Visual Behaviours for Binocular Tracking ⋆

Alexandre Bernardino , José Santos-Victor

Instituto de Sistemas e Robótica, Instituto Superior Técnico, 1096 Lisboa Codex, Portugal
{alex,jasv}@isr.ist.utl.pt

Abstract

This paper presents a binocular tracking system based on the cooperation of visual behaviours. Biologically motivated behaviours, *vergence* and *pursuit*, are integrated as parallel and complementary processes in the tracking system. Their low internal coupling simplify not only system design and control but also the acquisition of perceptual information. The use of a space variant image representation and low-level visual cues as feedback signals in a closed loop control architecture, allow real-time and reliable performance for each behaviour, despite the low precision of the algorithms and modeling errors. The overall system is implemented in a stereo head running at real-time (25 Hz), without any specific processing hardware. Results are presented for objects of different shapes and motions, illustrating that tracking can be robustly achieved by the cooperation of purposively designed behaviours, tuned to specific subgoals.

1 Introduction

Recent advances in robotics research have been reported to apply visual tracking as a mean to accomplish tasks, such as trajectory reconstruction, object recognition [4], navigation and egomotion estimation [9]. When objects have constrained shapes or motions, or move in simple backgrounds, some tracking systems have been successfully employed [12,1]. However, in complex and dynamic environments, most systems still lack robustness, reactivity and flexibility. In nature, biological systems interact constantly with unstructured environments, and show high reliability in executing many important tasks. This fact led some researchers to design their particular applications based on biological evidence [20].

The tracking system presented in this paper applies solutions inspired by the visual system of humans and other animals, that are advantageous over other more straightforward approaches:
– *Binocularity:* Although the tracking problem is often faced as a monocular one [15,16], binocularity is important because allows the recovery depth from the tracked object and simplifies figure-ground segmentation [7].
– *Ocular Movements:* The most influent ocular movements in biological sys-

⋆ Partially funded by projects JNICT-PBIC/TPR/2550/95 and PRAXIS/3/3.1/TPR/23/94.

tems are saccadic, vergence and smooth pursuit movements [19]. Saccadic and smooth pursuit movements use retinal position and velocity errors, respectively, to compensate lateral motion of the target. Vergence movements, in contrast, are mainly based on disparity cues to compensate motion in depth. This decomposition simplifies the perceptual processes used here.

– *Space Variant Image Resolution:* Although some work is being done on space variant image resolutions [5,13,24], most active vision systems use cartesian image representations. However, biological systems generally have a space variant density of photoreceptors in the retina, distinguishing between the central high resolution fovea and a low resolution peripheral zone [6]. This is one of the main reasons for the existence of ocular movements. Combining a wide field of view with a high resolution fovea and the ability to move the eyes, it is possible to perform the desired tasks, without considering all the data contained in uniformly sampled images and concentrates the processing effort in the fovea

Two main *visuomotor behaviours* are developed and integrated in the tracking system: the vergence and the pursuit behaviours. The focus of this paper is on the design, integration and test of these behaviours. The design includes perceptual and control aspects, but special attention is given to perception. Control issues are currently well established and can be formalized in several ways. The Visual Servoing approach [8] establishes a framework for motor control in the presence of well defined visual features of the target. However, under general conditions and complex environments, the estimation of target position and motion is not simple and many visual perception techniques often fail to provide adequate inputs to the control system.

2 Visuomotor Behaviours

When designing a general visual servoing system we are faced with two main problems:

- What is the relevant perceptual information? (the perceptual problem).
- How to use it for motion control? (the control problem).

