Short Papers

Active Vision During Coordinated Head/Eye Movements in a Humanoid Robot

Xutao Kuang, Mark Gibson, Bertram E. Shi, and Michele Rucci

***Abstract*—While looking at a point in the scene, humans continually perform smooth eye movements to compensate for involuntary head rotations. Since the optical nodal points of the eyes do not lie on the head rotation axes, this behavior yields useful 3-D information in the form of visual parallax. Here, we describe the replication of this behavior in a humanoid robot. We have developed a method for egocentric distance estimation based on the parallax that emerges during compensatory head/eye movements. This method was tested in a robotic platform equipped with an anthropomorphic neck and two binocular pan-tilt units specifically designed to reproduce the visual input signals experienced by humans. We show that this approach yields accurate and robust estimation of egocentric distance within the space nearby the agent. These results provide a further demonstration of how behavior facilitates the solution of complex perceptual problems.**

***Index Terms*—Active perception, behavior-based systems, biologically inspired robots, humanoid robots, 3-D vision.**

I. Introduction

Autonomous robots often need accurate estimation of the distance of objects and surfaces in the nearby space. Use of visual information would be ideally suited for this task. However, artificial 3-D vision systems remain highly sensitive to changes in their operating conditions, thus, seriously limiting the overall degree of robustness of the agent. Nature has faced a similar problem in the development of self-sufficient organisms. While stereopsis has become the predominant 3-D vision technique in robotics [1], [2], depth perception in humans and other species is an extremely robust process which relies on the analysis of multiple visual cues.

A highly informative cue used by many species is motion parallax, the displacement in the retinal position of the projection of an object as an agent moves through the environment [3]. In robotics, most of the work on this cue (a field known as depth from motion) has concentrated on large relocations of mobile platforms [4]–[12]. However, there is overwhelming evidence from biology that 3-D vision systems also benefit from the parallax resulting from more subtle movements, such as head and eye movements [13]–[16]. Before striking a prey, various types of insects, including the grasshopper and the mantis, perform peculiar peering head movements [17], which have been shown to contribute to perceptual judgments of distance. Similar exploratory head movements have also been observed in birds [18] and primates [19]. In humans, motion parallax resulting from head motion is used in various ways, including the control of body sway [20], and presentation on a 2-D display of stimuli that move by different amounts synchronously with the head induces a vivid 3-D percept [3].

Motion parallax resulting from head and eye movements is complementary to the stereo disparity cue usually considered by active binocular vision systems. By working simultaneously on each camera, this cue would increase the robustness of a binocular system, as it would continue to provide information even in the event of failure of one of the cameras. It would also enlarge the 3-D field of view, as it enables extraction of depth information in the regions that lack binocular overlap. Furthermore, it provides a controllable tradeoff between depth accuracy and computational complexity. In a stereo system, the disparity range over which to search for corresponding points is primarily determined by the distance between the two cameras (a fixed parameter, for a given robot), whereas with parallax, the search range varies with the amount of motion performed.

Here, we examine the 3-D information provided by motion parallax in an anthropomorphic robotic system that replicates the coordinated head/eye movements by which humans maintain normal fixation. Past work has considered only head or eye movements in isolation. For example, a recent study emulated the head movements of insects and birds for distance estimation [21]. Previous work from our group has shown that replication of human eye movements yields accurate estimation of egocentric distance within a range of nearby distances [22], [23]. However, no previous study has examined the impact of mutually compensatory head and eye movements similar to those continually performed by humans under natural viewing conditions. These movements maintain the attended objects at the center of the visual field, while providing reliable parallax.

The remainder of this paper is organized as follows. Section II describes a method for distance estimation based on the motion parallax caused by joint head and eye rotations. Section III examines its performance and robustness by means of computer simulations. Sections IV and V detail the results of robotic experiments. Section VI concludes this paper.

II. Distance Estimation Based on Head/Eye Motion Parallax

During fixation, humans use small head and eye movements to maintain the target within the fovea, the region on the retina with highest resolution. Since the focal nodal points of the eyes do not lie on the rotation axes of the head, this behavior yields a detectable parallax. Fig. 1 gives an example of this parallax. As the head rotates on the yaw axis (O) and the eyes compensate with pan rotations (axis C), the projection of the fixated object (F) remains relatively immobile at the center of the retina. However, every other object moves on the retina following a trajectory that depends on its position relative to the fixated point; objects closer and further than F will move in opposite directions. In the example in Fig. 1, object A projects onto the same retinal location as F at time t_1 and shifts to a different location during the movement.

In this study, we describe a method to estimate the position of an object in space based on its apparent motion during fixational head/eye coordination. We focus on yaw/pan rotations, as humans primarily rely on cues on the horizontal axis in 3-D judgments [24]. However, the method can be extended and applied to the parallax resulting from more complex coordinated rotations. Fig. 2 illustrates the approach

Manuscript received January 27, 2012; accepted June 1, 2012. This work was supported by the National Science Foundation under Grant CCF-0726901 and Grant BCS-1127216. This paper was recommended for publication by Associate Editor J. A. Castellanos and Editor D. Fox upon evaluation of the reviewers' comments.

X. Kuang, M. Gibson, and M. Rucci are with the Department of Psychology, Boston University, Boston, MA 02215 USA (e-mail: xtk@bu.edu; mggibson@bu.edu; mrucci@bu.edu).

B. E. Shi is with the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Kowloon, Hong Kong (e-mail: eebert@ee.ust.hk).

Digital Object Identifier 10.1109/TRO.2012.2204513

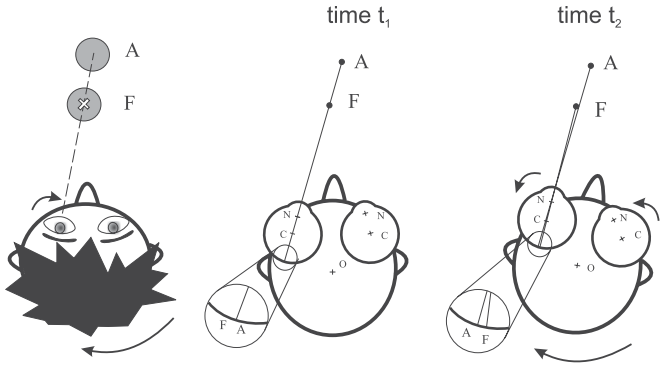

Fig. 1. Maintenance of fixation during head movement yields motion parallax. During normal fixation (left), small head movements are compensated by eye movements in order to maintain the fixation target F immobile on the retina. If the head rotates clockwise, the eyes need to rotate counterclockwise by a proportional amount in order to maintain fixation. Because of this coordinated motion, two objects that project onto the same retinal location at time t_1 (center) will project to different locations at a later time t_2 during fixation (right).

