Visually guided vergence in a new stereo camera system

Adam Schneider
Namrata Sharma
Bryan Tripp
University of Waterloo, ON, Canada

Abstract

People move their eyes several times each second, to selectively analyze visual information from specific locations. This is important, because analyzing the whole scene in foveal detail would require a beachball-sized brain and thousands of additional calories per day. As artificial vision becomes more sophisticated, it may face analogous constraints. Anticipating this, we previously developed a robotic head with biologically realistic oculomotor capabilities. Here we present a system for accurately orienting the cameras toward a three-dimensional point. The robot's cameras converge when looking at something nearby, so each camera should ideally centre the same visual feature. At the end of a saccade, we combine priors with cross-correlation of the images from each camera to iteratively fine-tune their alignment, and we use the orientations to set focus distance. This system allows the robot to accurately view a visual target with both eyes.

1 Introduction

We previously developed a robotic head [1, 2] with unique oculomotor capabilities (Figure 1). The main parameters (range of motion, stereo baseline, and saccade velocities) are all within human ranges. The robot uses liquid lenses to change focus distance within a few milliseconds, and we are in the process of developing foveated lenses to allow high acuity simultaneously with wide field of view. As far as we know, this is the only robotic head with oculomotor capabilities very similar to humans. Each of these capabilities is important to human vision. For example, saccades must be executed in about 100ms, because several saccades are performed per second, little vision occurs during the movement time, and visual inference after a saccade develops over 200ms.

Here we describe a system for accurately pointing both cameras to the same target. In contrast with standard stereo camera systems, humans orient both eyes toward the same target, resulting in convergence of up to five degrees or more. This allows detailed sensing and analysis of a selected visual feature, such as a word of printed text, in the fovea of each eye. Rapid saccades of both eyes in the same direction are followed by slower corrective periods in which vergence is fine-tuned [3, 4, 5, 6]. In this work we emulate this corrective phase, using the cross-correlation of images from both cameras, to precisely orient the cameras toward the same visual target.

When viewing objects at closer distances (such as objects within reach) , coordinated eye movements ensure binocular fusion and provide triangulation information essential for depth perception in visual scenes [7]. Maintaining minimal visual disparity in foveated points of fixation can facilitate fundamental visual functions such as depth estimation [8], and tracking of moving objects [9, 10].

A number of previous studies have investigated vergence control on robots [11, 12, 13, 14], and the cross-correlation has been frequently used [15, 16]. Several previous works used a log-polar mapping or Gaussian weighting of the correlation function to emphasize the central parts of the images [17, 18, 19]. In this study we apply this approach to a new robot. We also incorporate prior error probabilities, to achieve robust vergence to small foreground objects, and we use calculations from the camera angles to set the focus distance quickly and efficiently.

2 Methods

The goal of this work is to allow the robot to accurately look at a visual target with both cameras. This includes orienting the cameras so that the same feature is centred in each one, and also bringing the images into focus. The command that initiates this process is in the form of a three-dimensional target point. Selecting this target point is a task-specific process that is outside the current scope, but we assume that it corresponds to the estimated position of an object or feature of interest in the robot's surroundings.

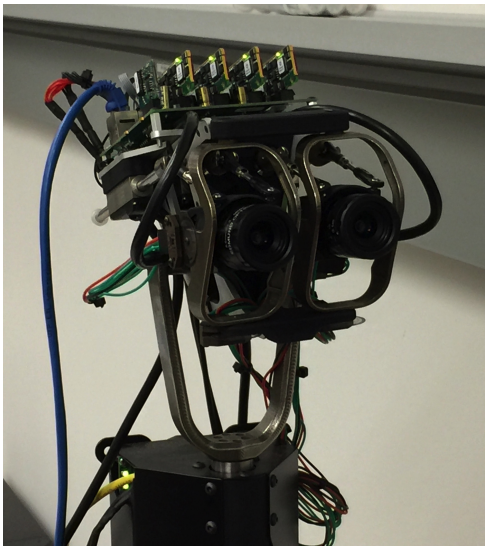

Fig. 1: This work develops a vergence controller for OREO (Open Robotic Eyes from Ontario), an open-hardware stereo camera system with seven degrees of freedom. Mechanically, OREO outperforms previous robotic heads [20, 21, 22], with saccade velocities and other oculomotor properties within human ranges.

