On Perceptual Advantages of Eye-Head Active Control

Enrico Grosso

DIST, Department of Communication, Computer and System Sciences
University of Genoa, Via Opera Pia 11a, 16145 Genoa, Italy

Abstract. The paper presents a theoretical study on the perceptual advantages related to the active control of a binocular vision system. In particular the presentation focuses on the process of *improving perception*, a task strategically important in humans and many vertebrates. The analysis is based on an anthropomorphic system; the sensitivity of the transformation from world to camera coordinates is used as a cost function for driving the movements of the eye-head system. The control strategy obtained in this way allows to formally motivate, outside of a purely behavioral context, some relevant aspects of the biological vision like fixation, vergence and eye-head compensation.
Keywords: robot vision, eye-head coordination, vision based control.

1 Introduction

The development of new paradigms for vision [2, 1, 3] is gradually demonstrating the intimate relationship existing between vision and movement. In many cases visual perception can take advantage of an active movement of the visual system. On one hand, movement can be conceived as a strategy to simplify and improve the efficiency of computational processes, in relation to the fact that some essential parameters can be easily estimated, and ill posed vision problems become well posed for a moving observer [1, 8]. On the other hand, moving and interacting with the environment can be considered in a behavioral context as a key issue to solve problems and satisfy specific purposes [3, 7].

In this paper we show that active gaze control is important in relation to the possibility of *ameliorating perception*, despite noise and uncertainties in the system parameters. In other words, it is argued that basic visual behaviors, like for instance fixation, admit a simple explanation in term of perceptual robustness, besides common computational or behavioral interpretations. This robustness issue is of course very important if visual feedback is used to control the movements of a mechanical device [11, 5] and can be related to the criteria usually adopted designing robotic heads.

A formal analysis is presented for a binocular anthropomorphic system. The transformation from world to camera coordinates is investigated and the sensitivity of this transformation is evaluated, both in relation to static and dynamic information. It is demonstrated that fixation, vergence and compensatory movements of the head can sensibly improve spatial perception. The importance of fixation is furthermore analyzed in the case of perception of relative distances.

This research has been supported by the VAP - ESPRIT Basic Research Project.

Lecture Notes in Computer Science, Vol. 801
Jan-Olof Eklundh (Ed.)
Computer Vision – ECCV '94

2 Preliminaries

The anthropomorphic system considered in this paper is composed by a moving head and two moving cameras. Figure 1 shows the kinematic structure of the system and the position of the left, right and central (head) frames. A generic inertial reference frame is denoted by $< e >$.

The notation used throughout this paper is basically that used in [9, 10]. We

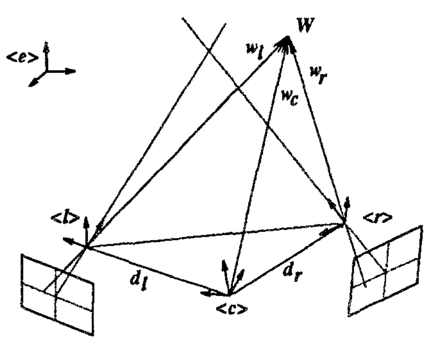

Fig. 1. Schematic representation of the head-eye reference frames.

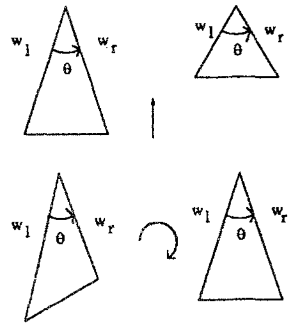

Fig. 2. Schematic picture showing the effect of rotation and translation of the central system on the vergence angle.

denote by ${}^a\boldsymbol{w} = ({}^a x,\ {}^a y,\ {}^a z)^t$ the projection of the vector $\boldsymbol{w}$ in the frame $< a >$ and by ${}^b_a\boldsymbol{R}$ the rotation matrix from the frame $< a >$ to the frame $< b >$.

