Vision-Based Assistance for Ophthalmic Micro-Surgery

Maneesh Dewan[1], Panadda Marayong[2], Allison M. Okamura[2], and Gregory D. Hager[1]

[1] Department of Computer Science
{maneesh,hager}@cs.jhu.edu
[2] Department of Mechanical Engineering
{panadda,aokamura}@jhu.edu
Johns Hopkins University, Baltimore, MD

Abstract. This paper details the development and preliminary testing of a system for 6-DOF human-machine cooperative motion using vision-based virtual fixtures for applications in retinal micro-surgery. The system makes use of a calibrated stereo imaging system to track surfaces in the environment, and simultaneously tracks a tool held by the JHU Steady-Hand Robot. As the robot is guided using force inputs from the user, a relative error between the estimated surface and the tool position is established. This error is used to generate an anisotropic stiffness matrix that in turn guides the user along the surface in both position and orientation. Preliminary results show the effectiveness of the system in guiding a user along the surface and performing different sub-tasks such as tool alignment and targeting within the resolution of the visual system.The accuracy of surface reconstruction and tool tracking obtained from stereo imaging was validated through comparison with measurements made by an infrared optical position tracking system.

1 Introduction

Age-related macular degeneration (AMD), choroidal neovascularization (CNV), branch retinal vein occlusion (BRVO), and central retinal vein occlusion (CRVO) are among the leading causes of blindness in individuals over the age of 50 [2, 15]. Current treatments include laser-based techniques such as photodynamic therapy (PDT) and panretinal laser photocoagulation (PRP). However, these treatments often result in high recurrence rate or complications in certain cases, leading to loss of sight [10]. Recently, alternative approaches employing direct manipulation of surgical tools for local delivery of a clot-dissolving drug to a retinal vein (vein cannulation) or chemotherapeutic drugs to destroy a tumor have been attempted with promising results. However, these procedures involve manipulation within delicate vitreoretinal structures. The challenges of small physical scale accentuate the need for dexterity enhancement, but the unstructured nature of the tasks dictates that a human be directly "in the loop." For example, retinal vein cannulation [17] involves the insertion of a needle of approximately 20-50 microns in diameter into the lumen of a retinal vein (typically

C. Barillot, D.R. Haynor, and P. Hellier (Eds.): MICCAI 2004, LNCS 3217, pp. 49–57, 2004.

100 microns in diameter or less–approximately the diameter of a human hair). At these scales, tactile feedback is practically non-existent, and depth perception is limited to what can be seen through a stereo surgical microscope.

In a recent series of papers [1,7,11], our group in the Engineering Research Center for Computer Integrated Surgical Systems (CISST ERC) has been steadily developing systems and related validation methods for cooperative execution of microscale tasks. The primary motivation for these techniques has been to develop assistance methods for microsurgery. The basis of these systems is a specific type of virtual fixture, which we term "guidance" virtual fixtures. Other virtual fixture implementations, often called "forbidden-region virtual fixtures," are described in [8,12,14]. Guidance virtual fixtures create anisotropic stiffness that promotes motion in certain "preferred" directions, while maintaining high stiffness (and thus accuracy) in orthogonal directions.

In this paper, we describe our recent progress at developing vision-based virtual fixtures that provide closed-loop guidance relative to (visually) observed surfaces as well as features on those surfaces. Although, in principle, a direct application of our previously developed guidance fixtures, the implementation of these methods for realistic situations requires attention to significant practical issues. In the remainder of this paper, we describe the structure of our testbed system, detail the vision and control algorithms used, and present data from preliminary demonstrations and validation of the system.

2 System Description

In our previous work, we have described several methods whereby it is possible to create guidance virtual fixtures from image data [4,11]. In this paper, we have chosen to employ algorithms that use relative measures derived from reconstructed 3D geometry. Our goal is to create fixtures that aid in moving tools relative to observed surfaces and surface features in a surgical environment.

2.1 Surgical Field Modeling

Accomplishing the aforementioned goal makes it necessary to detect and track both the underlying surface as well as the operative tool. For both problems, we assume a calibrated stereo camera system. We briefly describe how the vision system recovers both surface and tool geometry.

In our application, the surface is assumed to be smooth to the second order with no self-occlusions. The surface reconstruction is done by first modeling the disparity with third-order tensor B-splines using least squares minimization methods and then reconstructing the disparity surface to obtain the 3D reconstruction. For further details of the method, we refer to [13].

