Mirror Perspective-Taking with a Humanoid Robot

Justin W. Hart and **Brian Scassellati**

Department of Computer Science
Yale University
New Haven, CT, USA
justin.hart@yale.edu and scaz@cs.yale.edu

Abstract

The ability to use a mirror as an instrument for spatial reasoning enables an agent to make meaningful inferences about the positions of objects in space based on the appearance of their reflections in mirrors. The model presented in this paper enables a robot to infer the perspective from which objects reflected in a mirror appear to be observed, allowing the robot to use this perspective as a virtual camera. Prior work by our group presented an architecture through which a robot learns the spatial relationship between its body and visual sense, mimicking an early form of self-knowledge in which infants learn about their bodies and senses through their interactions with each other. In this work, this self-knowledge is utilized in order to determine the mirror's perspective. Witnessing the position of its end-effector in a mirror in several distinct poses, the robot determines a perspective that is consistent with these observations. The system is evaluated by measuring how well the robot's predictions of its end-effector's position in 3D, relative to the robot's egocentric coordinate system, and in 2D, as projected onto it's cameras, match measurements of a marker tracked by its stereo vision system. Reconstructions of the 3D position end-effector, as computed from the perspective of the mirror, are found to agree with the forward kinematic model within a mean of 31.55mm. When observed directly by the robot's cameras, reconstructions agree within 5.12mm. Predictions of the 2D position of the end-effector in the visual field agree with visual measurements within a mean of 18.47 pixels, when observed in the mirror, or 5.66 pixels, when observed directly by the robot's cameras.

Introduction

When we look into a mirror, the image that we see is a reflection of what actually exists in space. Objects in this reflection appear as if they exist on the other side of the mirror, opposite their real-world counterparts. If one were to naïvely reach towards these reflections, one's hand would hit the glass of the mirror, rather than the object being reached for. By understanding this reflection, however, one is able to use the mirror as an instrument to make accurate inferences about the positions of objects in space based on their reflected appearances. When we check the rearview mirrors on our cars for approaching vehicles or use a bathroom mirror to aim a hairbrush, we make such instrumental use of these mirrors.

The use of mirrors for spatial reasoning is a precursor to what is tested in the widely-known "Mirror Test," as originally proposed by Gallup (1970), which has become the classical test of self-awareness in humans and animals. After an animal is given time to acclimate to the presence of a mirror, it is anesthetized and marked on the face with odorless, non-tactile dye. The animal's reaction to their reflection is used as a gauge of their self-awareness, based on whether they inspect the mark on their own body, or react as if it does not appear on themselves, as in cases where they react as if it is a mark on another animal. The test has now been performed in many variations and on many animals (see Gallup, Anderson, and Shillito (2005) or Anderson and Gallup (2011) for reviews), but to date, only a few non-human species pass these tests, including some primates (Anderson and Gallup 2011), elephants (Plotnik, de Waal, and Riess 2006), and dolphins (Reiss and Marino 2001). Infants are unable to pass this test, developing the necessary skills by around 18 months (Bertenthal and Fischer 1978).

Tests have been devised to determine whether animals that are unable to pass the classical mirror test are able to use mirrors as instruments to solve spatial reasoning tasks. These tests have shown that there is a larger category of animals that are capable of such instrumental use. Infants who are too young to pass the mirror task can retrieve an object that is presented behind them in a mirror, demonstrating a self-centered awareness of space and reflectance (Bertenthal and Fischer 1978). Marmosets (which fail the classical test) are able to use a mirror to obtain food pellets that are visible only in a mirror reflection (Heschl and Burkart 2006). Using both mirrors and monitors displaying live video feeds of their arms, chimpanzees can overcome inversions and rotations of these images, manipulations which break the spatial relationship that can be established by looking into a mirror, using these images for spatial reasoning, thus demonstrating even more general spatial reasoning capabilities than mirror use (Menzel, Savage-Rumbaugh, and Lawson 1985).