Considering the pin hole model for the cameras [12] and denoting by (u, v) the image plane coordinates we can write:

$$\begin{cases} {}^l x_l + \bar{u}_l\, {}^l z_l = 0 \\ {}^l y_l + \bar{v}_l\, {}^l z_l = 0 \\ {}^r x_r + \bar{u}_r\, {}^r z_r = 0 \\ {}^r y_r + \bar{v}_r\, {}^r z_r = 0 \end{cases} \qquad \begin{matrix} \bar{u}_l = \frac{u_l - u_{0l}}{\alpha_{ul}} & \bar{v}_l = \frac{v_l - v_{0l}}{\alpha_{vl}} \\ \bar{u}_r = \frac{u_r - u_{0r}}{\alpha_{ur}} & \bar{v}_r = \frac{v_r - v_{0r}}{\alpha_{vr}} \end{matrix} \tag{1}$$

where the indices l and r stand for left and right camera, respectively. The parameters α and u are usually called "intrinsic" because they define the internal structure of the cameras. Referring to figure 1, the position of a generic point W in space is given by:

$$\begin{aligned} {}^l\boldsymbol{w}_l &= {}^l_c\boldsymbol{R}\, {}^c\boldsymbol{w}_l = {}^l_c\boldsymbol{R}\, (\, {}^c\boldsymbol{w}_c - {}^c\boldsymbol{d}_l\,) \\ {}^r\boldsymbol{w}_r &= {}^r_c\boldsymbol{R}\, {}^c\boldsymbol{w}_r = {}^r_c\boldsymbol{R}\, (\, {}^c\boldsymbol{w}_c - {}^c\boldsymbol{d}_r\,) \end{aligned} \tag{2}$$

Vectors ${}^c\boldsymbol{d}_l$ and ${}^c\boldsymbol{d}_r$ define the position of the optical centers of the cameras with respect to the central frame $< c >$. They are usually referred, together with the rotation matrices ${}^l_c\boldsymbol{R}$ and ${}^r_c\boldsymbol{R}$, as "extrinsic" parameters.

It is worth noting that the estimation of points and objects in the 3D space is an inverse problem typical in stereo vision [9]. This obviously depends on the model of the sensing system and on the uncertainties on the parameters of this model. However, it is very interesting to observe that stereo estimates are also *space dependent*. In other words, assuming different postures in space, the visual system can obtain different measures of the same physical quantity.

3 The Vision Problem

It is quite easy to show that equations (1) and (2) are sufficient to solve the inverse problem of locating a given point W with respect to the frame $< c >$ [9]. Assuming a perfect knowledge on both extrinsic and intrinsic parameters, we first define the matrices $\boldsymbol{J}_l = \begin{bmatrix} 1 & 0 & \bar{u}_l \\ 0 & 1 & \bar{v}_l \end{bmatrix}$ and $\boldsymbol{J}_r = \begin{bmatrix} 1 & 0 & \bar{u}_r \\ 0 & 1 & \bar{v}_r \end{bmatrix}$. Then, using equations (1) and (2), ${}^c\boldsymbol{w}_c$ can be computed solving the following over-determined set of linear equations:

$$\begin{bmatrix} \boldsymbol{J}_l & 0 \\ 0 & \boldsymbol{J}_r \end{bmatrix} \begin{bmatrix} {}^l_c\boldsymbol{R} & 0 \\ 0 & {}^r_c\boldsymbol{R} \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_3 \\ \boldsymbol{I}_3 \end{bmatrix} {}^c\boldsymbol{w}_c = \begin{bmatrix} \boldsymbol{J}_l & 0 \\ 0 & \boldsymbol{J}_r \end{bmatrix} \begin{bmatrix} {}^l_c\boldsymbol{R} & 0 \\ 0 & {}^r_c\boldsymbol{R} \end{bmatrix} \begin{bmatrix} {}^c\boldsymbol{d}_l \\ {}^c\boldsymbol{d}_r \end{bmatrix} \tag{3}$$

Remark 1 *It is straightforward to demonstrate that, for any bounded measurement, equation (3) always admits solution if and only if $\boldsymbol{w}_l$ and $\boldsymbol{w}_r$ are not parallel. The uniqueness of the solution derives from linear algebra.* □

