Proceedings of the 1997 IEEE
International Conference on Robotics and Automation
Albuquerque, New Mexico - April 1997

Visually Guided Microassembly Using Optical Microscopes and Active Vision Techniques

Barmeshwar Vikramaditya Bradley J. Nelson

Department of Mechanical Engineering
University of Illinois at Chicago
Chicago, Illinois 60607

Abstract

Robust micromanipulation strategies are needed in order to overcome a potential technology barrier to the commercial development of complex microelectromechanical (MEMS) devices. Development of robust strategies requires that two critical issues be addressed that are not applicable to the traditional macro manipulation domain. These two issues include the much higher relative positioning accuracy required, and the vastly different mechanics of manipulation. Both of these differences suggest the need to use micromanipulation strategies guided by real-time sensor feedback. In this paper, we present experimental results that investigate the use of high resolution optical systems with controllable intrinsic and extrinsic parameters for microassembly applications. Results of two active vision strategies, depth-from-defocus and visual servoing, using an optical microscope are presented. Depth-from-defocus results indicate a repeatability of 8μm (just one order of magnitude greater than the wavelength of light) is achievable. Visual servoing results demonstrate visually servoed motion over large scales of motion with submicron repeatability being achieved. The ability to servo over large scales is made possible by the use of a unique formulation of the image Jacobian for an optical microscope.

1. Introduction

If a microdevice must be made of different materials, has a complicated geometry, or is manufactured using incompatible processes, assembly of the device is required. Although automatic assembly has been a research issue for over two decades, almost all of this research has been in a macro domain. Important differences between manipulation in macro and micro domains exist and must be directly addressed in order to develop robust micromanipulation strategies.

The most obvious difference is the required positional accuracy of automatic assembly mechanisms. In the macro domain, accuracies in the range of 10^{-3} inches can be achieved with sensor-less manipulators. In the micro domain (part dimensions within micron to millimeter ranges), submicron (10^{-4} to 10^{-5} inches) accuracies are often required. These accuracies are beyond the calibration range of conventional open loop precision assembly devices used in industry [15].

The second major difference is in the mechanics of object interactions. In a micro domain, gravity is often negligible and electrostatic, surface tension, and Van der Waals forces dominate [1][2]. Currently, the mechanics of micromanipulation are poorly understood, and thus the results of simple sensorless manipulation strategies are unpredictable.

These two critical differences between macro and micro domains indicate that sensor-based manipulation strategies must be developed for even the simplest micromanipulation strategies, such as the alignment tasks that are prevalent in microassembly. In a macro domain, visually guided manipulation has been shown to effectively compensate for errors in the calibration of camera-lens systems, manipulators, and workspaces. Unfortunately (for visual servoing researchers), manufacturing engineers usually prefer the cost of strongly calibrated parts handling systems to the complexity of vision systems. However, in a micro domain precise calibration is highly dependent on precisely modeled kinematics which are subject to thermal growth errors. Two common techniques for compensating for thermal errors include either the use of expensive cooling systems, or waiting hours for the thermal equilibrium of the device to stabilize. Slocum [15] points out that "thermal growth errors are typically the most difficult to control and compensate [in precision machine design]." Because these types of factors greatly affect the cost and reliability of precision assembly machines, real-time vision feedback can be used effectively and economically as a component of a microassembly system. In addition to thermal growth, it should also be possible to compensate for some characteristics of poorly modeled micromanipulation mechanics through vision feedback, as is currently done with manual microassembly.

In this paper, we investigate active vision issues that apply to the use of optical microscopes for visually guided microassembly. A high resolution optical microscope is used to experimentally determine the performance of two active vision strategies, depth-from-defocus and visual ser-

voing. In Section 2, we discuss previous work in micromanipulation. In Section 3, we present our visual servoing formulation, and in Section 4 we present our depth-from-defocus formulation. Section 5 describes experimental results from the two strategies, and Section 6 summarizes our work.