Inspired by the classical Mirror Test (Gallup 1970), a number of projects have attempted to mimic the act of mirror self-recognition in robots. Michel, Gold, and Scassellati (2004) and Gold and Scassellati (2007) solved a task of im-

age segmentation, classifying pixels as either belonging to the robot ("self") or not ("other"), based on temporal correlations between changes in the visual field and the initiation of motor activity. However, this system is unable to pass the classical Mirror Test because it does not model the visual appearance of the robot. Takeno, Inaba and Suzuki (2005) observe a robot when imitating its reflected image in a mirror, versus when imitating another robot, to determine that the robot can distinguish between the two using a neural network architecture. The mirror behavior in this task, however, is based on flashing LEDs, and the robot performing this task has no way of interpreting visual self-image in the way that the Mirror Test requires. More recently, the Qbo robot was programmed to respond differently to images of itself (using first-person pronouns) rather than other objects, by using object recognition techniques (Ackerman 2011). This system is designed to respond differently to the specific image corresponding to that of the robot, by cueing the system with the phrase, "This is you, Qbo," during training. If it were trained to respond this way to a different image, then Qbo would respond to that object as if it was itself.

While all of these systems incorporate impressive technologies, none of them are able to pass Gallup's (1970) classical Mirror Test, nor are any of them capable of using a mirror for spatial reasoning. The present model is a component of an architecture that we are developing with the intention of building a robot that is capable of passing the classical Mirror Test. This component enables the robot to estimate a visual perspective that is consistent with observations of the position of its end-effector, as reflected in a mirror when the robot moves into different poses. In this way, self-knowledge regarding its kinematics and visual system enables the robot to use a mirror for spatial reasoning.

An Architecture for Mirror Self-Recognition

The overall goal of this project is to develop an architecture that allows a robot to pass the Mirror Test. The proposed architecture is composed of six models describing different forms of self-knowledge that we believe are sufficient to accomplish this task. They are the *End-Effector Model*, the *Perceptual Model*, the *Perspective-Taking Model*, the *Structural Model*, the *Appearance Model*, and the *Functional Model*. These are learned by the robot through observation, allowing for refinement and change over time. We propose that this process of self-observation will enable the system to pass the Mirror Test, as the system will detect differences between its expected appearance and its current one.

The present work develops a version of the Perspective-Taking Model, given a working implementation of portions of the End-Effector and Perceptual Models. This portion of the architecture will enable the robot to perform the spatial reasoning required to pass the Mirror Task. By allowing the robot to take the perspective of a mirror, the Perspective-Taking model allows the robot to compute a projection of a self-taught 3D visual representation of its appearance in the mirror, for comparison against the image of itself reflected in a mirror. The model should be computable in a small enough number of samples for inference to be practical, while being accurate enough to make meaningful comparisons between the robot's expected reflected appearance and its measured reflected appearance. As such, evaluation focuses on the number of samples required to train the model and the spatial accuracy of its reconstructions.

The End-Effector and Perceptual Models

The End-Effector Model, Figure 1a, describes the motion of the robot's end-effectors through space. Its kinematics are modeled using the Denavit-Hartenberg convention (Denavit and Hartenberg 1955). It is learned based on observations made by the robot's stereo vision system.

The Perceptual Model describes the robot's vision system using the common Pinhole Camera model (Hartley and Zisserman 2004). It is capable of both reconstructing a 3D point in space, given 2D coordinates in both of the robot's stereo cameras, and projecting a known 3D point to its corresponding 2D coordinates in each camera.

An important feature of these two models is that they are calibrated to each other. By learning the End-Effector Model through the stereo vision system, the samples used to reconstruct the robot's kinematic chains are expressed in the native coordinate system of the stereo vision system. The mounting of the robot's cameras with respect to its frame is known, as in Figure 1b. These subsystems are able to refine each-other's calibrations by minimizing the difference between the expected positions of the robot's end-effectors in each camera and their observed positions, utilizing models and methods that we developed in prior work (Hart and Scassellati 2011).

The Perspective-Taking Model

The Perspective-Taking Model, Figure 1c, is an extension of the Perceptual Model that allows the robot to model sensors from an external point of view. Though one could imagine social functions of this model, such as representing the visual perspectives of other actors in an interaction, the focus of the work presented in this paper is to allow the robot to take the perspective of a mirror in its current visual landscape.

The Structural and Appearance Models

The Structural Model will represent the robot's rigid, external 3D structure, as shown in Figure 1d. It will be computed by automatically, choosing features along the robot's frame and computing a model of their position using the techniques from the End-Effector Model. These points will become control points in splines approximating the robot's surface geometry. The Appearance Model will map surface properties, such as color, onto this geometry, as in Figure 1e.

