

# Active estimation of distance in a robotic system that replicates human eye movement

Fabrizio Santini\*, Michele Rucci

*Boston University, Department of Cognitive and Neural Systems, Boston, MA 02215, USA*

Received 1 April 2005; received in revised form 6 July 2006; accepted 6 July 2006

Available online 28 August 2006

## Abstract

In a moving agent, the different apparent motion of objects located at various distances provides an important source of depth information. While motion parallax is evident for large translations of the agent, a small parallax also occurs in most head/eye systems during rotations of the cameras. A similar parallax is also present in the human eye, so that a redirection of gaze shifts the projection of an object on the retina by an amount that depends not only on the amplitude of the rotation, but also on the distance of the object with respect to the observer. This study examines the accuracy of distance estimation on the basis of the parallax produced by camera rotations. Sequences of human eye movements were used to control the motion of a pan/tilt system specifically designed to reproduce the oculomotor parallax present in the human eye. We show that the oculomotor strategies by which humans scan visual scenes produce parallaxes that provide accurate estimation of distance. This information simplifies challenging visual tasks such as image segmentation and figure/ground segregation.

© 2006 Elsevier B.V. All rights reserved.

*Keywords:* Biomorphic robotics; 3-D Vision; Saccade; Stereopsis; Head/eye system; Depth from motion

## 1. Introduction

A critical function of a visual system is to estimate the distance of objects and surfaces with respect to the agent. Accurate estimation of distance is crucial for most operations involving interactions with the environment, as in the motor tasks of navigation and manipulation. Depth information also provides a fundamental contribution to the segmentation of images into their basic elements, one of the most serious challenges faced by machine vision systems. Since the process of reconstructing a 3-D scene from its projections on a 2-D sensor is inherently ambiguous, vision systems usually rely on cues that originate from the comparison of images taken either from different points of view (stereoscopic vision) [1–6] or in different instants of time (depth from motion) [7,2,3,8], as well as from a priori knowledge of the scene and its structure (depth from shading, size, occlusions, etc.) [9–12].

Not surprisingly, in nature, where greater accuracy of depth perception can mean the difference between capturing a prey

or failing to survive, many species exhibit a striking precision in estimating distance by means of vision [13]. In humans, depth perception relies on the integration of a variety of cues, each of them measured with exquisite accuracy. For example, stereoacuity, which is at the basis of the perception of depth from the slightly different images in the two eyes, reaches thresholds as low as 3" [14]. That is, the human brain compares features in both eyes with a precision of about one tenth of a photoreceptor.

As is the case with many other visual functions, the close interaction between motor and sensory processes is a critical feature of the way biological vision systems evaluate distance. Behavior contributes to depth perception in at least two ways. First, motor activity is known to facilitate the computations involved in 3-D vision. Several studies have shown that oculomotor activity simplifies the extraction and use of both binocular [15–17] and monocular depth cues [18,19]. Secondly, behavior directly contributes distance information by generating depth cues in different sensory modalities. For example, vergence eye movements and focus adjustment produce depth cues in the proprioceptive modality [20]. Similarly, as an agent navigates through the environment, cues

\* Corresponding author.

*E-mail addresses:* [santini@cns.bu.edu](mailto:santini@cns.bu.edu) (F. Santini), [rucci@cns.bu.edu](mailto:rucci@cns.bu.edu) (M. Rucci).

of distance in the visual modality emerge in the form of motion parallax, *i.e.* the apparent motion of stationary objects [21,22]. In humans, these cues generate a vivid depth percept as if the observer were looking with two eyes using binocular stereopsis.

In species with mobile eyes, an interesting contribution of behavior to depth perception is the oculomotor parallax generated by eye movements. Since in the eyes of most species the optical nodal points are not coincident with the center of rotation, during a relocation of gaze, the projection of an object on the retina moves by an amount that depends both on the amplitude of the rotation and on the distance of the object from the observer (see Fig. 1). Species as diverse as the chameleon and the sandlance, for which the optics of the cornea and lens maximize the distance between nodal points and the center of rotation, rely on this cue to judge distance [23,24]. A similar parallax is also present in the eyes of primates, and during the normal scanning of a visual scene it produces retinal shifts that are well within the range of human visual acuity [25–27]. Under natural viewing conditions, humans tend to shift their direction of gaze by small steps [28]. These small saccades may facilitate the analysis of the oculomotor parallax by producing little distortions in the retinal images.

This study investigates the use of the oculomotor parallax for the visual estimation of distance in robotics. In machine vision, use of a parallax for depth estimation is not novel, a tight link between motor and sensory processes has been identified as a key element for the development of more efficient algorithms [29–35]. Most studies have examined the parallax that emerges during large movements of a camera such as those produced by a vision system mounted on a mobile platform [36,37,7,38–40]. These conditions amplify the motion parallax. No previous study, however, has specifically focused on the parallax produced by rotating the camera of a stationary head/eye system. Camera rotations produce a parallax similar to that of the eye if the nodal points of the optical system do not lie on the axes of rotations. Such a misalignment occurs unless it is intentionally eliminated by careful specification of the system's optical and mechanical characteristics. Thus, an oculomotor parallax is present in virtually every pan/tilt unit used in robotics.

Since the 3-D information provided by this cue depends on the precise movements of the sensor that produced it, in this paper we focus on the oculomotor behavior of human subjects. In this study, we replicate the strategy by which humans scan complex visual scenes by using a robotic system specifically designed to reproduce the parallax present in the human eye. We show that the oculomotor parallax that emerges during the small relocations of gaze that characterize human oculomotor activity under natural viewing conditions provides reliable depth information within a range of nearby distances. This information offers a robust contribution to the processes of image segmentation and figure/ground segregation.

This paper is organized as follows. First, we analyze the information provided by the oculomotor parallax and describe an active approach to the estimation of distance based on rotations of the sensor. Both the cases of the human eye (spherical sensor) and a camera (planar sensor) are analyzed

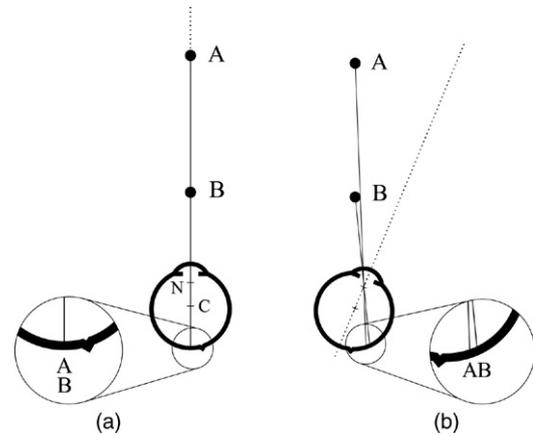


Fig. 1. Distance information provided by the oculomotor parallax. In the eye of many species including humans, the center of rotation  $C$  is not coincident with the position of the nodal point  $N$ . During a relocation of gaze, the rotation of the eye shifts the projection of an object on the retina by an amount that depends both on the rotation magnitude and on the distance of the object. (a) and (b) illustrate the retinal projections of the points  $A$  and  $B$  before and after a saccade. The zoomed panels show that the two projections are coincident before, but not after, the rotation. For clarity, a single nodal point is shown.

in Section 2. Section 3 describes the pan/tilt unit used in this study, its optical and mechanical properties, and the calibration process followed to precisely position the cardinal points. Section 4 reports results obtained with controlled camera rotations and with recorded sequences of human eye movements. A brief discussion of experimental results and possible extensions of the proposed approach concludes the paper.

## 2. The oculomotor parallax

The robotic system described in this paper was designed to replicate the oculomotor parallax present in the human eye. A critical difference between the eyes of primates and the cameras used in robotic vision systems is the geometrical arrangement of photoreceptors. Whereas retinal receptors in the eye cover a surface that can be interpolated by a portion of a sphere, transducing elements in cameras are usually organized to compose a planar array. This section analyzes the distance information provided by the oculomotor parallax in both spherical and planar sensors. As explained in the following section, comparison between these two systems was critical for the design and control of the robot.

### 2.1. Spherical sensor

The oculomotor parallax present in the human eye can be studied by means of Gullstrand's schematic eye model [41] (see Fig. 2(a)). To ensure that parallel rays entering the eye focus perfectly on the retina, this model assumes a spherical ocular bulb with radius of 11.00 mm, a lens with average index of refraction equal to 1.413, and an axial distance from the cornea to the retina of 24.40 mm [27]. The lens is characterized by two nodal points that, in the unaccommodated eye, are located at 6.07 mm and 6.32 mm from the center of rotation. For simplicity, in this section we analyze the oculomotor parallax of

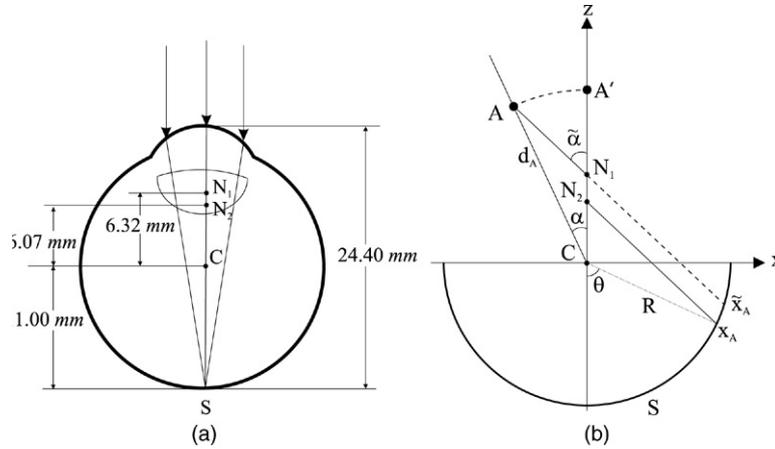


Fig. 2. Oculomotor parallax in a spherical sensor. (a) Gullstrand's three-surface reduced schematic eye model.  $N_1$  and  $N_2$  indicate the two nodal points of the optical system.  $C$  is the center of rotation.  $S$  is the retinal surface. (b) Symbol notation and conventions used in the geometrical analysis.  $\theta$  identifies the projection on the sensor of a point light source at distance  $d_A$  and eccentricity  $\alpha$ .

an infinitesimally small light source that produces a projection of a single point on the retina. As shown in Fig. 2(b), we assume this Point Light Source (PLS) to be located in front of the sensor at position  $A = (-d_A \sin \alpha, d_A \cos \alpha)$ .<sup>1</sup>