The tool tip position and orientation in 3D characterize the tool geometry, which is computed in a two-step tracking algorithm. The first step involves cue segmentation and maintaining robust cue statistics, similar to the one used in [18]. The second step localizes the tool in the segmented image. The algorithm

used for localization is discussed in detail in [5]. The tool is localized in both left and right images and finally reconstructed to obtain its 3D geometry.

2.2 Virtual Fixtures

The virtual fixtures (described in [4]) provide guidance by eliminating the component of the user's applied force in directions orthogonal to the preferred direction. Intuitively, the preferred direction is task-dependent. The preferred direction, $D(t)$, is a $6 \times n$, $0 < n < 6$, time-varying matrix representing the instantaneous preferred directions of motion. Introducing an admittance gain $c_\tau \in [0, 1]$ that attenuates the non-preferred component of the force input, we define a velocity controller which guides the user along the direction parallel to the preferred direction of motion as

$$\mathbf{v} = c(\mathbf{f}_D + c_\tau \mathbf{f}_\tau), \tag{1}$$

where $\mathbf{f}_D$ and $\mathbf{f}_\tau$ are the force components along and orthogonal to the preferred directions, respectively. By choosing c, we control the overall admittance of the system. Choosing $c_\tau < 1$ imposes the additional constraint that the robot is stiffer in the non-preferred directions of motion. Hard virtual fixtures guidance can be achieved by setting c_τ to zero, maintaining the tool motion strictly in the preferred direction.

In our implementation, we consider five different sub-tasks:

- **Free Motion:** The user can freely move the robot to any desired pose. This is equivalent to setting c_τ to 1.
- **Surface Following:** The tool is constrained to move along the surface with a specified offset and, optionally, the tool can maintain an orientation orthogonal to the surface.
- **Tool Alignment:** The user can orient the tool along a preferred direction which is the desired tool orientation.
- **Targeting:** The only motion allowed is toward a fixed target specified on the surface.
- **Insertion/Extraction:** The only motion allowed is along the axis of the tool. In this case, only a separate insertion mechanism of the robot is activated.

The tasks are designed to simulate the sub-steps in performing a more complicated procedure. For example, in a retinal vein cannulation procedure, the surgeon may want to move freely until the tool comes close to the retinal surface, follow with some safe distance along the surface, move to a specified target, align the needle to a desired orientation, then insert and extract the needle for drug delivery into the vein.

We compute the preferred direction by establishing the normal to the surface at the point closest to the tool tip on the reconstructed B-spline surface. The instantaneous preferred direction of motion is then along a plane determined by the span of orthogonal tangents. In order to implement this, the closest point, the normal, and the tangents to the surface at the closest point are determined from

the algorithm described in [3]. A relative position error and orientation error can be calculated from the tool tip position and tool orientation estimated from the tool tracker. Position error is a signed vector from tool tip to the intersection point on the surface. Orientation error between the tool orientation and the desired orientation is just their cross product. Hence, the preferred direction, D, and the control, u, can be written for the closed-loop velocity control law for each task (see [4] for detailed derivation).

3 Experimental Setup

Fig. 1. Experimental setup.

In retinal micro-surgical procedures, the surgeon views the retina and manipulates the tool inside the eye through a microscope. As a preliminary testbed for the application of virtual fixture guidance to real surgery, we developed a scaled experiment, shown in Figure 1. We used the Steady-Hand robot [16] with stereo cameras located approximately 0.6 m away from the robot. A tool is attached to the end-effector of the robot through a 6 DOF force/torque sensors. The user applies force directly at the tool handle. Two types of surfaces were tested, a slanted plane and a concave surface to simulate the back of the human retina. For the latter, we used a concave surface of approximately 15cm in diameter covered with an enlarged gray-scale image of a human retina. The stereo system was comprised of two digital color cameras with 12mm focal length. During manipulation, the user has a direct view of the surface. The user is allowed to change the state of manipulation (Free motion, Tool insertion/extraction, Tool alignment, Targeting, and Surface following) from the GUI. To test the effectiveness of the virtual fixtures, only hard constraints ($c_\tau = 0$) were implemented. The desired offset for the surface tracking task was set at a constant value of 3mm above the surface.