2. The Mechanics of Micromanipulation

The mechanics of manipulation in the macro world are predictable, to a degree. For example, when a gripper opens, forces due to gravity cause the part drop. This predictability has enabled the success of many complex sensorless manipulation strategies. In the micro world, forces other than gravity may dominate. For example, [2] describes a microassembly scenario in which electrostatic forces dominate and cause a part to "jump" into a gripper before contact actually occurs. As the gripper opens to place the micropart at its goal, the part may stick to the gripper fingers and fails to remain at the desired location. If humidity in the room happens to be high, surface tension effects can dominate gravitational forces, and the part would also remain stuck to the gripper [1]. It is estimated that for parts with major dimensions below 100μm, gravity is negligible and other forces dominate [1]. However, this is only a rough estimate and depends on several factors, such as mass density, surface roughness, humidity, part geometry, electrical grounding, etc. The more accurate story is that micromanipulation is a very complex process. Although particular forces can be defined, their effect on the process can only be roughly estimated. For this reason, the development of micromanipulation strategies must incorporate vision feedback to correct for poorly modeled micromanipulation mechanics.

Three forces, electrostatic, surface tension, and Van der Waals, tend to dominate interactions between microparts. Some of these forces may be used to advantage in micropart manipulation. For example, controlled variations in electric potential between objects can be used for grasping. Similarly, surface tension may be controlled through microtubes and microheaters. Various micromanipulation techniques have been proposed [1][2][12][17]. This paper focuses on alignment strategies using visual servoing and depth-from-defocus techniques that entail the use of an active optical microscope.

3. Visual Servoing with an Optical Microscope

3.1. Visual Servoing Formulation

In formulating the visual servoing component of our system, the Jacobian mapping from task space to sensor space is needed. We desire a Jacobian for the camera-lens system of the form

$$\dot{x}_S = J_v(\phi)\dot{X}_T \qquad (1)$$

where $\dot{x}_S$ is a velocity vector in sensor space; $J_v(\phi)$ is the image Jacobian matrix and is a function of the extrinsic and intrinsic parameters of the vision sensor as well as the number of features tracked and their locations on the image plane; and $\dot{X}_T$ is a velocity vector in task space.

3.2. Derivation of the Image Jacobian

Past work in visual servoing has assumed a pinhole camera model in deriving the image Jacobian [13][6][10] as shown in Figure 1. A feature on a manipulated object is observed on the image plane with respect to the camera frame {C} at ${}^{C}\boldsymbol{P}$, but is controlled with respect to the manipulator's task frame {T} at ${}^{T}\boldsymbol{P}$. This model is inappropriate for an optical microscope. Instead, the ray diagram for a typical optical microscope is shown in Figure 2.. The intermediate image is projected at a distance g behind the posterior principal focus of the objective. By similar triangles:

$$h' = (h_1 g)/f_o' \qquad (2)$$

where f_o' is the posterior objective focal length, and h_1 is the object size in the task frame.

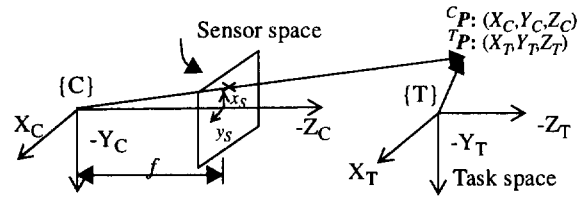

Figure 1. Task frame-camera frame definitions with a pinhole camera model.

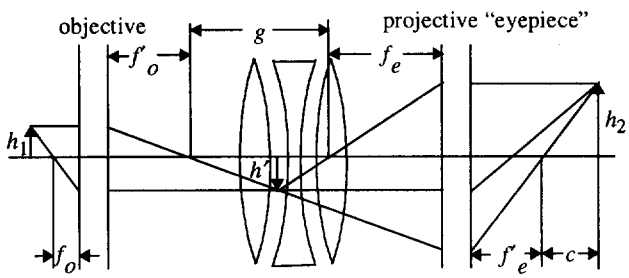

Figure 2. Ray diagram for the optical system.