In the case of an optical system with a single nodal point, the retinal projection of the PLS,  $\tilde{x}_A$ , is given by the intersection between the sensor surface and the line  $AN_1$ . This line verifies the equation:

$$x(d_A \cos \alpha - CN_1) + z d_A \sin \alpha - d_A CN_1 \sin \alpha = 0. \quad (1)$$

In a two-nodal-point system, as in the case of Gullstrand's eye model, a ray of light going through the first nodal point,  $N_1$ , with an angle  $\tilde{\alpha}$  exits the lens by the second nodal point,  $N_2$ , with the same angle [42]. Thus, the projection  $x_A$  of  $A$  on the sensor is given by the intersection of the sensor's surface and the line parallel to  $AN_1$  going through  $N_2$ . This line has the equation:

$$x(d_A \cos \alpha - CN_1) + z d_A \sin \alpha - d_A CN_2 \sin \alpha = 0. \quad (2)$$

In polar coordinates the retinal surface is represented by  $r = R$ . Thus, the intersection between the line in (2) and the sensor surface is given by:

$$R \sin(\theta + \tilde{\alpha}) = -\frac{|-d_A CN_2 \sin \alpha|}{\sqrt{CN_1^2 + d_A^2 - 2 CN_1 d_A \cos \alpha}} \quad (3)$$

where  $\theta$  is the angle that identifies the projection  $x_A$  of the PLS on the sensor, and  $\tilde{\alpha}$  is the angle between the line  $AN_1$  and the optical axis  $z$ :

$$\tilde{\alpha} = \arctan\left(-\frac{d_A \sin \alpha}{d_A \cos \alpha - CN_1}\right). \quad (4)$$

From (3) and (4) it is possible to obtain explicitly the projection angle  $\theta$  of the PLS on the sensor as a function of its distance  $d_A$

and eccentricity  $\alpha$ :

$$\begin{aligned} \theta &= f(d_A, \alpha) \\ &= \arcsin\left(\frac{|-d_A CN_2 \sin \alpha|}{R\sqrt{CN_1^2 + d_A^2 - 2 CN_1 d_A \cos \alpha}}\right) \\ &\quad + \arctan\left(\frac{d_A \sin \alpha}{d_A \cos \alpha - CN_1}\right). \end{aligned} \quad (5)$$

Eq. (5) shows that the projection  $\theta$  depends on the PLS distance  $d_A$  from the center of rotation  $C$ . The origin of the oculomotor parallax lies in this dependence. Indeed, this equation is equivalent to the oculomotor parallax generated by a rotation of the eye by an angle  $\alpha$  that brings the PLS from position  $A'$  on the optical axis to its current position  $A$  (see Fig. 2(b)).

Given a specific value  $\theta$ , there are an infinite number of possible PLS locations that verify (5). In fact, every point on the line  $AN_1$  yields the same projection, although separate points differ in their distances and eccentricities. Whereas the location of the retinal projection angle of the PLS,  $\theta$ , can be directly measured,  $d_A$  and  $\alpha$  are not known.

This paper proposes an active approach to disambiguating the position of an object in space. This method relies on the change in  $\theta$  following a rotation of the sensor. Fig. 3 explains the proposed method for a rotation  $\Delta\alpha$  around the pan axis. In this study we have focused on this kind of rotation, as there is ample evidence that humans rely mainly on horizontal disparities in their evaluation of distance [43]. However, since the cardinal points of the model are located on the optical axis  $z$ , the method can be applied to both the horizontal and vertical displacements produced by rotations along the pan and tilt axes.

Let  $\theta$  and  $\theta'$  indicate the angles of a projection of a PLS on the sensor before and after a rotation  $\Delta\alpha$ . For each of these two measurements, (5) establishes a relationship between possible values of the PLS eccentricity  $\alpha$  and distance  $d_A$ . Although this relationship cannot be expressed analytically, it can be shown to be continuous and monotonic. Fig. 3(a) provides an example of the two curves verifying (5) before and after a rotation of  $2^\circ$ .

<sup>1</sup> In this paper all distances are evaluated with respect to the center of rotation  $C$  of the eye or the camera.

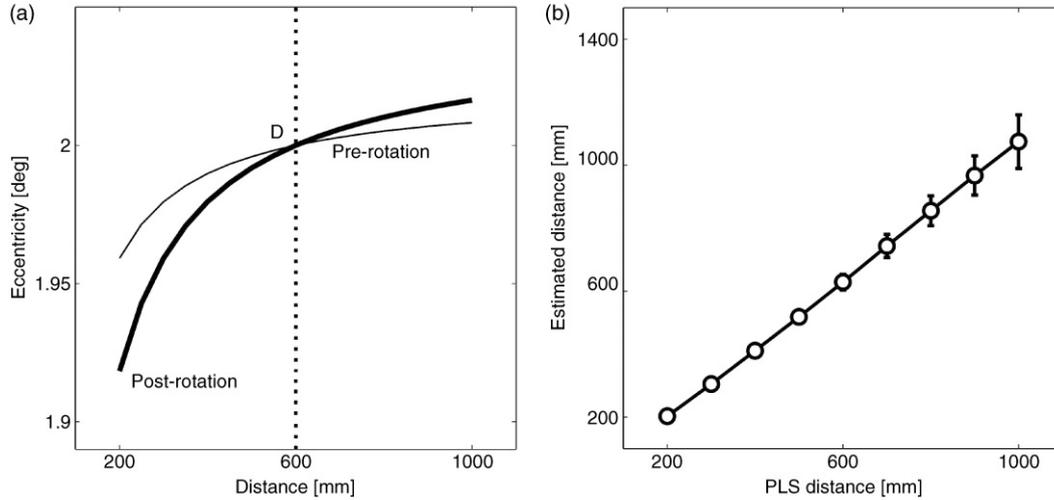


Fig. 3. Distance estimation on the basis of the oculomotor parallax in the human eye. (a) A rotation of the eye identifies the position of a target. The two curves represent the ensembles of possible target positions compatible with the projections measured before (thin line) and after (thick line) a 2° rotation. The point of intersection *D* gives the distance and eccentricity of the target. (b) Application of the method to estimating the distance of a Point Light Source (PLS). Data are the results of simulations in which Gullstrand’s eye model was rotated by 2°. Error bars represent one standard deviation.

These graphs were obtained by simulating the projection on an ideal sensor (a sensor with infinitesimally small receptors) of a PLS located at an eccentricity of 2° and a distance of 60 cm. Each curve represents all possible combinations of values of distance–eccentricity that produced either  $\theta$  or  $\theta'$ . Since a rotation around *C* does not change the distance between the camera and the PLS, the two curves intersect at a single point *D*. This point identifies the spatial coordinates of the PLS.

Fig. 3(b) shows the accuracy of the proposed method. The distance estimated by evaluating the intersections of the two curves in Fig. 3(a) is plotted as a function of the actual distance of the PLS. Data were obtained from simulations in which the sensor was rotated by 2°. To simulate measurement errors, a Gaussian noise with zero mean and 2'' standard deviation was superimposed on the shift in the PLS projection generated by the rotation of the sensor. Each data point summarizes averages obtained over 100 repetitions of the experiment. It is clear from these data that the proposed method provides accurate estimation of distance. Fig. 3(b) also shows that since the amplitude of the oculomotor parallax decreases with the target distance, measurement errors result in a larger variability of localization when the distance of the PLS increases. This variability limits the application of the method to distances shorter than a few meters. A variety of important tasks in robotics occur within this range.

2.2. Planar sensor

The dissimilar shape of receptor surface in the eye and in the camera causes the oculomotor parallax to be different in the two systems. This difference is minimal around the center of the sensor and increases with the eccentricity of the projection. Nonetheless, the same approach described in Fig. 3 can also be applied to the case of a planar sensor.

The geometry of the oculomotor parallax for the case of a camera is shown in Fig. 4. The sensor plane is located at  $z = -d_C$ . The point *A* at a distance  $d_A$  from the sensor indicates

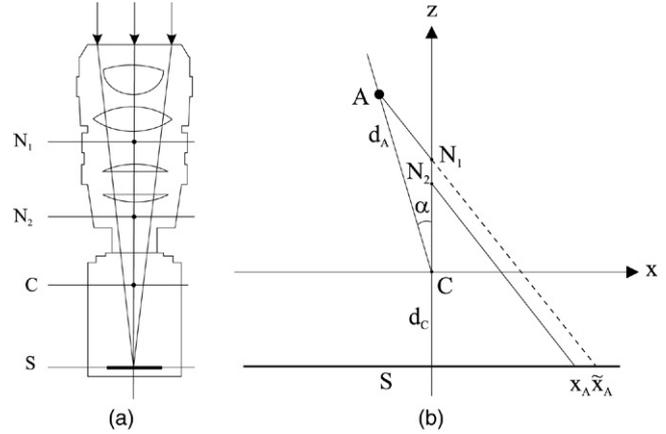


Fig. 4. Oculomotor parallax in a camera with a planar sensor. (a) Cardinal points of a system similar to the one used in the experiments.  $N_1$  and  $N_2$  represent the lens nodal points. *C* is the center of rotation. *S* represents the sensor plane. (b) Symbol notation and conventions used in the geometrical analysis.  $x_A$  identifies the projection on the sensor of a PLS at distance  $d_A$  and eccentricity  $\alpha$ .

the position of a PLS. As in the case of the eye (see (5)), the projection of the PLS on the receptor array,  $x_A$ , is given by the intersection of the sensor surface and the line parallel to  $AN_1$  going through  $N_2$ :

$$x_A = \frac{d_A f \sin \alpha}{d_A \cos \alpha - CN_1} \tag{6}$$

where  $f = d_C + CN_2$  represents the focal length of the lens.