We define a *visuomotor behaviour* as a process that deals with the perceptual problem and the control problem as a single system. Perceptual and control strategies are purposively designed to the task at hand and may suffer constraints from each other. Each *visuomotor behaviour* extracts only the relevant information and controls only the degrees of freedom needed for a particular purpose. A coordination module may be required to integrate the contribution of each behaviour (Figure 1). We describe a tracking system composed of two *visuomotor behaviours* motivated by biological visual systems, both in the perceptual and the motor parts : vergence and pursuit. The vergence behaviour acquires depth cues from the stereo images and controls the camera angles in order to keep the target in the center of the images, despite motions in depth. The pursuit behaviour extracts target position and velocity errors from the stereo images and controls the gaze direction. These behaviours are separa-

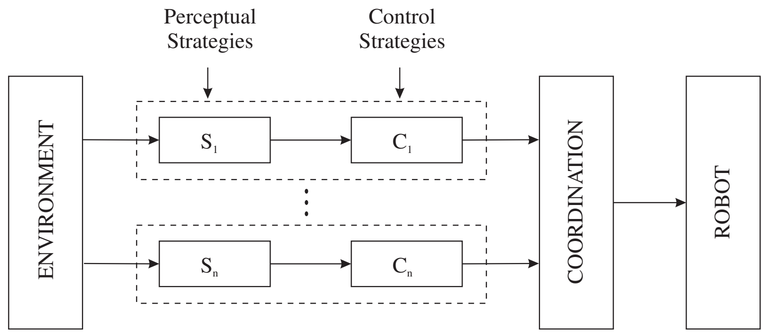

Fig. 1. The *visuomotor behaviours* (inside dashed boxes).

ble as they acquire different stimuli and control different motions but are highly coupled in the sense that each one depends on the performance of the other.

3 Space Variant Imaging

Most biological visual systems have retinas with a non-uniform distribution of photo-receptors. In particular, humans and other primates have a very high density of photo-receptors in the center of the retina [1] which decreases toward the periphery. This way, it is possible to have a wide field of view and, together with ocular movements, have a high resolution description of the environment. The log-polar mapping [22] provides a non-uniform resolution geometry which is similar to the distribution of photo-receptors in the human retina. Our perceptual strategies use this kind of mapping for image representation, with benefits both in algorithmic and computational aspects [2]. Let us define appropriate notation for the cartesian and log-polar coordinate systems:

- Let $p = (x, y)$ be the vector of cartesian coordinates.
- We will denote $q = (\xi, \eta)$ the vector of log-polar coordinates – ξ is the radial and η is the angular coordinate.
- Images defined over a continuous domain are denoted with arguments between brackets, such as $I(x, y)$ or $I(p)$, and discrete images have subscript arguments, e.g. $I_{x,y}$ or I_p.
- Log-polar images are written in caligraphic font, such as $\mathcal{I}(\xi, \eta)$.

The log-polar mapping (α) can be defined as:

$$\begin{cases} \alpha : \mathbb{R}^2 \backslash \{0\} \to \mathbb{R}_0^{+2} \\ q = \left(\log \sqrt{x^2 + y^2}\,,\ \arctan \frac{y}{x}\right) \end{cases} \tag{1}$$

The inverse mapping is given by:

$$\begin{cases} \alpha^{-1} : \mathbb{R}_0^{+2} \to \mathbb{R}^2 \backslash \{0\} \\ p = \left(\mathrm{e}^{\xi} \cos \eta\,,\ \mathrm{e}^{\xi} \sin \eta\right) \end{cases} \tag{2}$$

Let image $\mathcal{I}$ be the result of mapping image I to log-polar coordinates. The following relations hold:

$$\mathcal{I}(\xi, \eta) = I\Big(\alpha^{-1}(\xi, \eta)\Big) \tag{3}$$

$$I(x, y) = \mathcal{I}\Big(\alpha(x, y)\Big) \tag{4}$$

[1] the high resolution area in the center of the retina is usually called *the fovea*.