To begin looking at a target, the required actuator positions are calculated, and set using individual proportional-integrate-derivative (PID) controllers for each actuator. There is error in this process due to non-ideal mechanical properties, such as slight misalignment of encoders and play in the joints. So the cameras do not point exactly where intended. Also, in general, the position of the target will be estimated imperfectly. As a result of these two factors, the two cameras generally do not have quite the same features centred after this actuator-level PID phase.

To orient both cameras more precisely toward the same visual feature, we introduce a subsequent corrective phase that uses visual feedback. To begin this phase, a region in the centre of each image is cross-correlated with the other image, and the peak of the correlation function is found. Because the images are taken from different perspectives, and only the central part of one of the images is used in each case, the results are not symmetric in general (see Figure 2). The offsets of the correlation peaks from the image centres are used to calculate the horizontal and vertical translations of each image that would be required to make them roughly overlap. These image translations (in pixels) are converted to angles (in degrees), and each camera is then moved by these amounts. This constitutes a single step of the correction process. Multiple steps can be taken to iteratively improve the alignment, or the method can run continuously to account for possible motion of the target object in depth. We experimented with standard correlation and also with correlation weighted by a spatial Gaussian function to emphasize features in the centre of each image.

If there are multiple objects close to the centre of the field of view, but at different depths, the strongest correlation does not necessarily correspond to the desired target. For example, smaller objects create smaller peaks in the correlation function. Also, objects that lack rotational symmetry may create a smaller peak when they are close to the robot, due to differences in perspective between cameras. To allow robust vergence toward small, nearby objects, we incorporated the prior expectation that the desired correlation peak is close to the centre. This was done by scaling the correlation function point-wise by a Gaussian function. This approach is closely related to Bayesian incorporation of a prior probability distribution over vergence errors, but we did not explicitly use a statistical formulation. This approach could potentially be refined in future work by collecting histograms of misalignments and correlations.

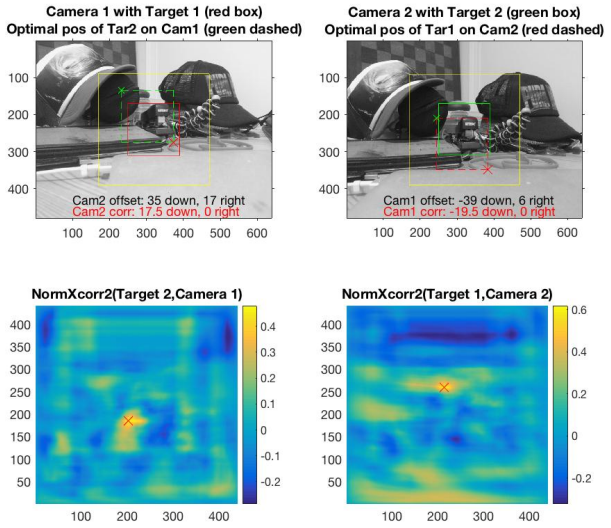

Fig. 2: Left Camera 1 image with centre Target 1 box in red, and right Camera 2 image with centre Target 2 box in green. The Target box from one camera image is moved around the other camera's image to find the shifted position in which they are best aligned, by using the normalized 2D cross-correlation function (plotted below) and choosing the shift corresponding to the peak value (red x). The optimal position of the Target 2 box on the Camera 1 image is shown by the dashed green box, and the optimal position of the Target 1 box on the Camera 2 image is shown by the dashed red box. The maximum shifts that are checked are indicated by the yellow boxes. The offset amounts are shown in black text. If the offsets from each perspective are in the opposite direction, then each camera should move by half that amount to meet in the middle. The amount each camera should be moved is in red text.