Robot-to-camera and stereo calibration are integral steps in the experimental setup. The resolution and accuracy of the calibration determines the accuracy of the virtual fixture geometry. The 3D coordinates of the surface and the tool pose are computed in the camera coordinate frame by the visual system. The preferred directions and errors are hence first calculated in the camera frame and then converted to the robot frame, where the virtual fixture control law is implemented to obtain the desired Cartesian tip velocity.

The Optotrak, an infrared optical position tracking system, was used for robot calibration and validation of the imaging process. The robot's fixed frame is computed in the Optotrak frame of reference by first rotating the robot about the fixed remote-center-motion point in both the X and Y axes and then translating it along the X, Y, and Z axes with a rigid body placed at the last stage of the joint. In order to obtain the robot-to-camera transformation, the transformation

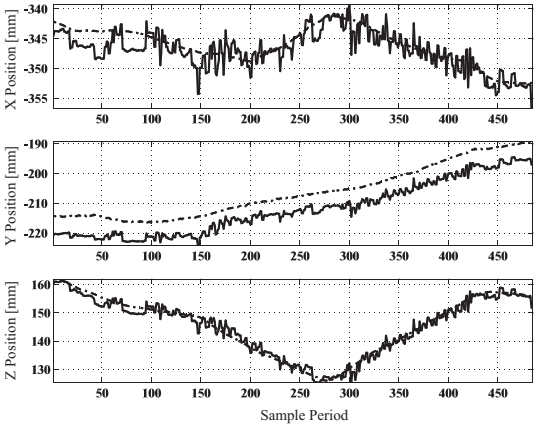

Fig. 2. X, Y, and Z positions of the tool tip estimated by the tool tracker (solid line) and estimated by the Optotrak (dashed line). Calibration error introduced a constant offset of 5 mm in the transformation as shown in the Y position plot.

between the Optotrak and the camera is computed by observing Optotrak LED markers rigidly attached to a planar rigid body in both the camera and the Optotrak frame of reference.

All transformations are computed with reference to an independent fixed rigid body in order to avoid any noise from the motion of the Optotrak. Using a least squares technique to compute rigid body transformations, the average error for the transformation was approximately 2mm with a standard deviation of +/- 1.5mm. We believe that the most significant source of error is due to difficulty of segmenting the centers of the LED markers in the images. However, it is known that slight errors in calibration do not affect the accuracy of the visual servoing system.

3.1 System Validation

First, the accuracy of the tool tracker was validated with the Optotrak. The tool tip positions obtained from the Optotrak were converted to the camera frame using the transformation between the Optotrak and the camera computed earlier. Figure 2 shows the comparison of the X, Y, and Z positions obtained by the Optotrak and the tool tracker. Note here that there is a constant offset of approximately 5mm in the Y coordinates. We believe that this constant offset is due to an error in the Camera-Optotrak calibration. The tool tracking has about 1 pixel error when the offset is accounted for. To validate the accuracy of the reconstructed surfaces (a plane and a concave surface), we compare the reconstructed surfaces with the ground-truth surfaces observed from the Optotrak estimated by tracing the surfaces with a pointed rigid body to obtain points on the surfaces expressed in the Optotrak reference frame. Figure 3 shows the overlay of the data points obtained from the Optotrak (black dots) on the reconstructed surfaces (with texture). The error in the plane reconstruction (left) is

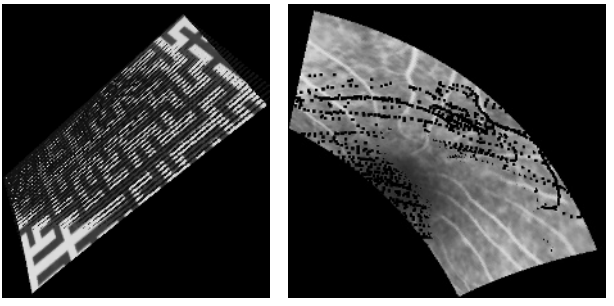

Fig. 3. Textured surface reconstruction overlaid with ground-truth surface data (black dots) obtained from the Optotrak. (Left) Slanted plane and (Right) Portion of the eye phantom.

quite low. For the concave surface (right) however, the reconstruction from the camera shows some errors, especially toward the boundary where the neighborhood information is sparse. The reconstruction covers an area of approximately 6.5cm x 17cm and 6.5cm x 11cm for the plane and the concave surface, respectively.