Similar to (5), the projection  $x_A$  given by (6) exhibits a dependence on the distance  $d_A$  of the PLS. This dependence is at the basis of the oculomotor parallax. In this case, however, the PLS distance  $d_A$  can be expressed explicitly as a function of the PLS eccentricity  $\alpha$  (the angle between  $AC$  and the optical axis  $z$ ) and its projection on the sensor  $x_A$ :

$$d_A = \frac{x_A CN_1}{x_A \cos \alpha - f \sin \alpha} \tag{7}$$

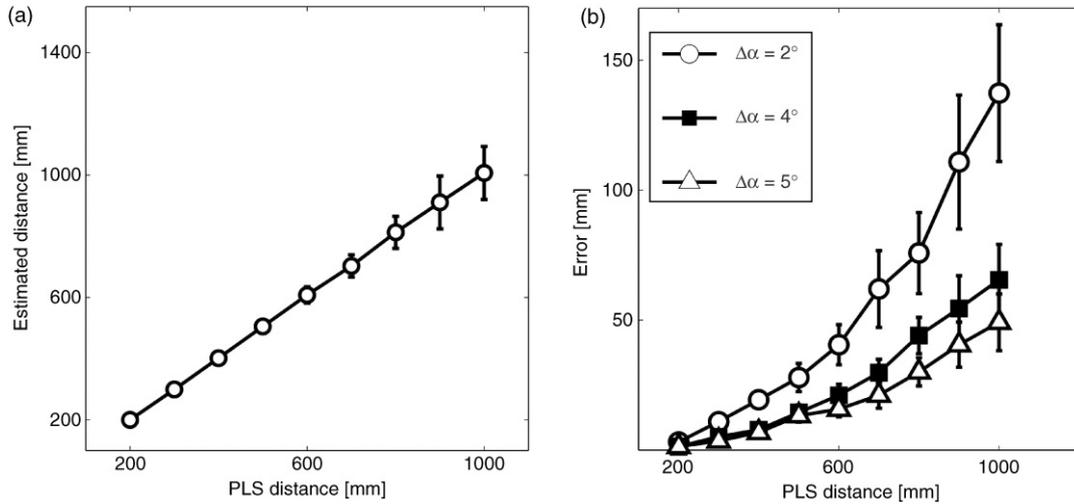


Fig. 5. Distance estimation on the basis of the oculomotor parallax. Case of a planar sensor. Data are results from simulations in which the distance of a PLS target was estimated by rotating the camera by variable amounts. (a) Results obtained with a 2° rotation. Error bars represent one standard deviation. Error bars are not visible for nearby targets due to the small variability of the measurements. (b) Mean estimation error for targets at various distances. Results are shown for rotations of three different amplitudes. Error bars represent half standard deviation.

This equation does not uniquely identify the position of the PLS, because any point located on the line  $AN_1$  produces the same projection  $x_A$ . To recover the eccentricity ( $\alpha$ ) and distance ( $d_A$ ) of the PLS, the same approach of the previous section can be followed. Since a rotation around  $C$  does not change the PLS distance  $d_A$ , the two estimates of distance obtained from (7) before and after a rotation of known amplitude  $\Delta\alpha$  can be equated:

$$\frac{x_A CN_1}{x_A \cos \alpha - f \sin \alpha} = \frac{x'_A CN_1}{x'_A \cos(\alpha + \Delta\alpha) - f \sin(\alpha + \Delta\alpha)} \quad (8)$$

where  $x'_A$  is the PLS projection on the sensor measured after the rotation. (8) yields an analytical expression of  $\alpha$  as a function of the rotation amplitude  $\Delta\alpha$  and the PLS projections:

$$\alpha = \arctan \left( \frac{x'_A [1 - \cos \Delta\alpha] + f \sin \Delta\alpha}{f \left[ \frac{x'_A}{x_A} - \cos \Delta\alpha \right] - x'_A \sin \Delta\alpha} \right). \quad (9)$$

Substitution in (7) of the value for  $\alpha$  obtained from (9) gives the distance of  $A$ .

Fig. 5 shows the accuracy of the method in simulations that replicated the robotic system used in our experiments. The projections of a PLS before and after a rotation of the camera were simulated by means of (6). The eccentricity and distance of the PLS were recovered by means of (9) and (7). To examine the impact of measurement errors, random Gaussian noise with zero mean and 5  $\mu\text{m}$  standard deviation was superimposed on the parallax. Each data point represents an average over 100 repetitions of the experiment.

Fig. 5(a) shows results obtained with a rotation of 2°. The estimated distance of the target is plotted as a function of its simulated distance. As shown by these data, the oculomotor parallax provides extremely accurate estimation of distance for targets located within a near range of distances. In these simulations, the estimates varied from  $20.0 \pm 0.4$  cm for a target located at 20 cm to  $102.4 \pm 18.9$  cm for a target at 100 cm.

That is, the mean percentage error in evaluating the absolute distance of a target at 100 cm was only 2%. As illustrated by Fig. 5(a), the range of applicability of the method depends on the accuracy with which the parallax can be measured. As in the case of the eye, the impact of measurement noise becomes more pronounced for distant targets that produce a small parallax.

Fig. 5(b) shows the mean estimation error defined as the mean absolute value of the difference between the estimated and actual distances of the target. Results are shown for PLSs at various distances and for rotations of three different amplitudes. Not surprisingly, the estimation error increased in conditions that produced a small parallax, such as with rotations of small amplitudes or with targets located at large distances. Yet, within a near range of distances, even small rotations produced accurate estimates. This result is important in real-world applications, where, unlike the simulations of Fig. 5, measurement errors do depend on the amplitude of rotation. Large rotations introduce a significant degree of distortion in the way objects appear in images acquired before and after the movements. This distortion complicates the identification of corresponding features and results in a high level of measurement noise. In practice, the range of rotation amplitudes that provide useful distance information is bounded by the resolution of the sensor on one side and the solution of the correspondence problem on the other. On one extreme, very small rotations produce parallaxes that are too tiny to be reliably detected. On the other, large rotations produce parallaxes that cannot be reliably measured.

### 3. An oculomotor robotic system

The experiments of this study were conducted by means of the robotic head/eye system shown in Fig. 6(a). This system, referred to in this paper as the “robotic oculomotor system”, was developed to model the visual input signals to the human retina during eye movements. It consisted of two mobile cameras, each with two degrees of freedom. Motion was provided

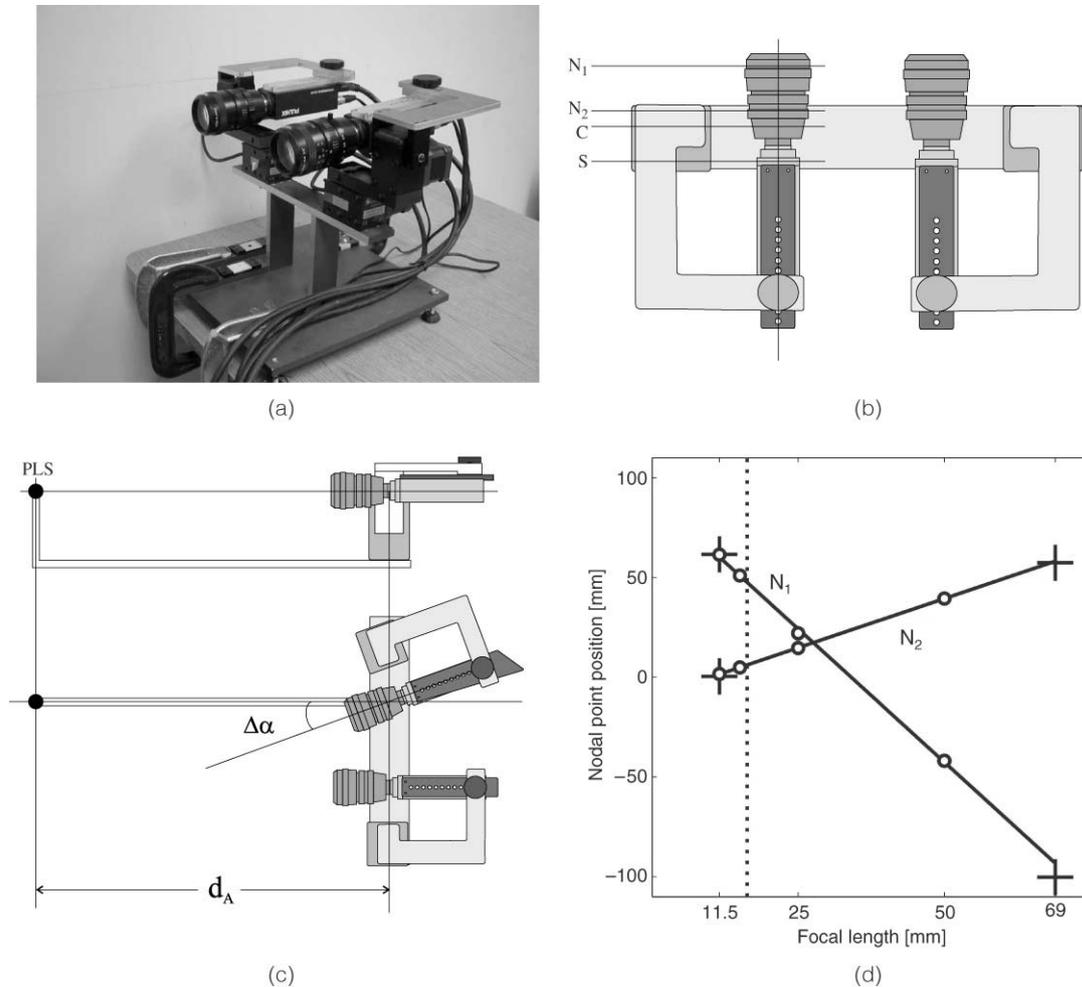


Fig. 6. The robotic oculomotor system used in the experiments. (a)–(b) A custom designed head–eye system replicated the oculomotor parallax present in the human eye. In this system, the relative positions of nodal points ( $N_1$  and  $N_2$ ), center of rotation ( $C$ ), and sensor surface ( $S$ ) followed the arrangement of the eye. (c) Selection of the lens focal length occurred by means of a preliminary calibration procedure. A PLS model was located at various distances from the robot. The shift in the projection of the PLS centroid in images acquired before and after a  $3^\circ$  rotations were measured and interpolated by a model of the planar sensor. (d) Positions of  $N_1$  and  $N_2$  for various focal lengths of the zoom lens mounted on the camera. Circles represent data points obtained from the calibration procedure. The two lines represent linear regressions of calibration data. The crosses mark the positions of  $N_1$  and  $N_2$  provided by the factory specifications. The dotted vertical line indicates the focal length value used in the experiments ( $f = 16$  mm).

by two pan–tilt units (Directed Perception, Burlingame, CA) supported by a custom developed frame. Units were digitally controlled by proprietary microprocessors that ensured movements with a precision higher than  $1'$ . This degree of accuracy is comparable to the spatial resolution of the eyetracker used to acquire the sequences of eye movements that controlled the robot in the experiments (see Section 4.2). By means of a nodal adapter, the two units were mounted so that the pan and tilt axes of rotation intersected at a point  $C$  (the center of rotation, see Fig. 6(b)). Aluminum wings specifically designed for this study enabled positioning of the center of rotation between the sensor plane  $S$  and the nodal points of the camera, as occurs in the human eye. Each of the two wings possessed a sliding bar with calibrated holes that allowed the distance  $d_C$  between the center of rotation  $C$  and the sensor  $S$  to be varied in discrete steps. In this study,  $d_C$  was set to 10.5 mm, which is close to the value of 11 mm measured in the human eye.

