2007 IEEE International Conference on
Robotics and Automation
Roma, Italy, 10-14 April 2007

Intuitive Command of Manipulators in Micro-scale Tasks

Andrew P. Shacklock, Member IEEE, Matthew C. Pritchard, Hong Luo, Etienne Burdet and Wei Lin, Member IEEE.

***Abstract*— Manipulation tasks performed under microscopes are fatigue inducing on account of the difficulty in visualisation of the scene and hand-eye coordination. This work concentrates on providing assistance to the operator in the form of a visual-haptic interactive system which allows intuitive command of the manipulator motion axes. Specific examples of visualisation tools are presented which compensate for problems such as: spatial disorientation, loss of depth perception, and occluded data.**

I. Introduction

The aim of the work presented in this paper is to give operators the freedom to select the most convenient mode of interaction when conducting tasks at the micro-scale. Obtaining a clear view of the task scene is often very difficult for many reasons such as: occlusions, limited depth-of-field, narrow field-of-view, poor contrast, etc. Even if a view is achieved, continuous operation induces fatigue and therefore operator errors. However, within the electronics and semiconductor industries for example, numerous repetitive tasks such as inspection and test are performed under microscopes. Beyond visual inspection, in the biomedical sector there are tasks that involve hand-eye coordination such as micro-assembly, micro-injection and micro-manipulation. There is a clear need to provide assistance to the operator to reduce fatigue, improve productivity and reduce errors.

The loss of depth perception and the problems of spatial disorientation make micro-manipulation tasks extremely challenging if vision is the predominant sensory mode, prompting the need for haptic feedback. However, little is known about designing an operator interface for different micro-scale tasks and how to configure it optimally to suit different operator needs. Given these numerous problems and task/operator variability, we have adopted the philosophy that the operator should be able to customise the system and switch modes of operation rapidly.

This work is part of a Collaborative Research Project (CRP) co-funded by the Agency for Science Technology and Research (A*Star) and the National University of Singapore.

A. Shacklock, M. Pritchard, H Luo and W. Lin are with the Singapore Institute of Manufacturing Technology (SIMTech). E. Burdet is with Imperial College London and the National University of Singapore.

II. Related work

This research builds on our earlier work with multi-scale, multi-view vision to include haptic-feedback motion control. This paper presents an experimental system which will be used to quantify the benefits of combined visual-haptic sensing in micro-assembly tasks. Although techniques from computer vision feature strongly, the details of these algorithms are not the prime focus (there is an existing body of literature on the subject of machine vision and guidance in micro-tasks [1-3]). Our current concern is how the data from sensors (such as cameras) should be presented to operators so that they can interact with a micro-scale environment in a comfortable and productive manner. Ferreira [4] has addressed similar problems at the micro- and nano-scale although much of this work is focused on virtual environments. Our emphasis is on real images for visualisation and employs image manipulation to enhance the understanding of degraded data; this is more akin to Augmented Reality (AR). Our work so far has concentrated on tasks performed under optical microscopes; applications using electron microscopes and scanning probe microscopes introduce many extra challenges and considerations, for a review of computer vision for nano-scale imaging see [5].

A key issue identified in this paper is the need to quantify the benefits of using human-machine interaction tools. At the micro-scale, there are many specific pitfalls that must be avoided but some of the general principles observed by Steinfeld et al on human-in-the-loop operations provide useful starting points [6]. Usability is critical to acceptance and so we present some real examples of our visualisation techniques that have been implemented for commercial systems. The aim is to demonstrate that the system's underlying principles are easy to understand and will permit straightforward implementation of new ideas.

III. Visual-Haptic System

The major components of the visual-haptic interface are shown in Fig. 1. At the centre is a standard desktop PC with video and sound outputs, perfectly adequate for the relatively modest run-time computations. This is attached to cameras, a

1-4244-0602-1/07/$20.00 ©2007 IEEE.

microscope, hand-held input devices and a motion stage via suitable interfaces. The video display is able to show single, multiple or fused views of the scene depending on the user's requirements. The options are:

- Microscope view.
- Perspective view of the 'global' workspace.
- Graphical model view.
- Mosaic image.
- Fused microscope/perspective view.