by considering, for simplicity, the case of an ideal point light source (PLS) that projects onto a single point in the image. The agent, in this case a model of the humanoid robot used in the experiments, maintains fixation on F, while the head rotates at a speed of $5^\circ/$s. Fig. 2(b) shows the projections of the objects in the images acquired by the camera at two different times separated by 1 s. At time t_1, both F and A lie on the horizontal line, which represents the locus of points composing line χ in Fig. 2(a), i.e., all the spatial locations that project onto the center of the image. Since many locations project onto the same pixel, estimation of distance based on a single image is obviously not feasible. However, during active fixation, when the eyes counterrotate to compensate for the movement of the head, objects at different distances translate on the sensor by different amounts. To illustrate this point, the dashed curve in Fig. 2(b) represents the new locations assumed at time t_2 (1 s later) by all the objects which initially projected at the center of the sensor. The displacement between the two images taken at different times t_1 and t_2, during head/eye movement, can be used to recover the distance of A.

Fig. 2(b) also shows a convenient way to quantify the distance information resulting from head/eye motion. As illustrated by the t_2 curve, the displacement of a target on the sensor converges asymptotically as the target's distance increases to infinity. That is, given two objects at a fixed separation, their relative displacement on the sensor will decrease the further away the objects are from the agent. Thus, the positions of distant objects cannot be well discriminated on the basis of fixational head/eye parallax. In the remainder of this paper, we will use the term target parallax (P_Δ) to indicate the displacement of a given PLS S_A, relative to a PLS S_∞ that projects onto the same pixel in the image acquired at t_1 but is located at infinite distance from the agent

$$P_\Delta = (x'_\infty - x_\infty) - (x'_A - x_A) = x'_\infty - x'_A \tag{1}$$

where x_A and x_∞ represent the projections of S_A and S_∞ at time t_1 ($x_A = x_\infty$), and x'_A and x'_∞, their projections at time t_2. The greater the $|P_\Delta|$, the more informative the movement.

The method for egocentric distance estimation described in this study relies on two images acquired at different times t_1 and t_2 during active fixation. Fig. 2(c) illustrates the geometry of the system. Let $\Delta\alpha$ indicate the rotation of the head in the interval (t_1, t_2), and $\Delta\beta$ be the corresponding rotation of the camera that is necessary to maintain the fixated point F at the center of the camera ($x_F = 0$). In the image acquired at time t_1, a PLS A located at distance d_A and eccentricity α_A will appear on the sensor at location

$$x_A = (CN_2 + d_C)\tan\phi. \tag{2}$$

Since ϕ is the angle between $\overline{AN_1}$ and $\overline{CF}$ [see Fig. 2(c)], we have

$$\tan\phi = \frac{d_A E(\alpha_A, \beta) + s\tan\beta}{d_A F(\alpha_A, \beta) - s - CN_1/\cos\beta} \tag{3}$$

where

$$E(\alpha_A, \beta) = \tan(\beta)\sin(\alpha_A) + \cos(\alpha_A) \tag{4}$$

$$F(\alpha_A, \beta) = \tan(\beta)\cos(\alpha_A) - \sin(\alpha_A) \tag{5}$$

and s represents the distance $\overline{CO}$ between the center of rotations of the camera and head. Thus, the projection of the PLS A in the image acquired at t_1 can be expressed as

$$x_A = f\left[\frac{d_A E(\alpha_A, \beta) + s\tan\beta}{d_A F(\alpha_A, \beta) - s - CN_1/\cos\beta}\right] \tag{6}$$

where f indicates the focal length of the lens $f = CN_2 + d_C$. Equation (6) gives a relationship between the distance of A, d_A, and its eccentricity α_A

$$d_A = \frac{x_A\,(CN_1/\cos\beta + s) + fs\tan\beta}{x_A F(\alpha_A, \beta) - fE(\alpha_A, \beta)}. \tag{7}$$

The infinite pairs (d_A, α_A) that satisfy (7) correspond to all the points composing the line $\overline{AN_1}$. All these points project onto the same location x_A in the image.

A similar locus of points can also be found using the image acquired at time t_2. In Fig. 2(c), a yaw rotation of the head is geometrically equivalent to a horizontal rotation of the scene around the agent by the same amount, but in the opposite direction. After this rotation, F no longer projects at the center of the sensor, and the camera needs to counterrotate by $\Delta\beta$ in order to maintain fixation on F. The geometrical relationships described previously still holds at time t_2, and the distance d_A can again be calculated as in (7) with the change of parameters $\alpha_A \to \alpha_A - \Delta\alpha$, $\beta \to \beta + \Delta\beta$, and $x_A \to x'_A$

$$d_A = \frac{x'_A\left[\frac{CN_1}{\cos(\beta+\Delta\beta)} + s\right] + fs\tan(\beta+\Delta\beta)}{x'_A F(\alpha_A - \Delta\alpha_A, \beta+\Delta\beta) - fE(\alpha_A - \Delta\alpha_A, \beta+\Delta\beta)}. \tag{8}$$

Equation (8) gives another ensemble of points in space (d_A, α_A), which project onto the same pixel x'_A in the image acquired at time t_2. When the motion produces nonzero target parallax ($P_\Delta \neq 0$), the new ensemble of points with projection x'_A will only share A in common with the previous ensemble. Thus, intersection of the two sets of points acquired at times t_1 and t_2 enables determination of A's position. By equating (7) and (8), we can explicitly calculate the eccentricity α_A

$$\alpha_A = \arctan\left[\frac{K_1(M_2\sin\Delta\alpha - N_2\cos\Delta\alpha) + K_2 N_1}{K_1(M_2\cos\Delta\alpha + N_2\sin\Delta\alpha) - K_2 M_1}\right] \tag{9}$$

where

$$\begin{aligned} K_1 &= x_A(s + CN_1/\cos\beta) + sf\tan\beta \\ M_1 &= x_A + f\tan\beta \\ N_1 &= f - x_A\tan\beta \end{aligned} \tag{10}$$

and

$$\begin{aligned} K_2 &= x'_A[s + CN_1/\cos(\beta+\Delta\beta)] + sf\tan(\beta+\Delta\beta) \\ M_2 &= x'_A + f\tan(\beta+\Delta\beta) \\ N_2 &= f - x'_A\tan(\beta+\Delta\beta). \end{aligned} \tag{11}$$

