

Acquiring Visual-Motor Models for Precision Manipulation with Robot Hands*

Martin Jägersand, Olac Fuentes, Randal Nelson

Department of Computer Science
University of Rochester
Rochester, N.Y. 14627
U.S.A.

{jag,fuentes,nelson}@cs.rochester.edu
<http://www.cs.rochester.edu/u/{jag,fuentes,nelson}/>

Abstract. Dextrous high degree of freedom (DOF) robotic hands provide versatile motions for fine manipulation of potentially very different objects. However, fine manipulation of an object grasped by a multifinger hand is much more complex than if the object is rigidly attached to a robot arm. Creating an accurate model is difficult if not impossible. We instead propose a combination of two techniques: the use of an approximate estimated motor model, based on the grasp tetrahedron acquired when grasping an object, and the use of visual feedback to achieve accurate fine manipulation. We present a novel active vision based algorithm for visual servoing, capable of learning the manipulator kinematics and camera calibration online while executing a manipulation task. The approach differs from previous work in that a full, coupled image Jacobian is estimated online without prior models, and that a trust region control method is used, improving stability and convergence. We present an extensive experimental evaluation of visual model acquisition and visual servoing in 3, 4 and 6 DOF.

1 Introduction

In an active or behavioral vision system the acquisition of visual information is not an independent open loop process, but instead depends on the active agent's interaction with the world. Also the information need not represent an abstract, high level model of the world, but instead is highly task specific, and represented in a form which facilitates particular operations. Visual servoing, when supplemented with online visual model estimation, fits into the active vision paradigm. Results with visual servoing and varying degrees of model adaption have been presented for robot arms [23, 4, 21, 15, 16, 13, 2, 11]². Visual models suitable for specifying visual alignments have also been studied [9, 1, 10], but it remains to be proven that the approach works on more complex manipulators than the serial link robot arm. In this paper we present an active vision technique, having interacting action (control), visual sensing and model building modules, which allows the simultaneous visual-motor model estimation and control (visual servoing) of a dextrous multi finger robot hand. We investigate experimentally how well our algorithms work on a 16 DOF Utah/MIT hand.

A combined model acquisition and control approach has many advantages. In addition to being able to do uncalibrated visual servo control, the online estimated models are useful for instance for: (1) Prediction and constraining search in visual tracking [16, 4] (2) Doing local coordinate transformations between manipulator (joint), world and visual frames [15, 16]. (3) Synthesizing views from other agent poses [14]. In robot

* Support was provided by the Fulbright Commission and ONR grant N00014-93-I-0221.

² For a review of this work we direct the reader to [15] or [2].

arm manipulation we have found such a fully adaptive approach to be particularly helpful when carrying out difficult tasks, such as manipulation of flexible material [15, 16], or performing large rotations for visually exploring object shape [17]. For the Utah/MIT dextrous robot hand the fully adaptive approach is appealing because precisely modeling manipulation of a grasped object is much harder than for a typical robot arm, where the object is rigidly attached to the end effector. The four fingers of the Utah/MIT hand have a total of 16 actuated joints. When grasping a rigid object the fingers form a complex parallel kinematic chain. Although a few experimental systems that demonstrate the capabilities of dextrous robot hands have been implemented (e.g. [18, 20, 22]), their potential for performing high-precision tasks in unmodelled environments has not been fully realized. We have proposed an approach to dextrous manipulation [7], which does not require a-priori object models; all the required information can be read directly from the hand's sensors. This *alignment virtual tool* [19] allows versatile manipulation as shown in fig. 1 by providing an approximate transformation between the 6 DOF pose space of the grasped object and the 16 DOF joint control space. But the transformation is only approximate, and there is no way to account for slips in the grasp and other inaccuracies, so some external, object pose based feedback is needed to achieve high precision object fine manipulation.

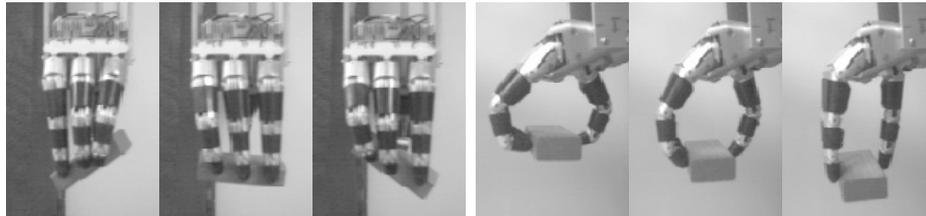


Fig. 1. Examples of manipulation with Utah/MIT hand.

