Hand-Eye Calibration of Robonaut

Kevin Nickels  
Engineering Science  
Trinity University  
San Antonio, TX 78212–7200  
Email: knickels@engr.trinity.edu

Eric Huber  
Metrica, Inc.  
Dexterous Robotics Laboratory  
NASA Johnson Space Center  
Houston, TX 77058  
Email: huber@roboteyes.com

Abstract—NASA's Human Space Flight program depends heavily on Extra-Vehicular Activities (EVA's) performed by human astronauts. EVA is a high risk environment that requires extensive training and ground support. In collaboration with the Defense Advanced Research Projects Agency (DARPA), NASA is conducting a ground development project to produce a robotic astronaut’s assistant, called Robonaut, that could help reduce human EVA crew and workload.

The project described in this paper designed and implemented a hand-eye calibration scheme for Robonaut, Unit A. The intent of this calibration scheme is to improve hand-eye coordination of the robot. The basic approach is to use kinematic and stereo vision measurements, namely the joint angles self-reported by the right arm and 3-D positions of a calibration fixture as measured by vision, to estimate the transformation from Robonaut's base coordinate system to its hand coordinate system and to its vision coordinate system.

Two methods of gathering data sets have been developed, along with software to support each. In the first, the system observes the robotic arm and neck angles as the robot is operated under external control, and measures the 3-D position of a calibration fixture using Robonaut's stereo cameras, and logs these data. In the second, the system drives the arm and neck through a set of prerecorded configurations, and data are again logged.

Two variants of the calibration scheme have been developed. The full calibration scheme is a batch procedure that estimates all relevant kinematic parameters of the arm and neck of the robot. The daily calibration scheme estimates only joint offsets for each rotational joint on the arm and neck, which are assumed to change from day to day. The schemes have been designed to be automatic and easy to use so that the robot can be fully recalibrated when needed such as after repair, upgrade, etc, and can be partially recalibrated after each power cycle.

The scheme has been implemented on Robonaut Unit A and has been shown to reduce mismatch between kinematically derived positions and visually derived positions from a mean of 13.75cm using the previous calibration to means of 1.85cm using a full calibration and 2.02cm using a suboptimal but faster daily calibration. This improved calibration has already enabled the robot to more accurately reach for and grasp objects that it sees within its workspace. The system has been used to support an autonomous wrench-grasping experiment and significantly improved the workspace positioning of the hand based on visually derived wrench position estimates.

I. INTRODUCTION

The Dexterous Robotics Laboratory (DRL) at NASA Johnson Space Center (JSC) has developed a ground-based prototype humanoid robot called Robonaut\(^1\), shown in Figure 1.

\(^1\)There are two versions of Robonaut, referred to as Unit A and Unit B. In this paper, we will use the name Robonaut for Unit A.

Fig. 1. Ground-based Robonaut system

Robonaut has been designed so that it could, for example, assist astronauts during EVA type tasks\([1]\). Its initial control has been by tele-operation, but the DRL is beginning to implement several semi-autonomous and fully autonomous controllers for Robonaut, necessitating improved hand-eye coordination for the system. This paper documents two methods for gathering kinematic and visual data, and two automatic hand-eye calibration schemes developed for Robonaut in support of these methods.

A. Prior Work

Much previous work has been done on the self-calibration of redundant manipulators using internal or external kinematics constraints. \([2], [3], [4], [5]\). Of particular note is the treatment of Bennett and Hollerbach of a vision or metrology system as an additional kinematic link \([5], [6]\), allowing us to treat a one-arm plus vision setup as a closed kinematic chain. This approach allows us to leverage works on the automatic self-calibration of closed kinematic chains, such as \([7]\).

One precondition of this approach is the accurate local-
ization of a point or points of the arm's kinematic chain in the coordinate system of the eyes. Several other calibration schemes utilize special visual markers [4] or LEDs to localize points; we opted for a spherical calibration fixture and visual measurements of this fixture. In order to accurately locate the spherical fixture in the image, a generalized Hough transform was used. The generalized Hough transform is described in [8].