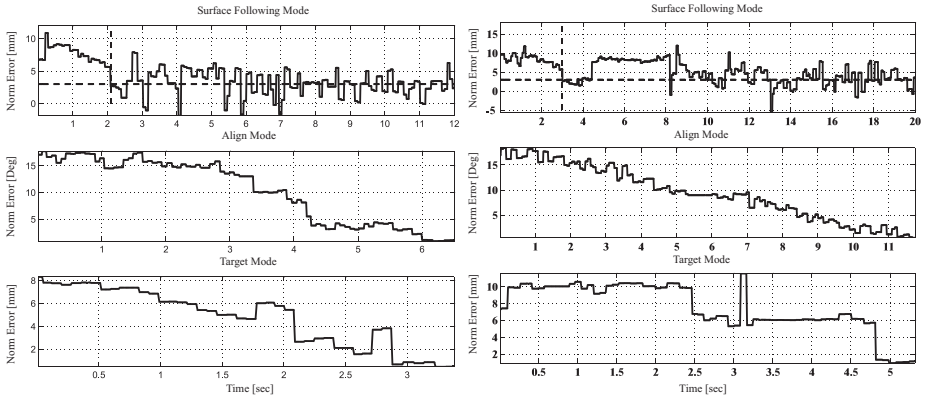

Fig. 4. The signed magnitude of error between the tool tip position and its intersection on a plane (left) and a concave surface (right).

4 Results and Discussion

Figure 4 shows the magnitude of position error over time for manipulation in Surface following, Alignment, and Targeting modes with hard virtual fixtures. Note here that the small steps shown on the plot are the result of the update-rate difference between the cameras and the robot. Sources of noise in the magnitude

Table 1. Resolution estimation of different visual systems.

Systems	Disparity Range [Pixel]	Resolution [mm/pix]
Sony X700 with 12mm lens	275-300	3.07-2.58
Endoscope (Olympus OTV-3D2)	76-86	0.49-0.389
Microscope (Zeiss OPMI1-H)	1992-2036	0.13-0.125

plot may be due to inaccuracy of the surface reconstruction, error in the estimation of tool tip position, and error in calibration. In addition, the width of the tool can introduce some error in the tool tracker's estimation of the tool center as the tool is viewed from different angles. In Surface following mode, we use 3mm as the desired offset above the actual surfaces. The average magnitude of error was approximately 3mm with a standard deviation of +/- 2mm, which is within the 1 pixel error envelope shown in Table 1. The dashed vertical and horizontal lines indicate the time when the tool reached the surface and the average (at 3mm), respectively. We believe that the few negative errors in the surface following error plots are the result of the noise in the system mentioned above. Note that the user started above the desired surface and then the error decreased when the tool moved closer to the surface. The decays of the error as the tool approaches a specified orientation (Alignment mode) and target (Targeting mode) can be clearly seen. In Alignment and Targeting modes, the virtual fixture control law allows movement along the direction of the error vector (forward and backward motion toward the desired orientation/target). However, gain tuning is used to minimize the motion in directions away from the desired pose. Similar results can be seen for the concave surface. We also performed the surface following task in free motion. The performance with virtual fixture guidance in the surface following and targeting tasks do not show significant improvement over free motion, especially with the level of noise and the resolution of the system. However, performance with virtual fixture guidance surpasses free motion in the alignment task.

The accuracy of the calibration and the limitations of the experimental setup are important factors affecting the system performance.Table 1 provides an estimate of the resolution of the current setup (Sony X700) along with an endoscope and a microscope system. Based on our implementation in the current setup and some preliminary experiments, we believe that the method is extendable to the high resolution systems to perform tasks at micro-scale. In order to achieve the required accuracy for vascular procedures in the eye, it appears that it will be necessary to compute tool and surface locations to sub-pixel accuracy. Recent results in this direction appear to promise depth resolution on the order of 1/10 pixel, which is sufficient for the applications described above [9]. Also we would like to incorporate fine motion tracking techniques like SSD and kernel-based methods to obtain subpixel accuracy for tool tracking.

One of the limitations of the system is the slow update rate of the cameras (10-15Hz) with respect to the robot (100Hz). To deal with this, we plan to explore the use of estimation techniques, such as Kalman filtering, in order to obtain smoother tool positions.

