Active Estimation of Distance in a Robotic Vision System that Replicates Human Eye Movement
Active estimation of distance in a robotic vision system that replicates human eye movement

Fabrizio Santini and Michele Rucci

January, 2005

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

Fabrizio Santini and Michele Rucci

Department of Cognitive and Neural Systems and Center for Adaptive Systems
Boston University

Submitted: January, 2005

All correspondence should be addressed to:

Prof. Michele Rucci
Active Perception Lab
Department of Cognitive and Neural Systems
Boston University
677 Beacon Street
Boston, MA 02215
Phone: (617) 358-1385
Fax: (617) 358-7755
E-mail: rucci@cns.bu.edu
Active estimation of distance in a robotic vision system that replicates human eye movement

Fabrizio Santini and Michele Rucci

Department of Cognitive and Neural Systems, Boston University, Boston, MA 02215.

Abstract

Many visual cues, both binocular and monocular, provide 3D information. When an agent moves with respect to a scene, an important cue is the different motion of objects located at various distances. While a motion parallax is evident for large translations of the agent, in most head/eye systems a small parallax occurs also during rotations of the cameras. A similar parallax is present also in the human eye. During a relocation of gaze, the shift in the retinal projection of an object depends not only on the amplitude of the movement, but also on the distance of the object with respect to the observer. This study proposes a method for estimating distance on the basis of the parallax that emerges from rotations of a camera. A pan/tilt system specifically designed to reproduce the oculomotor parallax present in the human eye was used to replicate the oculomotor strategy by which humans scan visual scenes. We show that the oculomotor parallax provides accurate estimation of distance during sequences of eye movements. In a system that actively scans a visual scene, challenging tasks such as image segmentation and figure/ground segregation greatly benefit from this cue.

Index Terms

Biomorphic robotics, saccade, stereopsis, head/eye system, parallax.

I. INTRODUCTION

A critical function of a visual system is to estimate the distance at which objects and surfaces are located with respect to the agent. Accurate estimation of distance is not only crucial for operations involving interactions with the environment, such as navigation and manipulation, it also provides a fundamental contribution to the segmentation of images into their basic elements, one of the most challenging tasks of machine vision.

The process of reconstructing a 3-D scene from its projections on a 2-D sensor is inherently ambiguous. To solve this ambiguity, vision systems usually rely on cues that originate from the comparison of images taken either at different points of view (binocular vision), or at different instants of time (target movements, egomotion, etc.) [1], [2], [3], as well as from a priori knowledge of the scene and its structure (size, occlusions, shadows, etc.) [4], [5], [6], [7], [8], [9].

Not surprisingly, in nature, where precise depth perception makes the difference between capturing a prey and failing to survive, many species exhibit a striking accuracy in estimating visual distance. In humans, depth perception relies on the integration of a variety cues, each of them measured with exquisite accuracy. For example, stereoaucuity, which is at the basis of the perception of depth from the slightly different views of a three-dimensional scene received by the two eyes, reaches thresholds as low as 3” [10], [11]. That is, the human brain compares the relative locations of features in both eyes with a precision of about one tenth of a photoreceptor diameter.

As it is the case for many visual functions, in biological vision systems an important feature of depth perception is the close interaction between motor and sensory processes. Behavior contributes to depth perception in several ways. It provides visual cues of distance, as it is the case for the motion parallax, i.e. the way the experienced relative positions of stationary objects change as an agent navigates through the environment (von Helmholtz, 1909/1962 [12], Gibson, 1950 [13]). It also provides proprioceptive cues, as in the case of vergence eye movements [14]. Furthermore, several studies have shown that oculomotor activity simplifies the visual computations involved with the extraction and use of both binocular [15], [16], [17] and monocular depth cues [18], [19].
In machine perception, a tight link between behavior and perception appears to be one of the key factors for the development of more efficient algorithms [20], [21]. Several studies in robotics have followed an active approach to the estimation of distance. While a number of techniques have been proposed [22], [23], most studies have focused on the parallax produced by egomotion, a method often indicated as depth from motion [3], [2], [24], [1], [25].