This image is subsequently viewed by the projection eyepiece and projects the image after second stage magnification onto the CCD receptor that is lying at a distance of c behind the posterior principal focal plane of the eyepiece. Thus, the linear magnification is given by similar triangles as:

$$h_2 = (h'c)/f_e' \qquad (3)$$

and the total linear magnification is given by

$$m = h_2/h_1 = (gc)/(f_o'f_e') \qquad (4)$$

where g is the optical tube length and is the distance between the posterior principal focal plane of the objective

and the anterior principal focal plane of the eyepiece. For typical microscopes g is a constant 160mm.

The coordinate frame for the camera is located at the center of the CCD array and the Z_C axis coincides with the optical axis of the lens system. Then we have

$$x_s = \left(\frac{gc}{s_x f_o' f_e'}\right) X_C \tag{5}$$

$$y_s = \left(\frac{gc}{s_y f_o' f_e'}\right) Y_C \tag{6}$$

where x_S and y_S are the projected coordinates in sensor space.

Mapping the motion in task space into sensor space optical flow we obtain by differentiating (5) and (6) with respect to time to get the optical flow equations

$$\dot{x}_S = \frac{gc}{s_x f_o' f_e'}\dot{X}_C + \frac{g\dot{c}}{s_x f_o' f_e'}X_C - \frac{gc}{s_x f_o' f_e'^2}X_C\dot{f}_e' - \frac{gc}{s_x f_o'^2 f_e'}X_C\dot{f}_o'$$

$$\dot{y}_S = \frac{gc}{s_y f_o' f_e'}\dot{Y}_C + \frac{g\dot{c}}{s_y f_o' f_e'}Y_C - \frac{gc}{s_y f_o' f_e'^2}Y_C\dot{f}_e' - \frac{gc}{s_y f_o'^2 f_e'}Y_C\dot{f}_o'$$

It can be seen that $c + f_e'$ is constant, and thus

$$\dot{c} = -\dot{f}_e' \tag{7}$$

Using (7) and the optical flow equations we get

$$\dot{x}_S = x_S\frac{\dot{X}_C}{X_C} - x_S\left[\frac{1}{f_e'} + \frac{1}{c}\right]\dot{f}_e' - x_S\frac{\dot{f}_o'}{f_o'} \tag{8}$$

$$\dot{y}_S = y_S\frac{\dot{Y}_C}{Y_C} - y_S\left[\frac{1}{f_e'} + \frac{1}{c}\right]\dot{f}_e' - y_S\frac{\dot{f}_o'}{f_o'} \tag{9}$$

If the camera frame undergoes translation and rotation with respect to the current frame and ${}^C V = \left[\dot{x}_C\ \dot{y}_C\ \dot{z}_C\right]^T$ and ${}^C\Omega = \left[\omega_{xC}\ \omega_{yC}\ \omega_{zC}\right]^T$ are the respective velocities then ${}^C\dot{P} = -{}^C V - {}^C\Omega \times {}^C P$ where ${}^C\dot{P}$ is the motion velocities due to camera frame motion. Expanding the expression gives

$$\dot{X}_C = -\dot{x}_C - \omega_{yC}Z_C + \omega_{zC}Y_C \tag{10}$$

$$\dot{Y}_C = -\dot{y}_C - \omega_{zC}X_C + \omega_{xC}Z_C \tag{11}$$

$$\dot{Z}_C = -\dot{z}_C - \omega_{xC}Y_C + \omega_{yC}X_C \tag{12}$$

Using (5), (6), (10), and (11) and substituting into (8) and (9) we get

$$\dot{x}_S = -\frac{m\dot{x}_C}{s_x} - \frac{m\omega_{yC}}{s_x}Z_C + \frac{s_y\omega_{zC}}{s_x}y_S - x_S\left[\frac{c + f_e'}{f_e' c}\right]\dot{f}_e' - \frac{x_s\dot{f}_o'}{f_o'}$$

$$\dot{y}_S = -\frac{m\dot{y}_C}{s_y} + \frac{m\omega_{xC}}{s_y}Z_C - \frac{s_x\omega_{zC}}{s_y}x_S - y_S\left[\frac{c + f_e'}{f_e' c}\right]\dot{f}_e' - \frac{y_s\dot{f}_o'}{f_o'}$$