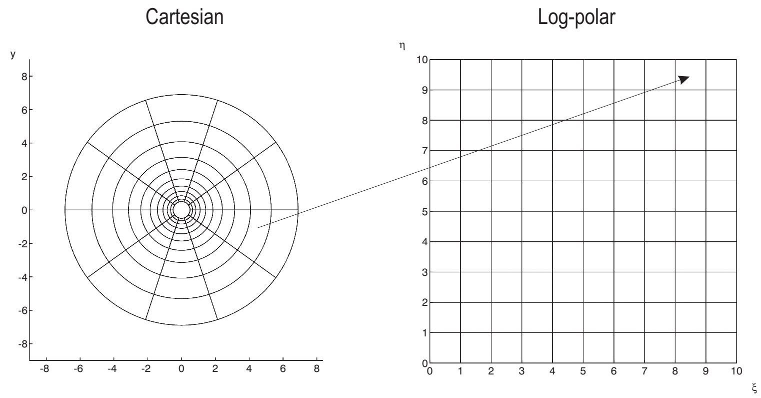

Fig. 2. Discrete log-polar coordinates (right) and corresponding cartesian grid (left).

In the discrete case, we can consider the log-polar mapping as the process of re-sampling the cartesian image according to a log-polar grid, as shown in Figure 2. Each cell of the grid over the cartesian image maps into a rectangular cell in the log-polar image. One advantage of the discrete version of the log-polar mapping is an important data reduction on the periphery of the visual field, which reduces the computational cost of the image processing algorithms. However, for tracking purposes, the greatest advantage comes from the implicit focus of attention implemented in the center of the visual field. Properties of the objects in the images center become dominant over other objects in the periphery and the background. Providing that we can keep the target in the center of the visual field, this fact improves the performance of the tracking algorithms relatively to the use of uniform images.

All image processing algorithms developed in this work use log-polar images. However, in some processes, the output is required to be expressed in cartesian coordinates. For that purpose we introduce the Jacobian matrices:

$$J(x,y) = \begin{bmatrix} \frac{x}{x^2+y^2} & \frac{y}{x^2+y^2} \\ -\frac{y}{x^2+y^2} & \frac{x}{x^2+y^2} \end{bmatrix} \tag{5}$$

and:

$$J^{-1}(\xi,\eta) = \begin{bmatrix} e^{\xi}\cos\eta & -e^{\xi}\sin\eta \\ e^{\xi}\sin\eta & e^{\xi}\cos\eta \end{bmatrix} \tag{6}$$

Their determinants[2] are given by:

$$|J(x,y)| = \frac{1}{x^2+y^2} \tag{7}$$

$$|J^{-1}(\xi,\eta)| = e^{2\xi} \tag{8}$$

4 Vergence Behaviour

The purpose of the vergence movements is to control the angle between the optic axes, such that a fixation point along some specified gaze direction is kept at the center of the visual field. When this happens, we say that the target is "foveated". Figure 3 illustrates the vergence angle θ_v, in a binocular system. As the vergence angle is directly re-

[2] also called *Jacobians*

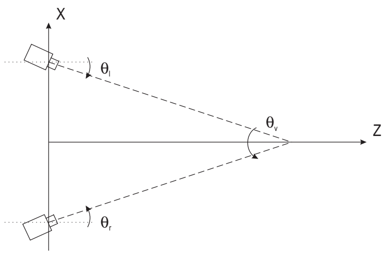

Fig. 3. Stereo cameras verging at a point.

lated to target distance, depth cues can be used to help the vergence system in compensating depth changes. Binocular disparity and accommodation are important inputs to the vergence control system, and humans rely also on motion and shading [6], among others.