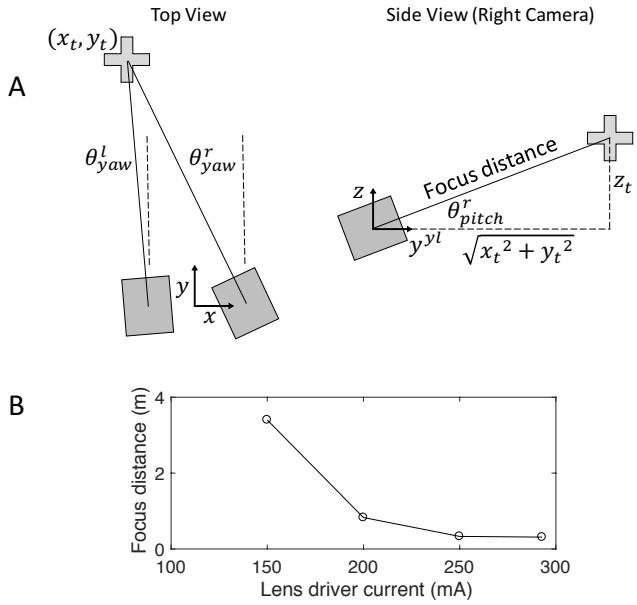

Fig. 3: A) After vergence, the distance to the target is estimated from each camera's yaw and pitch angles, which are measured by optical encoders. In the right diagram, the axis y^{yl} is the y-axis of the yaw link, which is rotated relative to the head by the yaw angle. B) The focus distance is controlled by applying current to an Optotune EL-10-30 electrically tunable lens. The plot shows focus distance vs. current for this lens in series with a manually tunable lens (Edmund Optics 58-000) that is focused at infinity. This plot shows means of three measurements per point. Optotune publishes nominal curves, but recommends manual calibration. Focus distance was measured only to about 3m.

The focus distance must also be adjusted to clearly image the target surface. OREO incorporates Optotune EL-10-30 electrically tunable lenses, which are capable of large changes in focus distance within a few milliseconds (10%-90% step in <2.5ms). There are a number of well-established algorithms for automatic focus correction [23]. A complication is that the blur is symmetric for focus distances that are too near and too far, so the sign of the required change can only be determined through exploratory changes. However, in our system the required focus distance is calculated from the camera angles. This allows us to find the appropriate focus distance in one step, using encoders on the camera joints. Figure 3 illustrates our approach. After vergence, the camera angles are used to calculate distance from each camera to the target surface, and this distance determines the appropriate input to the tunable lenses.

3 Results

Figure 4 shows stereo image pairs from an example vergence-correction sequence. Initially, after individual PID control of the actuators, the cameras are only roughly aligned (top image pair). The alignment improves with each correction step.

To improve robustness of vergence toward small foreground objects, we scaled the correlation functions point-wise with a Gaussian function centred at zero offset. This approach relied on fairly accurate initial saccades (in contrast with the inaccurate initial saccade in Figure 4), which required an additional calibration step with a visual target. Figure 5 shows examples of vergence to objects at different depths using this approach. The scene contains a small tool 0.5m from the robot and a spray can 1.5m from the robot. In the background there is a large map on the wall, 2.3m from the robot. Each panel shows overlaid images from both cameras. The left panels depict vergence toward the tool in the foreground, and the right panels depict vergence toward the spray can. Vergence to each target was achieved simply by saccading close to it, so that the corresponding peaks in the correlation functions were fairly close to centre. Without Gaussian scaling of the correlation functions, the robot verged toward the map in the background. This is because the map created a large peak in the correlation functions, as it covered many pixels and looked similar from the perspective of each camera.

4 Conclusion

This work allows the OREO robot to accurately orient to a visual target in three dimensions, including fine-tuning of vergence and focus distance. Incorporating prior probabilities allows reliable vergence to small foreground objects. Future work with the robot will focus on lower-level optimal control of coordinated head and eye movements, and higher-level task-related active vision.

References

[1] S. Huber, B. Selby, and B. P. Tripp, Design of a saccading and accommodating robot vision system, *Proc. CRV* (2016).

