. The convergence of the coordinates to their desired value demonstrates the correct realization of the task. The final errors on the coordinates is less than 1 pixel that corresponds to an positionning accuracy beyond $10^{-1}\mu m$ The computed control law is given in Figure 6(d).



(a) (b)

(c)

Figure 5: *Focusing the loop: (a) defocused image, (b) focused image, (c) Tenengrad measure versus time*

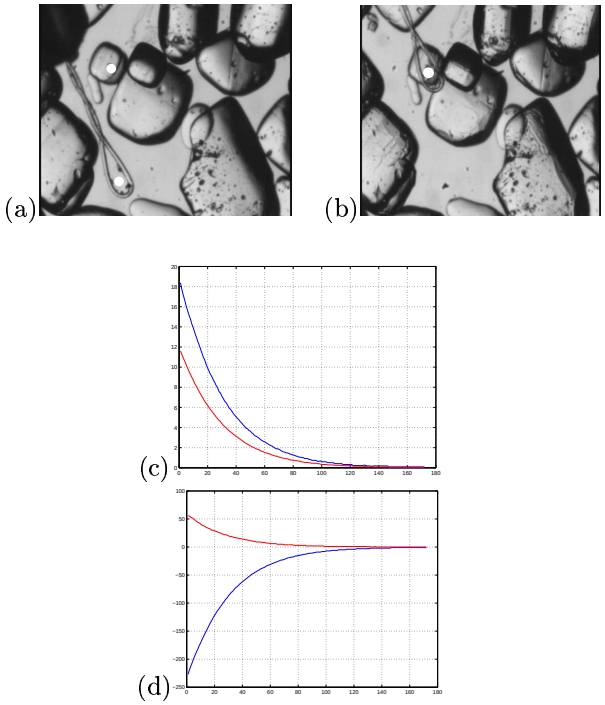


Figure 6: *(a) initial image, (b) final image, (c) control vector ($\mu m/s$), (d) error in the image space (pixel versus time)*

7 Conclusion

The development of visually guided strategies applied to high resolution optical systems show promise in overcoming a technology barrier to the automated micromanipulation of biological cell. In this paper, we have investigated and experimentally validated the use of uncalibrated visual servoing techniques to position a loop mounted on a micromanipulator with respect to a protein crystal observed by a camera mounted on an optical microscope. The experimental results show that the visual feedback can provide the required precision to achieve the fixed task. This work is the first step on the design of an integrated system that will be able to visually isolate individual proteins in a culture, to move customs instruments to contact the isolated proteins, to grasp the proteins and transport them for further processing.

Acknowledgement

This work was supported by a grand to the Northeast Structural Genomics Consortium from the Protein Structure Initiative of National Institutes of Health (P50 GM62413).

References

[1] P. K. Allen, A. Timcenko, Y. B., and P. Michelman. Automated tracking and grasping of a moving object with a robotic hand-eye system. *IEEE Trans. on Robotics and Automation*, 9(2):152–165, April 1993.

[2] F. Chaumette. Potential problems of stability and convergence in image-based and position-based visual servoing. *The Confluence of Vision and Control, D. Kriegman, G. Hager, A. Morse (eds), LNCIS Series, Springer Verlag*, 237:66–78, 1998.

[3] J. T. Feddema and T. R. Christenson. Parallel assembly of high aspect ratio microstructures. In *SPIE Conference on Microrobotics and Micromanipulation*, pages 153–164, Boston, USA, September 1999.

[4] J. T. Feddema, R. Simon, M. Polosky, and T. R. Christenson. Ultra-precise assembly of microelectromechanical systems MEMS components. April 1999.

[5] G. Flandin, F. Chaumette, and E. Marchand. Eye-in-hand / eye-to-hand cooperation for visual servoing. In *IEEE Int. Conf. on Robotics and Automation*, San Francisco, CA, April 2000.

[6] K. Hashimoto. *Visual Servoing : Real Time Control of Robot Manipulators Based on Visual Sensory Feedback.* World Scientific Series in Robotics and Automated Systems, Vol 7, World Scientific Press, Singapor, 1993.

[7] S. Hutchinson, G. Hager, and P. Corke. A tutorial on visual servo control. *IEEE Trans. on Robotics and Automation*, 12(5):651–670, October 1996.

[8] L. C. Martin. *The theory of the Microscope.* American Elsevier Publishing Compagny, 1966.

[9] P. Martinet, N. Daucher, J. Gallice, and M. Dhome. *Robot control using monocular pose estimation.* Worshop on New Trends in Image-based Robot Servoing, IROS'97, Grenoble, 1997.

[10] B. J. Nelson, Y. Zhou, and B. Vikramaditya. Sensor-based microassembly of hybrid mems devices. *IEEE J. of Control System*, pages 35–45, December 1998.

[11] V. L. Niennaber, P. Richardson, V. Klighover, J. Bouska, V. L. Giranda, and J. Geer. Discovering novel ligands for macromolecules using x-ray crystallographic screening. *Nat. Biotechnol.*, 18:1105–1108, 2000.

[12] R. F. Service. The automated approach to protein structure. *Science 285*, 1345, 1999.

[13] M. Subbarao, T. Choi, and A. Nikzad. Focusing techniques. Technical Report 92.09.04, Dept. of E.E, SUNY Stony Brook, 1992.

[14] C. L. Verlinde. Structure-based drug design: progress, results and challenges. *Structure 2*, pages 577–587, 1994.

[15] L. Weiss, A. Sanderson, and C. Neuman. Dynamic sensor-based control of robots with visual feedback. *IEEE Journal of Robotics and Automation*, 3(5):404–417, October 1987.

[16] W. Wilson, C. W. Hulls, and G. Bell. Relative end-effector control using cartesian position-based visual servoing. *IEEE Trans. on Robotics and Automation*, 12(5):684–696, 1996.