From the above equations the Jacobian can be formulated as

$$\dot{x}_S = J_v\left[{}^C V\ {}^C\Omega\ \dot{f}_o'\ \dot{f}_e'\right]^T$$

where $\dot{x}_S = \left[\dot{x}_S\ \dot{y}_S\right]^T$, J_v represents the Jacobian for the system and contains system parameters such as focal lengths, pixel spacing, optical tube length and total linear magnification. This is similar to the results for an extended Jacobian proposed in [6] for a zoom lens. The final image Jacobian matrix is given as

$$J_v = \begin{bmatrix} -\frac{m}{s_x} & 0 & 0 & 0 & -\frac{Z_C m}{s_x} & \frac{y_s s_y}{s_x} & -\frac{x_s}{f_o'} & -\left[\frac{c + f_e'}{f_e' c}\right]x_s \\ 0 & -\frac{m}{s_y} & 0 & \frac{Z_C m}{s_y} & 0 & -\frac{x_s s_x}{s_y} & -\frac{y_s}{f_o'} & -y_s\left[\frac{c + f_e'}{c f_e'}\right] \end{bmatrix} \tag{13}$$

Generally several features are tracked. Thus, for n feature points the Jacobian is of the form.

$$J_v = \left[J_1(t)\ \ldots\ J_{n(t)}\right]^T \tag{14}$$

where $J_i(t)$ is the Jacobian matrix for each feature given by (13).

3.3. Visual Servoing Controller

The state equation for the visual servoing system is created by discretizing (1) and rewriting the discretized equation as

$$\boldsymbol{x}(k+1) = \boldsymbol{x}(k) + TJ_v(k)\boldsymbol{u}(k) \tag{15}$$

where $\boldsymbol{x}(k) \in R^{2M}$ (M is the number of features being tracked); T is the sampling period of the vision system; and $\boldsymbol{u}(k) = \left[\dot{X}_T\ \dot{Y}_T\ \dot{Z}_T\ \omega_{X_T}\ \omega_{Y_T}\ \omega_{Z_T}\right]^T$ is the manipulator end-effector velocity. The Jacobian is written as $J_v(k)$ in order to emphasize its time varying nature due to the changing feature coordinates on the image plane. The intrinsic parameters of the camera-lens system are constant for the experimental results to be presented.

The visual servoing controller is formulated based on controlled active vision [13]. The control objective of the system is to control end-effector motion in order to place the image plane coordinates of features on the target at some desired position. The desired image plane coordinates could be constant or changing with time. The control strategy used to achieve the control objective is based on the minimization of an objective function that places a cost on errors in feature positions and a cost on providing control energy.

$$F(k+1) = [\boldsymbol{x}(k+1) - \boldsymbol{x}_D(k+1)]^T\boldsymbol{Q}[\boldsymbol{x}(k+1) - \boldsymbol{x}_D(k+1)] + \boldsymbol{u}^T(k)\boldsymbol{L}\boldsymbol{u}(k) \tag{16}$$

This expression is minimized with respect to the current control input $\boldsymbol{u}(k)$. The end result yields the following expression for the control input

$$\boldsymbol{u}(k) = -(TJ_v^T(k)\boldsymbol{Q}TJ_v(k) + \boldsymbol{L})^{-1}TJ_v^T(k)\boldsymbol{Q}[\boldsymbol{x}(k) - \boldsymbol{x}_D(k+1)] \tag{17}$$

The weighting matrices $\boldsymbol{Q}$ and $\boldsymbol{L}$ allow the user to place more or less emphasis on the feature error and the control input. Their selection effects the stability and response of the tracking system. The $\boldsymbol{Q}$ matrix must be positive semi-definite, and $\boldsymbol{L}$ must be positive definite for a bounded response. Although no standard procedure exists for

choosing the elements of Q and L, general guidelines for a macro system can be found in [13], along with system models and control derivations that account for system delays, modeling and control inaccuracies, and measurement noise.