Previous studies on depth from motion have selectively focused on large translations of a camera mounted on a mobile robot. However, a motion parallax is present also during rotations of a camera if rotation axes do not intersect the focal points. In a robotic head, unless special care is taken in the design of the optical and mechanical characteristics of the system, the projection of a target on the sensor moves by an amount that depends not only on the amplitude of the rotation, but also on the distance of the target with respect to the camera. In the eyes of many species, a similar parallax is produced by the rotations associated with eye movements (see Fig. 1). Species as diverse as the chameleon and the sandlance use this oculomotor parallax to judge distance [26]. In these species, the optics of the cornea and lens maximize the distance between nodal points and the center of rotation. Although smaller than in the chameleon, an oculomotor parallax is present also in the human eye and produces retinal shifts that are well within the range of visual acuity [27], [28], [29]. This study examines the use of the oculomotor parallax for the visual estimation of distance. We consider the case of a robotic system specifically designed to replicate the parallax present in the human eye during oculomotor activity. We show that when visual scenes are examined in a way that replicates the scanning strategy of human subjects, the emerging oculomotor parallax provides reliable 3D information within a range of nearby distances.

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 the robot camera (planar sensor) are analyzed in section II. Section III describes the oculomotor robotic system used in this study, its optical and mechanical properties, and the calibration process that was followed for precisely positioning the cardinal points. Section IV 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 follows.

![Diagram](image)

Fig. 1. Distance information provided by the oculomotor parallax. In the human eye, 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 inserts show that the two projections are coincident before but not after the rotation. For clarity the case of a simple lens with a single nodal point is shown.
II. THE OCOLOMOTOR 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 the 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.

A. Spherical sensor

This section describes an approach to the estimate of distance based on the parallax that emerges from a rotation of the eye along the pan axis. Since the cardinal points of the model described below are located on the visual axis $z$, the same method can be applied to vertical displacements produced by rotations along the tilt axis.

In the human eye, the oculomotor parallax can be studied by means of Gullstrand’s schematic eye model [30] (see Fig. 2a). To ensure that parallel rays entering the eye focus perfectly on the retina, this model assumes a spherical ocular bulb with radius of 11 mm, a lens with average index of refraction equal to 1.413, and an axial distance from the cornea to the retina of 24.4 mm [29]. 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, we consider the case of an infinitesimally small light source that produces a projection of a single point on the retina. As shown in Fig. 2b, we assume this Pinpoint Light Source (PLS) to be located in front of the sensor at position $A = (-d_A \sin \alpha, d_A \cos \alpha)$ [1]. In the case of an optical system with a single nodal point, the retinal projection of the PLS, $\hat{x}_A$, would be given by the intersection between the sensor surface and the line $AN_1$ that goes through $A$ and $N_1$. This line verifies the equation:

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

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 $\hat{\alpha}$ exits the lens by the second nodal point, $N_2$, with the same angle [31]. 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 equation:

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

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:

$$R \sin (\theta + \hat{\alpha}) = \frac{|-d_A \sin \alpha|}{\sqrt{CN_1^2 + d_A^2 - 2CN_1d_A \cos \alpha}} \quad (3)$$

where $\theta$ is the angle that identifies the projection of the PLS on the sensor, and $\hat{\alpha}$ (the angle between the line $AN_1$ and the axis $z$) determines the eccentricity of the PLS:

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

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

$$\theta = f(d_A, \alpha) = \arcsin \left( \frac{|-d_A \sin \alpha|}{R \sqrt{CN_1^2 + d_A^2 - 2CN_1d_A \cos \alpha}} \right) + \arctan \left( \frac{d_A \sin \alpha}{d_A \cos \alpha - CN_1} \right) \quad (5)$$

1 All distances are evaluated with respect to the center of rotation $C$ of the mobile camera.
(5) shows that the projection \( \theta \) of the PLS on the sensor depends on the PLS distance 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 brought the PLS from position \( A' \) on the optical axis to its current position \( A \) (see Fig. 2b).

![Diagram](image)

**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) Geometrical assumptions and axis orientation in a semi-spherical sensor.

Given a measured projection \( \theta \), there are an infinite number of possible PLS locations that verify (5). Every point on the line \( AN_1 \) yields the same projection, although separate points differ for their distances \( (d_A) \) and eccentricities \( (\alpha) \). Whereas the location of the retinal projection 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 the PLS in space. This method, explained in Fig. 3, relies on the change in \( \theta \) with a rotation of the sensor by an angle \( \Delta \alpha \).

Let \( \theta \) and \( \theta' \) indicate the projections of the PLS on the sensor before and after a rotation \( \Delta \alpha \). For each of the two measurements, (5) establishes a relationship between possible values of PLS eccentricity \( \alpha \) and distance \( d_A \). Although this relationship cannot be expressed analytically, it can be shown to be continuous and monotonic. Fig. 3a provides an example of the two curves verifying (5) before and after a rotation of 3°. These graphs were obtained by simulating the projection of a PLS located at an eccentricity of 2° and a distance of 600 mm from the center of rotation of the camera. Each curve represents all possible couple of values distance-eccentricity that would produce the measured projection either before or after the rotation of the sensor. 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 actual spatial location of the PLS.