The system was equipped with two digital cameras (Pulnix, Sunnyvale, CA) each with a 11.5–69 mm zoom lens. Cameras

possessed  $640 \times 484$  CCD sensors (photoreceptor size  $9 \mu\text{m}$ ) and allowed an acquisition rate of 120 frames/s. Images were acquired by means of a fast frame-grabber (Datacube, Danvers, MA). Only one of the two mobile cameras was used in the experiments described in this paper.

In a zoom lens, the positions of the nodal points depend on the focal length. By means of a preliminary calibration procedure, we examined the projections on the sensor of standard calibration images to verify that the locations of  $N_1$  and  $N_2$  matched the values provided by the manufacturer for various focal lengths. This procedure enabled selection of the focal length that most accurately reproduced the oculomotor parallax present in the human eye. As illustrated in Fig. 6(c), this calibration was performed by means of a real-world approximation of a PLS. This PLS model was composed of an ultra-bright LED (wavelength 660 nm, typical power 2000 mcd) covered by a pierced steel mask with a 3 mm hole. The mask minimized possible imperfections in the LED, as well as changes in the shape of the PLS in images taken from

different points of view. A calibrated aluminum bar allowed precise control of the distance at which the LED was located with respect to the center of rotation  $C$ . The PLS model was placed on the optical axis at various distances from the robot. For each distance, a rotation of  $3^\circ$  was performed, and the centroid of the PLS projection in the image was calculated. The centroid shifts in the images acquired before and after each rotation were measured for several focal lengths and compared to the predictions of the planar sensor model (see Section 2.2). For each considered focal length, the coordinates of  $N_1$  and  $N_2$  were estimated by interpolation of experimental data by means of regularized least squares.

Fig. 6(d) shows the positions of the two nodal points estimated by means of this calibration procedure. The distances  $CN_1$  and  $CN_2$  varied almost linearly with the focal length of the zoom lens. As shown by Fig. 6(d),  $N_1$  was closer to the sensor than  $N_2$  for focal lengths greater than 27 mm. This observation constrained the range of possible focal length values to be less than 27 mm, as in the human eye  $N_2$  is the nodal point closer to the retina. It should be observed that in both the eye and the camera, the projection of the stimulus on the receptor surface is strongly sensitive to the position of  $N_2$  and relatively unaffected by the location of  $N_1$ . Indeed, in both (5) and (6),  $CN_1$  appears at the denominator together with the much larger term  $d_A$ . For this reason, we chose the focal length that positioned  $N_2$  as close as possible to the location specified in Gullstrand's eye model of  $CN_2 = 6.03$  mm. Selection of a focal length  $f = 16$  mm yielded  $CN_2 = 5.72$  mm, thus producing an oculomotor parallax very similar to that present in the human eye.

#### 4. Experimental results

Two sets of experiments were conducted in this study. To rigorously examine the accuracy of the proposed method, we first analyzed the oculomotor parallax produced by rotations of the camera with predetermined amplitudes. As an example of application of the method, we show how a single camera rotation can disambiguate a camouflaged object from the background. A second series of experiments examined the depth information that emerges during sequences of human eye movements. We show that the oculomotor strategy by which humans scan the visual environment provides accurate estimation of distances of objects present in the scene.

##### 4.1. Distance estimation with a single camera rotation

In a first series of experiments, we used the oculomotor parallax produced by a single rotation of the camera to estimate the distance of a PLS target. Following the approach described in Section 2 and by means of the same experimental set up used in the preliminary calibration procedure, we measured the position of the PLS centroid in images acquired before and after a rotation  $\Delta\alpha$  (the motor command transmitted to the robot). The oculomotor parallax was then converted into an estimation of distance on the basis of the camera model provided by (7) and (9). Model parameters replicated the positions of the cardinal points in the robot.

Fig. 7 shows results obtained from these experiments. Data points represent average values evaluated over 25 measurements. Fig. 7(a) compares the actual distance of the PLS model to the distance estimated by means of rotations with amplitudes  $2^\circ$  and  $5^\circ$ . Fig. 7(b) shows the mean estimation error produced by rotations with three different amplitudes. Consistent with the results of computer simulations (see Fig. 5), the proposed method produced accurate spatial localization for objects located not far from the robot. As in the simulations, more accurate estimates were obtained with larger rotations and nearby targets, conditions that produced larger parallaxes. Nonetheless, good localization was also obtained with relatively small rotations. For example, the mean estimation error of a target at 92 cm obtained by means of a rotation as small as  $2^\circ$  was 11 cm, a percentage error smaller than 12%.

An important difference with respect to simulations was the impact of the quantization that originated from the finite dimension of the pixel in the camera's CCD sensor. Quantization errors were responsible for the staircase appearance of the  $2^\circ$  curve in Fig. 7(a). In the experiments with the robot, quantization errors became significant when the amplitude of the oculomotor parallax was of the order of a few photoreceptors. This could occur for large rotations if the target was sufficiently distant from the robot. Due to quantization, the system was unable to determine the position of a target farther than a limit distance, beyond which the oculomotor parallax was smaller than a single photoreceptor.

Fig. 8 examines the impact of quantization errors on estimating the distance of a PLS target. In Fig. 8(a), the oculomotor parallax measured with a rotation of  $3^\circ$  is compared to the parallax predicted by simulations of the planar model. No quantization error was present in the simulations, as the size of the simulated photoreceptor was infinitesimally small. As shown in Fig. 8(a), the effect of pixel quantization was more pronounced for objects located at larger distances from the robot where the discrete steps of single pixel increments were visible. Due to quantization, a specific parallax value corresponded to a range of distances, rather than a single distance as in the simulations. Thus, the method produced an uncertainty zone, which was determined by the set of distances that gave the same parallax. The size of this zone increased with the distance of the target and decreased with the amplitude of the rotation.

Fig. 8(b) shows the influence of pixel size on the limit distance, *i.e.* the distance at which the oculomotor parallax becomes smaller than the quantization error. Data are results from simulations of  $2^\circ$  and  $3^\circ$  rotations. The smaller the photoreceptor, the wider the range of distances for which the oculomotor parallax provided useful depth information. With a rotation of  $2^\circ$ , a photoreceptor of  $9\ \mu\text{m}$  (the size of the pixel in the cameras mounted on the robot) yielded a limit distance of 2 m. In contrast, the limit distance obtained with the same rotation amplitude for a receptor of  $2.5\ \mu\text{m}$  (the size of a cone in the fovea) was approximately 9 m. Thus, by virtue of the small retinal receptors (and without considering the phenomenon of visual hyperacuity), the human eye extends

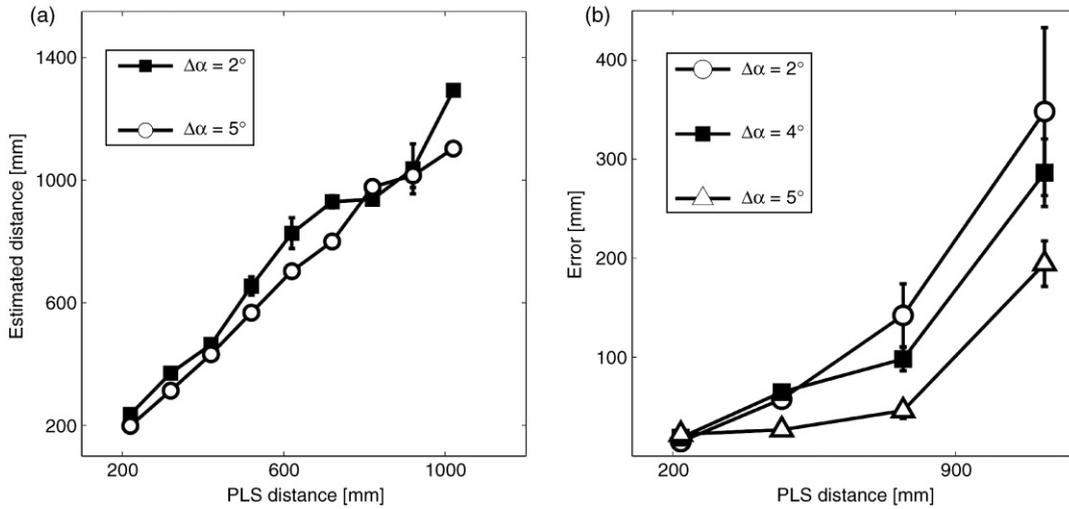


Fig. 7. Estimating the distance of a PLS target on the basis of the oculomotor parallax. Measurements were obtained with a single rotation of the camera. (a) Target localization obtained by means of rotations with  $2^\circ$  and  $5^\circ$  amplitudes. Error bars represent one standard deviation. (b) Mean estimation error for targets at various distances. Results are shown for rotations of three different amplitudes. Error bars represent half standard deviation.