The measurement of the motion of the features on the image plane must be done continuously and quickly. The method used to measure this motion is based on an optical flow technique called Sum-of-Squares-Differences (SSD). A more complete description of the algorithm and its implementation can be found in [10].

4. Depth-From-Defocus for Optical Microscopes

4.1. Depth of focus and Depth of field

Depth-from-defocus has been studied extensively as a technique for recovering depth estimates from the limited depth-of-field exhibited by optical lenses, for example [4][9][14]. In the past, active vision researchers have assumed that depth-of-field is determined by pixel size. This is because magnification is relatively low for macro-domain tasks. For example, in [16] the depth-of-field formulation is given as

$$\Delta = Daf\left(\frac{1}{af - p(D-f)} - \frac{1}{af + p(D-f)}\right) \tag{18}$$

where Δ is the depth-of-field, D is the focus distance, a is the lens aperture, f is the lens focal length, and p is the minimum of the x and y pixel dimensions on the CCD array.

The above formulation is valid for optical systems well approximated by a pinhole camera model. However, for low aperture systems, i.e. large magnification systems such as optical microscopes, diffractive effects between wavelengths of light dominate, and Δ is given as

$$\Delta = \frac{\lambda_0 n}{2A^2} + \frac{n}{7mA} \tag{19}$$

where λ_0 is the wavelength of light in a vacuum, n is the diffractive index of the lens, A is the numerical aperture of the lens systems, and m is the magnification of the optical system [8]. The significance of this equation is that extremely small depths-of-field result, i.e. Δ is on the order of the wavelength of light. This provides the opportunity to calculate depth from active microscope systems with accuracies approaching the wavelength of light, which can be used to advantage for microassembly.

4.2. Focus Measure

In order to estimate depth-from-defocus a focus measure is needed. Many focal measures have been proposed in the past, and an excellent comparison of many of these can be found in [7]. A technique based on an analysis of the frequency content of image patches is used in order to determine where the image is in maximum focus. This technique is described in detail in [11] and is based on an analysis of the Fourier Transform of image patches.

5. Experimental Results

5.1. Hardware Setup

Experiments were conducted with a Mitutoyo FS60 optical microscope and Wentworth MP950 IC probe station shown in Figure 3. The probe station was retooled for motion control using high precision Kollmorgen brushless DC motors. Image processing and visual servoing control calculations were performed with a vision system consisting of a digitizer and multiple TMS320C40 DSP's. The motors were controlled using a Delta Tau PMAC-PC servo card. Both the vision system and the PMAC communicate directly over a PC ISA bus. The servomotor PID loop is closed at 2.2KHz on the PMAC card using encoder feedback. The vision system is able to track up to 5 16x16 feature templates at 30Hz. The visual servoing loop is closed at this frequency.