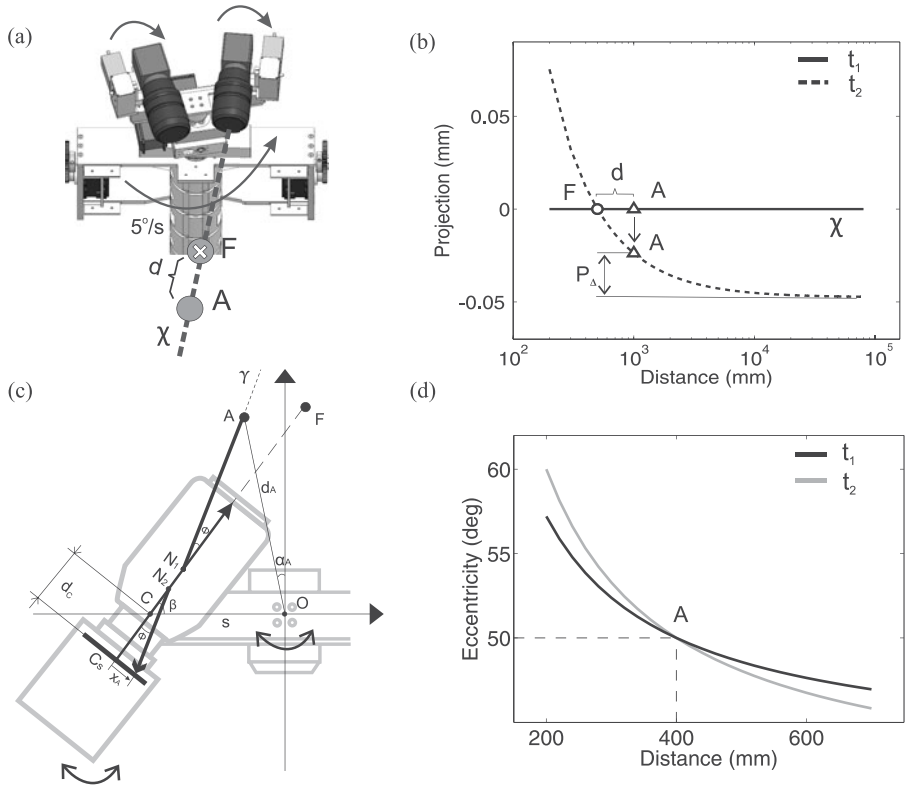

Fig. 2. Parallax emerging during head/eye fixational coordination. (a) Agent maintains fixation on the PLS F, while performing compensatory head and eye rotations on the yaw and pan axes. In this example, the head rotates at $5^\circ/\text{s}$. A second PLS A on the same projection line χ of F is also shown. (b) Projections of all the points on χ in two images acquired 1s apart. At t_2, F remains at the center of the image, while A translates by an amount which depends on its distance from the agent. P_Δ represents the target parallax. (c) Geometry of head/eye parallax. The camera is modeled as a two nodal point system, in which a ray of light that goes through nodal point N_1 with angle ϕ will exit from N_2 with the same angle. (d) Spatial locations of all the PLSs that project onto the same pixel of target A before (x_A at t_1) and after coordinated head/eye rotations (x'_A at t_2). The position of A in a head-centered frame of reference is uniquely determined by the intersection of the two curves.

The distance d_A can then be estimated by substituting α_A into either (7) or (8). An illustration of this method is given in Fig. 2(d), which plots the functions given by (7) and (8) before and after a rotation of the neck. The intersection of the two curves in the figure reveals the target's location.

This approach can be applied to any point in space, including the fixation point F. In this case, the target is at the center of the image both at t_1 and t_2 ($x_F = x'_F = 0$). From (9), we therefore obtain

$$\alpha_F = \arctan\left\{\frac{\tan(\beta+\Delta\beta) - E(-\Delta\alpha, \beta+\Delta\beta)\tan\beta}{F(-\Delta\alpha, \beta+\Delta\beta)\tan\beta - \tan(\beta+\Delta\beta)\tan\beta}\right\}. \tag{12}$$

The distance d_F can then be estimated from (7)

$$d_F = -\frac{\tan\beta}{(\cos\alpha_F + \tan\beta\sin\alpha_F)}s. \tag{13}$$

These equations confirm that knowledge of the rotation angles of the head and camera, together with the distance between their rotation axes s, is sufficient to localize the fixation point F.

III. Analysis of Localization Performance

To evaluate the robustness and accuracy of the method, we first examined its performance in numerical simulations of the APLab humanoid robot. Fig. 3 shows the estimated distance of a PLS as a function of its real distance for various head/eye rotations. In the simulation in Fig. 3(a), the neck rotated by $\Delta\alpha = 2^\circ$, while maintaining fixation on a point at distance $d_F = 500$ mm and eccentricity $\alpha_F = -10^\circ$. A second object (the target) was also located at eccentricity $\alpha_A = -10^\circ$, but its distance varied from 0.2 to 1 m. To model the impact of measurement errors, Gaussian noise with zero mean and standard deviation of 3 μm (approximately 0.7 pixel) was superimposed to the location of the PLS in the acquired image. It is clear from these data that the method yields accurate distance estimation within the space nearby the agent. The tendency to overestimate the distance of the target was caused by the combined effect of positional noise in the image and the highly nonlinear relationship between the location of the target on the image and its estimated distance. Fig. 3 shows that the accuracy of the estimate decreased with increasing distance of the object from the agent increased, and improved with increasing amplitude of neck rotation. These results are to be expected since the target parallax also decreases with the distance of the target and tends to increase with the amplitude of head/eye movements.