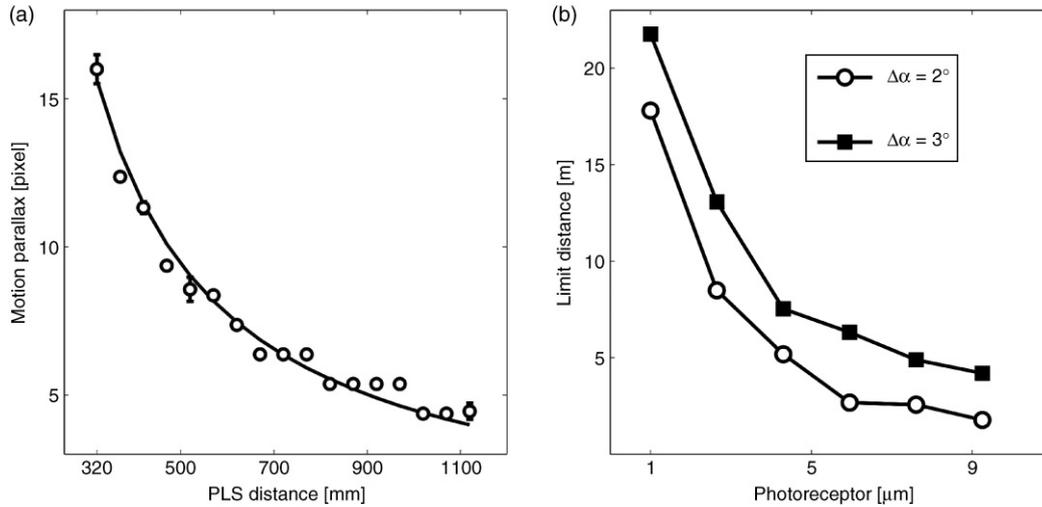


Fig. 8. Impact of quantization errors on system performance. (a) Oculomotor parallax measured with a rotation of  $3^\circ$ . Circles represent mean data obtained from the robot. The continuous line is the estimated oculomotor parallax obtained from simulations of the planar model. Note the discrete steps produced by quantization in the oculomotor parallax measured experimentally. (b) Maximum measurable distance as a function of pixel size. The curves represent the limit distances beyond which the oculomotor parallaxes measured with rotations of  $2^\circ$  and  $3^\circ$  fall below the quantization error.

the range of applicability of the method by almost a factor of 5 with respect to the robot. It is important to observe that the limit distance provides only a theoretical upper bound to the range of distances to which the method can be applied. In practice, sources of measurement error other than quantization further narrow this range. Conservatively, the experiments of this paper considered distances up to 1 m.

Fig. 9 shows an example of the application of the oculomotor parallax to a problem of figure/ground segregation. In this example, the depth information provided by the oculomotor parallax was used to detect an object camouflaged by a flat background. As illustrated in Fig. 9(a), a rectangular panel was positioned in between the robot and the background. The background portrayed a reproduction of Van Gogh’s painting “Irisles” located at a distance of 90 cm. The panel was located at a variable distance from the robot and covered with an enlarged

portion of the same painting. This enlargement was carefully designed to disguise the panel within the background, and both the panel and the background were kept in focus by means of a small lens aperture. In this way, as shown in Fig. 9(b), the edges of the panel could not be detected in the images acquired by the camera, and the panel was virtually invisible. In these experimental conditions, figure/ground segregation would fail with any visual cue other than the oculomotor parallax.

Fig. 9(c) shows the distance estimated on the basis of a rotation of  $3^\circ$ . In this example, the panel was located at 70 cm from the robot. The image acquired before the rotation was fragmented into an array of  $45 \times 45$  rectangular patches, each composed of  $14 \times 11$  pixels. Normalized cross-correlation of each patch with the image sampled after the rotation provided an estimate of the oculomotor parallax at the patch location. In this way, an estimate of distance was obtained for a number

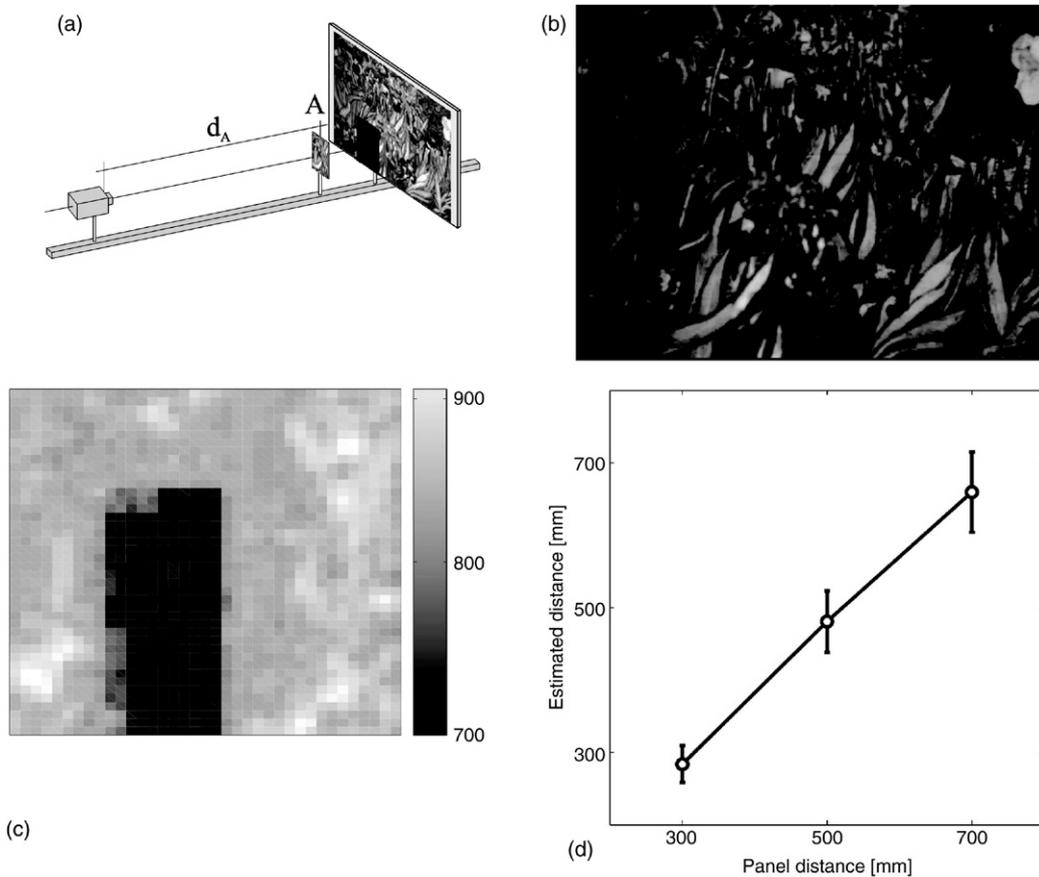


Fig. 9. Application of the oculomotor parallax to a figure/ground segregation task. (a) Experimental setup. A rectangular panel *A* is located in between the robot and a textured background. (b) The panel contained an enlarged replica of the background. The enlargement was tailored to achieve perfect camouflage. (c) Distance estimation on the basis of the oculomotor parallax produced by a rotation of  $3^\circ$ . Each point in the matrix represents the estimated distance of a  $14 \times 11$  pixel patch in the images acquired by the robot. The panel was at a distance of 70 cm. (d) Results obtained for panels at three different distances (30, 50 and 70 cm) with a  $3^\circ$  rotation. Each data point represents the average distance evaluated over 250 points on the panel. Error bars represent one standard deviation.

of equispaced locations in the image. As shown by Fig. 9(c), a single rotation of the camera was sufficient to expose the camouflaged panel. Pixels on the panel produced different distances than those on the background. Fig. 9(d) summarizes results obtained with panels located at three different distances. Each data point represents the average distance evaluated over 250 points on the panel. As shown by these data, a single small rotation was sufficient to produce a reliable estimate of the panel position.

#### 4.2. Distance estimation by replication of human oculomotor behavior

In a second series of experiments, we examined the depth information produced by sequences of camera rotations that replicated the oculomotor activity of human observers. Eye movements were recorded by means of a Dual Purkinje Image (DPI) eyetracker (Fourward Technologies, Buena Vista, VA). This high-resolution device estimates rotations of the eye by measuring differences in the first and fourth corneal reflections (the Purkinje images) of an infrared beam. It achieves spatial and temporal resolutions of  $1'$  and 1 ms, respectively [44]. As shown in Fig. 10, subjects viewed the scene with the right eye, while the left eye was covered by an opaque eye-patch. Eye

movement data were first low-pass filtered and then sampled at 1 kHz.

The high spatial sensitivity of the DPI eyetracker allowed discrimination of the small movements of fixation that are not visible with most other eye-tracking systems. Indeed, although we are not aware of them, small eye movements continuously occur even when the eye fixates on a target<sup>2</sup> [49,50]. Fig. 10(b) shows a trace of recorded eye movements. Both macroscopic saccades and fixational eye movements are present.

Fig. 11 shows an example of distance estimation obtained by replicating human oculomotor activity. In this experiment, the eye movements of a subject were recorded for a period of 10 s while viewing the same scene presented to the robot. As shown in Fig. 11(a), this scene was composed of 5 objects located at various distances. Visual occlusions, together with similarities in the colors and textures of the objects, made segmentation of the scene difficult. Recorded eye movements (shown in Fig. 13(b)) were used as motor commands for

<sup>2</sup> It is surprising that we perceive a stable world despite this constant jittering of the eyes. A striking demonstration of fixational instability is available on the World Wide Web at <http://visionlab.harvard.edu/Members/Alumni/ikuya/html/memorandum/VisualJitter.html> [45]. These movements have been proposed to play a fundamental role in the way visual information is encoded in the brain [46–48].

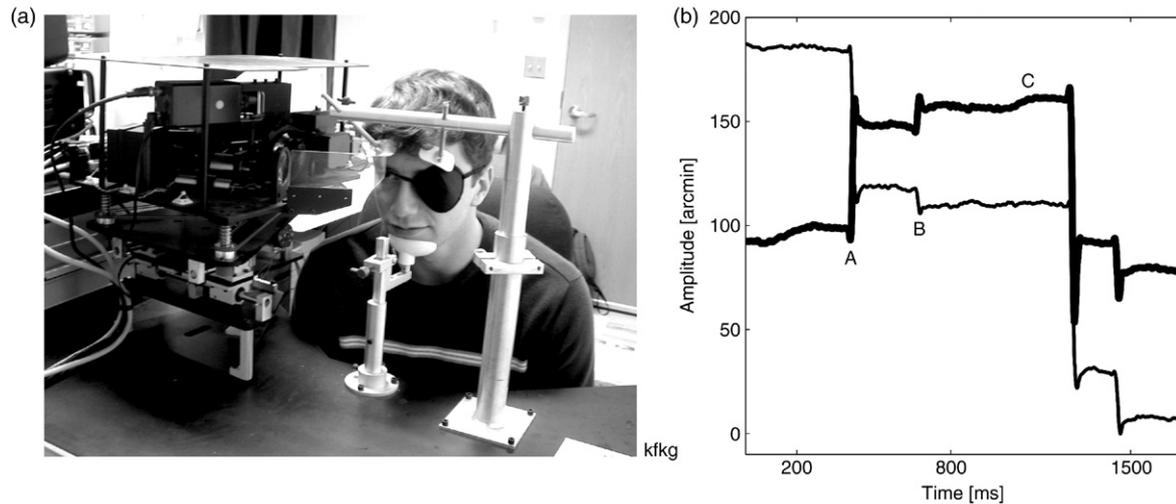


Fig. 10. Measuring the eye movements of human subjects. (a) A high-resolution Dual-Purkinje-Image eyetracker recorded oculomotor activity while subjects examined the same scene presented to the robot. The scene was viewed monocularly with the right eye, while the left eye was covered by an opaque eye-patch. Head movements were prevented by means of a chin-rest. (b) Example of recorded eye movements (thick line: horizontal displacement; thin line: vertical displacement). Letters mark various oculomotor events: A — saccade; B — fixational saccade; C — ocular drift. Both macroscopic and fixational saccades were used to estimate distance.