Fig. 3b shows the accuracy of the method. The distance estimated by evaluating the intersections of the two curves in Fig. 3a is plotted as a function of the actual distance of the PLS. Data are results of simulations in which the sensor was rotated by 3°. To simulate measurement errors, a Gaussian noise with zero mean and 0.5 mm std was superimposed to the shift in the PLS projection generated by the rotation of the sensor. Each data point summarizes averages obtained over 500 repetitions of the experiment. It is evident that the proposed method provides accurate estimation within a nearby range of distances. Since the amplitude of the oculomotor parallax decreases with the distance of a target, measurement errors resulted in a larger variability of localization as the distance of the PLS increased. This variability limited the application of the method to distances of up to 1 m, a range that finds application in a variety of important tasks in robotics.
Fig. 3. Distance estimation on the basis of the oculomotor parallax. Case of a spherical sensor. (a) A rotation of the sensor disambiguates the target position. The two curves represent the ensembles of possible spatial positions of the target that produce the projections measured in the images acquired before (thick line) and after (thin line) a 3° rotation. The point of intersection D gives the distance and eccentricity of the target. (b) Accuracy of the method in simulations that rotated the sensor by 3°. Data were obtained with the parameters of Gulstrand's eye model. Error bars represent one standard deviation.

B. Planar sensor

The dissimilar geometry 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. The geometry of the oculomotor parallax for the case of a planar sensor 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 the position of a PLS. Similar to the case of the spherical sensor (see (2)), the PLS projection on receptor array, $x_A$, is given by the intersection of the sensor plane and the line parallel to $AN_1$ going through $N_2$:

$$x_A = \frac{d_A f \sin \alpha}{d_A \cos \alpha - CN_1}$$  \hspace{1cm} (6)

where $f = d_C + CN_2$ represents the focal length of the lens. As shown by (6), the projection $x_A$ exhibits a dependence on the distance $d_A$ of the PLS. This dependence is at the basis of the oculomotor parallax.

In the case of a planar sensor, the PLS distance $d_A$ can be extracted from (6) and explicitly expressed as a function of the PLS projection $x_A$ on the sensor and the eccentricity $\alpha$ (the angle between $AC$ and the optical axis):

$$d_A = \frac{x_A CN_1}{x_A \cos \alpha - f \sin \alpha}$$  \hspace{1cm} (7)

Similar to the case of a spherical sensor, the position of the PLS cannot be recovered from a single image, as any point located on the line $AN_1$ would produce the same projection $x_A$. To recover the eccentricity $\alpha$ and distance $(d_A)$ of the PLS, we can follow the same approach used in the previous section. Since a rotation around $C$ does not change the PLS distance $d_A$, we can equate the two estimates of distance obtained from (7) before and after a rotation of known amplitude $\Delta \alpha$:

$$\frac{x_A CN_1}{x_A \cos \alpha - f \sin \alpha} = \frac{x'_A CN_1}{x'_A \cos(\alpha + \Delta \alpha) - f \sin(\alpha + \Delta \alpha)}$$  \hspace{1cm} (8)

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

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

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

Fig. 4. Oculomotor parallax in a planar sensor. (a) Cardinal points of the oculomotor system. \( N_1 \) and \( N_2 \) represent the position of the two nodal points in the lens. \( C \) is the center of rotation. \( S \) represents the sensor plane. (b) Geometrical assumptions and axis orientation for a planar sensor. \( A \) represents the PLS position. It is located at a distance \( d_A \) from the center of rotation \( C \) and an eccentricity \( \alpha \). \( x_A \) is the location of the PLS projection on the sensor.

Fig. 5 shows the results of simulations in which this method was applied to estimating the distance of a PLS. The two panels show results obtained for rotations with two different amplitudes (1° and 3°). In these simulations, given a PLS position in space, its projection on the sensor before and after the rotation was measured by means of (6). The PLS distance and eccentricity were estimated by means of (9) and (7). To examine the impact of measurement errors on distance estimation, random Gaussian noise with zero mean and 0.5 mm std was superimposed to the shift in the projection produced by the rotation. Each data point summarizes averages obtained over 500 repetitions of the experiment. To replicate the oculomotor parallax present in the human eye, parameters of the simulation were \( CN_1 = 49.65 \text{ mm} \), and \( CN_2 = 5.72 \text{ mm} \). The data in Fig. 5 shows that the oculomotor parallax provided accurate distance estimation for object located up to 1 m from the camera even with relatively small rotations.