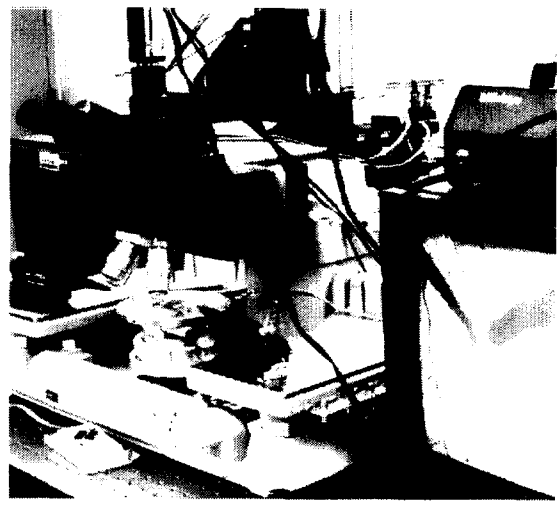

Figure 3. Experimental setup

5.2. Visual Servoing Results

The step response of the visual servoing system with four different objective lenses was tested. Overall results are shown in Table 1. Images of a miniature scanning electron microscope electrostatic column [3] which was servoed by the positioning stage of the probe station are shown in Figure 4. Settling time for each of the four trials was quite similar, approximately 0.1s, as shown in Figure 5. This demonstrates that the model of the microscope developed is an accurate representation of the optics. Optimal performance was achieved by tuning the values of the diagonal terms in the control gain matrix Q in (17). Relatively small adjustments in Q between magni-

fications were required in order to achieve optimal performance. Critically damped response was easily achieved.

It should be noted that, although the 500X lens has a greater magnification than the 200X lens, they both have identical numerical apertures. Therefore, the resolution with which objects can be positioned relative to one another is of a similar magnitude. This resolution constraint was manifested in the deadband that was purposely introduced in the vision tracker in order to reduce oscillations in commanded motion due to image tracking noise.

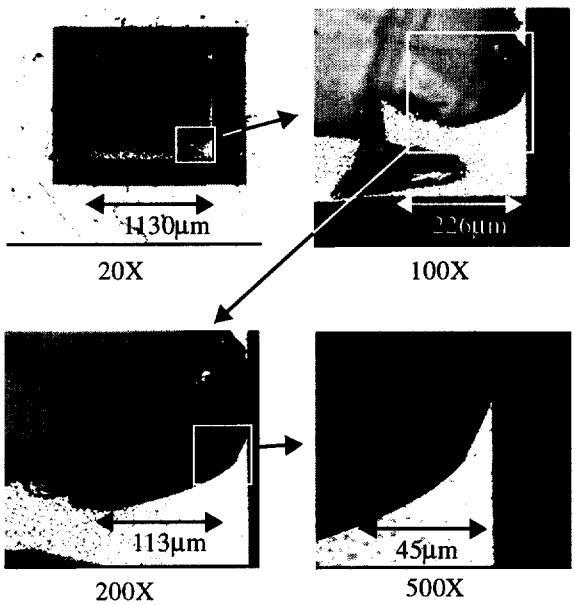

Figure 4. Images of miniature scanning electron microscope electrostatic column [3] used for tracking at four different magnifications.

TM	NA	Gain	R (μm)	Δ (μm)	Step (pix)	Step (μm)	Rep. (μm)
20X	0.055	90	5.0	220.8	50	218.75	8.75
100X	0.28	75	1.0	8.6	50	43.21	1.73
200X	0.42	100	0.7	3.3	50	21.875	0.875
500X	0.42	75	0.7	2.2	50	8.75	0.35

Table 1: Visual Servoing Results with Four Different Objective Lenses. TM:total magnification, NA:numerical aperture, Gain:Experimentally tuned value of diagonal elements of Q, R:resolution, Δ:depth-of-field, Step:commanded step size in pixels and microns, Rep:estimated repeatability of positioning system based on controller deadband.

5.3. Depth-from-Defocus Results

Results from the depth-from-defocus strategy are shown in Figure 6. The depth was determined by moving the probe stage along Z_C in steps of 8.6μm. This step size was