In this paper we show how to learn online, and adaptively refine a local estimate of a visual-motor model, and to use that estimate for stable, convergent control even in difficult non-linear cases. Our objectives are to completely do away with prior calibration, either experimental “test movements” or model based, and to improve the convergence to allow solving manipulation tasks in larger workspaces, and/or more difficult cases than previously possible. The difficulty in the problem we solve lies much in the - for manipulator control purposes - extremely low sampling frequency imposed by the video rate of cameras. With very few samples, we want to control over large portions of the visual-motor work space and deal with large initial deviations between desired and actual states. This is different from what we feel mainstream control theory is meant to solve. In the paper we instead cast the estimation and control problem in an optimization theory framework, where we can draw on previous experience on estimating with very sparse data, and obtaining convergence despite large initial errors and large step sizes.

The overall structure of our vision based manipulation system is shown in fig. 2. In the two following sections we describe the online model acquisition. In sections 4 and 5 we describe the visual and motor control of the hand. Visual goal assignment, trajectory generation and what kind of visual features to use are task dependent. In [16] we describe the high (task) level parts of the system and how to solve real world manipulation problems with this system, such as solving a shape sorting puzzle, handling flexible materials and exchanging a light bulb. In section 6 we study low level aspects of visual control, and use simple random trajectories to quantitatively evaluate positioning and model estimation.

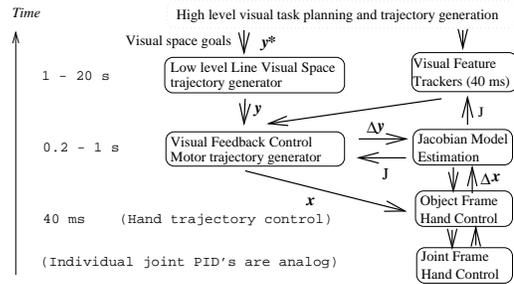


Fig. 2. Structure of the system, and the time frames each part runs in. Arrows indicate information exchange.

2 Viewing Model

We consider the robot hand³ as an active agent in an unstructured environment. The hand manipulates a grasped object in approximate motor pose space⁴ $\mathbf{x} \in \mathfrak{R}^n$ through the transformations derived in section 5, and observes the results of its actions as changes in a visual perception or “feature” vector $\mathbf{y} \in \mathfrak{R}^m$. Visual features can be drawn from a large class of visual measurements[23, 15], but we have found that the ones which can be represented as point positions or point vectors in camera space are suitable[16]. We track features such as boundary discontinuities (lines, corners) and surface markings. Redundant visual perceptions ($m \gg n$) are desirable as they are used to constrain the raw visual sensory information.

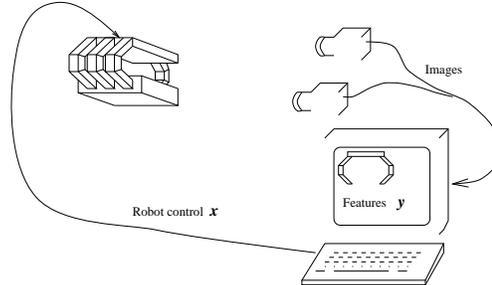


Fig. 3. Visual control setup using two cameras.

The visual features and the agent’s actions are related by some visual-motor model f , satisfying $\mathbf{y} = f(\mathbf{x})$. The goal for the control problem is, given current state \mathbf{x}_0 and \mathbf{y}_0 , the desired final state in visual space \mathbf{y}^* , to find a motor command, or sequence thereof, $\Delta \mathbf{x}_k$ s.t. $f(\mathbf{x}_0 + \sum_k \Delta \mathbf{x}_k) = \mathbf{y}^*$. Alternatively, one can view this problem as the minimization of the functional $\phi = \frac{1}{2}(\mathbf{f} - \mathbf{y}^*)^T(\mathbf{f} - \mathbf{y}^*)$. In a traditional visual control setup we have to know two functions, camera calibration h and robot kinematics g , either a priori, or through some calibration process. The final robot positioning accuracy

³ The visual feedback methods derived here also apply to other holonomic manipulators.