Fig. 5. Distance estimation on the basis of the oculomotor parallax. Case of a planar sensor. Data are results from simulations in which a camera was rotated by 1° (a)) and 3° (b). See text for details.
III. AN OCULOMOTOR ROBOTIC SYSTEM

The experiments of this study were executed by means of the robotic head/eye system shown in Fig. 6a. This system, indicated in the following of this paper as the robotic oculomotor system, was developed to replicate the visual input signals that occur in the human retina during eye movements. It consisted of two mobile cameras, each with two degrees of freedom. Motion was provided by two pan-tilt units (Directed Perception Inc. - CA) supported by a custom designed aluminum frame. Units were digitally controlled by proprietary microprocessors that ensured movements with a precision higher than 1 arcmin. This degree of accuracy is of the same order of the spatial resolution of the eye-tracker that was used to acquire the sequences of eye movements that controlled the robot (see Sec. IV-B). By means of a nodal adapter, units were mounted so that the two axes of rotation, pan and tilt, intersected at a point C (the center of rotation, see Fig. 6b). Specifically designed aluminum wings enabled positioning of the center of rotation between the sensor plane S and the nodal points of the camera as it occurs in the human eye. Each of the two wings possessed a sliding bar with calibrated holes that allowed the distance between the center of rotation C and the sensor S to be varied in discrete steps. The system was equipped with two digital cameras (Pulnix Inc. - CA) each with a 11.5 - 69 mm zoom lens. Cameras possessed a 640 x 484 CCD sensors and allowed an acquisition rate of 120 frames/s. The dimension of the photoreceptors on the sensor was 9 µm. Images were acquired by means of a fast frame-grabber (Datacube Inc. - MA). Only one of the two mobile cameras was used in the experiments described in this paper.

In a zoom lens, the exact positions of the nodal points depend on the focal length. To replicate the oculomotor parallax present in the human eye, we selected a focal length value that positioned the nodal points in a way that precisely replicated the arrangement present in the eye. Selection of this focal length was achieved by means of a preliminary calibration procedure, in which we fit the model described in Sec. II-B to a set of data acquired by the system. For several focal lengths of the zoom lens, a target was placed at systematically varied distances and rotations of various amplitudes were performed. The shifts in the projections of the target in images acquired before and after the rotations were measured. Model fitting was performed by regularized least squares interpolation using (6). This led to an estimate of the coordinates of N1 and N2 for each considered focal length.

As illustrated in Fig. 6c, the calibration procedure 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 with a typical power emission of 2000 mcd) covered by a pierced steel mask with a 3 mm hole, in order to minimize bright aberrations due to possible LED cap imperfections as well as distortions of the PLS shape with the point of view. The PLS model was mounted on an calibrated aluminum bar that allowed precise control of the distance at which the LED was located with respect to the center of rotation C. In the calibration procedure, the PLS distance was varied systematically between 270 mm and 1120 mm, with intervals of 50 mm. The lens focus was kept at infinite as in the model of Gullstrand's eye.

Fig. 6d shows the positions of the two nodal points estimated by means of this calibration procedure. The distances CN1 and CN2 varied almost linearly with the focal length of the zoom lens. The focal length used in the experiments was selected on the basis of these data. Fig. 6d shows that N1 was closer to the sensor than N2 for focal lengths greater than 27.7 mm. Since in the human eye N2 is the nodal point closer to the retina, we constrained the focal length to be smaller than 27.7 mm. In both semi-spherical and planar sensors, the projection of the stimulus on the receptor surface is strongly sensitive to the position of N2 and relatively unaffected by the location of N1 (notice how in both (6) and (5), CN1 appears at the denominator together with the much larger term dα). For this reason, we set the location of N2 as close as possible to the position CN2 = 6.03 mm specified in Gullstrand's eye model. Selection of a focal length f = 10.24 mm yielded CN2 = 5.72 mm thus producing an oculomotor parallax very similar to that present in the human eye (CN1 was 49.65 mm).