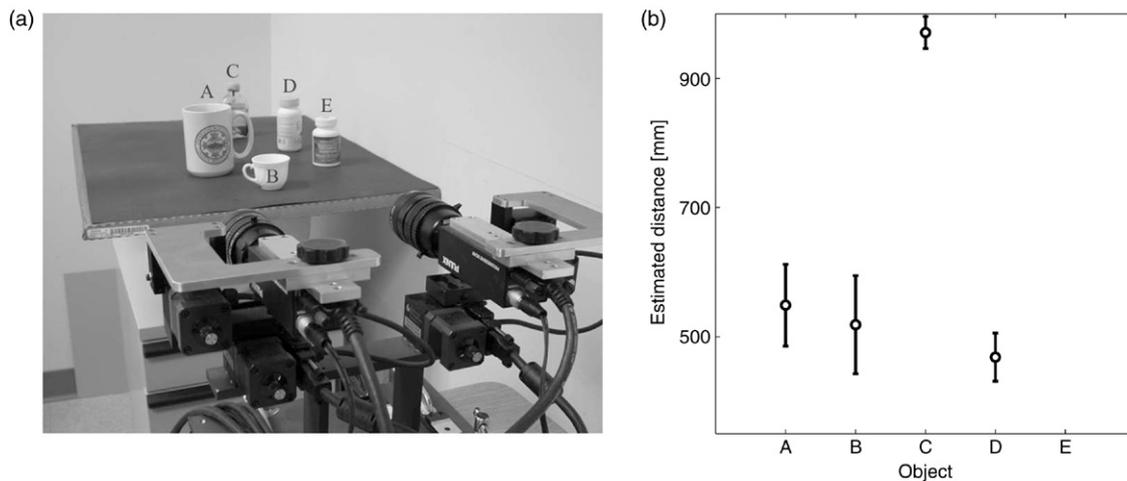


Fig. 11. Estimating distance with sequences of human eye movements. Case of macroscopic saccades (saccades larger than  $1^\circ$ ). (a) Experimental setup. Several objects were placed at various distances in front of the robot to compose a complex scene. The distances of the various objects were: A = 59 cm; B = 43 cm; C = 88 cm; D = 74 cm; E = 64 cm. (b) Mean estimated distance of the pixels composing the objects. Error bars represent one standard deviation. The distance of E could not be estimated with the recorded trace of eye movement, as it did not appear within the field of view of the camera (see Fig. 13(b)).

the robot. In transforming the oculomotor signals into motor commands, the velocity of movement was reduced by properly expanding the time axis. In this way, it was possible to operate the robot within a range of velocities that could be achieved reliably, without jeopardizing the spatial accuracy with which eye movements were reproduced. A preliminary calibration in which the subject fixated on a number of predetermined points in the scene was used to find a linear correspondence between voltages generated by the eyetracker and motor signals fed to the robot. This calibration ensured that the images acquired by the camera were centered on the points fixated by the subject.

Fig. 11(b) shows estimates of distance obtained on the basis of the oculomotor parallax measured during macroscopic saccades (saccades larger than  $1^\circ$ ). The mean amplitude of these saccades in the recorded sequence was  $4.76 \pm 2.60^\circ$  ( $N = 9$ ). Similar to the data analysis for Fig. 9, the oculomotor

parallax at various locations in the scene was evaluated on the basis of normalized cross-correlation of rectangular  $32 \times 24$  pixel patches. Every sample of the oculomotor parallax was converted into an estimate of distance by means of the robot model. Data points in Fig. 11(b) represent the mean distance estimated for each object. Since the oculomotor parallax could be measured only for saccades that maintained an object within the field of view of the camera, averages were evaluated over a different number of measurements for each object. Means were calculated both over space (the patches composing each object) and time (the 9 saccades of the sequence). The total numbers of measurements were 8357 for object A, 2461 for object B, 981 for object C, and 259 for object D. Given the limited field of view of the camera, the recorded sequence of eye movements did not allow estimation of the oculomotor parallax for object E, which was located at the margins of the scene. The data

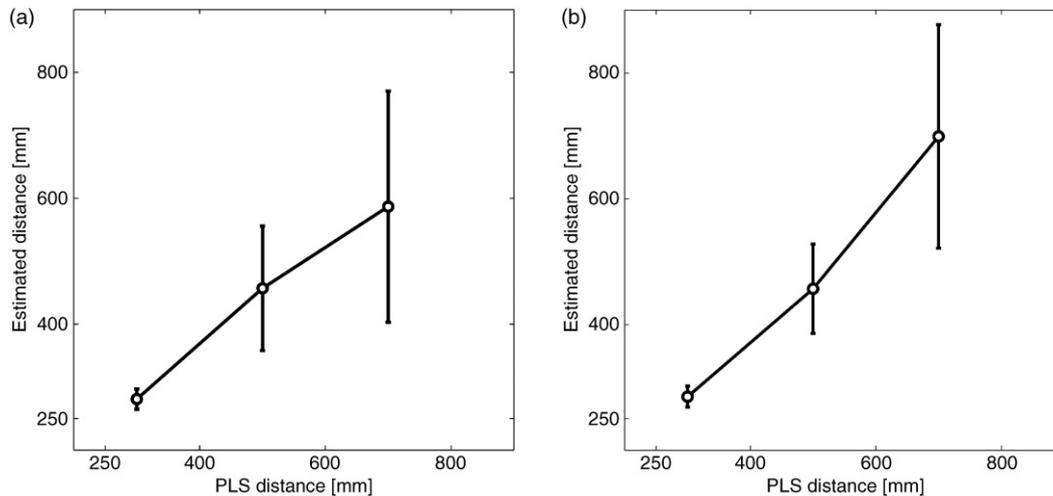


Fig. 12. Estimating distance by means of fixational saccades. Also the small saccades of visual fixation produce parallaxes that provide useful depth information. The two panels show results obtained with traces of eye movements from two subjects (Left — subject MV, Right — subject AR). The same sets of fixational saccades were repeated for each distance of the PLS model. Error bars represent half standard deviation.

in Fig. 11(b) show that the method produced good estimates of distances when a sufficient number of measurements was available. The higher error in estimating the distance of object D originated from the limited number of parallax measurements available for this object.

The results of Fig. 11 were obtained using relatively large saccades. These eye movements did not include the smaller saccades that often occur during visual fixation. Although small in amplitude, fixational saccades move the stimulus on the retina by tens to hundreds of photoreceptors and produce parallaxes that are well within the resolving power of the visual system. Due to their small amplitudes, fixational saccades have two important properties: (a) they introduce little distortion in the way objects appear in images acquired before and after rotations; and (b) they limit the areas in which corresponding features can be found in the two images. These two properties greatly simplify the correspondence problem that underlies the extraction of the oculomotor parallax.

Replicating fixational saccades by means of a robotic system is a challenging task. In the absence of dedicated image processing, the smallest parallax that can be reliably discriminated in the images acquired by the robot is determined by the size of photoreceptors in the camera. On the contrary, the resolving power of the visual system goes far beyond the limit set by the size of retinal receptors. Humans can reliably detect misalignments as small as  $2''$ – $5''$  even though the theoretical resolution given by receptor spacing in the eye is approximately  $50''$  [14]. Enlarging fixational eye movements by a factor that compensates for this difference between the camera and the eye gives an amplification factor  $A_f = 135$ . That is, to discriminate a fixational saccade of  $30''$  with the same resolving power present in the human eye, it is necessary to rotate the camera by  $67^\circ$ ! Such a high amplification factor clearly defeats the purpose of replicating fixational saccades in the robot, as it transforms these movements into large rotations.

The previous observations clarify the need for using high-resolution cameras in order to replicate small fixational

saccades. Nonetheless, although far from reproducing the depth information present in the eye, even a moderate amplification of fixational instability, one that does preserve the small rotations of fixational eye movements, yields 3-D information that can be reliably exploited by the average resolution cameras commonly used in robotics. Fig. 12 shows the results of an experiment in which fixational saccades, amplified only by a factor  $A_f \approx 2$ , were used to estimate the distance of the PLS model. The oculomotor activity of two subjects (MV and AR) was recorded while they maintained visual fixation for a period of 3 s. Ten traces were acquired for each subject. Fixational saccades were identified, amplified, and used to control the robot. The mean amplitudes of fixational saccades in the recorded traces were  $0.15 \pm 0.05^\circ$  for subject MV (30 saccades) and  $0.47 \pm 0.19^\circ$  for subject AR (29 saccades). As shown by the data in Fig. 12, the oculomotor parallaxes produced by these small rotations were sufficient to infer the distance of the PLS model.

Macroscopic saccades and fixational eye movements can be used together to obtain more reliable estimates of distance. Fig. 13 shows results obtained by using all the saccades that were present in the sequences of eye movements recorded while viewing the scene of Fig. 11(a). In this experiment, the camera followed the recorded trajectories switching among fixation points through large saccades. At each fixation point, fixational saccades were amplified by a factor  $A_f \approx 2$ , as in the experiment of Fig. 12. This motor strategy was a trade-off between precisely replicating the trajectory of the eye and compensating for the lower resolution of the camera with respect to that of the human eye.