[2] S. Huber, B. Selby, and B. Tripp, OREO: An Open-Hardware Robotic Head That Supports Practical Saccades and Accommodation, *IEEE Robotics and Automation Letters, 3(3)*, pp. 2640-2645 (2018).

[3] J. Enright, Monocularly programmed human saccades during vergence changes? *The Journal of Physiology*, 512(1), 235-250 (1998).

[4] J. Samarawickrama, S. Sabatini, Version and vergence control of a stereo camera head by fitting the movement into the Hering's law. In *Fourth Canadian conference on computer and robot vision*. CRV (pp. 363-370). IEEE (2007).

[5] X. Zhang, A. L. T. Phuan, A physical system for binocular vision through saccade generation and vergence control. *Cybernetics and Systems: An International Journal*, 40(6), 549-568 (2009).

[6] W. Muhammad, M. Spratling, A neural model of binocular saccade planning and vergence control. *Adaptive Behavior*, 23(5), 265-282 (2015).

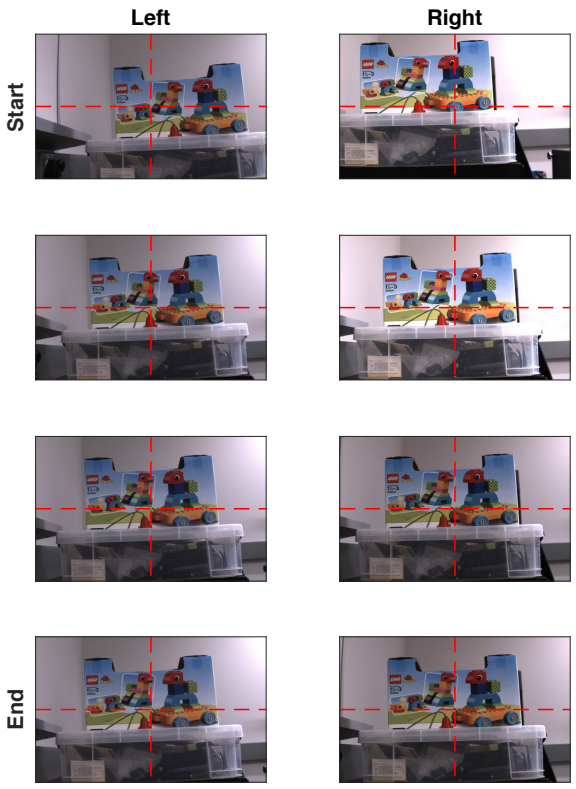

Fig. 4: A sequence of stereo images taken during vergence correction. Red dashed lines mark the centre of each figure. Initially, both cameras are approximately oriented toward a point 0.8m in front of the robot, via individual PID control of each actuator. However, as seen in the top pair of images, this is insufficient to point both cameras to the same target, due to non-ideal mechanical factors, and also because the target object is not exactly 0.8m from the cameras. In fact, a calibration step was skipped in this example, to make the misalignment artificially large for illustrative purposes. Each lower pair of images follows corrective steps based on visual feedback. After three steps (bottom) the left and right images are well aligned at the centre. The images do not align perfectly because they are taken from different perspectives.

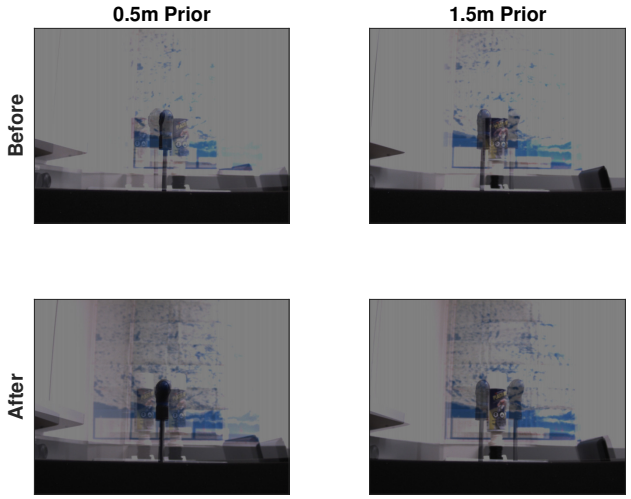

Fig. 5: Examples of vergence in which the correlation functions are scaled to account for prior probabilities, allowing robust orientation to small objects at different depths. On the left, the robot initially saccades to a point 0.5m away, close to the small tool in the foreground, and then verges accurately toward it. On the right, the robot saccades close to the spray can 1.5m away, and then verges accurately toward it. Without this scaling, the robot orients to the map in the background in each case.