5 Conclusion

This work is, to our knowledge, the first example of a human-machine cooperative system guided entirely based on a visual reconstruction of the surrounding environment. This general approach of creating systems that are able to sense and react to the surgical environment is central to our goal of creating effective human-machine systems.

The preliminary demonstration outlined in this paper will be further refined and ported to work with a stereo endoscope (courtesy of Intuitive Surgical, Inc) and a stereo microscope available in our laboratory. Our objective is to demonstrate the system working at scales of 10's to 100's of microns on deforming biological surfaces. Once this is accomplished, the four-step procedure for retinal cannulation will be implemented using our task control architecture [6] and tested on an eye phantom to determine its efficacy for that procedure.

Acknowledgment. We would like to thank Anand Viswanathan and William Lau for their help. This material is based upon work supported by the National Science Foundation under Grant Nos. IIS-0099770, IIS-0205318, and EEC-9731478.

References

1. A. Bettini, P. Marayong, S. Lang, A. M. Okamura, and G. D. Hager. Vision assisted control for manipulation using virtual fixtures. *IEEE ITRA*. To appear.
2. N. Bressler, S. Bressler, and S. Fine. Age-related macular degeneration. *Survey of Ophthalmology*, 32(6):375–413, 1988.
3. J. Corso J. Chhugani and A. Okamura. Interactive haptic rendering of deformable surfaces based on the medial axis transform. *Eurohaptics*, pages 92–98, 2002.
4. G. D. Hager. Vision-based motion constraints. *IEEE/RSJ IROS, Workshop on Visual Servoing*, 2002. http://www.cs.jhu.edu/CIRL/new/publications.html.
5. G. D. Hager and K. Toyama. The "XVision" system: A general purpose substrate for real-time vision applications. *CVIU*, 69(1):23–27, January 1998.
6. D. Kragic and G. Hager. Task modeling and specification for modular sensory based human-machine cooperative systems. *IEEE/RSJ IROS*, 4:3192–3197, 2003.
7. R. Kumar, G. D. Hager, P. Jensen, and R. H. Taylor. An augmentation system for fine manipulation. In *MICCAI*, pages 956–965. Springer-Verlag, 2000.
8. F. Lai and R. D. Howe. Evaluating control modes for constrained robotic surgery. In *IEEE ICRA*, pages 603–609, 2000.
9. W. Lau, N. A. Ramey, J. J. Corso, N. Thakor, and G. D. Hager. Stereo-based endoscopic tracking of cardiac surface deformation. *MICCAI*, 2004. To appear.
10. M.P.S. Group. Argon laser photocoagulation for neovascular maculopathy. 5 yrs results from randomized clinical trial. *Arch Ophthalmol*, 109:1109–14, 1991.
11. P. Marayong, M. Li, A. Okamura, and G. Hager. Spatial motion constraints: Theory and demonstrations for robot guidance using virtual fixtures. *IEEE ICRA*, pages 1954–1959, 2003.
12. S. Payandeh and Z. Stanisic. On application of virtual fixtures as an aid for telemanipulation and training. *Symposium on Haptic Interfaces For Virtual Environments and Teleoperator Systems*, pages 18–23, 2002.

13. N. Ramey. Stereo-based direct surface tracking with deformable parametric models. Master's thesis, Dept. of Biomedical Engineering, Johns Hopkins University, 2003.
14. L. Rosenberg. Virtual fixtures: perceptual tools for telerobotic manipulation. *IEEE Virtual Reality International Sympsoium*, pages 76–82, 1993.
15. I. Scott. Vitreoretinal surgery for complications of branch retinal vein occlusion. *Curr Opin Ophthalmol*, 13:161–6, 2002.
16. R. Taylor, et al. Steady-hand robotic system for microsurgical augmentation. *IJRR*, 18(12):1201–1210, 1999.
17. J. N. Weiss. Injection of tissue plasminogen activator into a branch retinal vein in eyes with central retinal vein occlusion. *Ophthalmology*, 108(12):2249–2257, 2001.
18. C. R. Wren, A. Azarbayejani, T. Darrell, and A. Pentland. Pfinder: Real-time tracking of human body. *IEEE PAMI*, 19(7):780–785, 1995.