To fully explore performance during fixation at different locations, Fig. 4 shows the target parallax, while maintaining fixation on six different points in space. The point of fixation varied both in distance (either near at 200 mm or far at 800 mm) and eccentricity (left at 30°, center at 0°, or right at -30°). For each of these six positions, target parallax was estimated for objects at every possible location in the range 0–1 m from the robot and within $\pm 60^\circ$ around the fixation point. For each given eccentricity of the fixation point, comparison between the top and bottom rows in Fig. 4 shows that similar patterns of parallax were obtained with near and far fixations. Thus, the head/eye parallax was relatively little influenced by the actual distance of the fixation point. In contrast, the parallax map changed considerably with

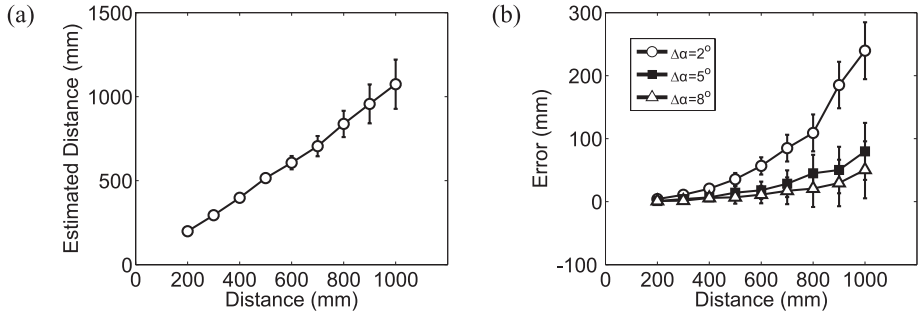

Fig. 3. Distance estimation in simulations of the APLab humanoid robot. A PLS target was located at -10° eccentricity and variable distance within the 200–1000-mm range. The fixation point was at distance 500 mm and -10° eccentricity. (a) Results obtained with neck rotation of 2°. (b) Errors in distance estimation obtained with neck rotations of 2°, 5°, and 8°. The parameters of the simulation modeled the robotic system used in the experiments. The focal length was $f = 19$ mm; the first nodal point was at 8 mm in front of the center of rotation of the eye; the distance between pan and yaw rotation axes was 76 mm. Data represent means over $N = 30$ repetitions for each measurement. Error bars represent standard deviations.

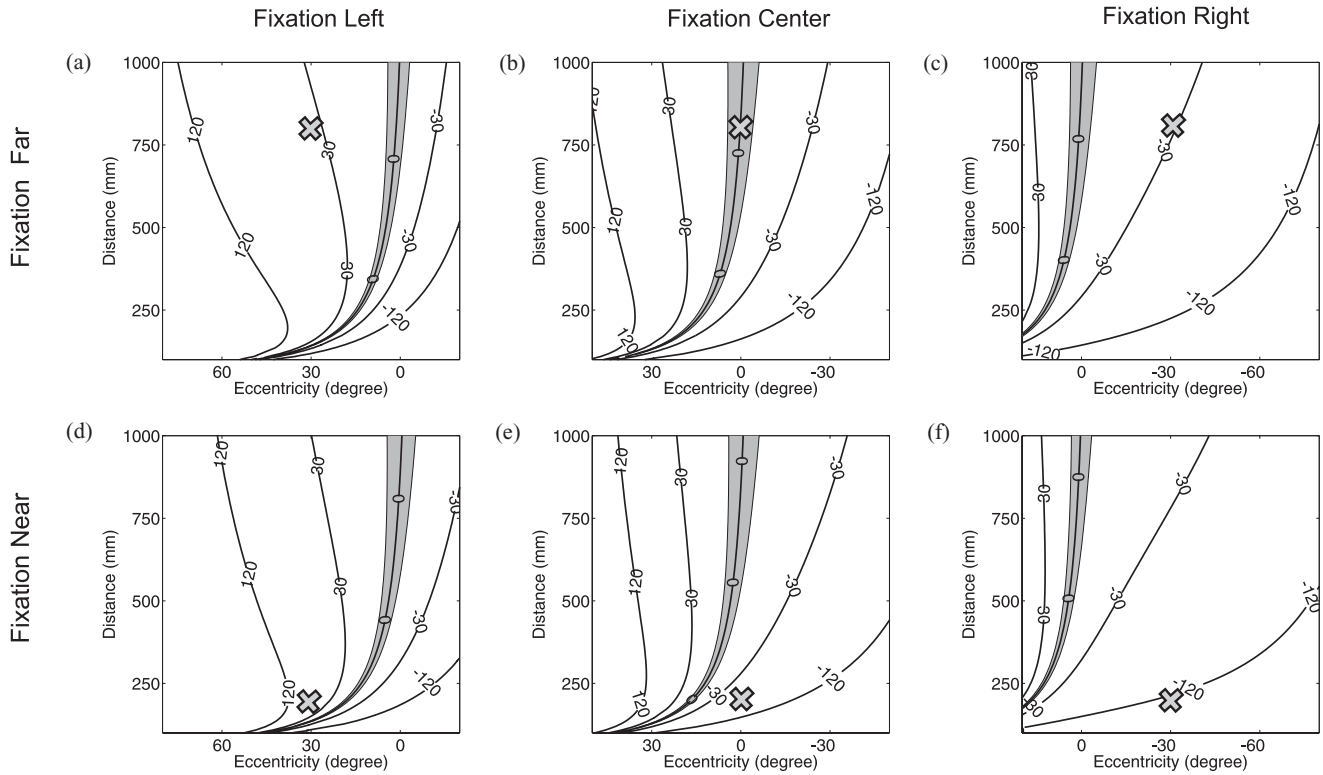

Fig. 4. Target parallax for different positions of the visual field during a 2° neck rotation. Each panel represents data obtained during fixation at a different location (the cross in the panel). The point of fixation was at a distance of either 800 mm (top row) or 200 mm (bottom row) and at eccentricity of 30° (left column), 0° (middle column) and -30° (right column). Lines represent isoparallax contours (μm). The shaded area marks the region with parallax smaller than 4.65 μm, the width of the pixel in the robot's camera.

the eccentricity of fixation. When the fixation point was not directly in front of the agent, the gradient of the parallax surface decreased significantly. For example, during fixation to the right by 30°, a change in parallax from -30 to -120 μm spanned a much larger visual angle compared with when the fixation point was at 0°. Thus, the use of fixational head/eye parallax is optimized with fixation in the frontal space. In all conditions, however, head/eye parallax gave discriminable values in almost all the examined regions of space. The gray regions in Fig. 4 mark the areas in which the target parallax was below 4.65 μm, the size of each pixel in the robot's camera. A target located within this region gave parallax too small to be detected without the development of subpixel hyper-resolution methods. This occurred when the nodal point N_1 lay at t_2 on the same line determined by the target and the position of N_1 at time t_1 [the line γ in Fig. 2(c)]. Fig. 4 shows that the region of undetectable parallax occupied a very narrow area of the space surrounding the robot so that accurate distance estimation could be achieved almost everywhere. Furthermore, the data in Fig. 4 also show that the location of the region of undetectable parallax was only mildly affected by the position of the fixation point. Thus, this region can be easily predicted in advance in applications of the method, as shown later in Section V.