IV. EXPERIMENTAL RESULTS

Two sets of experiments were conducted in this study. First, to rigorously examine the accuracy of the proposed method, we analyze the estimation of distance provided by rotations of the camera with predefined amplitudes. As an example of application of the approach, we show how a camouflaged object can be disambiguated from the background with a single rotation of the sensor. A second series of experiments examined the depth information provided by sequences of eye movements recorded from human subjects. We show that the range of amplitude
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 (N1 and N2), center of rotation (C) and sensor surface (S) followed the arrangement of the eye. The distance between C and S was adjusted by proper positioning of the cameras. Careful selection of the lens allowed precise tuning of the positions of N1 and N2. (c) Selection of the lens focal length. In a preliminary calibration procedure, the shifts in the projection of a PLS model located at various distances were measured for rotations of different amplitudes $\Delta \alpha$. (d) Positions of N1 and N2 for various focal lengths of a 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 provided by factory specifications.

of the saccades that characterize natural viewing provide ideal conditions to estimate the distances of objects composing a complex scene.

A. Distance estimation with a single camera rotation

We first examined the oculomotor parallax provided by rotations of the sensor with predetermined amplitudes. These experiments were conducted using the same experimental set up of the preliminary calibration procedure. The PLS model was placed a various distances from the robot. The distance of the PLS was estimated on the basis of its oculomotor parallax following the approach described in Sec. II. Gray-level images acquired before and after the rotation of the camera were transformed into binary values by means of thresholding, and the positions of the PLS centroid were calculated. The centroid locations in the two images, together with the rotation amplitude $\Delta \alpha$, were used to estimate the distance of the PLS by means of (7) and (9).

Fig. 7 shows typical results obtained in these experiments. The PLS distance estimated on the basis of the oculomotor parallax is plotted as a function of its actual distance for rotations of amplitude 2.5° and 5°. Data
represent the mean distance estimated over 25 measurements. Similar to the results of simulations (see Fig. 5), accurate localization was obtained for PLS distances up to 1 m. Again, the estimation error increased with the PLS distance and decreased with the rotation amplitude. The main difference with respect to simulation results was the staircase appearance of the graphs, which originated from the quantization error introduced by the finite size of pixels in the CCD sensor of the camera.

![Graphs showing distance estimation](image)

**Fig. 7.** Distance estimation on the basis of the oculomotor parallax. Data are estimates obtained with a single rotation of the camera. The rotation amplitude was different in the two panels: (a) 1°; (b) 3°. Error bars represent one standard deviation.

To examine the impact of quantization errors on system performance, Fig. 8a compares the displacement in the position of the PLS centroid measured in the experiments to the shift in target projection predicted by simulations of the planar model. No quantization error was present in the simulations, since the size of the simulated photoreceptor was infinitesimally small. Measurements obtained for PLSs at various distances and a rotation of 3° are shown. As illustrated by these data, the effect of pixel quantization became more pronounced for objects located at larger distances from the robot where the discrete steps of single pixel increments are clearly visible. Due to this error, the estimation of distance based on oculomotor parallax cannot be applied to distant objects, for which the oculomotor parallax moves the centroid by less than a photoreceptor. Fig. 8b shows the accuracy of localization obtained in simulations that modeled cameras with two different pixel dimensions, 9 μm and 0.3 μm. As for Fig. 8a, results refer to a rotation of 3°. It is clear that the smaller the pixel size, the wider the range of distances for which the oculomotor parallax provides accurate estimation. A receptor of 0.3 μm replicates the acuity of human visual system, which goes far beyond the actual size of the retinal receptor [32], [33]. As shown by Fig. 8b such a small receptor extends the range of distance to which the method can be applied to a factor of 23.

**Fig. 9.** Example of application of the distance information provided by the oculomotor parallax. In this example, a rotation of the camera was used to disambiguate an object camouflaged with a background. As illustrated in Fig. 9a, a rectangular panel was positioned in between the robot and a flat background. The background portrayed a reproduction of Van Gogh’s painting “Irises” located at distance 900 mm. 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 with the background. A bright illumination allowed the use of a small aperture in the camera’s shutter thus keeping both the panel and the background in focus. In this way, as shown in Fig. 9b, the panel edges were almost completely absent and the panel was virtually invisible in the images acquired by the camera.