Let us now assume to be able to measure the time derivatives of the terms in (1). Differentiating equations (1) and (2) and eliminating ${}^l\dot{\boldsymbol{w}}_l$ and ${}^r\dot{\boldsymbol{w}}_r$ we obtain:

$$\begin{bmatrix} -\frac{1}{{}^l z_l}\boldsymbol{J}_l & 0 \\ 0 & -\frac{1}{{}^r z_r}\boldsymbol{J}_r \end{bmatrix} \begin{bmatrix} {}^l_c\boldsymbol{R} & 0 \\ 0 & {}^r_c\boldsymbol{R} \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_3 \\ \boldsymbol{I}_3 \end{bmatrix} {}^c\dot{\boldsymbol{w}}_c = \begin{bmatrix} \dot{\bar{u}}_l \\ \dot{\bar{v}}_l \\ \dot{\bar{u}}_r \\ \dot{\bar{v}}_r \end{bmatrix} +$$

$$\begin{bmatrix} -\frac{1}{{}^l z_l}\boldsymbol{J}_l & 0 \\ 0 & -\frac{1}{{}^r z_r}\boldsymbol{J}_r \end{bmatrix} \begin{bmatrix} {}^l_c\boldsymbol{R} & 0 \\ 0 & {}^r_c\boldsymbol{R} \end{bmatrix} \begin{bmatrix} \boldsymbol{C}({}^c\boldsymbol{w}_c - {}^c\boldsymbol{d}_l) & 0 \\ 0 & \boldsymbol{C}({}^c\boldsymbol{w}_c - {}^c\boldsymbol{d}_r) \end{bmatrix} \begin{bmatrix} {}^l\boldsymbol{\Omega}_l \\ {}^r\boldsymbol{\Omega}_r \end{bmatrix} \tag{4}$$

where $\boldsymbol{\Omega}_l$ and $\boldsymbol{\Omega}_r$ are the angular velocities of the frames $< l >$ and $< r >$, respectively, and $\boldsymbol{C}(\boldsymbol{u})$ is the operational matrix of the vector product.

Remark 2 *Equations (3) and (4) represent our complete inverse model. They give an interesting example of the extent of the problem which involves a large number of parameters and position/velocity measures. In particular, it is clear that equation (4) depends upon the solution of equation (3), since there explicitly appears ${}^c\boldsymbol{w}_c$.* □

4 Sensitivity Analysis

Both (3) and (4) are overdetermined sets of linear equations in the unknowns ${}^c\boldsymbol{w}_c$ and ${}^c\dot{\boldsymbol{w}}_c$. Since in practice noise and parameters uncertainties perturb the solution it is interesting to investigate the effect of these perturbations [6]. Consider first the system (3) and write it in the canonical form:

$$\boldsymbol{A}\,\boldsymbol{B}\,\boldsymbol{x} = \boldsymbol{A}\,\boldsymbol{b} \tag{5}$$

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{J}_l & 0 \\ 0 & \boldsymbol{J}_r \end{bmatrix} \begin{bmatrix} {}^l_c\boldsymbol{R} & 0 \\ 0 & {}^r_c\boldsymbol{R} \end{bmatrix} \in \Re^{4x6} \quad \boldsymbol{B} = \begin{bmatrix} \boldsymbol{I}_3 \\ \boldsymbol{I}_3 \end{bmatrix} \in \Re^{6x3} \quad \boldsymbol{b} = \begin{bmatrix} {}^c\boldsymbol{d}_l \\ {}^c\boldsymbol{d}_r \end{bmatrix} \in \Re^6 \tag{6}$$

Assume that the system admits a solution $\boldsymbol{x}_0$, that is $\|\boldsymbol{A}\,\boldsymbol{B}\,\boldsymbol{x}_0 - \boldsymbol{A}\,\boldsymbol{b} = 0\|$, and that this solution is unique. Then, following the rationale reported in [6] and provided that $||\delta A|| \leq \frac{1}{2}\, \sigma_{min}(AB)$ the perturbation can be bounded by:

$$\frac{||\delta x||}{||x_0||} \leq \left\|(A\ B)^{\dagger}\right\|\ ||\delta A|| \left(||B|| + \frac{||b||}{||x_0||}\right) + \left\|(A\ B)^{\dagger}\right\|\ ||A||\ \frac{||\delta b||}{||x_0||} \quad (7)$$

Note that in the above equations $M^{\dagger}$ indicates the pseudo-inverse of the matrix M while $\sigma_{min}(M)$ denotes the minimum singular value.