IV. APLab Humanoid Robot

In order to examine the use of head/eye parallax in real-world applications, we conducted a series of experiments with the APLab

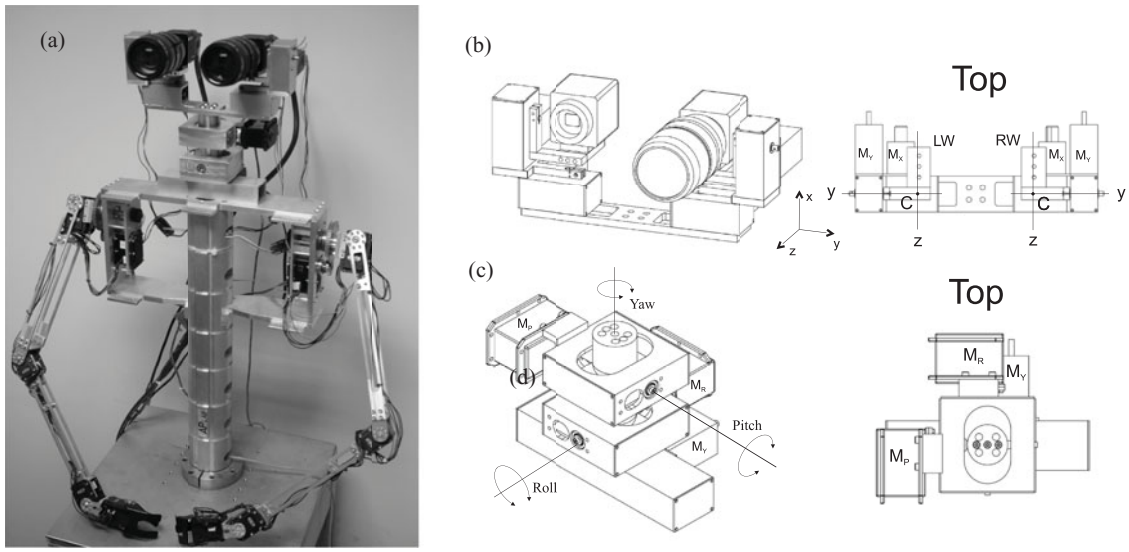

Fig. 5. APLab humanoid robot, a system developed to reproduce the visual input signals normally experienced by humans during behavior. (a) This robotic platform consists of a torso equipped with a head/eye system and two arms. (b) Pan-tilt motion of each camera was provided by two stepper motors M_x and M_y. The cameras were mounted on two aluminum wings (LW and RW) which allowed precise control of the parallax. (c) Anthropomorphic neck-enabled replication of head movements. Rotations were controlled by two servo motors (M_P and M_R for pitch and roll motion) and one stepper motor (M_Y for yaw motion).

humanoid robot. This robotic platform, which is shown in Fig. 5, was specifically designed to model the sensory inputs experienced by humans during behavior [25]. It consists of a torso equipped with a 3-degree-of-freedom (DOF) neck, two pan-tilt cameras, and two 5-DOF arms. As shown in Fig. 5(c), motion in the neck was provided by two servo motors (Dynamixel RX-64, Robotis, Fullerton, CA) and one stepper motor (HT11-012D, Applied Motion, Watsonville, CA), with each motor responsible for controlling rotations around one axis (yaw, pitch, and roll). Although this system does not allow replication of head translations, it enables camera trajectories very similar to those followed by human eyes during normal fixation. Cameras were moved by two pairs of stepper motors (HT11-012D, Applied Motion, Watsonville, CA) controlled by proprietary microprocessors. This system was designed to yield position accuracy of $1.8'$ and maximum velocity of $240^\circ/s$. These specifications enable reasonable replication of human eye movements [22]. All motors were controlled via a multichannel servo and stepper motor controller (PCI-7350, National Instruments, Austin, TX).

The vision system of the APLab robot was specifically designed to replicate the head/eye parallax present in the human eye. Following Gullstrand's eye model [26], the unaccommodated eye can be modeled as a sphere with radius $d_C = 11$ mm, which rotates around its center of rotation C. The image on the retina is formed by a lens with two nodal points N_1 and N_2 located at distances 6.32 and 6.07 mm from the center of rotation C, respectively. The focal length is given by $f = d_C + CN_2 = 17$ mm. To recreate a parallax similar to that experienced by humans, the key parameters are the focal length f and the distance CN_1, i.e., the distance of the first nodal point from the center of rotation. Tuning of these parameters was achieved by means of a zoom lens with adjustable focal length and an aluminum wing that allowed positioning of the camera's sensor at a desired distance from C. In the experiments, f was 18.9 mm and CN_1 was 8.2 mm. These values well approximated the optical properties of the human eye. The rotation center C was at 76.2 mm from the neck yaw axis. Images were acquired at a rate of 30 frames/s (PCI-1428, National Instruments) from two high-resolution (1392×1040) monochrome cameras (AccuPiXEL TM-1402CL, Pulnix Inc., San Jose, CA). Pixels in these cameras were almost as small as the cones in the human retina

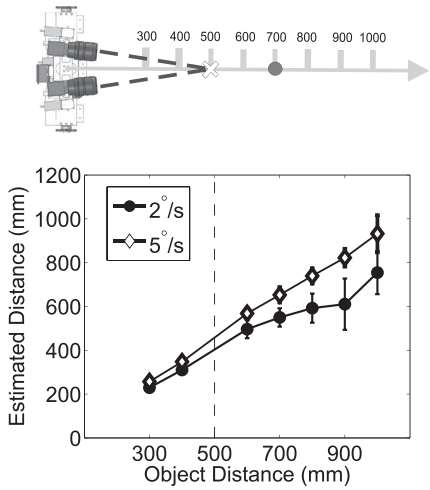

Fig. 6. Distance estimation for an isolated target. The robot maintained fixation on an LED at 500 mm and estimated the position of a second LED at variable distance. Both LEDs were mounted on a calibrated aluminum bar placed in front of the robot. Distance estimation was conducted for two speeds of neck rotation: 2 and $5^\circ/s$. Data points represent means $\pm$ standard deviations over 30 trials. Each measurement is based on two images acquired after an interval of 1 s.

(4.65 μm). Only the images acquired by the left camera were used in the experiments.

V. Robotic Experiments

To measure the accuracy of the method, we first estimated the distance of an isolated small target, like in the simulations in Fig. 3. In this experiment, the robot maintained fixation on an LED, while performing compensatory head/eye rotations. Fixation was achieved in real time by continuously centering the projection of the fixation point in the acquired images. To this end, the camera continuously rotated with angular velocity proportional to the offset of the centroid of the fixation point from the center of the image. Pairs of images acquired at intervals of 1 s were used to estimate the distance of a second LED (the target).