Fig. 9c shows the estimated distance for a number of points in the image following a camera rotation of Δα = 3°. In this example, the panel was located at a distance of 700 mm from the robot. The image taken before the rotation was decomposed into an array of 45 × 45 patches, each of 14 × 11 pixels. The oculomotor parallax of each patch was estimated by means of cross-correlation with the image acquired after the camera
rotation. For each image patch, an estimate of distance was obtained from the measured oculomotor parallax. As shown by Fig. 9c, a single rotation of the camera was sufficient to expose the camouflaged panel. Points of the image corresponding to the panel gave different distances than background pixels. Fig. 9d summarizes the results obtained with panels located at three different distances. Each data point represents the average distance evaluated over 10 \times 25 patches, each of 14 \times 11 pixels, composing the panel. As shown by these data, accurate identification of the panel distance was obtained with a 3° rotation.

B. Distance estimation with sequences of eye movements

In a second series of experiments, we examined the distance information provided by the oculomotor parallax in sequences of rotations that replicated traces of eye movements. The eye movements of two human subjects (MV and AR) were recorded by means of a Dual Purkinje Image (DPI) eye-tracker (Fourward Technologies, Inc., CA). This high-resolution device estimates rotations of the eye by measuring differences in the first and fourth reflections (the Purkinje images) of an infrared beam on the eye. It achieves a spatial and temporal resolution of 1 arcmin and 1 ms [34], [35]. As shown in Fig. 10, subjects viewed the scene monocularly with the right eye, while the left eye was covered by an opaque eye-patch. Eye movements were recorded for a period of 10 s while subjects observed the same scene that was later presented to the robot. Eye movement data were first low-pass filtered, and then sampled at 1 kHz and recorded by an analog/digital sampling board (Measurement Computing Corp., MA).

Traces of recorded eye movements were later used as motor commands for the robot. A preliminary calibration in which subjects were asked to fixate on a number of predetermined points in the scene was used to find a linear correspondence between voltages provided by the eye tracker and motor signals fed to the robot. This calibration ensured that the camera aimed in such a way that the center of the image corresponded to the point fixated by the subject. In the experiments, the speed of the eye movements was reduced in order to operate within the range of velocities that the robot could achieve reliably, without jeopardizing the spatial accuracy with which eye movements were reproduced.

The high spatial sensitivity of the DPI eye-tracker 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
Fig. 9. Application of the oculomotor parallax to a figure/ground segregation task. (a) Experimental set up. A rectangular panel O is located in between the robot and a textured background. (b) The panel contained an enlarged patch of the same texture of the background. The enlargement was carefully made to eliminate edges and achieve a perfect camouflage. (c) Distance estimation on the basis of the oculomotor parallax. Each point represent the estimated distance of a $14 \times 11$ pixels patch in the images acquired by the robot. The panel was at a distance of 700 mm. (d) Results obtained for distances of the panel. Each data point represents the average distance of the patches composing the panel. Error bars represent one standard deviation.

movements continuously occur even when the eye is fixating on a target\textsuperscript{2} [36], [37]. Fig. 10b shows a trace of recorded eye movements. Both macroscopic saccades and fixational eye movements are present.

Fig. 11 shows an example of distance estimation based on the oculomotor parallax with real sequences of eye movements. In this experiment, the robot replicated the eye movements recorded from the two subjects while viewing the scene shown in Fig. 11a. This scene was composed of 5 objects located at various distances (see Fig. 11b). The similar colors and textures of several of the objects as well as the occlusions present from the point of observation made segmentation of the scene difficult. Only macroscopic saccades were used in this experiment. That is, recorded eye traces were filtered to eliminate all non-saccadic movements and all saccades smaller than $1^\circ$. This led to the selection of, respectively, 11 and 9 saccades for subject MV and AR. The mean amplitude of saccades was $1.82^\circ \pm 1.05^\circ$ for subject MV and $4.76^\circ \pm 2.60^\circ$ for subject AR. Selected saccades were used as motor commands for the robot and pairs of images were acquired before and after each saccade.

\textsuperscript{2}We are usually not aware of the eye movements that occur during visual fixation. A striking demonstration of fixational instability is available on the World Wide Web at http://www.visionlab.harvard.edu/Members/kuyu/html-memorandum/VisualJitter.html (Murakami and Cavanagh, 1998).
Figs. 11c and 11d show results obtained with sequences of eye movements from the two subjects. As before, the oculomotor parallaxes of selected points in the scene were evaluated by normalized cross-correlation of 14 x 11 pixel patches. The measured oculomotor parallax was then converted into an estimate of distance. As shown by these data, different objects produced different parallaxes, and the distances of the various objects were accurately estimated.