⁴ Vectors are written bold, scalars plain and matrices capitalized.

depends on the accuracy of both of these functions ($f^{-1} = g(h(\cdot))$), since feedback is only performed over the joint values.

In our uncalibrated visual servoing, the visual-motor model is unknown and at any time k we estimate a first order model of it, $f(\mathbf{x}) \approx f(\mathbf{x}_k) + J(\mathbf{x}_k)(\mathbf{x} - \mathbf{x}_k)$. The model is valid around the current system configuration \mathbf{x}_k , and described by the “image”[2] or visual-motor Jacobian defined as

$$(J_{j,i})(\mathbf{x}_k) = \frac{\partial f_j(\mathbf{x}_k)}{\partial x_i} \quad (1)$$

The image Jacobian not only relates visual changes to motor changes, as is exploited in visual feedback control but also highly constrains the possible visual changes to the $n-D$ subspace $\mathbf{y}_{k+1} = J\Delta\mathbf{x} + \mathbf{y}_k$ of $\mathbf{y}_{k+1} \subset \mathfrak{R}^m$ (remember $m \gg n$). Thus the Jacobian J is also a visual model, parameterized in exactly the degrees of freedom our system can change in. A collection of such Jacobians represents a piecewise linear visual-motor model of the part of the workspace the agent has explored so far. In our case f is linearized on an adaptive size mesh, as explained in section 4.

3 Visual-Motor Model Estimation

In previous visual servoing work a Jacobian has been either (1) derived analytically, (2) derived partially analytically and partially estimated (eg. [4, 21]), or (3) determined experimentally by physically executing a set of orthogonal calibration movements \mathbf{e}_i (eg. [1, 9]) and approximating the Jacobian with finite differences:

$$\hat{J}(\mathbf{x}, \mathbf{d}) = (f(\mathbf{x} + d_1\mathbf{e}_1) - f(\mathbf{x}), \dots, f(\mathbf{x} + d_n\mathbf{e}_n) - f(\mathbf{x}))D^{-1} \quad (2)$$

where $D = \text{diag}(\mathbf{d})$, $\mathbf{d} \in \mathfrak{R}^n$.

We instead seek an online method, which estimates the Jacobian by just observing the process, without a-priori models or introducing any extra “calibration” movements. In observing the object movements we obtain the changes in visual appearance $\Delta\mathbf{y}_{measured}$ corresponding to a particular controller command $\Delta\mathbf{x}$. This is a secant approximation of the derivative of f along the direction $\Delta\mathbf{x}$. We want to update the Jacobian in such a way as to satisfy the most recent observation (secant condition): $\Delta\mathbf{y}_{measured} = \hat{J}_{k+1}\Delta\mathbf{x}$. The above condition is under determined, and a family of *Broyden* updating formulas can be defined [5, 13]. In previous work we have had best results with this asymmetric correction formula [15, 16, 13]:

$$\hat{J}_{k+1} = \hat{J}_k + \frac{(\Delta\mathbf{y}_{measured} - \hat{J}_k\Delta\mathbf{x})\Delta\mathbf{x}^T}{\Delta\mathbf{x}^T\Delta\mathbf{x}} \quad (3)$$

This is a rank 1 updating formula in the *Broyden* hierarchy, and it converges to the Jacobian after n orthogonal moves $\{\Delta\mathbf{x}_k, k = 1 \dots n\}$. It trivially satisfies the secant condition, but to see how it works let’s consider a change of basis transformation P to a coordinate system O' aligned with the last movement so $\Delta\mathbf{x}' = (\Delta x_1, 0, \dots, 0)$. In this basis the correction term in eq. 3 is zero except for the first column. Thus our updating schema does the minimum change necessary to fulfill the secant condition (minimum change criterion). Let $\Delta\mathbf{x}_1 \dots \Delta\mathbf{x}_n$ be n orthogonal moves, and $P_1 \dots P_n$ the transformation matrices as above. Then $P_2 \dots P_n$ are just circular permutations of P_1 , and the updating schema is identical to the finite differences in eq. 2, in the O'_i frame.