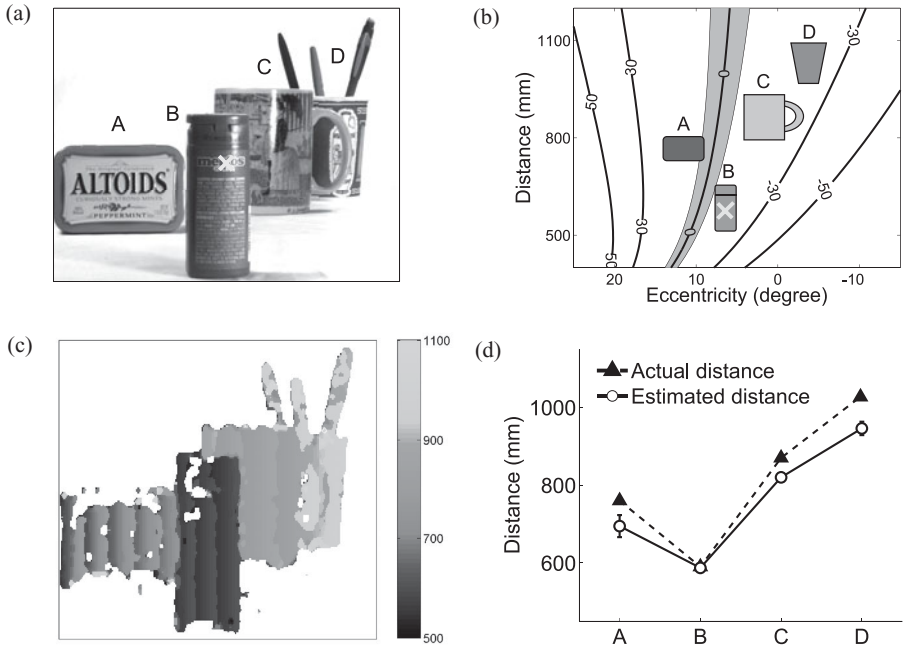

Fig. 7. Image segmentation based on head/eye parallax with a single fixation. (a) Scene with four objects was examined, while maintaining fixation (the marker) during $5^\circ/s$ head rotation. (b) Object positions relative to isoparallax contours. With the chosen fixation point, no object fell within the zero-parallax region. (c) Map of egocentric distance estimation on the basis of head/eye parallax. The periodicity in the pattern originated from pixel quantization. (d) Accuracy of distance estimation for the four individual objects. Data points represent the means $\pm$ standard deviations over all measurements on each object. The estimated distance is compared with the actual distance of the object from the robot.

Fig. 6 shows results obtained when fixation was maintained at 500 mm and -10° eccentricity. The two curves in each panel refer to data acquired when the head rotated at two different speeds: 2 and $5^\circ/s$. The target was also located at eccentricity -10°, and its distance varied from a minimum of 30 cm to a maximum of 1 m. Following the method described in Section II, we recorded the camera angles at time t_1 and t_2 (β and β') and computed the position of the target's centroid in the images acquired at these two times (x_A and x'_A). The data in Fig. 6 are similar to those of the simulations in Fig. 3. Accurate localization was achieved for neck rotations of just $2^\circ/s$, a speed comparable with the range measured in humans [27]. As expected, the accuracy of estimation improved with the extent of the rotations and the proximity of the target, as the target parallax also increased. With head rotation of 5°, the localization error for an object located at 1000 mm was only 5%. The underestimation of the target distance with a $2^\circ/s$ rotation was caused by imprecision in maintaining fixation during head rotation.

Having assessed the accuracy of the method with an isolated object, in successive experiments, we investigated whether the 3-D information provided by head/eye parallax could facilitate segmentation of complex scenes. Fig. 7(a) shows a scene presented to the robot, in which four objects partially occluded each other. Similarities in the luminance and texture of different objects made segmentation of individual images into the constituent objects extremely difficult. Scenes of this kind often occur in the real world, and depth information could greatly improve their segmentation. Fixation was maintained on object B, while the neck rotated at $5^\circ/s$. As shown in Fig. 7(b), all four objects fell in regions with sufficiently large parallax, in which estimation of their distances is possible.

During viewing of a single object, like in the experiment in Fig. 6, estimation of the translation of the object in the two images is straightforward. In a complex scene, however, the translation of each individual element in the scene needs to be determined, a task similar to the correspondence problem faced by stereo vision algorithms. To this end, the image acquired at t_1 was fragmented into a tessellation of squared patches, each 20 pixels wide. Normalized cross correlation was then used to find the position of each patch in the image acquired at time t_2. This led to the estimation of egocentric distance at 279×208 equispaced locations in the scene. Fig. 7(c) shows that the estimated distance map gives an accurate segmentation of the scene into the regions corresponding to the four objects. The white regions in the map correspond to areas in which the algorithm failed to identify a correspondence for the considered patch (i.e., the correlation coefficient fell below a threshold of 0.05). This occurred almost exclusively in the uniform luminance regions of the background, thus, effectively enabling figure-ground segregation, another vexing computer vision problem. Fig. 7(d) shows that mean distance estimated over all the patches composing each object was very close to the real distance of the object from the robot.

The results in Fig. 7 were obtained while the robot maintained precise fixation on a single point in the scene. However, when observing an object, humans often make small saccades [28], [29]. Fig. 8 summarizes results obtained by replicating this behavior. The eye movements of a human observer were recorded by means of a high-resolution eyetracker (Fourward Technologies, Gallatin, MO) during examination of the same scene presented to the robot. The recorded trajectory was then used to control the position of the robot cameras. Fig. 8(a) shows an example of recorded eye movements overlapped to the image viewed by the observer. This trace contained 14 small saccades. Fixational head/eye coordination was introduced during each of the 15 intersaccadic periods of fixation while the neck rotated at $5^\circ/s$.