The results of Fig. 11 were obtained using relatively large saccades. These eye movements did not include the smaller saccades that occur during visual fixation. Although small, fixational saccades move the stimulus on the retina by tens to hundreds of photoreceptors. These are shifts of substantial amplitude in consideration that human acuity goes below a single photoreceptor [10], [38]. By virtue of their small amplitudes, fixational saccades introduce little distortion in the appearance of object features. This, together with the bound in the search space, facilitates the solution of the correspondence problem, i.e. the matching of features in images acquired before and after each rotation. For these reasons, in a separate experiment, we examined the information of distance provided by small fixational saccades.

Since the relatively large dimension of the pixel in the cameras CCD sensor did not allow a fine discrimination of oculomotor parallaxes, fixational saccades were amplified by a factor that compensated for the difference between the size of the pixel and human visual acuity. With this application factor, the stimulus moved by the same amount of photoreceptors in the human eye and in the camera:

$$A_f = \frac{S_p}{S_x} \left[ \frac{(R + N_{2a})^2}{R (d_C + N_{2a})} \right]$$

(10)

where $S_x$ and $S_p$ are, respectively, the dimension of the photoreceptor on the semi-spherical sensor and the dimension of the photoreceptor cell on the planar sensor; $N_{2a}$ and $N_{2a}$ are the position of the second nodal point in the optical system of the camera and the position of the second nodal point in Gullstrand's eye model (see Figure 4a); $R$ and $d_C$ are the radius of the semi-spherical sensor and the distance of the planar sensor with the center of rotation of the camera. Use $S_p = 9 \ \mu m$ and $S_x = 7.93 \ \mu m$ (the dimension of a virtual photoreceptor that takes in account hyperacuity) together with the values reported for the Gullstrand's model (see Paragraph II-A) produced an amplification ratio $A_f = 1.89$.

Fig. 12 shows results obtained by replicating fixational saccades. Traces of oculomotor activity were recorded while the two subjects maintained visual fixation for a period of 3 s. Ten traces were acquired for each subject. Fixational saccades were identified, amplified by the amplification ratio $A_f$, and used to control the robot. The total number of fixational saccades were 30 (subject MV) and 29 (subject AR). Their amplitude was
Fig. 11. Estimating distance with sequences of eye movements. Case of macroscopic saccades (saccades larger than 1 deg). (Top Row) Experimental setup. (a) Several objects were placed at various distances in front of the robot to compose a complex scene. (b) Top view of the scene. The distances of the various objects were: A = 630 mm; B = 590 mm; C = 640 mm; D = 740 mm; E = 880 mm. (Bottom Row) Mean estimated distance of the pixels composing the objects. The two panels show results obtained with sequences of eye movements recorded from subject MV (c) and AR (d). Error bars represent one standard deviation. The distance of C could not be estimated with AR’s eye-movement trace as it did not appear within the field of view of the camera (see Fig. 13(c) and (e)).

0.15° ± 0.05° for subject MV and 0.47° ± 0.19° for subject AR. In this experiment, fixational saccades were used to estimate the distance of the PLS model. For each PLS distance, the same sequence of saccades was executed. As illustrated by the data in Fig. 12, also the small eye movements of visual fixation produce an oculomotor parallax that is sufficient to accurately estimate distance.

Macroscopic and fixational saccades can be used together to obtain more reliable measurements 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. 11b (29 saccades for the subject MV and 29 saccades for the subject AR). In this experiment, the camera followed the recorded trajectories switching among fixation point through large saccades. At each fixation point, fixational saccades were amplified by using the amplification ratio $A_f$. This motor strategy was a trade-off between the need of carefully following the sequences of fixation points selected by the subjects and compensating for the lower resolution of the camera with respect to that of the human eye. As shown in Fig. 13, the accuracy of distance estimation improved when both large and small saccades were used, and measurements resulted in a smaller standard deviation. This improved level of accuracy led to accurate segmentations of the visual scene into its constituent objects (see Figs. 13d and 13f).
V. CONCLUSIONS

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 behavior and perception may be one of the key factors for the development of more efficient vision algorithms [20], [21]. According to this proposal, the proper coordination of motor behavior and visual processes could greatly simplify visual computations. Indeed, the close interaction between motor and perceptual processes is one of the most striking features of biological vision systems. These systems have not only developed to control behavior. They have actually 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 [39] to birds [40] and humans [41].

In the case of depth perception, several studies have shown clear involvements of oculomotor activity in the generation and extraction of distance information [14], [15], [16], [17]. As an agent moves through the scene, images acquired at successive times differ for the relative positions of objects in the landscape. These changes provide accurate information on the relative distance between the objects and the agent, and many species use them in their analysis of a 3D scene [13], [42]. In humans, motion parallax generates a vivid, unambiguous depth percept as if the observer were looking with two eyes using binocular stereopsis [12].

