Calibrating the Eye Motion of an Humanoid Robot

Justin Hart, Brian Scassellati, and Steven Zucker
justin.hart@yale.edu, scaz@cs.yale.edu, steven.zucker@yale.edu

In the human visual system, the projective relationship between the images seen in each eye with each other changes with their motion as the viewer attends to different points in space. The active vision heads built for many humanoid robots approximate human gaze behavior and share this property. Knowledge of this projective relationship is used in stereo vision tasks and is captured entirely by the position and orientation of the cameras with respect to each other and properties intrinsic to the individual cameras. In this work we present a method for inferring a kinematic model of a robots active vision system and use it to estimate of the system's epipolar geometry as it changes when the cameras undergo motion. This kinematic model constitutes a model of the self in terms the visual system.

Epipolar geometry describes the projective relationship between the two cameras in a stereo vision system. We model these cameras using the *pinhole camera model*. In a classical pinhole camera, we are able to take pictures because a pinhole in the center of the camera allows only a narrow ray of light to pass through, carrying the color information of the object that it bounced off of prior to entering the camera. Therefore, where a single pixel imaged by a pinhole camera resides in 3-space is constrained to a single ray of light. In a stereo pair of cameras, this ray of light will lie along a single line in the other image. This is referred to as the *epipolar constraint*, so named because all such *epipolar lines* pass through the *epipole*, the image of the other camera's *camera center*[1]. Estimating this relationship is referred to simply as the *estimation of epipolar geometry*, whereas estimating those parameters intrinsic to the camera is referred to as *camera calibration*.

An *active vision system* is one in which either the cameras are able to move or in which they include manipulators that are able to interact with the environment. An example of such a vision system is a camera attached to a movable arm, referred to as a *hand camera*. *Hand-Eye* and *Head-Eye* calibration both refer to the process of inferring the mounting of a camera to an underlying system with known kinematics. If these kinematics are unknown, the process of inferring them is *kinematic calibration*.

Nico is an upper-torso humanoid robot modeled after the fiftieth percentile 12-month-old male infant. The head comprises an active vision system with 6 mechanical degrees of freedom. The pitch angle of the eyes is mechanically coupled, while their yaw is independent of one another. The capability of this head to move its eyes independently of each other gives rise to the afforementioned changes in their epipolar geometry. These changes are analagous to changes in the human visual system under similar conditions. One notable difference between the human visual system and Nico's is that the human visual system exhibits *cyclovergence*, in which the eye rotates torsionally during its motions.

In computer vision, it is common to utilize the epipolar constraint in order to restrict the search for *stereo matches*, pairs of pixels in the stereo pair that image the same 3D point, to a region indicated by the epipolar lines. Using dynamic random element stereograms, Stevenson and Schor demonstrated that the human visual system does not restrict stereo matches to those regions indicated by epipolar lines[2]. Matches and accurate distance estimations can be made in a regions significantly wider than is indicated by the epipolar line, by up to 45 arcmin of visual angle. Schreiber et al used cyclorotated stereograms to demonstrate that scans for stereo matches occur over fixed regions of the retina. Instead of changing the regions of retina that are scanned for matches, the eye uses cyclovergence to assure that the epipolar lines lie within this scanned region[3]. Analagously, we can use the kinematics of an active vision system to update our estimate of the epipolar geometry as the system undergoes motion.

We present two algorithms. The first is an algorithm that builds on existing computer vision techniques to yield a kinematic model of the visual system in terms of its camera center, the orientation of the camera, and a joint behind each eye. The second allows us to compute the system's epipolar geometry after moving these joints.

In order to compute the kinematics of the visual system, we estimate the epipolar geometry of the cameras pointed in multiple orientations. Because the only joints in the head that affect the epipolar geometry are the yaw joints behind the eyes, we will concentrate on them. Our procedure is as follows. First, we point the eyes in an arbitrary direction, then, we pick an angle for each eye, and turn it to that angle along its yaw joint. We estimate the epipolar geometry between the two cameras in the first orientation, and, for each camera, between its two orientations. Looking at the case of the single camera, we now have two camera centers and two rotation matrices describing orientation. Using this data, and knowing that the camera faces directly away from its center of rotation, we can determine the location of the center of rotation, as this data fully describes an isoceles triangle from which we can compute all of the parameters of our kinematic system. Combining this with the estimate of epipolar geometry between the two cameras in their first view gives us sufficient information to fully describe the kinematic system in terms of the epipolar geometry. Computing the updated epipolar geometry as the system undergoes motion is a matter of adjusting the angle of rotation in two rotation matrices that describe the rotational component of each joint in this kinematic system. This angle can be retrieved from the robot's encoders. The remainder of the computation breaks down into several matrix multiplications that can be carried out in time linear in the number of joints modeled, assuming that we are modeling only the yaw joints in a stereo active vision head, this is constant time.

Joints that the camera does not face directly away from can be modeled by performing an analagous computation for three orientations. The three camera centers lie along a circle from which we can compute the kinematics of the system. The camera orientation with respect to this system can be computed using the rotation matrices from the estimate of epipolar geometry. Additionally, linkages observed in the visual field can be similarly modeled by tracking the motion of points along their arms as measured by the vision system. We can learn the kinematics of systems with multiple linkages by moving both sets of joints to uncover their mountings with respect to each other. Construction of a system that uses these techniques is in progress. Understanding this relationship brings us to the key insight of this work. Systems that perform hand/head-eye calibration compute the mounting of cameras with respect to a system of known kinematics. However, we are able to model these kinematics directly by realizing that for many tasks we only care about the kinematics insomuch as it yields changes that are directly witnessable in the vision system. For the estimation of epipolar geometry and camera orientations we only care about changes in artifacts of the vision system, such as the position of the camera center and orientation of the camera, which can be retrieved using existing computer vision techniques and used

in our kinematic computations.

The model presented in this paper constitutes a self-model of the imaging properties of the visual system, as well as its kinematics and layout in the head. Our first work in the area of self-modeling was in the form of self-other discriminiation [4], in which Nico learned to discriminate itself from other objects reflected in a mirror. The development of self-awareness and self-other discrimination is an active area of interest in developmental psychology and ethology. The process by which infants are able to distinguish the limits of their own bodies from the outside world is a rich area for computational modeling because empirical testing of hypotheses of this sort is extremely difficult [5]. Much like puppies, robots need to not chase their own mechanical tails. This work is an important step towards our overall dual goals of developing robots that model themselves and better understanding how humans do the same.

REFERENCES

[1] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, ISBN: 0521540518, 2004.

[2] S. S.B., "Human stereo matching is not restricted to epipolar lines," *Vision Research*, vol. 37, pp. 2717–2723(7), October 1997.

[3] K. Schreiber, D. J. Crawford, M. Fetter, and D. Tweed, "The motor side of depth vision," *Nature*, vol. 410, pp. 819–822, 4 2001.

[4] K. Gold and B. Scassellati, "A bayesian robot that distinguishes self from other," in *Proceedings of the 29th Annual Meeting of the Cognitive Science Society (CogSci2007)*, Nashville, Tennessee, 2007.

[5] P. Rochat and T. Striano, "Who's in the mirror? selfother discrimination in specular images by four- and nine-month-old infants," *Child Development*, vol. 73, pp. 35–46, January/February 2002.