The Functional Model

The Functional Model will allow the robot to determine the effect that its actions have on objects in its environment. It will enable the robot to infer causal relationships, such as that between enacting a motor command and changes in the visual field, and will be based on the methods presented by Michel et al. 2004, Gold and Scassellati 2007, and Stoytchev 2007, for related tasks.
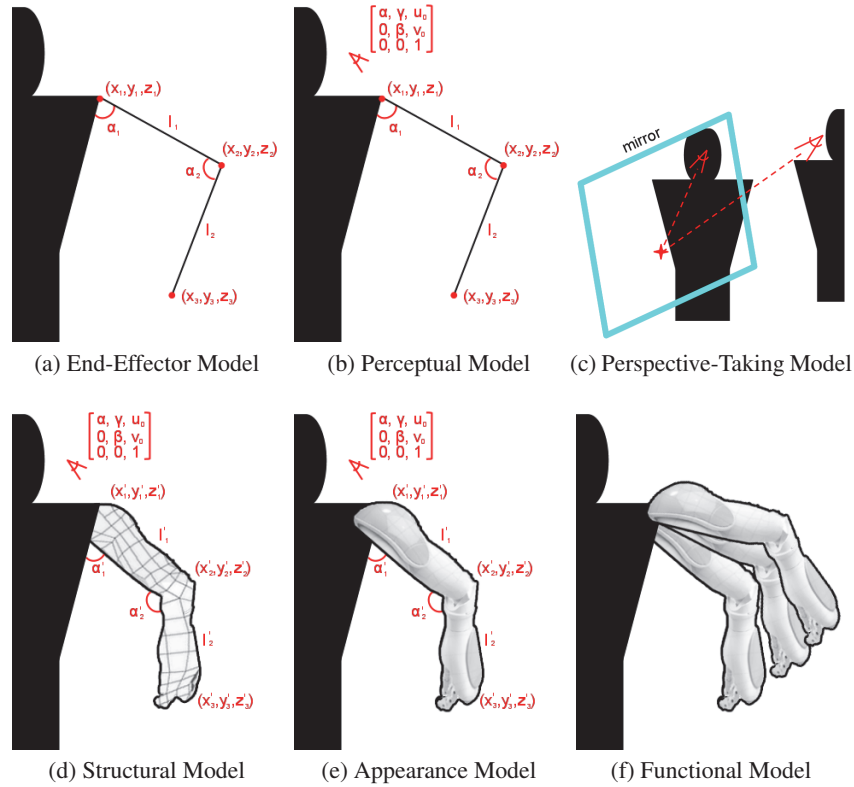
(a) End-Effector Model   (b) Perceptual Model   (c) Perspective-Taking Model

(d) Structural Model   (e) Appearance Model   (f) Functional Model

Figure 1: Diagrams describing the six of the basic components of the proposed architecture for mirror self-recognition.

The Mirror Perspective Model

Consider the scenario in Figure 2. The robot is only able to observe the reflection of its end-effector in the mirror, as it is not directly in its visual field. Reconstructions of the reflected end-effector's position in space will place it behind the plane of the mirror, rather than in front of the mirror, where it actually is. In order to accurately reconstruct positions reflected in the mirror, they must be computed from the reflection of the camera's perspective, as in Figure 1c.

The Mirror Perspective Model allows the robot to estimate this visual perspective. In order to do so it leverages the robot's Perceptual and End-Effector Models, allowing the robot to compute a virtual calibration for the perspective of each of its stereo cameras, for objects that they witness reflected in the mirror. We will call each of these cameras a *Mirror-Perspective Camera*. The basic method for calibrating these cameras is for the robot to move into several poses, yielding a known set of 3D points in space, and their corresponding 2D images. In this way, the technique in this section is a form of photogrammetric calibration (Faugeras 1993), with the robot acting as its own calibration target.

Because this model deals with mirrors, it will frequently be the case that variables are related based on their disposition with respect to the mirror. Variables referring to quantities based on reflections, rather than the original, physical properties of the robot, are marked with a caret. For example, whereas J_i is the position of the robot's end effector, $\hat{J}_i$

is the reconstruction of its reflection in the mirror. Because the robot samples many poses of its arm, the subscript i is used to refer to a set of variables describing a single pose.

The procedure is as follows:

1. Sample reflected end-effector images - Record three versions of the end-effector's position:

   J_i Predicted by the End-Effector Model and appearing in front of the mirror.