Let $I^r(x,y)$ and $I^l(x,y)$ be the left and right cartesian images. One point in space projects in the right and left camera coordinates (x^r, y^r) and $\left(x^l, y^l\right)$, respectively. We define the horizontal disparity as $d = x^l - x^r$. Being the most important component for vergence purposes, the horizontal disparity is the only measure currently used to control the vergence angle. Consider that, for some vergence angle θ_v^0, the target is centered in both images ($d^0 = 0$). If we use symmetrical vergence ($\theta_r = \theta_l$), linearization of the kinematic relations give:

$$d = G_v \cdot \left(\theta_v - \theta_v^0\right) \qquad (9)$$

where the sensitivity term G_v depends on system calibration parameters and is constant for the whole workspace. Traditional algorithms to calculate explicit disparity measures include cepstral filtering [17] and phase correlation [10], but these techniques are very time consuming, and therefore, not very adequate for real-time applications. By relying on closed loop control, it is still possible to achieve vergence with low precision algorithms and coarse system calibration.

4.1 Disparity Estimation

To obtain estimates of target disparity, we use correlation between the stereo image pair. The *Sum of Squared Differences* is a frequently used measure of similarity for discrete images, which is minimal for the higher similarity between images:

$$\mathcal{S}\left(I^r_{x,y}, I^l_{x,y}\right) = \sum_{x,y} \left(I^r_{x,y} - I^l_{x,y}\right)^2$$

where the sum extends for all image points. If d_0 is the average amount of horizontal disparity between images, then we expect the correlation function:

$$\mathcal{C}\left(I^r, I^l; d\right) = \mathcal{S}\left(I^r_{x+d,y}, I^l_{x,y}\right)$$

to have its minimum for $d = d_0$. Thus, disparity estimation can be considered as the search for the value of d that minimizes the correlation function. This search can be restricted to a discrete set of disparities, $D = \{d_1, ..., d_n\}$, whose cardinality depends on the requirements of precision versus computation time for the desired application. Figure 4 shows a block diagram of the perceptual process, composed by a set of channels that translate the current images by a pre-determined disparity and compute the correlation value for those images. The estimated disparity is given by the channel whose output is minimal. Assuming smooth trajectories for the target motion, disparity remains small most of the time, and the set D should cover small disparities in a dense manner. To cope with faster

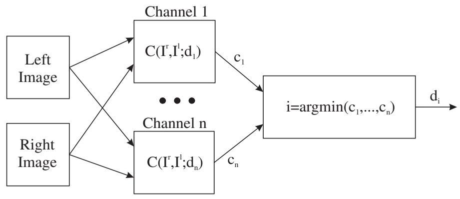

Fig. 4. Vergence perceptual strategy.

displacements, large disparities are also considered but in a coarse manner to keep the required computational power within limits.

The decomposition into channels tuned to specific disparities is coherent with some theories of human vergence behaviour. For instance, [18] proposes a theory based on a discrete set of parallel channels, sensitive only to certain values of disparity. Each channel, when activated, produces a motor response with a low-pass time characteristic. However, the motor response of each channel has different time constants and gains. Our control strategy considers the same dynamics for all channels (see Section 4.2).

When using log-polar images, the major difference in the analysis is that a pure horizontal translation affects both radial and angular components of the log-polar coordinate system. Thus, a simple translation in cartesian coordinates becomes a complex warping in log-polar images. For each channel d_i we must compute the correspondence map (w) between pixels in the right and left image:

$$\begin{aligned}\left(\xi^l, \eta^l\right) &= w\left(\xi^r, \eta^r; d_i\right)\\ &= \alpha\Big(\alpha^{-1}\left(\xi^r, \eta^r\right) + \left(d_i, 0\right)\Big)\end{aligned}$$

Therefore, the correlation function for log-polar images is given by:

$$\mathcal{C}_l\left(\mathcal{I}^r, \mathcal{I}^l; d\right) = \mathcal{S}\left(\mathcal{I}^r_{w(\xi,\eta;d)}, \mathcal{I}^l_{\xi,\eta}\right)$$