Unlike the previous experiment, fixation was here located on object C. Based on the robot model (see Fig. 4), it can be predicted that this location of fixation will cause object A to fall within the region of undetectable parallax. Indeed, the mean target parallax measured for

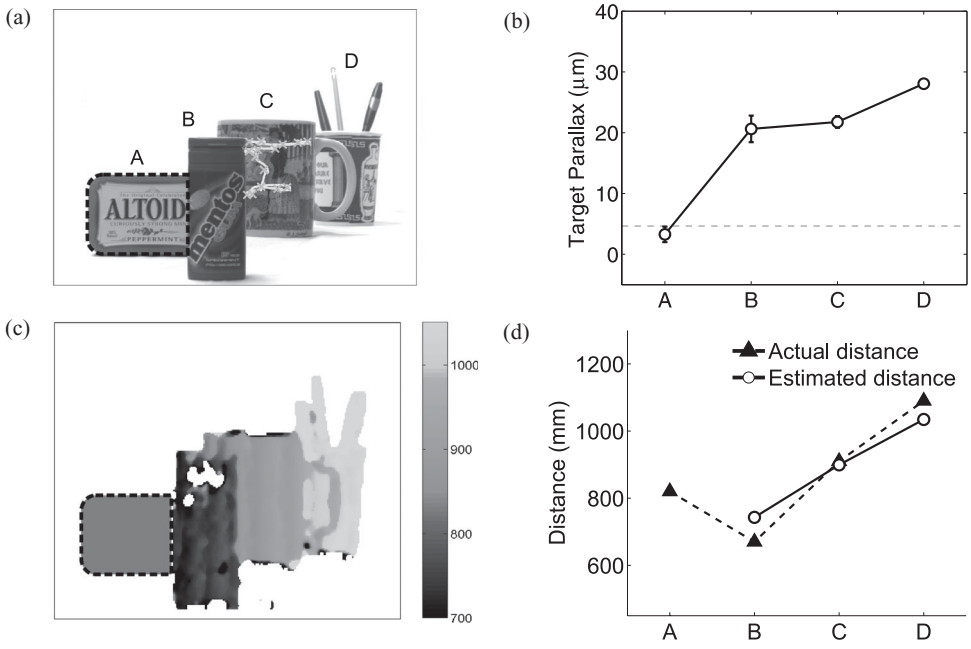

Fig. 8. Image segmentation based on head/eye parallax with multiple fixations. (a) Recorded trace of eye movements during fixation on object C. The head rotated at $5^\circ/s$ during each intersaccadic period (the crosses on the trace). (b) Mean target parallax over all fixation periods for the four objects in the scene. The dashed line marks the size of the pixel in the camera. (c) Map of egocentric distance. Data represent averages over 15 fixation periods. The gray region with dashed outline corresponds to object A, for which distance estimation was not performed. (d) Estimated distance of each object. Each data point represents the average value over all patches belonging to an individual object. Standard deviations were too small to be shown.

object A was smaller than the quantization of the sensor [see Fig. 8(b)]. Thus, this region was not further considered in the analysis.

Fig. 8(c) shows the segmentation map obtained in the remaining portion of the scene. Each point in the map represents the average distance estimated over all the periods of intersaccadic fixation. Integration of information acquired across multiple fixations improved segmentation by yielding a more accurate and smoother map of distance. The regions corresponding to the three objects are clearly distinct in the resulting map. Furthermore, the method gave very accurate estimation of each object's distance [see Fig. 8(d)]. Thus, replication of the head/eye fixation strategy adopted by humans yielded accurate segmentation of the scene.

VI. Conclusion

The results of this study show that small coordinated head/eye movements similar to those performed by humans during fixation yield parallax sufficient to accurately localize objects in the space nearby the agent. With neck rotations of approximately 2°, localization was possible for targets located up to 2 m from the robot. Many important operations, such as reaching and manipulation, occur within this range of distances. Head/eye parallax can thus contribute to the establishment of 3-D representations and the guidance of behavior in robots, as it does in many biological species.

The head/eye coordination investigated in this study is inspired by the sensorimotor strategy humans follow to maintain fixation on a target. Similar strategies have been observed in many species. Owls, for example, are capable of determining the relative depth of surfaces solely on the basis of the motion parallax resulting from side-to-side head movements [15]. When walking, pigeons and other birds exhibit a rhythmic forward–backward head motion, which provides useful parallax [18]. Many insects perform peculiar peering head movements before jumping to a target [16]. In humans, head movements contribute to depth perception [3], as well as the small involuntary movements that occur during fixation yield parallax well above detection thresholds [30]. While in machine vision depth from motion is typically studied during navigation, our results show that more subtle camera movements generate highly useful parallax as well.

This study builds upon our previous work on replicating human eye movements in a robot [22], [23]. These previous studies demonstrated that in an anthropomorphic pan/tilt unit, even pure camera rotations yield resolvable parallax because of the distance between optical nodal points and rotation axes. Here, we have shown that when fixational head movements are also present, the resulting 3-D information becomes highly reliable, enabling accurate estimation of egocentric distance and image segmentation. It is important to note that this 3-D cue is complementary to stereo disparity. It can be used independently on each camera of a stereo system, improving robustness and enabling 3-D vision in a broader field than the binocular region. Furthermore, unlike stereo vision, this approach allows closed-loop recursive refinement of 3-D information by controlling the delay in the acquisition of pairs of images to be compared.

Head/eye parallax provides an example of the perceptual consequences of behavior. Although pioneering studies have long observed that active control of input sensory signals simplifies many perceptual tasks [31]–[34], computer vision algorithms are often designed without taking advantage of the agent's behavior. This approach contrasts with the way perception functions in biological systems. Organisms are not passively exposed to the incoming flow of sensory data. Instead, they actively seek useful information by coordinating sensory processing with motor activity. In humans, behavior is a necessary ingredient for the working of the visual system, and is always present, even when it is not immediately obvious. Visual perception ceases to function, and the scene fades away, when behavior is completely eliminated [35].

Even when head and body movements are prevented, microscopic eye movements continuously occur. These movements do not simply refresh neural responses; they are a critical component of the way visual information is acquired and encoded [36]–[38].

While our study gives a clear demonstration of the value of head/eye parallax, the specific algorithm that we used to estimate distance could be further improved. The approach described in this study relies on two images acquired at different times. This method requires the solution of a correspondence problem similar to the one faced by stereopsis. An alternative approach could use the optic flow present in sequences of images continuously acquired during head/eye coordination. This approach would present two main advantages over our algorithm. First, it would enable adoption of standard methods for the estimation of optic flow. Second, it would eliminate the problem of the small parallax region discussed in Section III. The existence of this region is a direct consequence of the use of only two images. Further work is needed to investigate the application of optic flow techniques to depth from motion during head/eye coordination.

REFERENCES

[1] N. Ayache, *Artificial Vision for Mobile Robots: Stereo Vision and Multisensory Perception*. Cambridge, MA: MIT Press, 1991.

[2] O. Faugeras, *The Geometry of Multiple Images: The Laws that Govern the Formation of Multiple Images of a Scene and Some of Their Applications*. Cambridge, MA: MIT Press, 2001.