The developed system is a closed-loop system. It reduces the errors between visually and kinematically derived predictions, but does not necessarily adjust these parameters to match the workspace. If the visual system is not well calibrated, via for example, the procedures described in [9], [10], [11], [12], the adjustments performed by this process will not cause the predictions to correlate well with workspace positions.

B. Task Background

Robonaut has historically been operated by a human teleoperator. The DRL is increasing the autonomy level of the tasks performed by Robonaut [13], [14].

When Robonaut Unit A is operated with only the relative joint encoders, as the arm is powered down in the evening and restarted the next morning, the position of all joints on the 7DOF right arm and the 2DOF neck can change, leading to errors in self-reported joint angles. Errors in these angles, as well as uncertainty in the as-built kinematic parameters of the arm, have lead to workspace errors of up to 10-15 cm in various situations. While human teleoperators are very good at correcting for this type of systematic error, it is unacceptable for the degree of autonomy now being required of Robonaut.

C. Kinematic Model

Robonaut's arm is a redundant manipulator with 7 degrees of freedom. This manipulator can be described by 7 homogeneous transformations \( A_j \) from link \( j \) to link \( j - 1 \) as defined by the Denavit-Hartenberg (DH) convention.

Each transformation is defined as

\[
A_j = \text{Trans}(x'_j, a_j)\text{Rot}(x'_j, a_j)\text{Trans}(z_j, d_j)\text{Rot}(z_j, \theta_j),
\]

where Rot implies a rotation about an axis and Trans implies a translation along an axis [15]. The position and orientation of the last link can be computed by a sequence of DH transformations defining the kinematic model

\[
T_c = A_1 A_2 A_3 \ldots A_{nf}
\]

where \( nf \) is the number of degrees of freedom.

Both the 7DOF transformation from the chest coordinate system to the hand coordinate system and the 2DOF transformation from the chest coordinate system to the eye coordinate system are parameterized in this way.

2For simplicity, and to allow for automatic calibration of the helmet-camera transform, we use 3 degrees of freedom in the chest-head transform. On Unit A, the joint angle will always be zero for the third DOF; but Unit B has active head roll as well as pitch and yaw.

II. CALIBRATION PROCEDURES

We have developed two methods for gathering calibration data from the robot. In the first, the robot is observed under external control and data are logged. In the second, the robot is actuated to each of a set of prerecorded target configurations, and data are again logged.

We have developed two methods to establish a set of DHP values from these data. In a daily calibration, estimates are generated only for the joint angle offsets. In a full calibration, estimates are generated for all relevant DHPs.

A. Supporting Programs

We have developed several computer programs to support these calibration methods. This section reviews the programs and foreshadows their use in the calibration. Throughout the user interfaces, the results of visual measurements are shown in red (or gray, if this paper is printed in grayscale). Blue or dark gray markings indicate kinematically derived predictions using the previous calibration, and are shown in the left image of a split display. Green or light gray markings indicate kinematically derived predictions using the updated calibration, and are shown in the right image of a split display.

The GenVisCalData (Generation of Visual Calibration Data) program, whose display is shown in Figure 2, is used to observe the robot under external control, and to log relevant kinematic and visual data. The program continually queries the robot for the current joint angles, allowing a live qualitative review of the quality of the current calibration.

The SumVisCal (Summary of Visual Calibration) program, whose display is shown in Figure 3, is used to automatically cycle through each configuration of the robot in a data set and update all measurements in the set. It will also summarize a calibration set, and is used to estimate the optimal joint angle offsets for a given data set (a daily calibration).

For each robot configuration in the set, this display shows the correlation between the visually-located and kinematically-predicted locations of the calibration fixture. If the updated calibration exactly predicted the visual measurements, the pairs of dots in the right hand frame would coincide in all cases.

The RoboVisCal (Robotic Visual Calibration) program is used to review a stored calibration set frame-by-frame. It is
also used either alone or as a slave to SumVisCal to actuate the robot to a particular recorded configuration and update the kinematic and visual measurements for this point in the data set. Figure 4 presents RoboVisCal’s display showing a typical configuration that is a member of one of these sets. The graphical conventions are the same as for GenVisCalData.