For small objects, the existing background mismatches between the two images can lead to low values of similarity even in correct vergence situations. The non-uniform resolution representation is an elegant solution to this problem. Having higher resolution in the center of the images, where the target is expected to be, the areas belonging to the target are dominant in the computation of the correlation index, reducing the negative influence of background elements [3]. This fact can be clearly understood by expressing the relation between correlation on cartesian and log-polar images. This can be done considering continuous domains for both geometries. Let us define the continuous version of the *Sum of Squared Differences*, applied to log-polar images:

$$\mathcal{D}\left(\mathcal{I}^r, \mathcal{I}^l\right) = \iint_q \left(\mathcal{I}^r - \mathcal{I}^l\right)^2 \mathrm{d}q \quad (10)$$

where integration is over the whole images. It can be interpreted as a "distance" in the space of continuous images. Applying the inverse log-polar mapping and changing the integration variable to cartesian coordinates, we obtain:

$$\mathcal{D}\left(\mathcal{I}^r, \mathcal{I}^l\right) = \iint_p \left(I^r - I^l\right)^2 |J| \,\mathrm{d}p$$

Replacing the jacobian expression, given by eq. (7), we have the following equality:

$$\mathcal{D}\left(\mathcal{I}^r, \mathcal{I}^l\right) = \mathcal{D}\left(\frac{I^r}{\sqrt{x^2+y^2}}, \frac{I^l}{\sqrt{x^2+y^2}}\right)$$

According to this, we can consider the correlation of log-polar images as the correlation of spatially weighted cartesian images. The weighting function is given by the inverse of the distance to the image center. Therefore, objects far from the center, have a reduced influence in the perceptual process, and will not perturb the tracking system.

4.2 Control Strategy

The disparity estimates obtained by the perceptual strategy are integrated in a standard feedback architecture. They are acquired at each time step as input to a control strategy. Its output is used as position commands to motor servoing control. A simple proportional controller is used and the integration of both perceptual and control strategies is illustrated in Figure 5. Remember

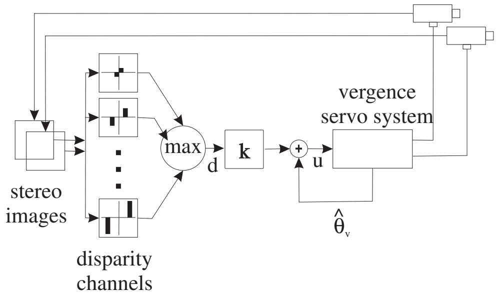

Fig. 5. Vergence visuomotor behaviour.

that we defined a set of disparity channels with a high number of channels for small disparities and fewer channels for large disparities. Therefore, for large displacements, convergence is achieved in two phases: initially, the vergence angle is roughly attracted to the correct angle and after that, a fine adjustment is performed. Interestingly, this control strategy is similar to the human vergence eye movements. The "dual-mode" theory of vergence control [23] considers two phases: a initial coarse and fast transient response with a step-like behavior and a late more accurate component that brings the eyes to their final position. One important difference in our approach is that closed-loop control is used all the time while the "dual-mode" theory argues that the initial phase is a pre-programmed open-loop response.

5 Pursuit Behaviour

We have seen how fixation depth along the gaze direction is controlled by the vergence behaviour. Now we are concerned in controlling the pan (θ_p) and tilt (θ_t) angles which correspond to the gaze azimuth and elevation. Figure 6 illustrates this geometry. Consider a tar-

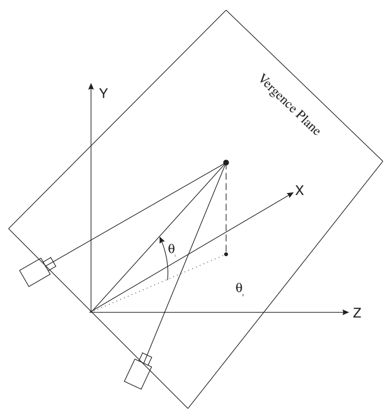

Fig. 6. Gaze direction is controlled by pan and tilt motors

get centered in both images for some

pan and tilt angles, θ_p^0 and θ_t^0. The linearization of the kinematic relations between joint angles and image plane coordinates of the target, results in the following approximate model:

$$\begin{cases} x = G_p\left(\theta_t^0, \theta_v^0\right) \cdot \left(\theta_p - \theta_p^0\right) \\ y = G_t\left(\theta_t^0, \theta_v^0\right) \cdot \left(\theta_t - \theta_t^0\right) \end{cases} \quad (11)$$

and, assuming that the target is static:

$$\begin{cases} v_x = G_p\left(\theta_t^0, \theta_v^0\right) \cdot \omega_p \\ v_y = G_t\left(\theta_t^0, \theta_v^0\right) \cdot \omega_t \end{cases} \quad (12)$$

where v_x and v_y are image plane velocities and ω_p and ω_t are pan and tilt angular velocities. Notice that the sensitivity terms, G_p and G_t, now depend on the current configuration of the stereo head, which did not happen for the vergence case. This is taken into account when the head configuration changes, by recomputing the sensitivity terms.

In biological visual systems, pursuit movements are accomplished by saccadic and smooth pursuit movements that respond to target position and velocity errors in the retina [6]. Similarly, we develop a perceptual module that extracts target position and velocity estimates from the images.

5.1 *Target Segmentation*

To measure target position and velocity in the image planes, the figure-ground segmentation process must be addressed. We should be able to distinguish between target points the background or other objects in the visual field. When the target is verged, its location in each of the stereo images is approximately the same, and the segmentation problem can be addressed by zero disparity filtering [7]. Algorithms to extract zero disparity points in the log-polar images were developed. Figure 7 shows the output of the segmentation process in a pair of real images.

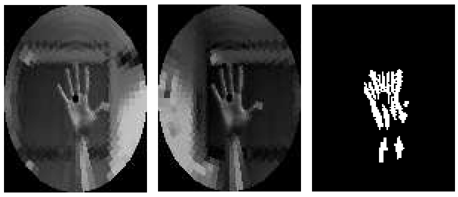

Fig. 7. Figure-ground segmentation: left image, right image and filter output.

5.2 *Motion estimation*

Once the target points have been extracted, we can estimate its motion. Position and velocity estimates are taken from the segmented areas and, although obtained from log-polar images, cartesian coordinates should be used for their representation, because pan and tilt angles depend directly in the horizontal and vertical cartesian coordinates of target motion, as expressed by eq. (11) and (12).

Position Error The position error is defined by the difference between the position of the object and the coordinates of the center of the image. To compute this error, we use the centroid of the object points, which in a continuous cartesian representation is given by:

$$\bar{p} = \frac{\iint_p I(p)\ p\ \mathrm{d}p}{\iint_p I(p)\ \mathrm{d}p}$$

where integration is done for all image domain. The computation, done on log-polar coordinates, is:

$$\bar{p} = \frac{\iint_q \mathcal{I}(q)\ \alpha^{-1}(q)\ |J^{-1}(q)|\ \mathrm{d}q}{\iint_q \mathcal{I}(q)\ |J^{-1}(q)|\ \mathrm{d}q} \tag{13}$$

Velocity Error Velocity estimation could be simply done by differencing the position data, but position estimates are rather noisy. Other alternative is the optical flow [14] which provides velocity estimates for each point in the image. The integration of the normal flow vectors [3] for all the target points, provides a stable velocity measure and good noise rejection. Thus, our perceptual strategy relies on computing the normal flow vectors in the segmented area, $v = (v_x, v_y)$, followed by averaging. To compute the log-polar normal flow vectors, $z = (v_\xi, v_\eta)$ we consider flow vectors as time derivatives of position vectors, $v = \frac{\mathrm{d}p}{\mathrm{d}t}$ and $z = \frac{\mathrm{d}q}{\mathrm{d}t}$, and by simple differential calculus, we obtain the following relation:

$$v(p) = J^{-1}\Big(\alpha(p)\Big) \cdot z\Big(\alpha(p)\Big)$$

where J^{-1} is the inverse log-polar mapping jacobian matrix, given by eq. (6). The average target velocity is given by:

$$\bar{v} = \frac{\iint_p J^{-1}\Big(\alpha(p)\Big) z\Big(\alpha(p)\Big)\,\mathrm{d}p}{\iint_p \mathrm{d}p}$$

[3] Projection of the optical flow vector in the direction of the image gradient. It is the only observable component of the optical flow due to the aperture problem [14].