Note that the estimation in eq. 3 accepts movements along arbitrary directions $\Delta\mathbf{x}$ and thus needs no additional data other than what is available as a part of the manipulation task we want to solve. In the more general case of a set of non-orthogonal moves $\{\Delta\mathbf{x}_k\}$ the Jacobian gets updated along the dimensions spanned by $\{\Delta\mathbf{x}_k\}$.

4 Trust Region Control

While previous work in visual servoing have mostly used simple proportional control, typically with slight modifications to account for some dynamics, we have found that for hand manipulation a more sophisticated approach is needed. We use two ideas from optimization: (1) A trust region method [3] estimates the current model validity online, and controller response is restricted to be inside this “region of trust”. (2) A homotopy or path following method [8] is used to divide a potentially non-convex problem into several smaller convex problems by creating subgoals along trajectories planned in visual space. For details see [16], in which we describe a method for high level visual space task specification, planning, and trajectory generation. For details on the low level control properties see [13]. In addition to allowing convergent control, these two methods serve to synchronize the model acquisition with the control, and cause the model to be estimated on an adaptive size mesh; dense when the visual-motor mapping is difficult, sparse when it is near linear.

The trust region method adjusts a parameter α so that the controller never moves out of the validity region of the current Jacobian estimate. To do that we solve a constrained problem for $\|\delta_k\| < \alpha$ instead of taking the whole Newton step $\Delta \mathbf{x}$ of a proportional controller.

$$\min_{\|\delta_k\| < \alpha_k} \|\mathbf{y}_k - \mathbf{y}^* + \hat{J}_k \delta_k\|^2 \quad (4)$$

Define a model agreement as $d_k = \frac{\|\hat{J}\delta\|}{\|\Delta \mathbf{y}_{measured}\|}$ and adjust α according to (d_{lower} and d_{upper} are predefined bounds):

$$\alpha_{k+1} = \begin{cases} \frac{1}{2}\alpha_k & \text{if } d_k \leq d_{lower} \\ \alpha_k & \text{if } d_{lower} < d_k \leq d_{upper} \\ \max(2\|\delta\|, \alpha) & \text{if } d_k > d_{upper} \end{cases} \quad (5)$$

5 Non-Model-Based Dexterous Manipulation

We propose a non-model-based approach to manipulation for the Utah/MIT hand, a 16 DOF four-fingered dextrous manipulator [12]. The method does not require any prior information about the object; all the required information can be read directly from the hand’s sensors. The method allows arbitrary (within the robot’s physical limits) translations and rotations of the object being grasped. We rely on the compliance of the Utah/MIT hand to maintain a stable grasp during manipulation, instead of attempting to compute analytically the required forces.

We assume that prior to manipulation the object has been stably grasped. The basic idea of our technique is that the commanded fingertip positions (setpoints) define a rigid object in three-dimensional space, the *grasp tetrahedron*. The contact points of the fingers on the object define another object that we refer to as the *contact tetrahedron*. Because of the compliant control system, we can model the statics of the situation by regarding each vertex of the contact tetrahedron as being attached to the corresponding vertex of the grasp tetrahedron by a virtual spring. In the case of an object free to move, net wrench is constrained to be zero. Moreover, for a well conditioned grasp, the wrench zero coincides with a deep local minimum of the spring energy function on a manifold defined by the rigid displacements of the grasp tetrahedron with respect to the contact tetrahedron. This means that if we treat the fingertip contacts as fixed points on the object, then the object can be rotated and translated by executing the desired rigid transformation on the grasp tetrahedron. Although, due to non-zero finger size and other effects, the assumption of fixed contact points is not strictly correct, our experimental results show that the compliance of the Utah/MIT hand compensates for the errors introduced by the use of this assumption.