B. Daily Calibration

A daily calibration method has been developed using these software modules. First, a prerecorded set of configurations is loaded into the system. The SumVisCal program is used to drive the robot to each configuration and to update the kinematic and visual measurements. As this happens, the red or gray dots shown in Figure 3 will dynamically update, and the average distance error between the visual measurements and the kinematically derived predictions will be dynamically updated. This process takes approximately 10 minutes for a 65-element calibration set, primarily time to move the robot. A dataset could also be captured using GenVisCalData as described in Section II-A, but this is more time-consuming.

An optimization algorithm is then used to estimate a set of joint angle offsets \( \theta_i, \ i = 0 \ldots 10 \) for the arm and neck based on the current set of visual measurements. This process takes approximately 5 seconds per iteration, and can be done repeatedly to improve the estimate. As this is an iterative search with a random initial value, repeated optimizations on the same data may improve the results. A daily calibration thus consists of updating a set of visual measurements, followed by estimating the joint angle offsets. Currently, these estimates are manually input to Robonaut’s control system.

C. Full Calibration

A full calibration method has also been developed using these software modules, in conjunction with some Matlab code. An updated set of kinematic and visual measurements is taken as described above. This data set is saved to a text file and taken to Matlab, where an optimization algorithm is used to find a set of DHPs that explain best this set of measurements. This process takes from 25-120 minutes, depending mostly on the computational hardware. The results from this search are a full set of DHPs that can be used in Robonaut’s control software, as well as in the visual cortex, to accurately map between the manipulator workspace and the visual workspace.

D. Calibration Fixture

While hand-eye calibration could be performed using visual measurements of any point on the kinematic chain (or many points on the chain), we designed the calibration fixture shown in Figure 4 for several reasons. A visual measurement point distant from the wrist axes gives good observability for motions in the wrist roll and yaw axes. This particular fixture does not give good observability of wrist pitch. The center of a sphere is observable and well defined regardless of the relative pose between the cameras and fixture. The fixture also has a hand-guard to ensure a relatively repeatable grasp. This prototype fixture should be replaced with a more robust fixture that exhibits a very repeatable grasp and significant distance from the wrist axes in each of the wrist DOFs.

III. THEORY

This section describes the theoretical underpinnings of the above methods and presents the algorithms used in the calibration. First, the Sphere Hough Transform and its use in locating the calibration fixture in the eye coordinate system are described. Then, the setup for the nonlinear optimization at the heart of the hand-eye calibration system is described.

A. Finding a Sphere in a depth dataset

Central to this task is the accurate localization of the calibration fixture, shown in Figure 4, in the visual coordinate system. We utilize the existing depth-from-disparity stereo algorithms developed by the DRL and perform a search for a sphere-shaped object in the depth map (a 2D array of depths measured from the visual coordinate system origin).

The BallFinder algorithm begins with a seed location. This location is currently set to the kinematically derived prediction of the calibration fixture location, expressed in the visual coordinate system.

Points outside a large spherical region centered at this location are rejected from consideration. This pruning step rejects distant points, such as the floor, from further consideration as possible members of the sphere surface. Next, a
minimal surface area test is performed on all surviving points. Based on the distance of the seed location from the camera, the expected number of points on the sphere's surface is computed. Locations under consideration that are not members of a contiguous set of some fraction of this size are rejected from consideration. This pruning step eliminates small isolated regions. All remaining points participate in a voting scheme based on the generalized Hough Transform described below.

The Hough Transform is a classic computer vision algorithm in which lines are located in an image by allowing each point that is a member of a line to vote for some set of M lines that could have created this point. Lines which truly exist in an image will accrue more votes, and the top vote getters are very good candidates for lines in an image. See [8] for more detail on the standard Hough Transform. This algorithm can be extended to describe many types of parameterized shapes, such as circles or spheres. In the Sphere-Hough Transform, each point $P_{ss}$ that survives the pruning algorithms described above votes for a set of M spheres (centered at $P_{ss} , i = 1 \ldots M$). Points randomly sampled from the surface of a sphere centered at $P_{ss}$ (of which this point could be a surface point. Each of these points $P_{sc}$ represents one vote, in the Hough Transform paradigm, for a sphere centered at point $P_{sc}$. In our case, the voting is in Cartesian space, the radius of our calibration fixture is known. The location with the most votes is deemed the most likely to contain the actual sphere center. Figure 5 depicts a slice of the voting results from an example image (the depth slice that contains the winning vote) on the right, and on the left the input image with the winning 3-D location projected into it using the current camera calibration.