Converting the integral to log-polar coordinates we obtain:

$$\bar{v} = \frac{\iint_q J^{-1}(q)\, z(q)\, |J^{-1}(q)|\,\mathrm{d}q}{\iint_q |J^{-1}(q)|\,\mathrm{d}q}$$

5.3 Pursuit Controller

Target position and velocity estimates are input to a dynamic controller. The motors are controlled in velocity mode to have smoother trajectories. For each joint j ($j \in \{\text{pan}, \text{tilt}\}$) we assume a simplified second order dynamics state-space model:

$$\Theta'_{\mathbf{j}}(t) = \begin{bmatrix} 0 & 1 \\ 0 & -\tau_j^{-1} \end{bmatrix} \Theta_{\mathbf{j}}(t) + \begin{bmatrix} 0 \\ \tau_j^{-1} \end{bmatrix} u_j(t)$$

where $\Theta_{\mathbf{j}} = (\theta_j, \omega_j)$ is the state variable composed by motor position and velocity, u_j is the velocity command and τ_j is the motors time constant. Starting from this model and using the kinematics relations described in eq. (11) and (12), we design a state estimator and regulator using standard techniques of state-space control theory [11]. In the present case, we used a simple pole-placement method.

6 Coordination

We have presented the *visuomotor behaviours* that compose the tracking system. The vergence and pursuit behaviours extract different image measures and control different motor actions, requiring a low coordination effort. The two behaviours run in parallel most of the time, and the only dependency arises when the vergence behaviour can no longer guarantee correct vergence in the target. In this situation, the segmentation process is not reliable and motion estimates cannot be extracted – the vergence behaviour inhibits the pursuit behaviour. Despite this low internal dependency, vergence and pursuit are highly coupled in an external sense. First, pursuit brings the target to the central area of the images, enhancing the performance of the vergence behavior. Second, vergence provides binocular fusion on the target, enabling good target segmentation and motion estimation, required for pursuit. In this sense, the tracking problem is an example of how separable behaviours, tuned to different goals, cooperate towards a common purpose.

7 Results

The system is implemented on the stereo head Medusa [21] and runs at 25 Hz (full video rate). All processing is done on a PENTIUM-200 computer. To illustrate the reliability and generality of the system we show results for targets with different shapes and motions. In a first set of experiments, the target moved with constant velocity in a linear trajectory. The position of all joints was collected for all time steps and the target trajectory reconstructed in a qualitative way [4]. Figures 8 and 9 show the 3D trajectory reconstruction for a motion in depth and perpendicularly to the initial gaze direction. In all cases, the reconstructed trajectory is qualitatively as expected. A second set of experiments was made with unknown target motion. Figure 10 shows the tracking of a human hand and the output of the segmentation process. In the beginning, fixation is stable in the door at the back of the room (images 1 and 2). Once the hand gets close to the gaze direction, the system starts verging and tracking the target. Notice that the hand is always kept very close to the center of the images despite background motion and target rotation and scaling.

[4] Notice that system calibration and target motion were made in a coarse fashion.

8 Conclusions

Visual tracking is an important subject in robotics research. This paper presented a real-time tracking system which copes with objects of different shapes and motions. The approach is centered in the cooperation of two biologically motivated *visuomotor behaviors* (vergence and pursuit). Vergence performs tracking in depth along the current gaze direction and pursuit controls the gaze direction using motion estimates from the verged object.