   $\hat{J}_i$ Reconstructed by the Perceptual Model from the point of view of the robot's cameras and appearing behind the mirror.

   $\hat{j}_i$ **&** $\hat{j'}_i$ Two dimensional positions of the end-effector in both cameras, appearing as reflections in the mirror.

2. Compute an initial estimate of each mirror perspective camera, based on the plane in which the mirror lies.

3. Nonlinear refinement of the pair of mirror perspective cameras.

Background

Homogeneous Coordinates   The homogeneous representation of a point in space is expressed as a vector. A 2D point is expressed as a vector with 3 elements $< x, y, w >$, a 3D point is expressed as a vector with 4 elements $< x, y, z, w >$. The Cartesian equivalent of a homogeneous
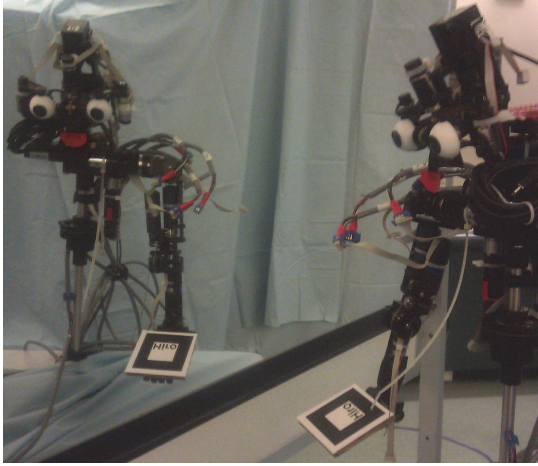
Figure 2: The humanoid robot, Nico, as configured for evaluation of the system.

coordinate is computed as a ratio of its elements, as in $(x : w, y : w)$.

The homogeneous representation of a plane is expressed as a vector, $< \Pi_1, \Pi_2, \Pi_3, \Pi_4 >$. The first three elements correspond to the vector perpendicular to the plane. When normalized such that $< \Pi_1, \Pi_2, \Pi_3 >$ is a unit vector, $-\Pi_4$ corresponds is the Euclidean distance to the plane from the origin. The dot product of the homogeneous representations of a point and a plane corresponds to the distance between the two. This model will also make use of the fact that the intersection of 3 planes defines a point.

The Pinhole Camera Model The Pinhole Camera Model (Hartley and Zisserman 2004) is a convention commonly used to describe a camera's projection. The *camera intrinsic matrix* is expressed as a 3×3 matrix as in Equation 1. It describes parameters which are intrinsic to the camera, independent of its position and orientation. The parameters α and β describe focal length, their ratio accounting for non-square pixels. The parameters (u_0, v_0) describe the *principal point*, where a ray perpendicular to the camera's image plane runs through the camera center. The parameter γ is a skew factor accounting for non-rectangular pixels. The extrinsic parameters R and C describe the rotation of imaged points about the coordinate frame centered at the camera, and the camera's position in space, respectively. Expressed as in Equation 2, the camera's extrinsic parameters yield the projection of an *ideal camera*, whose camera intrinsic matrix is identity. Multiplying this projection by K yields the *camera projection matrix*, Equation 3. In this implementation, radial lens distortion is corrected for using the first two terms of a commonly used model, as found in Zhang (2000).

Three-dimensional points can be projected to their imaged equivalents as in Equation 4, where J is the 3D point, and j its image. Given a point imaged in a stereo pair of cameras, j and j', and the calibrated parameters describing the projections of both cameras, it is possible to reconstruct the

position of that point in 3D space.

$$K = \begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \qquad (1)$$

$$O = [R| - RC] \qquad (2)$$

$$P = K[R| - RC] \qquad (3)$$

$$j = K[R| - RC]J \qquad (4)$$

Kinematics This system uses the *Denavit-Hartenberg parameters* (Denavit and Hartenberg 1955) to model kinematic chains. Each joint is represented by four parameters which describe the transformation performed by a joint in the chain, three geometric ones which describe the characteristics of the joint, and θ, the joint angle. For each joint in a kinematic chain containing n joints, a matrix M can be computed which describes the transformation performed by the joint. The matrix M_0 represents the transformation to the first joint, with M_1 representing the first joint itself. The 3D position of the end effector J_E, then, can be computed by Equation 5, using the robot's joint angles. This is frequently referred to as the robot's *forward-kinematic model*.