![Fig. 5. Sphere-Hough Transform Results](image)

B. Optimization

The daily and full calibrations described above differ only in which parameters are optimized. In this section, we will describe how this calibration is posed as an optimization problem. As described above, SumVisCal or GenVisCalData generates a set of joint measurements and visual measurements of the calibration fixture. For each configuration $i$ in the calibration set, the kinematic model is used to predict the location of the calibration fixture in the chest coordinate system. This is a function of the Denev-V'Hartenberg Parameters (DHPs) as well as the joint angles of the arm:

$$P_{ei} = A_1(DH, q_i)A_2(DH, q_i) \ldots A_7(DH, q_i)P_e,$$

where $P_e$ is the (fixed) position of the calibration fixture in the hand coordinate system, DH contains the DHPs for the arm and neck, and $q_i$ contains the joint angles for the arm and neck. The kinematic model for the neck is used to predict the transformation from the chest coordinate system to the eye coordinate system in the same way:

$$T_{ce,i} = A_{1N}(DH, q_i)A_{2N}(DH, q_i)A_{3N}(DH, q_i).$$

These transformations are used to create a kinematic estimate of the position of the calibration fixture in the eye coordinate system: $P_{ve,i} = (T_{ce,i})^{-1}P_{ei}$. We also have for each configuration $i$ in the calibration set the visual measurement of the 3-D position of the calibration fixture, also in the eye coordinate system, that we call $P_{v,i}$.

The optimization attempts to minimize the difference between $P_{v,i}$ (fixed) and $P_{ve,i}$ (function of DHPs) over all $i$ in the calibration set by search in DHP space. Our objective function (the function to minimize) for this search is the sum of the distances between point pairs in our calibration set. We currently use a Nelder-Mead simplex method [16] to minimize this function by search in the DHP space. For daily calibration, the joint angle offsets ($\theta_i = 0 \ldots 10$) are optimized. For a full calibration all nonzero (and non-$\pi/2$) DHPs are optimized. Several pairs of offsets, designed to be symmetric, are constrained to be equal and only contribute one dimension to the DHP search space.

IV. RESULTS

Four experiments were performed to validate our calibration procedures. In Section IV-A, the mean error between a set of visual observations and kinematically derived predictions is compared with the existing and revised DHPs. In Section IV-B, the updated DHPs derived from the data above are used in conjunction with a daily calibration on a different day. In Section IV-C, a daily calibration is performed on one-half of a dataset, and the errors in both this set and the half of the dataset not used for training are evaluated. Finally, in Section IV-D, our calibration is observed to cause an improvement in Robonaut's performance on an autonomous wrench-grasping experiment.

A. Effect of Full Calibration

The as-designed DHPs for the arm and neck are presented in Table I. A set of 67 robot configurations were chosen, and the reported joint angles and visual measurements logged. The mean distance between the kinematically derived prediction for these measurements and the actual visual measurements was 13.75cm.

A full calibration was performed on this data set. The DHPs shown in Table II were found. For the same set of 67 configurations, the mean distance between the kinematically derived prediction (using the updated DHPs) for these measurements and the visual measurements was 1.85cm. These data are summarized in Figure 3.

For convenience, we take $q_i = [q1,arm\ldots q7,arm_q1,neck_q2,neck_q3,neck]^T$, and similarly concatenate the DHPs.
B. Effect of Daily Calibration

The DHPs shown in Table II were used to predict the location of the calibration fixture in a set of 150 unique configurations, with an average error of 7.94 cm. This was several days (and several power cycles) after the experiment described in Section IV-A, so we expect that the reported joint angles deviated from the actual joint angles by different amounts than estimated in Table II. A daily calibration was used to compute the updated joint angle offsets shown in Table III. The remainder of the DHPs were as shown in Table II. The average error was reduced to 2.02 cm over this dataset.