Remark 3 *From the expression above we derive two important hints for the solution of the stereo vision problem. First of all it is important to minimize the effect of* $||b||$ *and* $||\delta b||$ *with respect to* $||x_0||$. *We will skip over a detailed analysis of this aspect but it is clear that the position of the central system can be constrained in this sense. Then, it is important to understand the structural properties of* A, δA *and* $(A\ B)^{\dagger}$. □

From a general point of view we have to consider three sources of uncertainty. The first is due to uncertainties in the intrinsic parameters and it obviously depends on the camera calibration procedure. In our analysis it appears in the matrices J_l and J_r and then in A. The second is due to uncertainties in the extrinsic parameters, essentially related to the mechanical calibration of the system. It appears in the matrices ${}^c_l R$ and ${}^c_r R$ and in the vectors ${}^c d_l$ and ${}^c d_r$. Finally we have to consider measurements errors hidden in J_l and J_r. Following sections will analyze in detail equation (7), describing the contribution of the various terms.

4.1 Explaining Fixation

Consider first the contribution of the term $||\delta A||$. We have immediately:

$$||\delta A|| \leq \left\|\begin{matrix} J_l & 0 \\ 0 & J_r \end{matrix}\right\| \left\|\begin{matrix} \delta^l_c R & 0 \\ 0 & \delta^r_c R \end{matrix}\right\| + \left\|\begin{matrix} \delta J_l & 0 \\ 0 & \delta J_r \end{matrix}\right\| \doteq ||J|| \left\|\begin{matrix} \delta^l_c R & 0 \\ 0 & \delta^r_c R \end{matrix}\right\| + ||\delta J|| \quad (8)$$

The perturbation of the rotational components depends only on the calibration errors; this means that we have simply to compute the norm of the matrices J and δJ:

$$||J|| = max\left(\sqrt{1 + (\bar{u}_l^2 + \bar{v}_l^2)}\ ,\ \sqrt{1 + \left(\bar{u}_r^2 + \bar{v}_r^2\right)}\right)$$
$$||\delta J|| = max\left(\sqrt{\delta\bar{u}_l^2 + \delta\bar{v}_l^2}\ ,\ \sqrt{\delta\bar{u}_r^2 + \delta\bar{v}_r^2}\right)$$

Remark 4 *On the base of the above discussion we can affirm that for bounded calibration errors* $||\delta A||$ *is minimized by* $\bar{u}_l = \bar{v}_l = \bar{u}_r = \bar{v}_r = 0$.
The above result defines an important specification for the movements of the eyes. In fact it states that w_l *and* w_r *must be preferably aligned with the optical axes of the cameras. In other words, fixation makes minimum the measurement error and, as a consequence, ameliorates 3D perception of the external world.* □

It is worth noting that, due to the structure of the matrix A, we have immediately $||A|| \leq ||J||$. Then, the bound of $||A||$ is minimum under the same assumptions of remark 4.

4.2 Explaining vergence and head compensation

The term $\left\| (A\ B)^{\dagger} \right\|$ in equation (7) is very important because it has the effect of multiplying the perturbations on measures and system parameters. We have:

$$\left\| (A\,B)^{\dagger} \right\| = max_i \sqrt{\sigma_i \left((A\,B)^{\dagger}\ (A\,B)^{\dagger^t} \right)} = \left(min_i \sqrt{\sigma_i \left(B^t\ A^t A\,B \right)} \right)^{-1} \quad (9)$$

The minimization of expression (9) is based on linear algebra and can be found in [4]. We give here the main result, which depends on the relative rotation between the optical axes of the two cameras:

$$\sigma_{min}(B^t\ A^t A\ B) = 1 - |\cos\theta|$$

where, by definition, $\cos\theta = \frac{w_l \cdot w_r}{\|w_l\| \cdot \|w_r\|}$

Remark 5 *From equation (9) and the above result follows that the norm is minimized when* $|\cos\theta|$ *is minimum. This happens in particular in two cases:*

1. *Approaching the object while fixating; this implies a translational motion of the central system and makes the expression* $|\cos\theta|$ *decreasing.*
2. *Rotating the baseline, so as to bring the cameras in a symmetric position with respect to the target point (figure 2).* □

In summary, we can affirm that the analysis presented in this section demonstrates the importance of gaze control. From a more anthropomorphic perspective it explains two interesting behaviors. First of all it motivates *vergence movements*, in the sense that a symmetric movement of the eyes maintains the position at "minimum norm" once reached. Secondly, it explains compensatory movements of the head. These movements could be specifically devoted to maximize the vergence angle, independently from the fixation task performed by the eyes.