The panels (c) and (d) in Fig. 13 show results obtained by means of the two sequences of eye movements displayed on the top row. These two traces were recorded from the two subjects and both included 29 saccades. The mean amplitude of saccades was  $1.15 \pm 1.11^\circ$  (subject MV) and  $2.12 \pm 2.75^\circ$  (subject AR). As shown by these data, both traces led to a good estimation of the distances of the objects present in the camera's field of view. In this experiment, averages were calculated over 9485

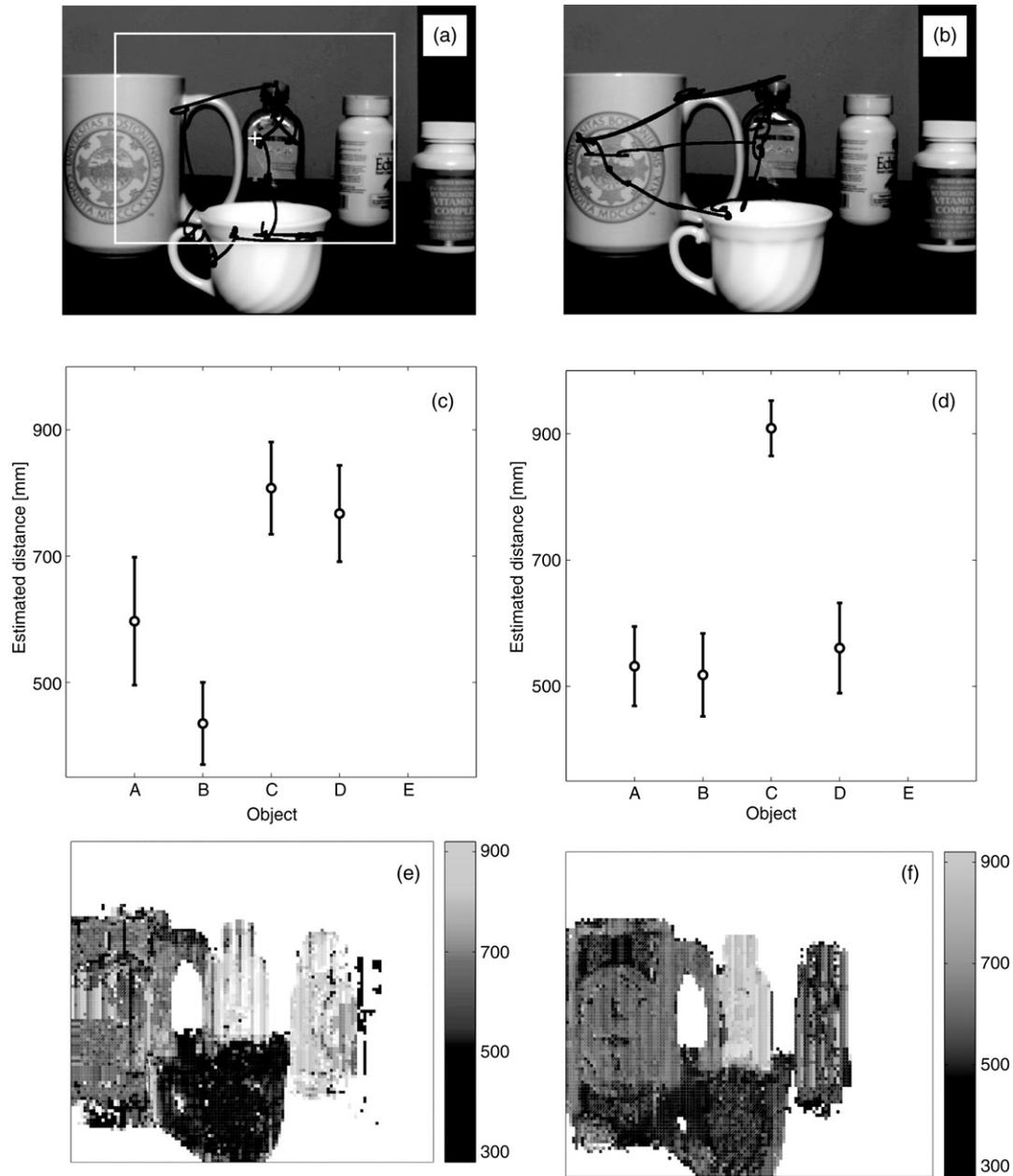


Fig. 13. Estimating distance with sequences of human eye movements. Both macroscopic and fixational saccades were used. The two columns show results with sequences of eye movements recorded from different subjects (left MV; right AR). (Top row) Traces of recorded eye movements (dark lines). The white rectangle represents the field of view of the robot camera. (Middle row) Mean estimated distance of the pixels composing the objects. Error bars represent one standard deviation. (Bottom row) Segmentation of the visual scene based on information of distance.

(AR) and 3549 (MV) parallax measurements for object A, 3067 (AR) and 5324 (MV) for object B, 990 (AR) and 2079 (MV) for object C, and 288 (AR) and 2918 (MV) for object D. The distance of object E on the far right of the scene could not be estimated with either of the two traces, since no saccade maintained this object within the field of view of the camera. These results show that fixational saccades are beneficial to distance estimation. Notice the improvements in measuring the distances of objects C and D in Fig. 13(d) with respect to the case of Fig. 11(b) in which only macroscopic saccades were used.

The panels in the bottom row of Fig. 13 illustrate the contribution of the oculomotor parallax to the segmentation of

the visual scene. Each pixel in Fig. 13(e) and (f) represents the mean distance estimated on the basis of the oculomotor parallax at a specific location of the scene. Averages at each location were estimated over all saccades that allowed measurement of the oculomotor parallax. Since objects at different distances produce different parallaxes, the pixels composing the various elements of the scene are immediately obvious in the images of Fig. 13(e) and (f). The white areas in these images correspond to the uniform surfaces of the table and the background, which did not produce measurable parallax. These data show that the small relocations of gaze that characterize the normal oculomotor activity of human subjects produce depth information that facilitates the visual segregation

of individual objects. The segmentation of a visual scene into its constituent objects is one of the hardest tasks of machine vision. In a system that actively scans a visual scene, the local estimates of distance provided by the oculomotor parallax can be directly integrated with processes based on other visual cues to produce robust algorithms of image segmentation.

## 5. Conclusions

One of the most striking features of biological vision systems is the close interaction between motor and perceptual processes. These systems have evolved in moving agents, and motor activity appears to be a necessary ingredient for their proper functioning. Motor contributions to visual computations have been shown in many species, ranging from insects [51] to birds [52] and humans [53]. A large body of evidence shows involvements of motor activity in the generation and extraction of depth information [22,20,25,27,15–17].

Despite the tremendous improvements in computational power of recent years, computer vision algorithms are still far from replicating the efficiency, robustness, and speed of biological systems. It has been observed that a tight link between motor and perceptual processes may be a key element for the development of more efficient vision algorithms [30–32]. Following these observations, many studies in robotics have exploited the agent's motor activity to extract depth information (see [8,5,54] for reviews). The work presented in this paper builds upon this rich literature.

Two important features differentiate this study from previous research on depth from motion. A first novel aspect regards the specific cue used to estimate distance. While previous publications have analyzed the motion parallax that emerges from large relocations of a vision system, this is the first study to explicitly focus on the parallax produced by camera rotations. This is an important difference, given that such a parallax is present in virtually every head/eye system used in robotics. The results of this paper show that this parallax provides accurate depth information within the space in proximity to the agent. This cue can be reliably used to control local motor interactions with the scene, as in the operations of object reaching and grasping.

A second critical difference with respect to previous studies is the emphasis of this research on emulating human oculomotor behavior. While a motion parallax occurs in most head/eye systems, this study analyzed the case of a robot specifically designed to reproduce the oculomotor parallax present in the human eye. Control of this system with recorded traces of eye movements enabled direct replication of human oculomotor activity. Under natural viewing conditions humans tend to relocate their gaze by small shifts. Most saccades have amplitudes within a few degrees [28]. Even though we are not aware of them, small saccades also occur during the periods of fixation. This study shows that such a scanning strategy produces oculomotor parallaxes that can be reliably measured and used to predict the distances of objects and surfaces. It is often debated whether a dynamic relocation of the direction of gaze presents advantages over the static analysis provided by a stationary camera with a wide field of view.

As shown by this study, an active system can exploit the 3-D information produced by the oculomotor parallax to simplify challenging visual tasks such as figure-ground segregation and image segmentation.

By rigorously analyzing the parallax provided by rotations of a mobile camera, the goal of this paper is to provide the foundations, as well as the motivation, for using the oculomotor parallax in robotics. There are several aspects of this research that are beyond the scope of this paper, but which deserve further investigation in future studies. A first issue regards the solution of the temporal correspondence problem posed by the extraction of the oculomotor parallax, *i.e.* the identification of corresponding features in images acquired at different times. The cross-correlation method used in this study is not suited for real-time performance. It would be interesting to extend to this task more elaborate methods developed in the context of depth from motion and stereopsis.

A second aspect that deserves an in-depth study concerns the transformation of the oculomotor parallax into an estimate of distance. In this study, distance estimation relied on a model of the robot tuned on the basis of a preliminary calibration. A more robust approach would make use of learning to enable autonomous and continuous calibration of the oculomotor parallax. Since this cue provides useful depth information for a proximal range of distances, self-calibration could be accomplished by using proprioceptive knowledge of the position of a robotic arm, an approach that replicates the way the brain learns about visual space.

The robotic oculomotor system described in this paper was developed as part of a wider program of research focused on integrating robotic systems with computational models of the brain [55,56]. Given the complexity of the human oculomotor system, this robot lacks several characteristics of the eye, including the coexistence of different receptors, their space-variant distribution, and the presence of torsional rotations. Nonetheless, the acquired images were sufficiently accurate replicas of retinal signals for the purpose of this study. Careful positioning of the cameras, selection of the optics, and specification of the motor characteristics ensured that the separation between nodal points and the center of rotation precisely replicated the oculomotor parallax present in the human eye.

This paper provides an example of an interdisciplinary approach to the study of perception, which establishes a direct link between the natural and engineering sciences. Studies in computational neuroscience often rely on simulations of simplified environments that bear little resemblance to the real world. Robotic systems provide access to the sensory inputs occurring during motor activity. Quantitative analysis of these signals can drive psychophysical and neurophysiological investigations. For decades, researchers in artificial intelligence and robotics have looked at biology as a source of inspiration. The use of robotic systems has recently opened a new way in which engineering contributes to neuroscience. Efforts like the one presented in this study carry the potential not only for innovative engineering applications, but also for fostering our understanding of the way sensory information is processed in the brain.