C. Effect on Non-Training Data

In this experiment, half a data set was used to tune the DHPs, and the power of these parameters to predict the position of the calibration fixture in the other half of the data set was tested. The DHPs shown in Table II were used to predict the location of the calibration fixture in a set of 75 unique configurations, the first half of the configurations from Section IV-B, with an average error of 7.87 cm. This was several days (and several power cycles) after the experiment described in Section IV-A, so we expect that the reported joint angles deviated from the actual joint angles by different amounts than estimated in Table II. A daily calibration was used to compute the updated joint angle offsets shown in Table III. The remainder of the DHPs were as shown in Table II. After this calibration, the average error was reduced to 2.04 cm over this dataset. This set of DHPs was then used with no further optimization to predict the position of the calibration fixture in the 75 configurations that had not been used in training, the second half of the configurations from Section IV-B. Over this dataset, the DHPs from Tables II and IV produced an average prediction error of 2.35 cm.

D. Wrench-Grasping Experiment

As an example of the types of tasks that the DRL is demanding of Robonaut, this section presents the contribution of visual calibration to an experiment run by a team from Vanderbilt University on autonomous wrench-grasping. In this experiment, a teleoperator is observed grasping wrenches in nine different workspace locations. Figure 6 shows the physical setup for this experiment. Robonaut’s vision system is used to observe the wrench in a unique location, and a learning algorithm [13] is used to grasp the wrench in this location.

In this experiment, described in more detail in [14], a 6DOF Cartesian-space vision-workspace correction was implemented. This correction was a linear interpolation between vision/workspace mismatches recorded at several locations using teleoperator data. This workspace correction reduced vision/kinematic mismatches, but not enough to enable Robonaut to grasp the wrench. The updated DHPs shown in Table II were experimentally placed into the inverse kinematics procedures for Robonaut, and the workspace correction was removed. The system was immediately able to grasp wrenches at several different positions in the workspace.

V. DISCUSSION AND CONCLUSIONS

Closed-loop self-calibration of the combined kinematic and visual systems for Robonaut Unit A has been performed. This calibration does not explicitly register the visual or the kinematic system with ground-truth, but modifies the perceptions associated with the kinematic movements to match
the perceptions of the vision system. Procedures and algorithms have been developed that will enable the robot to be recalibrated when necessary. These procedures have reduced vision-kinematic mismatch from 13-15 cm to 2-3 cm in various situations, and have enabled the DRL team to continue increasing the autonomous capability of Robonaut.

There are several directions in which this work could be improved. The most obvious is to do a careful extrinsic calibration of Robonaut's vision system so that this closed-loop procedure will more accurately reflect distances and rotations in the workspace. Also useful would be to systematically study the number of measurements required to calibrate the system, both in the reduced and full cases. The system should also be extended to calibrate the left arm of Unit A and each arm of Unit B. The method described in this paper should extend to these situations in a very straightforward manner. With some extension, this method could be extended to the simultaneous calibration of the vision system and both arms of Robonaut.

ACKNOWLEDGMENT

The authors would like to thank NASA, DARPA, and the Center for Intelligent Systems at Vanderbilt University for their contributions to this work.

FIG. 6. Vanderbilt wrench-grasping experiment

### TABLE III

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_j$</td>
<td>-1.70</td>
<td>0.243</td>
<td>-0.149</td>
<td>2.323</td>
<td>3.810</td>
<td>-90.634</td>
<td>9.340</td>
<td>0.601</td>
<td>-61.249</td>
</tr>
</tbody>
</table>

### TABLE IV

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_j$</td>
<td>0.186</td>
<td>0.153</td>
<td>0.404</td>
<td>3.198</td>
<td>2.716</td>
<td>-88.780</td>
<td>8.946</td>
<td>-0.121</td>
<td>-62.990</td>
</tr>
</tbody>
</table>

### REFERENCES