Remark 6 *Using equation (4) it is possible to demonstrate that also in the dynamic case good perception is guaranteed by using fixation. In other words, using a static vision system to deal with dynamic quantities is always a bad choice, independently from the ability of estimating intrinsic and extrinsic parameters. However, note that in the dynamic case it is impossible to derive a simple result on the vergence angle, equivalent to that presented in this section.* □

5 Measuring relative distances

In this section we briefly analyze the problem of estimating relative distances by using relative information on the image plain. Using the notation of section 4 and denoting by the indices 1 and 2 the quantities related to two different points in space we can write in canonical form:

$$A_1\ B\ x = A_1\ b \qquad\qquad A_2\ B\ (x + \Delta x) = A_2\ b$$

Denoting by ΔA the difference between A_2 and A_1 we obtain, by difference:

$$A_2\ B\ \Delta x = \Delta A\ (b - B\ x)$$

and, applying sensitivity analysis:

$$\frac{\|\delta\Delta\boldsymbol{x}\|}{\|\Delta\boldsymbol{x}_0\|} \leq \left\|(\boldsymbol{A}_2\ \boldsymbol{B})^{\dagger}\right\| \left(f\left(\|\delta\Delta\boldsymbol{A}\|,\ \|\Delta\boldsymbol{A}\|\right) + \|\delta\boldsymbol{A}_2\|\ \|\boldsymbol{B}\|\right) \tag{10}$$

It is clear that a good observation strategy consists again in fixating one of the two points (the second one), minimizing $\left\|(\boldsymbol{A}_2\ \boldsymbol{B})^{\dagger}\right\|$ and $\|\delta\boldsymbol{A}_2\|$. The remaining terms (here generically denoted by f) depend on the relative quantities measured on the images and can be only marginally influenced by an active control of the eye-head system.

6 Conclusions

In this paper the problem of the coordinated movement of an eye-head system has been faced. The solution outlined is based on the fact that in an active system the position of the visual sensors can be controlled in order to gather optimal measurements, and this independently from the knowledge of the intrinsic and extrinsic parameters of the system itself. The analysis demonstrates the importance of fixation and eye-head compensation estimating dynamic quantities or relative distances between points in space.

References

1. J. Aloimonos, I. Weiss, and A. Bandyopadhyay. Active vision. *International Journal of Computer Vision*, 1(4):333–356, 1988.
2. R. K. Bajcsy. Active perception vs passive perception. In *Proc. Third IEEE Computer Society Workshop on Computer Vision: Representation and Control*, pages 13–16, Bellaire, MI, 1985.
3. D.H. Ballard and C. M. Brown. Principles of animate vision. *CVGIP*, 56(1):3–21, July 1992.
4. G. Cannata and E. Grosso. Active eye-head control. Technical Report LIRA-Lab TR-4-93, University of Genova, Genova, Italy, October 1993.
5. B. Espiau, F. Chaumette, and P. Rives. A new approach to visual servoing in robotics. *IEEE Transactions on Robotics and Automation*, 8(3):313–326, June 1992.
6. G. H. Golub and C. F. Van Loan. *Matrix Computations.* The John Hopkins University Press, Baltimore, USA, 1989.
7. E. Grosso and D.H. Ballard. Head-centered orientation strategies in animate vision. In *3rd Internatiional Conference on Computer Vision*, Berlin - Germany, 1993.
8. E. Grosso, M. Tistarelli, and G. Sandini. Active/dynamic stereo for navigation. In *Proc. of Second European Conference on Computer Vision*, S. Margherita Ligure, Italy, May 1992.
9. B. K. P. Horn. *Robot Vision.* MIT Press, Cambridge, USA, 1986.
10. A.J. Koivo. *Fundamentals of Robotic Manipulators.* Wiley, New York - NY, USA, 1989.
11. N.P. Papanikolopoulos, P.K. Khosla, and T.Kanade. Visual tracking of a moving target by a camera mounted on a robot: A combination of control and vision. *IEEE Trans. Robotics Automation*, 9(1):14–35, 1993.
12. G. Toscani and O.D. Faugeras. The calibration problem for stereo. In *IEEE Conference on Comp. Vision and Pattern Recognition*, Miami - FL, USA, 1986. IEEE Comp. Soc.