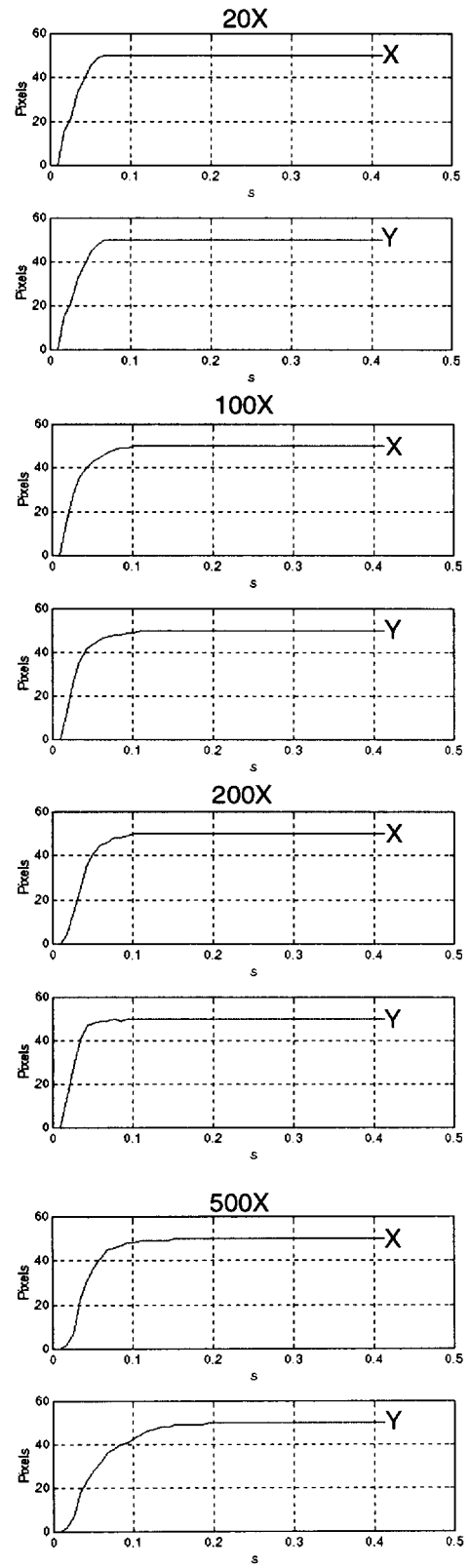

Figure 5. Step response of micro-visual servoing system at four different magnifications.

determined to be the depth-of-field according to (19) when the lens configuration results in a total magnification of 100x. A frequency based focal measure was used to determine dense depth maps for images of a 6 in. mil file shown below. A single serration on the file (approximately 220μm in height) was viewed under the microscope and multiple 400x6 pixel size images were taken at varying levels of focus. The depth map demonstrates the applicability of this technique for relative positioning of objects along the optical axis. This adds another degree of freedom to the sensor-based positioning system.

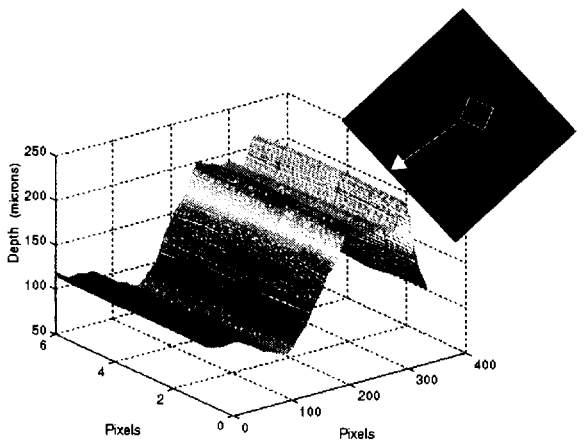

Figure 6. Depth map for a single serration (shown in the white box) of the 6 in. mil file pictured in the inset.

6. Conclusion

Active vision strategies applied to high resolution optical systems show promise in overcoming a technology barrier to the automated microassembly of hybrid MEMS devices. In this paper, we have theoretically and experimentally investigated the use of active vision techniques with an optical microscope. Visual servoing results demonstrate robust control down to submicron repeatability. This highly precise repeatability was achieved with visually servoed motion at relatively high speeds in the mm/s range. Depth-from-defocus techniques demonstrate that it is possible to create depth maps with a repeatability of approximately 8μm, and therefore possible to use this technique for aligning objects along the optical axis in real-time. The goal of this work is to develop microassembly strategies for hybrid MEMS devices that compensate for modeling uncertainties inherent in the micro-domain, such as thermal growth, humidity effects, electrostatic forces, etc.