$$J_E = M_0 \ldots M_n[0, 0, 0, 1]^T \qquad (5)$$

The Perceptual and End-Effector Models Using techniques described in prior work (Hart and Scassellati 2011), the End-Effector and Perceptual Models are calibrated to each other. This allows a prediction of a 2D image of the robot's end effector's position in its cameras, j_E to be determined by Equation 6.

$$j_E = K[R| - RC]M_0 \ldots M_n[0, 0, 0, 1]^T$$

Estimating the Mirror-Perspective Camera

Because an object's image in a mirror is a reflection of its real-world counterpart, its position in space can be correctly interpreted from the perspective of a camera whose position and orientation have been reflected with respect to the plane of the mirror. The goal of this model is to determine the parameters describing the mirror-perspective camera, $\hat{P}$, of a real-world camera, P, observing objects reflected in a mirror. The mirror-perspective can be determined by reflecting the camera's position and orientation across the plane of the mirror, the intrinsic parameters for the camera and its reflection are the same, requiring only the position $\hat{C}$ and orientation $\hat{R}$ of the mirror-perspective camera to be estimated.

Sample Reflected End-Effector Images First the robot moves its end-effector into a set of random poses that can be witnessed in the mirror. It records J_i, $\hat{J}_i$, and $(\hat{j}_i, \hat{j'}_i)$ for each pose. J_i is the position of the end-effector computed by the robots End-Effector Model. The coordinates $(\hat{j}_i, \hat{j'}_i)$ are the images of the end-effector's reflection in the mirror. $\hat{J}_i$ is reconstructed from $(\hat{j}_i, \hat{j'}_i)$ by the robot's perceptual model.

Compute Initial Estimate To simplify the process of computing the estimate of mirror-perspective camera position and orientation, we assume that the camera is situated at

the origin with $R = \mathbb{1}$. Because the robot's cameras are calibrated, this can be accomplished by transforming the sampled J_i's and $\hat{J}_i$'s into the camera's coordinate frame prior to computing the mirror plane, and transforming $\hat{R}$ and $\hat{C}$ back after they are computed.

Mirror plane estimation Because J_i and $\hat{J}_i$ should lie symmetrically about the plane, for each arm pose, the plane in which the mirror lies can be approximated as follows.

First the vector perpendicular to this plane is computed. This is computed as the mean vector from the J_i's to the $\hat{J}_i$'s, using their Euclidean representations, as shown in Equation 6.

$$< \Pi_1, \Pi_2, \Pi_3 > = \frac{\sum_{i=1}^{n} \hat{J}_i - J_i}{n} \qquad (6)$$

The plane corresponding to the correct orientation, centered at the origin is then computed by Equation 6. The distance of the J_i's and $\hat{J}_i$'s from this plane can then be used to compute Π_4, as in Equation 8, placing the mirror plane equidistant to the two sets of points.

$$Q_{origin} = < \Pi_1, \Pi_2, \Pi_3, 0 > \qquad (7)$$

$$\Pi_4 = -\frac{\sum_{i=1}^{n} Q_{origin} \cdot \hat{J}_i + \sum_{i=1}^{n} Q_{origin} \cdot J_i}{2n} \qquad (8)$$

$$Q = < \Pi_1, \Pi_2, \Pi_3, \Pi_4 > \qquad (9)$$

Estimating mirror-perspective camera position and orientation Figure 3 provides a diagram of the position and orientation of the real camera and the mirror-perspective camera with respect to the mirror. The reversal of mirror images is accounted for by having the mirror-perspective camera oriented such that it is looking away from the mirror, as if points are being imaged from behind it.
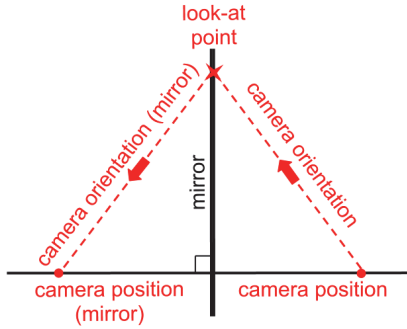

Figure 3: Diagram of the position and orientation of the real camera and mirror-perspective camera with respect to the mirror.