The operator can switch instantly between these options or open multiple windows so as to obtain the best view for navigating around the workspace. Any of the views can be used to issue a motion command in image coordinates by using the input device to position the cursor within the viewing window and clicking on a desired new position. The choice of input device is flexible, for example: standard mouse, joystick, play station controller, touch screen, and customised haptic devices. The game controller is particularly useful because the decoupling of two joysticks is ideal for many microassembly and insertion tasks which require 2D alignment followed by controlled motion along a third axis; the multiple 'fire' buttons allow easy switching between interface modes and options.

Camera specifications are not stringent, although synchronisation of images from multiple cameras is beneficial. In the laboratory development system, a Matrox Solios frame grabber was used to control the timing of analogue cameras. More recently, digital cameras on firewire connections have proven to be easy to synchronise and install, avoiding the need for a frame grabber. The quality of the camera attached to the microscope is of most significance; a progressive scan device with controllable shutter speed is advantageous because the image velocity in the microscope view can be very high, causing image blur and interlacing artefacts. Better quality results are obtained with monochrome cameras but in commercial applications

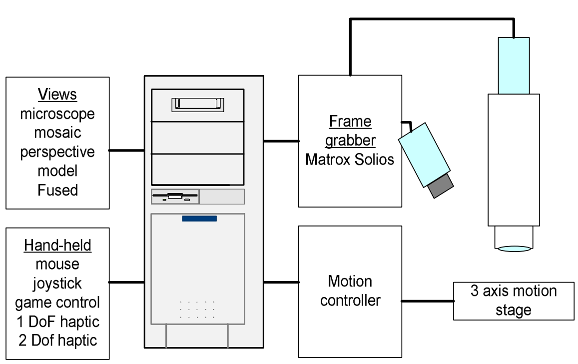

Fig. 1. Major components of the visual-haptic interaction system for microscope based tasks

Fig.2. Interface to teleoperate and provide haptic feedback in micromotion.

end-users prefer colour images despite the compromise on resolution (with single chip sensors) and the threefold increase in computation and memory loading. Although computation time is important, a more important (but related) consideration is the effect of latency on overall system performance.

Two customised haptic input devices have been built: a 2DoF planar device (Fig. 2) conceived to control horizontal motion and force and a single axis rotary device. The reasons for this decoupling are the same as the advantages cited for the game controller configuration: assembly tasks are frequently decomposed into alignment followed by insertion. This choice of a two-handed configuration is based on intuition but the collection of quantitative evidence to test this assumption is the focus of an ongoing collaborative research programme. The rotary device allows fine fingertip control combined with unlimited travel at high speeds. The planar device is dedicated to handle micro-motion, based on a design provided by McGill University [7]. Our device departs from the original McGill concept because it will be used for teleoperative micromotion rather than rendering of object surfaces. This interface produces a peak force of 500mN, a continuous force of 100mN and controls motion at a precision of 10µm. It is able to generate various haptic effects and dynamic environments [8].

IV. PRINCIPLES OF OPERATION

The visualisation system allows the user to express a motion command in the image coordinates of any of the viewing windows. These commands can be mapped to any of the other coordinate systems using simple planar projections

(2D homographies). The technique exploits the essentially planar nature of micro-scale tasks and has many advantages, notably:

- Seemingly complex sequences of transformations are combined into a single operation using the group theoretic properties of homographies.
- A mapping from one view to the next is easily solved by (initially) linear methods without knowledge of intrinsic or extrinsic camera parameters.
- Most importantly, the system self-calibrates so the operator is free to reposition the cameras at will.

Calibration involves matching corresponding features in different views and then solving the transformations. This process can be extended beyond the camera to solve the transformations required to map image motion to manipulator axis motion. In general, the transformations are projective so that the manipulator can be commanded from an oblique perspective view.

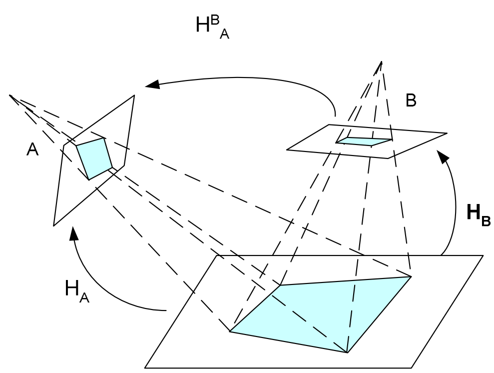

Fig. 3. Image transformations can be considered instances of a general linear projection. The reference plane induces a homography $H^B{}_A$ between the two image planes A and B.