Let $\mathbf{c} = [c_x, c_y, c_z]^T$ be the coordinates of the centroid of the grasp tetrahedron. We define an object-centered frame of reference C , with its axes parallel to the fixed hand-centered reference frame and with origin at \mathbf{c} . Let $\mathbf{q}_0, \dots, \mathbf{q}_3$ be the coordinates of the vertices of the grasp tetrahedron; given the manipulation command $\mathbf{x} = \langle x, y, z, \alpha, \beta, \gamma \rangle$, where x, y and z represent displacements of the object with respect to its initial position and α, β, γ are interpreted as sequential rotations about the x, y and z axes of C , the resulting setpoints $\mathbf{q}'_0, \dots, \mathbf{q}'_3$ are given by:⁵

$$\begin{bmatrix} q'_{ix} \\ q'_{iy} \\ q'_{iz} \\ 1 \end{bmatrix} = \begin{bmatrix} c\alpha c\beta c\alpha s\beta s\gamma - s\alpha c\gamma c\alpha s\beta c\gamma + s\alpha s\gamma & x + c_x \\ s\alpha c\beta s\alpha s\beta s\gamma + c\alpha c\gamma s\alpha s\beta c\gamma - c\alpha s\gamma & y + c_y \\ -s\beta & c\beta s\gamma & c\beta c\gamma & z + c_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} q_{ix} - c_x \\ q_{iy} - c_y \\ q_{iz} - c_z \\ 1 \end{bmatrix}$$

Addition of a Cartesian Controller At the lowest level, each joint in the Utah/MIT hand is controlled using a standard PD controller. In principle, the errors in object position and orientation can be reduced by increasing the gains in these controllers. However, this would also result in a decrease in compliance. As a way to reduce position errors without sacrificing compliance, we implemented a higher-level PID Cartesian controller to correct errors directly in the 6-dimensional position-orientation space, using an estimate of the errors in $\langle x, y, z, \alpha, \beta, \gamma \rangle$ as input. For details see [6].

6 Experimental Results

The Utah/MIT hand is a relatively imprecise and difficult to control manipulator. Each joint is actuated by a pair of pneumatically-driven antagonist tendons. This results in a compliant response, but it also makes reliable and consistent positioning difficult to achieve due to hysteresis and friction effects. Furthermore, especially during long manipulation sequences, slipping in the grasp points add to these errors.

We tested repeatability under closed loop visual control and compared the results to traditional joint control. The test object was a piece of standard 2 by 4 wood construction stud, approximately 2 inches long. Positions were measured through a 0.001" accuracy dial meter. Accurate visual measurements were provided by tracking special surface markings in the first experiments, see fig. 4. The markers could be placed arbitrarily on the object, since the servoing algorithm is self calibrating. Placement near the corners gives better conditioned visual measurements with respect to rotations. In the second experiment, for improved visual accuracy, we track LED's mounted on wood screws. Two cameras were positioned approximately 90 degrees apart. 1 pixel corresponds approximately to a 0.3mm object movement. Repeatability was measured by moving the object on a random trajectory towards the dial meter until it was in contact, reading the dial value and visual or joint values, then retreating on a second random trajectory, and then trying to reach the visual or joint goal. In fig. 4 let x be the optic ray, y the image row and z the column. Around the x, y and z axis respectively, a grasped object can be translated about 40mm, 60mm and 30mm, and rotate 70, 45 and 90 degrees (see fig. 1). We tested but did not find any correlation between trajectory length and positioning error.

6.1 Repeatability in 3 and 4 DOF

In table 1 we compare repeatability for four different positioning methods, based on 50 trials with each method. The 3 DOF visual feedback controls only translations of

⁵ we use $c\theta$ and $s\theta$ as shorthand for $\cos(\theta)$ and $\sin(\theta)$, respectively.

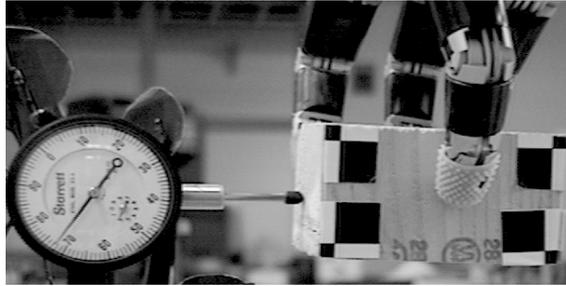


Fig. 4. Setup for hand fine manipulation repeatability experiments

the object. The 4 DOF visual feedback controls the translations and rotation around the optical axis of fig 4. Visual feedback in both cases consist of $m = 16$ feature values from 8 markers, and 2 cameras. The two joint feedback methods differ in that one only uses joint feedback, while in the Cartesian one, the error vector is projected onto the 6DOF space defined by the grasp tetrahedron. As seen in table 1 visual servoing performs on the average a little better than joint feedback methods. When studying maximum errors visual servoing is significantly better. This is because the visual feedback can compensate for slip in the grasp contact points.