Computing the mirror-perspective camera's position Because Q is expressed in the camera's coordinate frame, $< \Pi_1, \Pi_2, \Pi_3 >$ is the vector perpendicular to the mirror from the camera's position. Normalizing Q such that $< \Pi_1, \Pi_2, \Pi_3 >$ is a unit vector allows the position of the mirror-perspective camera to be computed by Equation 10.

$$\hat{C} = -2\Pi_4 < \Pi_1, \Pi_2, \Pi_3 > \qquad (10)$$

Computing the mirror-perspective camera's orientation Camera Projection Matrices can be interpreted as sets of three planes from which the distance of a 3D point is computed in order to determine its projection. Relatedly, the first two rows of the matrix O, from Equation 2, describe the XZ and YZ planes of the camera's coordinate system. Knowing that three planes meet at a single point, the intersection of the camera's z-axis with the mirror plane, L, can be computed according to Equation 11. The z-axis of the mirror-perspective camera, then, can be computed according to Equation 12. Its rotation, $\hat{R}$, is the transpose of the rotation from the canonical z-axis ($< 0, 0, 1 >$) to the mirror-perspective camera's z-axis, computed as a rotation about the axis perpendicular to both.

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \Pi_1 & \Pi_2 & \Pi_3 & \Pi_4 \end{bmatrix} L = 0 \qquad (11)$$

$$\hat{Z} = \hat{C} - L \qquad (12)$$

Nonlinear Refinement The estimate yielded by the previous step can be refined by minimizing the distance between estimated projections of the robot's end-effector position and their imaged equivalents for m samples, according to Equation 13. Optimizations in the presented results use LevMar (Lourakis Jul 2004), an implementation of the Levenberg-Marquardt algorithm in C++.

$$f(R, C) = \sum_{i=1}^{m} \| K[\hat{R}| - \hat{R}\hat{C}] J_i - \hat{j}_i \|^2 \qquad (13)$$

Evaluation

Setup

The system was implemented and evaluated on the upper-torso humanoid robot, Nico, seen in Figure 2. The robot includes a stereo vision system with two 640×480 resolution cameras. The evaluation utilized four degrees of freedom (DOFs) in the robot's right arm, which is composed of two main linkages, with pairs of joints mounted at the shoulder and elbow. The linkage from the shoulder to the elbow is 130mm long, and from the elbow to the end-effector is 127mm. The end-effector is tracked by the vision system through the use of fiducial markers implemented using AR-ToolKit (Kato and Billinghurst 1999). For this experiment the tracker was modified to provide more accurate 2D positioning of the fiducial marker in the visual field. These 2D positions are provided to the robot's stereo vision system to be reconstructed into 3D positions, rather than inferred from the image of the fiducial marker by ARToolKit.

The End-Effector and Perceptual Models were calibrated in the following way. The stereo vision system was first calibrated using Zhang's method (Zhang 2000), then refined via bundle adjustment (Hartley and Zisserman 2004). Known-good intrinsic parameters were then substituted for the estimates yielded by the calibration process, which then re-performed the bundle-adjustment procedure, pinning the intrinsic parameters, in order to derive an accurate estimate
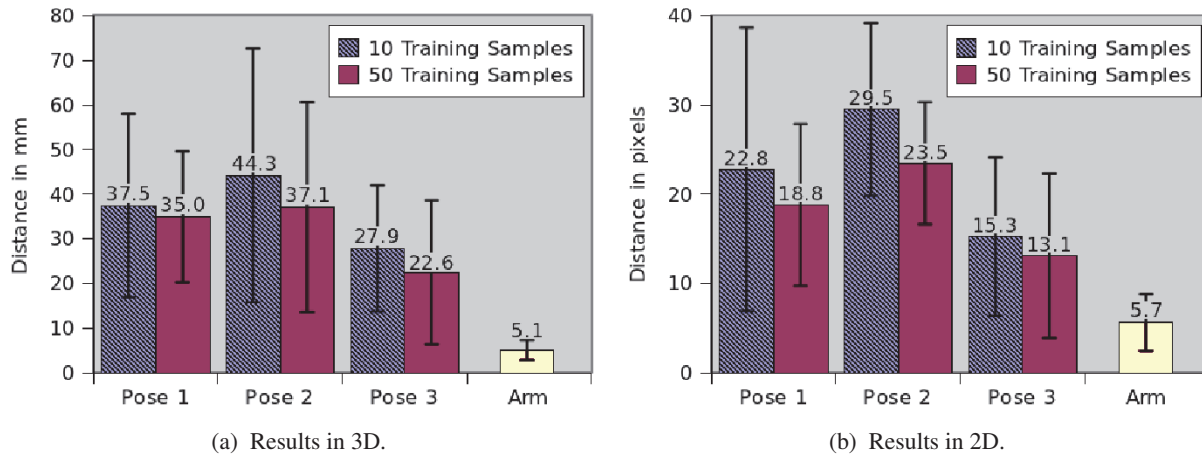
(a) Results in 3D.