In robotics, where many operations require an accurate estimation of distance, depth from motion has been an active area of research. Using various approaches, previous studies have analyzed the motion parallax that occurs during translations of a vision system. While a motion parallax occurs in most head/eye systems, this study has focused on the case of a robotic head specifically designed to replicate the oculomotor parallax present in the human eye. The motor characteristics of this system were designed to ensure that two axes of rotation would intersect in space. Careful positioning of the cameras and selection of the optics ensured that the separation between nodal points and the center of rotation precisely matched the distance present in the eye. Use of this system has enabled a direct implementation of the strategy by which humans examine 3D scenes.

Under natural viewing conditions, humans tend to relocate their gaze by small shifts, and most saccades have amplitudes within few degrees. Small saccades occur even within the periods of fixation even though we are not aware of them. This study shows that the scanning strategy followed by human observers produces oculomotor parallax that can be successfully measured and used to predict the distances of objects and surfaces. It is often debated whether examination of a scene by actively relocating the direction of gaze presents advantages with respect to the static analysis provided by a camera with a wider field of view. As shown by this study, the computations involved in image segmentation are greatly simplified in an active system that exploits the oculomotor parallax produced by relocations of gaze.
Fig. 13. Estimating distance with sequences of eye movements. Case of all saccades. (Top Row) Mean estimated distance of the pixels composing the objects. The two panels show results obtained with sequences of eye movements recorded from subject MV (left) and AR (right). Error bars represent one standard deviation. (Middle and Bottom Rows) Trace of recorded eye movements (left) and segmentation of the scene based on the estimation of distance (right). The results for subjects MV and AR are shown in the middle and bottom rows, respectively.
The implications of using a robot that closely replicates aspects of the human visual system are not limited to the field of robotics. Direct access to the oculomotor parallax present in the eye during sequences of eye movements enables a more quantitative evaluation of the oculomotor strategies followed by human observers. In particular, we have shown that reliable depth information is available when fixational saccades are amplified to achieve in the robot a resolution comparable to human visual acuity. This information could potentially be exploited by the brain. This observation raises the hypothesis that also fixational saccades might contribute to depth perception. Fixational saccades provide ideal conditions for estimating depth on the basis of the oculomotor parallax. They are sufficiently large to supply reliable distance information, while small enough to preserve the shape of features almost unchanged. This preservation of feature appearance facilitates the identification of corresponding points of the scene in images acquired before and after a rotation.

By rigorously analyzing the parallax provided by camera rotations, the goal of this paper has been to provide the foundations, as well as the motivation, for using the oculomotor parallax in robotics. There are several aspects of this work that need to be investigated in further detail in future studies. A first issue regards the aspects of this work that need to be investigated in further detail in future studies. A first issue regards the solution of the correspondence problem, the matching between features in images acquired at different moments in times. In its current format, the simple cross-correlation algorithm used in this study is not suited for real-time performance. It would be interesting to apply more elaborate methods developed in the areas of depth from motion and stereopsis to the estimation of the oculomotor parallax.

Furthermore, in a system that operates in the real world, depth cues acquire informative value only in the light of the actual physical and functional characteristics of the system. In nature, all organisms have to tune their evaluation of sensory cues while interacting with complex environments. In this study, estimation of distance relied on a model of the robot that was tuned on the basis of a preliminary calibration. A more robust approach would use learning to enable autonomous and continuous calibration of the oculomotor parallax. Since this cue provides useful information for a nearby range of distances, self-calibration of the oculomotor parallax can be obtained by using proprioceptive knowledge of the position of a robotic arm, an approach that replicates the way the brain learns about visual space.

The oculomotor system described in this paper was developed as part of a wider program of research that focuses on integrating robotic systems with computational models of the brain. This interdisciplinary approach establishes a direct link between the natural and engineering sciences. Instead of relying on simulations of simplified visual environments as in many studies in computational neuroscience, this methodology brings simulations of neuronal responses one step closer to their biological counterparts, as it provides access to realistic sensory inputs. This approach carries the potential not only for innovative engineering applications, but also for fostering our understanding of how sensory information is processed in the brain.

ACKNOWLEDGMENT

This work was supported by the National Science Foundation grants BIC-0432104 and CCF-0130851.

REFERENCES