A point **x** = $[x\ y\ 1]^T$ on a plane is transformed to a new point using a homography H. A subscript is used to indicate the point's coordinate system. The homography takes the point from its superscript to subscript coordinate system (see Fig 3).

$$x_b = H_b^a x_a \tag{1}$$

These principles are simple but in practice great care must be taken because the views have very different scales. This creates two major problems: very few or no shared features may be available in two views to solve the transformations; and there is potential for huge errors if a projection is extrapolated beyond the range of the original input data. The solution, which has been presented in previous publications [9,10], also exploits the properties of homographies; in brief, the system builds large high resolution mosaic images to create an image space in which data association can be performed. The mosaics are constructed by tracking the image motion of large sets of micro-scale artefacts. Eventually, these mosaics are large enough to detect features seen at the next scale up.

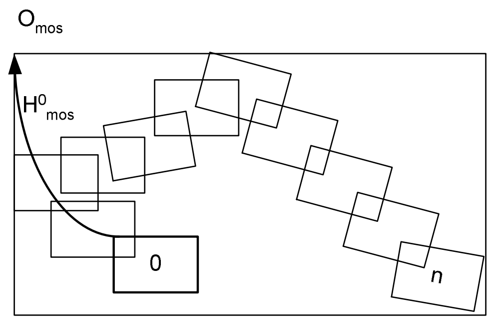

Fig. 4. As a sample moves under a microscope it can be tracked by analysing the image motion. This information is encoded in the transformations of the image reference frames and later can be used to reconstruct a mosaic image. $H^0{}_{mos}$ converts between the first image coordinate system and the coordinates of the mosaic image.

Fig. 4 summarises the principles of the mosaic space and the images from a microscope sequence. Equations 2 and 3 state that a transformation between two frames can be composed from the transformations in the chain of frames that link them. Note that the origin of a mosaic may drift with respect to the reference frame as the mosaic grows and so we need to keep track of this origin to construct a transformation $H^0{}_{mos}$.

$$H_0^n = H_0^1 \circ H_1^2 \cdots H_{n-1}^n = \prod_{i=1}^{n} H_{i-1}^i \tag{2}$$

$$H_{mos}^n = H_{mos}^0 \circ H_0^n \tag{3}$$

A graphical model is not only useful for visualisation and navigation, but also for data association: accumulating partial data and hypothesising about missing information. The model does not have to be known *a priori* – it can be generated from image data – nor does it have to be faithful to the sample's Euclidean geometry; we can accept projective distortion. In fact, considerable uncertainty in the geometry can be accommodated by starting from a hierarchical topological description of features and gradually replacing the links with transformations as coordinates are discovered. This is similar to solving a jigsaw puzzle. Although it is not a requirement, making the model a scaled version of the mosaic (microscope) coordinate system permits relation by similarity. Again, the similarity transformation is solved by matching features in the model to those extracted from the mosaic image. We can now transform coordinates from any microscope frame to the model space using eqn. 4:

$$H_m^n = H_m^{mos} \circ H_{mos}^0 \circ H_0^n \quad (4)$$

V. Key System Functions

Some simple examples will serve to illustrate various features of the system.

A. Virtual navigation using the mosaic or model

When the user clicks a location in the model window, a point x_m is generated in the model coordinate system. This point is transformed into a point x_{mos} in the mosaic system and highlighted on the mosaic image (Eqn. 5).

$$x_{mos} = H_{mos}^m x_m = \left(H_m^{mos}\right)^{-1} x_m \quad (5)$$

If so desired, the relevant portion of the mosaic surrounding x_{mos} can be cropped and expanded to fill the display as if it were digitally zoomed. Another user option is to create an artificial trajectory within the mosaic and reconstruct a 'virtual microscope' tour using a sequence of cropped regions of the mosaic. A limitation of this technique (in live navigation) is that the user is only able to view data that has been previously covered by the microscope: a wider view or map is needed so that the operator can identify new territory to explore, as explained in the next section.