(b) Results in 2D.

Figure 4: Distance between end-effector position predicted by the End-Effector Model and as tracked in the visual field over three mirror views. "Arm" denotes the robot's ability to perform this task on the arm when witnessed directly in the visual field.

of the extrinsic parameters. Known-good radial distortion parameters were used throughout this process. Kinematic parameters that had been estimated using the method from Hart and Scassellati (2011) were provided to the system to initialize the End-Effector model. The arm was then moved into 100 new, unique poses in order to re-calibrate the End-Effector and Perceptual Models to changes in the pose of the eye cameras, zeroing of the robot's encoders, and a new fiducial marker. This larger fiducial marker had to be used, because the robot's cameras were not of sufficient resolution to find the one used in Hart and Scassellati (2011) when reflected in the mirror, due to the larger apparent distance from its cameras.

Testing was performed over datasets containing 150 poses of the robot's arm, 50 of which were used for training, 100 for testing. Three such datasets were sampled with the robot observing its arm in a mirror, which appeared in a different position and orientation for each dataset, dominating the robot's field of view. Though the system was tested by batch-processing these datasets, for efficiency, no apparent technical barriers exist to developing a version of this system which operates in real-time, and efforts to do so have already commenced.

To measure system performance, the mean distance between predictions of end-effector position and measured end-effector position in 3D and 2D are reported. This provides an estimate of how well the Mirror-Perspective Model has been measured with respect to the robot's existing End-Effector and Perceptual Models, though it has the shortcoming that the end-effector will appear more distant in the mirror and, thus, measurements are inherently less accurate. It relates well, however, to the goal of passing the classical Mirror Test. The main difference is that robot makes predictions regarding the position of its end-effector in the mirror based on self-knowledge, rather than predictions regarding its appearance in the mirror. This test is also a form of instrumental mirror use, in that the robot compares predictions

of its end-effector position based on its forward-kinematic model, and measured positions based on observations made in the mirror, in its egocentric frame. The system was trained on datasets of varying lengths in order to establish the number of samples required to adequately train the system.

Data Collection

A mirror was mounted to a moveable, tiltable whiteboard, and placed into the field of view of the robot. Three positions were chosen for the mirror. For each pose a set of 50 training and 100 test samples of with the arm in various poses, imaged as reflected in the mirror, was collected.

Results

As can be seen in Figures 4a and 4b, the system performs well even after training on only 10 arm poses. While the robot is able to predict the position of the end-effector viewed directly in its visual field much better than it is able to in the mirror, it still outperforms competing systems that only attempt to predict end-effector position directly in their visual field, when doing so. Recent such systems include ones presented by Hersch, Sauser, and Billard (Hersch, Sauser, and Billard 2008) and Cantin-Martinez, Lopes, and Montesano (Cantin-Martinez, Lopes, and Montesano 2010), who both report performance to be within 5cm, and attempt neither the task of predicting end-effector position in pixels, nor the task of predicting end-effector position in a mirror.

Part of the system's degrade in performance when performing this task in the mirror can be attributed to the apparent distance of the end-effector when viewed in the mirror. The apparent distance of the object, when viewed in the mirror, combines the distance of the robot from the mirror and the distance of the object from the mirror. As a result, the view of the object is much farther away. Reconstructions of the tracked point are subject to a higher degree of error due to this, leading to a greater degree of disagreement, as the

same area of visual angle contains a greater physical area. This is consistent with the fact that performance in pixels is more similar between the arm in the visual field and the arm in the mirror, than performance in millimeters. Because the mirror-perspective cameras are optimized independently from each other, it is possible for the system to estimate positions and orientations for these cameras which changes their position and orientation with respect to each other. This also contributes to error. By optimizing the position of the mirror, and computing the mirror-perspective cameras from this position, we should be able to improve performance. This is saved for future work.