Error	Visual feedback		Joint feedback	
	3 DOF	4 DOF	Joint	Cartesian
Mean	0.3mm	0.4mm	0.9mm	0.5mm
Max	1.0mm	1.3mm	2.5mm	2.9mm

	Controlled DOF's			
	3 DOF	4 DOF	5 DOF	6 DOF
κ	1.6	2.6	11	30

Table 1. Left: Measured repeatability of the Utah/MIT hand fine manipulating a rigid object under visual and joint feedback control. **Right:** Condition numbers κ of different degree of freedom (DOF) Jacobians.

Visual feedback gets more difficult in high DOF systems. One reason is that the number of parameters estimated in the Jacobian increases. Another is that high DOF tasks are often more ill conditioned. Particularly, there is a difference between pure translations (3 DOF) and combinations of translations and rotations (4-6 DOF). The condition numbers of the 3 to 6 DOF estimated Jacobian are shown in table 1. We tried 5 and 6 DOF visual servoing using the setup described above, but could not get reliable performance. For 5 DOF 37 % of the trials failed (diverged), and for 6 DOF nearly all failed. This is not surprising given the mechanical difficulties in controlling the hand. When trying to make the small movements needed for convergence the actual response of the manipulator has significant random component. The signal to error ratio in visual control space is then further decreased by the bad condition number of the transformation between motor and visual space. The high condition numbers in the 5 and 6 DOF cases stem from two distinct problems. With the two camera setup we use, the rotation around the z-axis is (visually) ill conditioned (differentially similar to the translation along the optic axis.) This problem occurs with a two camera setup, when the points tracked in each camera are planar. The other problem is that two of the manipulations are differentially very similar in Cartesian motor space. By taking

out the two problematic DOF's we get the low condition numbers of the 3 and 4 DOF manipulations.

6.2 6 DOF visual feedback control of the Utah/MIT hand

In order to evaluate visual servoing in the full 6 DOF pose space we improved the visual measurements by tracking LED's mounted on adjustable screws attached to the test objects. The LED's facilitate more precise tracking than the surface markings, and using the screws we adjusted them into a non-coplanar configuration, avoiding the singularity we got from tracking the planar surfaces of the object. With this setup we were able to do both 5 and 6 DOF visual servoing, achieving accurate end positions in over 90% of the trials, and only diverging completely in two out of 67 trials.

We performed the same positioning experiment as described in section 6.1, but here in the full 6 DOF of the rigid object. Positioning errors were measured along the 3 major axes. For the Y and Z axes the measuring axis coincides with the center of rotation, so the errors measured are purely translational. For the X axis, the fingers are in the way, and we used a measuring point below the fingers, and thus measure the sum of rotational and translational error components.

The results are shown in table 2. The positioning error along the X-axis is much worse than the two others. We propose two reasons for this: (1) The difference in the measuring location discussed above and (2) along the Y and Z axes the translation is effected by all fingers moving in parallel, which tend to average random disturbances in an individual finger, while along the X axis only the thumb opposes the force of the three other fingers, so positioning is no more precise than for one finger.

Visual feedback improves the positioning in all cases. The improvement is significant along the Y and Z axes, but only marginal along the X axis. This result is consistent with our earlier observations that visual feedback improves an already precise manipulator more than an imprecise one. We don't have a good analysis of the exact reasons for this. In part it can probably be attributed to the visual feedback needing a monotonic response to very small movements, while in an imprecise manipulator the response is random. For a more precise manipulator the response may be monotonic, but non-linear, which affects only open loop accuracy.

Axis:	Visual feedback			Cartesian joint feedback		
	X	Y	Z	X	Y	Z
Mean error:	1.0mm	0.2mm	0.2mm	1.1mm	0.7mm	0.6mm
Max error:	2.4mm	0.5mm	0.5mm	2.9mm	4.9mm	2.3mm

Table 2. Measured repeatability of the Utah/MIT hand fine manipulating a rigid object under 6 DOF visual and Cartesian joint feedback control.