B. Using the live-fused view

Navigating around a sample with a microscope is disorienting, especially when the sample is large with repetitive structures – imagine walking around a city whilst staring through a telescope. The system's perspective view helps to find points of reference but the resolution is too poor to conduct most tasks efficiently. The user has the option of switching between the microscope view and perspective view, or zooming in and out with the microscope view, to locate points of reference. Both of these methods are tiresome and unproductive.

A fused view (see Fig. 5) allows the user to combine surrounding detail and 3D cues with the finer resolution of the microscope image. Zooming is software generated (so instantaneous); switching from the fused view to either of its component views is not disorienting on account of the shared visual cues.

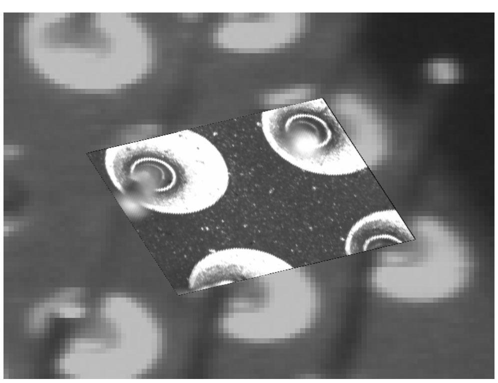

Fig. 5. Live fusion of a microscope image and a wide view. Note the differences in resolution and perspective.

The live video stream from the microscope camera is warped (transformed) so that it fits into its projected boundary on the reference plane of the perspective view. This view is combined with the live stream from the perspective camera to produce a synchronised composite image. Digital zoom is very simple as it only requires the modification of the warping transformation H^n_{pn} (a transformation from microscope frame n to its corresponding perspective frame n) and a shift and resizing of the perspective image. The on-line computations are extremely simple and fast so live fusion is achieved at standard frames rates even when using a modest PC specification. The transformation from a microscope pixel to its position on the reference plane of the perspective view is given by:

$$H_{p_n}^n = H_p^m \circ H_m^{mos} \circ H_{mos}^0 \quad (6)$$

This is valid if a) the transformation H_p^m (a linear transformation with 8 degrees of freedom) was solved for the same time instance as H_{mos}^0 and b) the relative imaging geometry remains fixed. These are reasonable constraints and it is safe to assume that $H_{p_n}^n$ remains constant for many tasks. In actuality, we require the inverse of this transformation because we need to map from each pixel on the perspective image x_p^i (within the projected boundary) to a location x'^i on the microscope image. (The intensity of this pixel is determined by interpolation of the pixels surrounding x'^i). With the addition of digital zoom the mapping is given by Eqn. 7.

$$x'^i = \left(H_z H_{p_n}^n\right)^{-1} x_p^i \quad (7)$$

The fused view seems somewhat strange because each image has a slightly different perspective, yet users show little difficulty in interpreting the information.

C. Recovering lost depth perception

Perhaps the most significant difficulty of using a microscope is the loss of depth perception. There is no perspective (the best cue is defocus), thus slowing down tasks that require contact, such as probing. The live-fused view has clear benefits because as an object approaches the region of interest it becomes visible in both views but has different projections (see Fig. 6)

This effect is an example of parallax and the disparity between the two projections is indicative of distance away from the plane of reference. Only if a point lies on the reference plane will its image coincide in both views and this helps the user gauge distance and guide the object to the plane. The effect is much more pronounced when viewing

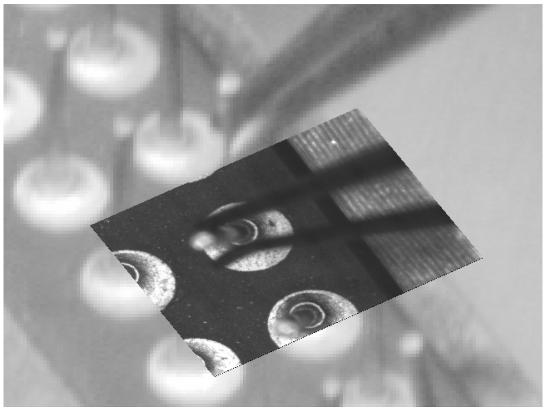

Fig. 6. The fused view can inform users of the position and height of the tool based on the parallax of the two images.

moving images, especially if the task involves hand-eye coordination.