Discussion

In this paper, an architecture is proposed to enable a robot to pass the classical test of self-awareness, the Mirror Test (Gallup 1970). This architecture proposes to learn a model of the structure of the body and senses that is sufficient to make predictions about the appearance of the robot as reflected in a mirror. In this work, we have developed a model that allows the robot to determine a perspective that is consistent its point of view when looking into a mirror. To do so, it uses self-knowledge about its body and senses in the form of kinematic and visual calibration information. To our knowledge, this is the first robotic system to attempt to use a mirror in this way, representing a significant step towards a cohesive architecture that allows robots to learn about their bodies and appearances through self-observation, and an important capability required in order to pass the Mirror Test.

Acknowledgments

References

Ackerman, E. 2011. Qbo robot passes mirror test, is therefore self-aware. Retrieved from http://spectrum.ieee.org/automaton/robotics/artificial-intelligence/qbo-passes-mirror-test-is-therefore-selfawar.

Anderson, J., and Gallup, G. J. 2011. Which primates recognize themselves in mirrors? *PloS Biology* 9(3).

Bertenthal, B. I., and Fischer, K. W. 1978. Development of self-recognition in the infant. *Developmental Psychology* 14(4):44–50.

Cantin-Martinez, R.; Lopes, M.; and Montesano, L. 2010. Body schema acquisition through active learning. In *IEEE Intl. Conf. on Robotics and Automation*.

Denavit, J., and Hartenberg, R. S. 1955. A kinematic notation for lower-pair mechanisms based on matrices. *Trans. of the ASME J. of Applied Mechanics* 23:215–221.

Faugeras, O. 1993. *Three-Dimensional Computer Vision: a Geometric Viewpoint*. Cambridge, Massachusetts: MIT Press.

Gallup, G. J.; Anderson, J.; and Shillito, D. 2005. Dynamic ambient paradigms. In Doe, A., ed., *Paradigm Gems 2*. Addison Wesley. 223–233.

Gallup, G. G. 1970. Chimpanzees: Self-Recognition. *Science* 167(3914):86–87.

Gold, K., and Scassellati, B. 2007. A bayesian robot that distinguishes "self" from "other". In *Proceedings of the 29th Annual Conference of the Cognitive Science Society*.

Hart, J., and Scassellati, B. 2011. A robotic model of the ecological self. In *Proceedings of the 11th IEEE-RAS International Conference on Humanoid Robots (HUMANOIDS)*.

Hartley, R. I., and Zisserman, A. 2004. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition.

Hersch, M.; Sauser, E.; and Billard, A. 2008. Online learning of the body schema. *Intl. J. of Humanoid Robot.* 5(2):161–181.

Heschl, A., and Burkart, J. 2006. A new mark test for self-recognition in non-human primates. *Primates* 47(3):187–198.

Kato, H., and Billinghurst, M. 1999. Marker tracking and hmd calibration for a video-based augmented reality conferencing system.

Lourakis, M. Jul. 2004. Levmar: Levenberg-marquardt nonlinear least squares algorithms in C/C++. [web page] http://www.ics.forth.gr/~lourakis/levmar/. [Accessed on 31 Jan. 2005.].

Menzel, E. J.; Savage-Rumbaugh, E.; and Lawson, J. 1985. Chimpanzees (pan troglodytes) spatial problem solving with the use of mirrors and televised equivalents of mirrors. *Journal of Comparative Psychology* 99(2):211–217.

Michel, P.; Gold, K.; and Scassellati, B. 2004. Robotic self-recognition. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*.

Plotnik, J.; de Waal, F.; and Riess, D. 2006. Self-recognition in an asian elephant. *Proceedings of the National Academy of Science* 103(45):17053–17057.

Reiss, D., and Marino, L. 2001. Mirror self-recognition in the bottlenose dolphin: A case of cognitive convergence. *Proceedings of the National Academy of Science* 98(10):5937–5942.

Stoytchev, A. 2007. *Robot Tool Behavior: A developmental approach to autonomous tool use*. Ph.D. Dissertation, Georgia Institute of Technology, Atlanta, Georgia, USA.

Takeno, J.; Inaba, K.; and Suzuki, T. 2005. Experiments and examination of mirror image cognition using a small robot. In *Proceedings of the 6th International Symposium on Computational Intelligence in Robotics and Automation (CIRA)*.

Zhang, Z. 2000. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* 22:1330–1334.