[3] B. Rogers and M. Graham, "Motion parallax as an independent cue for depth perception," *Perception*, vol. 8, pp. 125–134, 1979.

[4] G. Sandini and M. Tistarelli, "Active tracking strategy for monocular depth inference over multiple frames," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 1, pp. 13–27, Jan. 1990.

[5] J. Weng, T. S. Huang, and N. Ahuja, *Motion and Structure From Image Sequences*. New York: Springer-Verlag, 1993.

[6] Y. Aloimonos and Z. Duric, "Estimating the heading direction using normal flow," *Int. J. Comput. Vis.*, vol. 13, pp. 33–56, 1994.

[7] A. K. Dalmia and M. Trivedi, "High-speed extraction of 3D structure of selectable quality using a translating camera," *Comput. Vis. Image Understanding*, vol. 64, no. 1, pp. 97–110, 1996.

[8] Y. S. Hung and H. T. Ho, "A Kalman filter approach to direct depth estimation incorporating surface structure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 6, pp. 570–575, Jun. 1999.

[9] A. J. Davison and D. W. Murray, "Simultaneous localization and map-building using active vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 865–880, Jul. 2002.

[10] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, "MonoSLAM: Real-time single camera SLAM," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 1052–1067, Jun. 2007.

[11] G. Zhang, J. Jia, T.-T. Wong, and H. Bao, "Consistent depth maps recovery from a video sequence," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 6, pp. 974–988, Jun. 2009.

[12] M. Ramachandran, A. Veeraraghavan, and R. Chellappa, "A fast bilinear structure from motion algorithm using a video sequence and inertial sensors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 1, pp. 186–193, Jan. 2011.

[13] K. Kral and M. Poteser, "Motion parallax as a source of distance information in locusts and mantids," *J. Insect Behav.*, vol. 10, no. 1, pp. 145–163, 1997.

[14] L. W. Casperson, "Head movements and vision in underwater-feeding birds of stream, lake and seashore," *Bird Behav.*, vol. 13, pp. 31–46, 1999.

[15] R. F. van der Willigen, B. J. Frost, and H. Wagner, "Depth generalization from stereo to motion parallax in owl," *J. Comp. Physiol. A*, vol. 187, pp. 997–1007, 2002.

[16] K. Kral, "Behavioral-analytical studies of the role of head movements in depth perception in insects, birds and mammals," *Behav. Process.*, vol. 64, pp. 1–12, 2003.

[17] M. V. Srinivasan, "Distance perception in insects," *Curr. Dir. Psychol. Sci.*, vol. 1, no. 1, pp. 22–26, 1992.

[18] R. Necker, "Head-bobbing of walking birds," *J. Comp. Physiol. A*, vol. 193, pp. 1177–1183, 2007.

[19] E. W. Sargent and G. D. Paige, "The primate vestibulo-ocular reflex during combined linear and angular head motion," *Exp. Brain Res.*, vol. 87, pp. 75–84, 1991.

[20] M. Guerraz, V. Sakellari, P. Burchill, and A. M. Bronstein, "Influence of motion parallax in the control of spontaneous body sway," *Exp. Brain Res.*, vol. 131, no. 2, pp. 244–252, 2000.

[21] A. Bruckstein, R. J. Holt, I. Katsman, and E. Rivlin, "Head movements for depth perception: Praying mantis versus pigeon," *Auton. Robot.*, vol. 18, pp. 21–42, 2005.

[22] F. Santini and M. Rucci, "Active estimation of distance in a robotic system that replicates human eye movements," *Robot. Auton. Syst.*, vol. 55, no. 2, pp. 107–121, 2007.

[23] F. Santini, R. Nambisan, and M. Rucci, "Active 3D vision through gaze relocation in a humanoid robot," *Int. J. Humanoid Robot.*, vol. 6, no. 3, pp. 481–503, 2009.

[24] B. G. Cumming, "An unexpected specialization for horizontal disparity in primate primary visual cortex," *Nature*, vol. 418, pp. 633–636, 2002.

[25] M. Rucci, D. Bullock, and F. Santini, "Integrating robotics and neuroscience: Brains for robots, bodies for brains," *Robot. Auton. Syst.*, vol. 21, no. 10, pp. 1115–1129, 2007.

[26] A. Gullstrand, *Appendices to Part 1: The Optical System of the Eye*. Hamburg, Germany: Voss, 1909, pp. 350–358.

[27] B. T. Crane and J. L. Demer, "Human gaze stabilization during natural activities: Translation, rotation, magnification, and target distance effects," *J. Neurophysiol.*, vol. 78, no. 4, pp. 2129–2144, 1997.

[28] R. M. Steinman, G. M. Haddad, A. A. Skavenski, and D. Wyman, "Miniature eye movement," *Science*, vol. 181, pp. 810–819, 1973.

[29] H. K. Ko, M. Poletti, and M. Rucci, "Microsaccades precisely relocate gaze in a high visual acuity task," *Nature Neurosci.*, vol. 13, pp. 1549–1553, 2010.

[30] M. Aytekin and M. Rucci, "Motion parallax from microscopic head movements during visual fixation," *Vision Res.*, in press.

[31] R. A. Jarvis, "A perspective on range-finding techniques for computer vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-5, no. 2, pp. 122–139, Mar. 1983.

[32] Y. Aloimonos, I. Weiss, and A. Bandyopadhyay, "Active vision," *Int. J. Comput. Vis.*, vol. 2, pp. 333–356, 1988.

[33] R. Bajcsy, "Active perception," *Proc. IEEE*, vol. 76, no. 8, pp. 996–1005, Aug. 1988.

[34] D. Ballard, "Animate vision," *Artif. Intell.*, vol. 48, pp. 57–86, 1991.

[35] R. W. Ditchburn and B. L. Ginsborg, "Vision with a stabilized retinal image," *Nature*, vol. 170, pp. 36–37, 1952.

[36] M. Rucci, R. Iovin, M. Poletti, and F. Santini, "Miniature eye movements enhance fine spatial detail," *Nature*, vol. 447, pp. 852–855, 2007.

[37] M. Rucci, "Fixational eye movements, natural image statistics, and fine spatial vision," *Network*, vol. 19, no. 4, pp. 253–285, 2008.

[38] X. Kuang, M. Poletti, J. D. Victor, and M. Rucci, "Temporal encoding of spatial information during active visual fixation," *Curr. Biol.*, vol. 22, no. 6, pp. 510–514, 2012.