D. Viewing through occluding structures

Another serious and common problem in micro-scale tasks is occlusion. To address this, in the standard device configuration a parallel projection is superimposed on an oblique view. This view from ‘above’ allows the user to see through 3D objects that occlude the oblique view of the work scene. For example, the microscope lens body can block many viewing angles. Fig. 7 presents two frames from a live sequence: the full view above a zoomed version. Note that zooming is not the same as conventional digital zoom because increasing zoom delivers increased resolution from the microscope.

E. Commanding the motion axes

The virtual navigation tasks are useful for reviewing data but in most productive manipulation tasks we need to see live views and move the sample by actuating a motion stage. The usual procedure is to use precision motion stages and calibrate these with respect to rigidly fixed cameras.

Our system relaxes these constraints, allowing lower cost equipment and freedom to place the cameras in the most convenient orientation. For the operator to command intuitively, image coordinates need to be converted into the manipulator’s coordinate system, therefore a calibration must be performed. This is achieved by commanding the axes to move each axis in turn by a set amount whilst recording the image stream from the microscope view. The streams of image data are solved to give a sequence of transformations. Comparing a frame taken before motion commenced and a frame taken after motion ceased enables the definition of a vector in image space.

VI. Verification and Further Work

The purposes of the visualisation tools created for this system are to make control of manipulation tasks easier and intuitive. This means that the user enters commands in a natural way without needing to understand the kinematics of the manipulators or the relationships between the various coordinate systems: the user simply clicks on a desired end position without having to perform any mental rotations or other complex spatial reasoning. However, the benefits of these visualisation features are also intuitive and difficult to quantify. This quantification of benefits needs to be carried out (and not just for scientific completeness). If the techniques are to be implemented on industrial equipment (and achieve commercial success) the production gains must be clear and outweigh any extra costs. This is not a simple matter because the usefulness of any technique depends on the individual user and the nature of the task. This problem is being addressed by commissioning a demonstrator that will be configured for a selection of representative industrial tasks including wafer inspection and test probing.

However for the use of haptics at the micro-scale, there remain certain fundamental issues (such as the effects of latency, the degrees of freedom required, and how many axes of force sensing are required) which will be addressed through a series of investigative experiments. The first aim of the experiments is to study the importance of certain negative factors and how they can degrade overall performance. The second aim is to study the relative contribution of sensory cues. Factors such as latency and

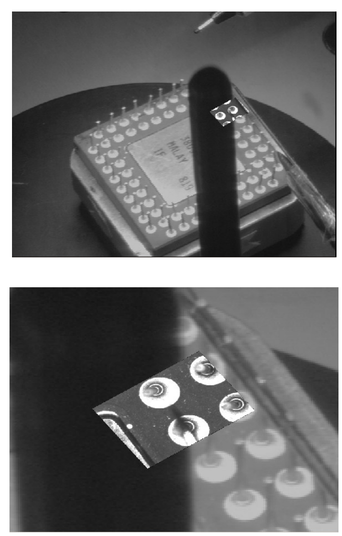

Fig. 7. Viewing through a structure. The lower figure is a zoomed version of the upper figure and the transition can be achieved instantly.

haptic noise will be quantified by artificially increasing their influence and quantifying the effect on task metrics. To satisfy the second aim, different modes will be simply switched on or off and/or the quality varied. The quality and

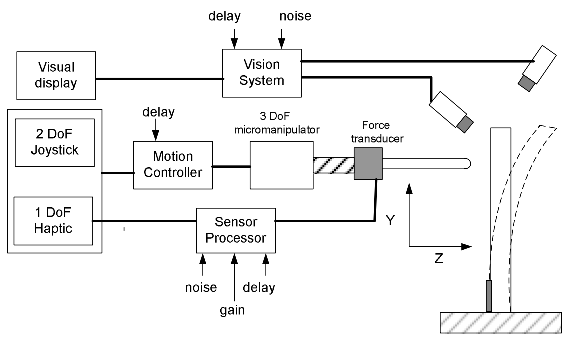

Fig. 8. An additional implementation of the visual haptic system is used to investigate fundamental issues that effect system performance. The rig is easily reconfigured to

influence of a mode can be varied by changing its strength and/or resolution. To allow meaningful study of complex problems, the experiments will commence on primitive tasks with haptic feedback along a single axis. The initial tasks will study

- Contact (e.g. probing)
- Peg-in-hole insertion (e.g. micro-assembly).