## Acknowledgment

This work was supported by the National Science Foundation grant CCF-0130851.

## References

- [1] N. Ayache, *Artificial Vision for Mobile Robots: Stereo Vision and Multisensory Perception*, MIT Press, Cambridge, MA, USA, 1991.
- [2] T. Huang, A. Netravali, Motion and structure from feature correspondences: A review, *Proceedings of the IEEE* 82 (2) (1994) 252–268.
- [3] A. Mitiche, *Computational Analysis of Visual Motion*, Plenum Press, New York, London, 1994.
- [4] A. Mitiche, J. Letang, Stereokinematic analysis of visual data in active convergent stereoscopy, *Robotics and Autonomous Systems* 25 (1998) 43–71.
- [5] O. Faugeras, *The Geometry of Multiple Images: The Laws that Govern the Formation of Multiple Images of a Scene and Some of Their Applications*, MIT Press, Cambridge, MA, USA, 2001.
- [6] R. Hartley, A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, Cambridge, England, 1994.
- [7] J. Weng, T. Huang, N. Ahuja, *Motion and Structure From Image Sequences*, Springer-Verlag, New York, USA, 1993.
- [8] J. Oliensis, A critique of structure-from-motion algorithms, *Computer Vision and Image Understanding* 80 (2) (2000) 172–214.
- [9] A. Pentland, A new sense for depth of field, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 9 (1987) 523–531.
- [10] B. Super, A. Bovik, Shape from texture using local spectral moments, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 17 (4) (1995) 333–343.
- [11] R. Zhang, P. Tsai, J. Cryer, M. Shah, Shape from shading: A survey, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21 (8) (1999) 690–706.
- [12] A. Torralba, A. Oliva, Depth estimation from image structure, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24 (9) (2002) 1226–1238.
- [13] I. Howard, B. Rogers, *Seeing in Depth*, I Porteous, Thornhill, Ontario, Canada, 2002.
- [14] G. Westheimer, S. McKee, Spatial configurations for visual hyperacuity, *Vision Research* 17 (1977) 941–947.
- [15] Y. Trotter, S. Celebrini, B. Stricanne, S. Thorpe, M. Imbert, Neural processing of stereopsis as a function of viewing distance in primate visual cortical area V1, *Journal of Neurophysiology* 76 (5) (1996) 2872–2885.
- [16] K. Schreiber, J.D. Crawford, M. Fetter, D. Tweed, The motor side of depth vision, *Nature* 410 (6830) (2001) 819–822.
- [17] K.M. Schreiber, D.B. Tweed, Influence of eye position on stereo matching, *Strabismus* 11 (1) (2003) 9–16.
- [18] M. Nawrot, Viewing distance, eye movements, and the perception of relative depth from motion parallax, *Investigative Ophthalmology & Visual Science* 41 (4) (2000) S45.
- [19] M. Nawrot, Eye movements provide the extra-retinal signal required for the perception of depth from motion parallax, *Vision Research* 43 (14) (2003) 1553–1562.
- [20] J.M. Foley, W. Richards, Effects of voluntary eye movement and convergence on the binocular appreciation of depth, *Perception Psychophysics* 11 (1972) 423–427.
- [21] H. von Helmholtz, in: J.P.C. Southall (Ed.), *Treatise on Physiological Optics* 1909, Dover, New York, USA, 1962.
- [22] E. Gibson, J. Gibson, O. Smith, H. Flock, Motion parallax as a determinant of perceived depth, *Journal of Experimental Psychology* 58 (1959) 40–51.
- [23] M. Land, Fast-focus telephoto eye, *Nature* 373 (1995) 658–659.
- [24] J. Pettigrew, S. Collin, M. Ott, Convergence of specialised behaviour, eye movements and visual optics in the sandlance (teleostei) and the chameleon (reptilia), *Current Biology* 9 (8) (1999) 421–424.
- [25] I. Hadami, G. Ishai, M. Gur, Visual stability and space perception in monocular vision: Mathematical model, *Journal of the Optical Society of America A* 70 (1980) 60–65.
- [26] A. Mapp, H. Ono, The rhino-optical phenomenon: Ocular parallax and the visible field beyond the nose, *Vision Research* 26 (1986) 1163–1165.
- [27] G. Bingham, Optical flow from eye movement with head immobilized: Ocular occlusion beyond the nose, *Vision Research* 33 (5–6) (1993) 777–789.
- [28] A. Bahill, D. Adler, L. Stark, Most naturally occurring human saccades have magnitudes of 15 degrees or less, *Investigative Ophthalmology* 14 (6) (1975) 468–469.
- [29] R. Jarvis, A perspective on range-finding techniques for computer vision, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 5 (1983) 122–139.
- [30] Y. Aloimonos, I. Weiss, A. Bandyopadhyay, Active vision, *International Journal of Computer Vision* 2 (1988) 333–356.
- [31] R. Bajcsy, Active perception, *Proceedings of the IEEE* 76 (8) (1988) 996–1005.
- [32] D. Ballard, Animate vision, *Artificial Intelligence* 48 (1991) 57–86.
- [33] W. Klarquist, A. Bovik, FOVEA: A foveated vergent active stereo vision system for dynamic three-dimensional scene recovery, *IEEE Transactions on Robotics and Automation* 14 (5) (1998) 755–770.
- [34] M. Bjorkman, J. Eklundh, Real-time epipolar geometry estimation of binocular stereo heads, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24 (3) (2002) 425–432.
- [35] A. Rajagopalan, Depth estimation and image restoration using defocused stereo pairs, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26 (11) (2004) 1521–1525.
- [36] G. Adiv, Determining three-dimensional motion and structure from optical flow generated by several moving objects, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 7 (4) (1985) 384–401.
- [37] G. Sandini, M. Tistarelli, Active tracking strategy for monocular depth inference over multiple frames, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12 (1990) 13–27.
- [38] Y. Aloimonos, Z. Duric, Estimating the heading direction using normal flow, *International Journal of Computer Vision* 13 (1994) 33–56.
- [39] A. Dalmia, M. Trivedi, High-speed extraction of 3D structure of selectable quality using a translating camera, *Computer Vision and Image Understanding* 64 (1) (1996) 97–110.
- [40] Y. Hung, H. Ho, A Kalman filter approach to direct depth estimation incorporating surface structure, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21 (6) (1999) 570–575.
- [41] A. Gullstrand, *Appendices to Part 1: The Optical System of the Eye*, Voss, Hamburg, Germany, 1909, pp. 350–358.
- [42] P. Mouroulis, J. MacDonald, *Geometrical Optics and Optical Design*, Oxford University Press, Oxford, England, 1997.
- [43] B. Cumming, An unexpected specialization for horizontal disparity in primate primary visual cortex, *Nature* 418 (2002) 633–636.
- [44] H. Crane, C. Steele, Generation V dual-purkinje-image eyetracker, *Applied Optics* 24 (4) (1985) 527–537.
- [45] I. Murakami, P. Cavanagh, Stabilizing the visual world: A jitter aftereffect reveals motion-based correction, *Nature* 395 (1998) 798–801.
- [46] M. Rucci, G. Edelman, J. Wray, Modeling LGN responses during free-viewing: A possible role of microscopic eye movements in the refinement of cortical orientation selectivity, *Journal of Neuroscience* 20 (12) (2000) 4708–4720.
- [47] M. Rucci, G. Desbordes, Contributions of fixational eye movements to the discrimination of briefly presented stimuli, *Journal of Vision* 3 (11) (2003) 852–864.
- [48] M. Rucci, A. Casile, Decorrelation of neural activity during fixational instability: Possible implications for the refinement of V1 receptive fields, *Visual Neuroscience* 21 (2004) 725–738.
- [49] R. Steinman, G. Haddad, A. Skavenski, D. Wyman, Miniature eye movement, *Science* 181 (1973) 810–819.
- [50] R. Ditchburn, *Eye-movements and Visual Perception*, Oxford University Press, Oxford, England, 1973.
- [51] M. Lehrer, M.V. Srinivasan, Active vision in honeybees: Task-oriented suppression of an innate behaviour, *Vision Research* 34 (4) (1994) 511–516.
- [52] M.S. Dawkins, A. Woodington, Pattern recognition and active vision in chickens, *Nature* 403 (6770) (2000) 652–655.
- [53] J. Findlay, Active vision: visual activity in everyday life, *Current Biology* 8 (18) (1998) 640–642.

- [54] Y. Lu, J. Zhang, Q. Wu, Z. Li, A survey of motion-parallax-based 3-D reconstruction algorithms, *IEEE Transactions on System, Man and Cybernetics C* 34 (4) (2004) 532–548.
- [55] M. Rucci, G. Edelman, J. Wray, Adaptation of orienting behavior: From the barn owl to a robotic system, *IEEE Transactions on Robotics and Automation* 15 (1) (1999) 96–110.
- [56] M. Rucci, J. Wray, G. Edelman, Robust localization of auditory and visual targets in a robotic barn owl, *Robotics and Autonomous Systems* 1–2 (2000) 181–194.



**Fabrizio Santini** received his M.Sc. degree in Computer Science Engineering from the University of Rome “La Sapienza”, Italy in 1999, and his Ph.D. degree in Computer Science and Robotics from the University of Florence, Italy in 2004. In 2002 he joined the Active Perception Laboratory at Boston University, first as a Visiting Scholar and then as a Research Associate. At Boston University, Dr. Santini has been working on the development of a humanoid robot with a vision system that models the brain of primates. Dr. Santini’s research

interests include biomimetic robotics, machine learning, and evolutionary hardware.



**Michele Rucci** is a faculty member of the Department of Cognitive and Neural Systems at Boston University. He received his Laurea and Ph.D. degrees in bioengineering from the University of Florence and the Scuola Superiore S. Anna in Pisa. Before joining Boston University, he was a Fellow in Computational Neuroscience at the Neurosciences Institute in San Diego. At Boston University, Dr. Rucci founded and directs the Active Perception Laboratory, a facility dedicated to the analysis of the interactions between perception and behavior. Dr. Rucci’s research follows an interdisciplinary approach that integrates experiments in visual psychophysics with computational models of the brain and the embodiment of neuronal models in robotic systems. This research has raised specific hypotheses regarding the influences of eye movements during visual development, has produced a theory of the role of fixational eye movements in the neural encoding of visual information, and has led to the development of robots directly controlled by models of the brain.