[7] M. Hansen, G. Sommer, Active depth estimation with gaze and vergence control using Gabor filters. In *Proceedings of the 13th international conference on pattern recognition* (Vol. 1, pp.287-291) (1996). doi:10.1109/ICPR.1996.546035.

[8] Theimer, Wolfgang M., and Hanspeter A. Mallot. Phase-based binocular vergence control and depth reconstruction using active vision. *CVGIP: Image Understanding* 60.3: 343-358 (1994).

[9] D. Coombs, C. Brown, Real-time binocular smooth pursuit. *International Journal of Computer Vision*, 11(2), 147-164 (1993).

[10] A. Dankers, N. Barnes, A. Zelinsky, MAP ZDF segmentation and tracking using active stereo vision: Hand tracking case study. *Computer Vision and Image Understanding*, 108(1), 74-86 (2007).

[11] K. C. Kwon, Y. T. Lim, N. Kim, Y. J. Song, and Y. S. Choi, Vergence control of binocular stereoscopic camera using disparity information, *J. Opt. Soc. Korea*, vol. 13, no. 3, pp. 379-385 (2009).

[12] Y. Wang, and B. E. Shi. Improved binocular vergence control via a neural network that maximizes an internally defined reward. *IEEE Transactions on Autonomous Mental Development* 3.3: 247-256 (2011).

[13] F. Barranco, J. Diaz, A. Gibaldi, S. P. Sabatini, and E. Ros, Vector disparity sensor with vergence control for active vision systems, *Sensors*, vol. 12, no. 2, pp. 1771-1799 (2012).

[14] A. Gibaldi, M. Vanegas, A. Canessa, and S. P. Sabatini, A Portable Bio-Inspired Architecture for Efficient Robotic Vergence Control, *Int. J. Comput. Vis.*, vol. 121, no. 2, pp. 281-302, (2017).

[15] T. J. Olson, D. J. Coombs, Real-time vergence control for binocular robots. *International Journal of Computer Vision*, 7(1), 67-89 (1991).

[16] J. H. Piater, A. G. Roderic, and K. Ramamritham. Learning real-time stereo vergence control." *Proc. IEEE International Symposium on Intelligent Control/Intelligent Systems and Semiotics*. (1999).

[17] C. Capurro, P. Francesco, and G. Sandini, Dynamic vergence using log-polar images. *International Journal of Computer Vision* 24.1: 79-94 (1997).

[18] R. Manzotti, A. Gasteratos, G. Metta, and G. Sandini, Disparity estimation on log-polar images and vergence control, *Comput. Vis. Image Underst.*, vol. 83, no. 2, pp. 97-117 (2001).

[19] X. Zhang and L. P. Tay, A spatial variant approach for vergence control in complex scenes, *Image Vis. Comput.*, vol. 29, no. 1, pp. 64-77 (2011).

[20] H. Christensen, K. Bowyer, and H. Bunke, *Active Robot Vision: Camera Heads, Model Based Navigation and Reactive Control*. Singapore: World Scientific, 1993, vol. 6.

[21] R. Beira et al., Design of the robot-cub (iCub) head, *Proc. IEEE ICRA*, pp. 94-100 (2006).

[22] T. Asfour, K. Welke, P. Azad, A. Ude, and R. Dillmann, The Karlsruhe humanoid head, *Proc. 8th IEEE-RAS Int. Conf. Humanoid Robots*, pp. 447-453 (2008).

[23] F. C. A. Groen, I. T. Young, and G. Ligthart, A comparison of different focus functions for use in autofocus algorithms, *Cytometry* 6(2), pp. 81-91 (1985).