A description of an example contact task and reference to Fig. 8 is sufficient to illustrate the important principles of the experimentation. The subject (operator) is requested to make contact with a flexible beam without inflicting excessive force/deflection. The probe can be aligned in the X-Y plane and the axis for making contact is in the Z direction. The operator will have to use visual and/or haptic feedback (on the Z-axis) to decide when contact has been made. The success of the task will be measured in terms of time to complete and the force/deflection exerted on the beam.

Fig. 8 presents a schematic diagram of a suitable configuration of the experimental rig. This set-up is easily reconfigured to cater for more complex tasks and additional axes of haptic control. Many replicated tests will be run as well as variations in viewing angles and software configuration. So far the experimental rigs have been used to demonstrate concepts and explore phenomena. The next step is to take an independent assessment of the technology from research teams at Imperial College and the National University of Singapore who have a specific interest in human factors engineering.

VII. Conclusion

This work started with the premise that many operator controlled micro-scale tasks will benefit from the additional perceptive feedback afforded by haptic devices and certain visualisation tools. A visual-haptic micromanipulation experimental system commissioned in our laboratory was presented. We believe that the use of multi-view multi-scale visualisation will also alleviate particular problems of micro-scale navigation. Examples were presented of applications where problems of occluding structure, depth perception were addressed and we demonstrated how commands are issued in intuitive image space representations.

The paper concludes that objective tests involving typical users are essential for the verification of this technology. Further fundamental work needs to be conducted to verify various premises and assumptions underlying the design concepts of the 'intuitive' interfaces.

Acknowledgment

We thank Peter Preisig for the planar haptic device which was developed in the context of his diploma project. We acknowledge the valuable contributions of Teo Chee Leong in many aspects of the research described in this paper.

References

[1] H. Yamamoto and T. Sano, "Study of micromanipulation using stereoscopic microscope" *IEEE Transactions on Instrumentation and Measurement*, vol. 51, pp 182-187, Apr 2002.

[2] J. Wedekind, M. Boissenin, B.P. Amavasai, F. Caparelli and J.Travis, "Object Recognition and Real-time Tracking in Microscope Imaging" , *IMVIP 2006*, September 2006, Dublin, Ireland.

[3] Yesin, K.B.; Nelson, B.J., "Robust CAD model based visual tracking for 3D microassembly using image space potentials," *Robotics and Automation, 2004. Proceedings. ICRA '04. 2004 IEEE International Conference on* , vol.2, pp. 1868- 1873, April 26-May 1, 2004.

[4] A. Ferreira and C. Mavroidis, "Virtual reality and Haptics for Nanorobotics", *Robotics & Automation*, vol 13, no. 3, September 2006.

[5] E. Ribeiro and M. Shah, "Computer Vision for Nanoscale Imaging", *Machine Vision and Applications*, vol 17, pp147-162, 2006.

[6] A.M. Steinfeld, T.W. Fong, D. Kaber, M. Lewis, J. Scholtz, A. Schultz, and M. Goodrich, "Common Metrics for Human-Robot Interaction", *Human-Robot Interaction Conference*, ACM, March, 2006.

[7] Campion, G., Wang, Q., and Hayward, V. "The Pantograph Mk-II: A Haptic Instrument". *Proc. IROS 2005, IEEE/RSJ Int. Conf. Intelligent Robots and Systems,* pp. 723-728, 2005.

[8] Preisig, P. "Development of a Haptic Interface for 3 DoF Microassembly of Scaffolds for Tissue Engineering*", diploma thesis, Swiss Federal Institute of Technology (ETH-Zurich),* 2006.

[9] Shacklock, A.; Wenting Sun, "Ephemera: Tracking of Motion in Microscope Views," *Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on*., pp. 1835- 1840, 18-22 April 2005.

[10] Shacklock, A.; Wenting Sun, "Integrating microscope and perspective views" *Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on,*.pp. 454- 459, 18-22 April 2005.