Error distributions for the 6 DOF positioning experiment are shown in fig. 5. For open loop joint feedback, the object positioning errors are of two distinct types. The first is due to the inaccurate response of the fingers themselves, the second is due to slippage in the grasp points. The mode around 0.5 mm is due to manipulator inaccuracies. The errors caused by slippage are represented in the mode around 2 mm.

In the visual servoing behavior we noted that trials fell into three categories. About half the trials converged very fast. Most of the rest converged, initially fast, but much slower towards the end. Last a small percentage did not converge at all (2 out of 67 visual feedback trials). We have not observed this difference in convergence speed when using

the same visual feedback controller on robot arms. We think a partial explanation may be that mechanical phenomena in the hand and remotizer linkage cause some positions to be harder to reach than others. We have also observed that when the visual servo controller is turned off, the hand often does not remain in the same position, but drifts significantly, despite the joint setpoints remaining the same.

Not all visual servoing trials converged satisfactory. Specifically along the x axis errors were sometimes as large as 2mm (see table 2). The controller convergence criterion is partly causing this. When convergence is unsatisfactory over several sample intervals, the controller adjusts the expected positioning accuracy down, and also the derived gain. This is to stop long oscillatory behaviors near the goal. In some cases the adjustment is too large, impairing accuracy.

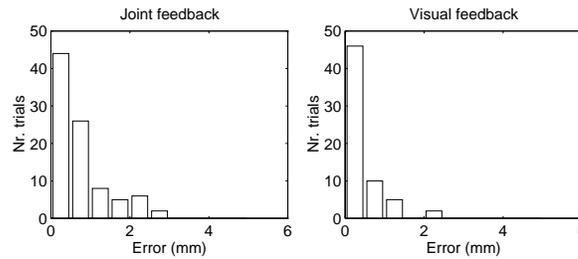


Fig. 5. Distribution of translational positioning errors for joint and visual feedback positioning of the Utah/MIT hand.

7 Discussion

We have been doing visual servo control on robot arms in high DOF spaces for several years[15, 16, 13]. Getting it to work well on a multi finger gripper like our Utah/MIT dextrous hand was much harder than we had expected. For simple robot arm control (eg. move along a line in 3 DOF) we have found that almost any controller based on the Newton method will work[15, 16]. There are also many accounts of success in practical application for low DOF manipulation in the literature [4, 21, 2, 11]. Robot arm control is also fairly insensitive to somewhat ill conditioned visual measurements, such as caused by bad camera placement, and the particular choice of visual features to track.

To successfully apply visual servo control on the hand we found that we had to be more careful in choosing the camera placement, and which features to track. Since it is the 6 DOF pose of the object we are controlling, rather than the individual fingers, we have to find viewpoints where the object is relatively unobstructed by the fingers so we can reliably see object features, and track them possibly through large rotations. The relative inaccuracy of the hand responses to motion commands, together with a, for 6 DOF manipulations, often somewhat ill-conditioned visual-motor Jacobian demand some special measures in the controller. The restricted step in our trust region controller proved to be crucial. Intuitively, both the trust region controller and the visual space trajectory planning serve to synchronize model acquisition with the actions, so that the actions never run too far ahead into a region where the model estimate is no longer accurate. Without the step restriction, and when the system is in a near singular state (which is not uncommon, see section 6.1), the controller may make a move of uncontrolled length and speed. This often causes the grasp to shift unfavorably, the object to be dropped, or the visual trackers to loose track. With the step restriction, the

movement is still erroneous, but small. In most cases, this small movement is enough to get out of the singularity, so that normal servoing can be resumed.

The estimation and control algorithms we developed have strong theoretical properties (see [13]). It is still very important to experimentally evaluate how they perform in real environments, with real process disturbances/noise and manipulators with real mechanically caused errors and finite precision. Our results show that visual feedback control yields significant repeatability improvement (5 times) in an imprecise PUMA robot arm[15], while only a smaller improvement in the Utah/MIT hand. This is explained by the different characteristics of the manipulators. The inaccuracies in positioning the PUMA arm stem mainly from backlash, which the visual feedback can correct, and the model adaption is made robust to. In the hand a combination of sticktion and flexibility makes it much harder to control accurately. When trying to make small movements, the hand will initially not move at all, and the controller ramps up the signal. When the hand finally unsticks it overshoots the intended goal. This makes it very hard to control with a feedback controller, and even harder to estimate the visual-motor Jacobian.