Acknowledgments

This research was supported by the National Science Foundation through Grant Numbers IRI-9612329 and CDA-9616757. We would like to thank Alan Feinerman for providing a miniature SEM assembly.

References

[1] F. Arai, D. Ando, T. Fukuda, "Micro Manipulation Based on Micro Physics-Strategy Based on Attractive Force Reduction and Stress Measurement," *Proc. 1995 IEEE/RSJ Int. Conf. on Intelligent Robots and Sys. (IROS95)*, 2:236-241, Pittsburgh, August 5-9, 1995.

[2] R.S. Fearing, "Survey of Sticking Effects for Micro Parts Handling," *Proc. 1995 IEEE/RSJ Int. Conf. on Intelligent Robots and Sys. (IROS95)*, 2:212-217, Pittsburgh, August 5-9, 1995.

[3] A.D. Feinerman, D.A. Crewe, D.C. Perng, S.E. Shoaf, and A.V. Crewe, "Sub-centimeter micromachined electron microscope," J. Vac. Sci. Technol. A, 10(4), pp. 611-616, 1992.

[4] P. Grossman, "Depth from Focus," *Pattern Recognition Letters*, 5, pp. 63-69, 1987.

[5] B.K.P. Horn, *Robot Vision*, MIT Press, 1986.

[6] K. Hosoda, H. Moriyama, M. Asada, "Visual Servoing Utilizing Zoom Mechanism," *Proc. 1995 IEEE Int. Conf. on Robotics and Automation*, pp. 178-183, 1995.

[7] E. Krotkov, "Focusing," *Int. J. Computer Vision*, 1, pp. 223-237, 1987.

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

[9] S.K. Nayar, Y. Nakagawa, "Shape from Focus," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 16, pp. 824-831, 1994.

[10] B. Nelson, N.P. Papanikolopoulos, and P.K. Khosla, "Visual servoing for robotic assembly," *Visual Servoing—Real-Time Control of Robot Manipulators Based on Visual Sensory Feedback*. ed. K. Hashimoto. River Edge, NJ:World Scientific Publishing Co. Pte. Ltd. pp. 139-164, 1993.

[11] B. Nelson and P.K. Khosla, "Integrating Sensor Placement and Visual Tracking Strategies," *Experimental Robotics III: The Third International Symposium, Kyoto, October 28-30 1993*, eds. T. Yoshikawa and F. Miyazaki, Springer-Verlag London Ltd., London, pp.169-181, 1994.

[12] K. Ozawa, S. Ogawa, H. Ishida, and Y. Hattori, "High-speed Measuring Equipment of Fiber Core Position of Optical Fiber Array Using Piezo Actuator," *Proc. 1995 IEEE Int. Conf. on Robotics and Automation*, pp. 672-678, 1995.

[13] N.P. Papanikolopoulos, Nelson, B. and Khosla, P.K., "Full 3-d tracking using the controlled active vision paradigm," *Proc. 1992 IEEE Int. Symp. on Intelligent Control (ISIC-92)*, pp. 267-274, 1992.

[14] A. Pentland, "A New Sense of Depth of Field," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 9(4), pp. 523-531, 1987.

[15] A.H. Slocum, *Precision Machine Design*, Prentice Hall, 1992.

[16] K. Tarabanis, R.Y. Tsai, and P.K. Allen, "Satisfying the Resolution Constraint in the "MVP" Machine Vision Planning System," in *Proc. of the 1990 Darpa Image Understanding Workshop*, 850-860, 1990.

[17] Y. Yamagata and T. Higuchi, "A Micropositioning Device for Precision Automatic Assembly using Impact Force of Piezoelectric Elements," *Proc. 1995 IEEE Int. Conf. on Robotics and Automation*, pp. 666-671, 1995.