References

1. R. Cipolla, P. A. Hadfield and N. J. Hollinghurst "Uncalibrated Stereo Vision with Pointing for a Man-Machine Interface" In *Proc of IAPR workshop on Machine Vision*, Tokyo, 1994.
2. P. I. Corke. *High-Performance Visual Closed-Loop Robot Control*. PhD thesis, University of Melbourne, 1994.
3. G. Dahlquist and A. Björck. *Numerical Methods*. Prentice Hall, 199x, Preprint.
4. J. T. Feddema and C. S. G. Lee. Adaptive image feature prediction and control for visual tracking with a hand-eye coordinated camera. *IEEE Tr. Sys, man and Cyber*, 20(5), 1990.
5. R. Fletcher. *Practical Methods of Optimization*. Chichester, 1987.
6. O. Fuentes and R. C. Nelson. Experiments on dextrous manipulation without prior object models. TR 606, Computer Science, U of Rochester, 1996.
7. O. Fuentes and R. C. Nelson. Morphing hands and virtual tools (or what good is an extra degree of freedom?). Technical Report 551, Computer Science U of Rochester, 1994.
8. Garcia and Zangwill. *Pathways to solutions, fixed points, and equilibria*. Prentice-Hall, 1981.
9. G. Hager. Calibration-free visual control using projective invariance. In *ICCV 1995*.
10. M. Harris. Vision guided part alignment with degraded data. TR-AI 615, U Edinburgh 93.
11. K. Hosoda, M. Asada. Versatile visual servoing without knowledge of true jacobian. In *Proc of IROS*, 1994.
12. S. Jacobsen, E. Iversen, D. Knutti, R. Johnson, and K. Bigger. Design of the Utah/MIT Dextrous Hand. In *Proc ICRA 1986*.
13. M. Jägersand. Visual servoing using trust region methods and estimation of the full coupled visual-motor jacobian. In *Proc of IASTED Conf Applications of Robotics and Control*, 1996.
14. M. Jägersand. *Model Free View Synthesis of an Articulated Agent*, Technical Report 595, Computer Science Department, University of Rochester, Rochester, New York, 1995.
15. M. Jägersand and R. C. Nelson. Adaptive differential visual feedback for uncalibrated hand-eye coordination and motor control. TR 579, Computer Science U of Rochester, 1994.
16. M. Jägersand and R. C. Nelson. Visual space task specification, planning and control. In *Proceedings of the 1995 IEEE Symposium on Computer Vision*, 1995.
17. K. Kutulakos and M. Jägersand. Exploring objects by purposive viewpoint control and invariant-based hand-eye coordination. In *Workshop on Vision for Robots, IROS*, 1995.
18. P. Michelman and P. Allen. Complaint manipulation with a dexterous robot hand. In *Proc 1993 IEEE Int. Conference on Robotics and Automation*, pages 711-716, 1993.
19. R. Nelson, M. Jagersand, O. Fuentes *Virtual Tools: A Framework for Simplifying Sensory-Motor Control in Complex Robotic Systems* TR 576 University of Rochester, 1995.
20. W. Paetsch and G. von Wichert. Solving insertion tasks with a multifingered gripper by fumbling. In *Proc IEEE Int. Conference on Robotics and Automation*, pages 173-179, 1993.
21. N. Papanikolopoulos and P. Khosla. Adaptive robotic visual tracking: Theory and experiments. *IEEE Transactions on Automatic Control*, 98(3), 1993.
22. T. H. Speeter. Primitive based control of the Utah/MIT dextrous hand. In *Proc IEEE International Conference on Robotics and Automation*, pages 866-877, Sacramento, 1991.
23. L. Weiss, A. Sanderson, and C. P. Neuman. Dynamic Sensor-Based Control of Robots with Visual Feedback. *IEEE Journal of Robotics and Automation*, RA-3(5), October 1987.