Each behaviour is composed by perceptual strategies and control strategies. Perceptual strategies use log-polar geometry to represent images, which proved adequate to extract visual stimuli for tracking purposes. This geometry focus attention in the center of the images, where the target is usually located, and distracting elements in the periphery become less influent in the overall system performance. The

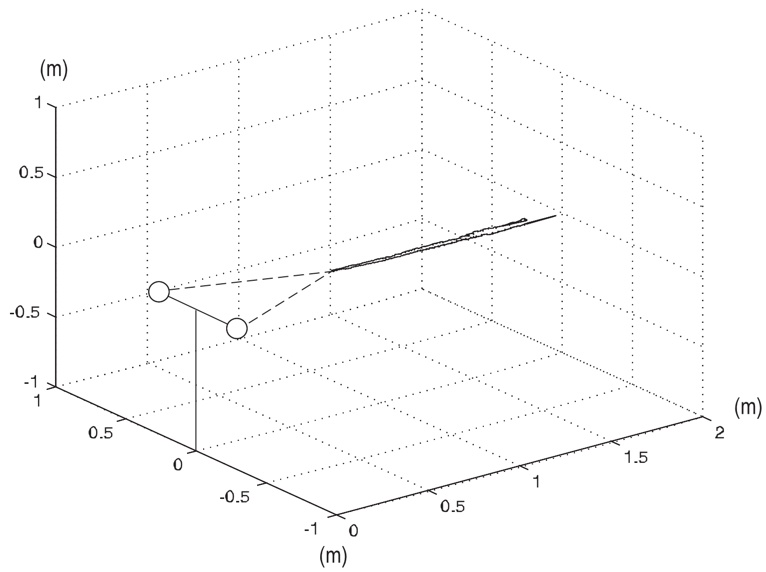

Fig. 8. Depth tracking.

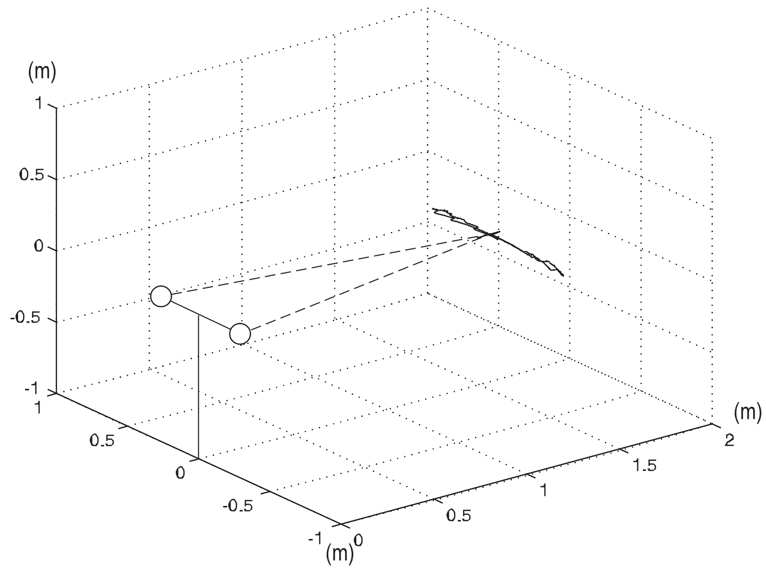

Fig. 9. Fronto-parallel tracking.

perceptual algorithms, based on image correlation, optical flow and disparity filtering, do not assume any previous knowledge about target shape or motion. Furthermore, log-polar mapping reduces the size of the images, allowing faster algorithms and real-time functionality. Motor strategies are based on simplified behaviour models and a feedback control architecture, which compensates for the low precision of the perceptual algorithms and modeling errors, and produces smooth motions even in the presence of considerable amounts of sensor noise.

Integration and coordination of the visual behaviours is simplified by their separable nature. Separate stimuli and motions are addressed by each behaviour allowing low coordination effort.

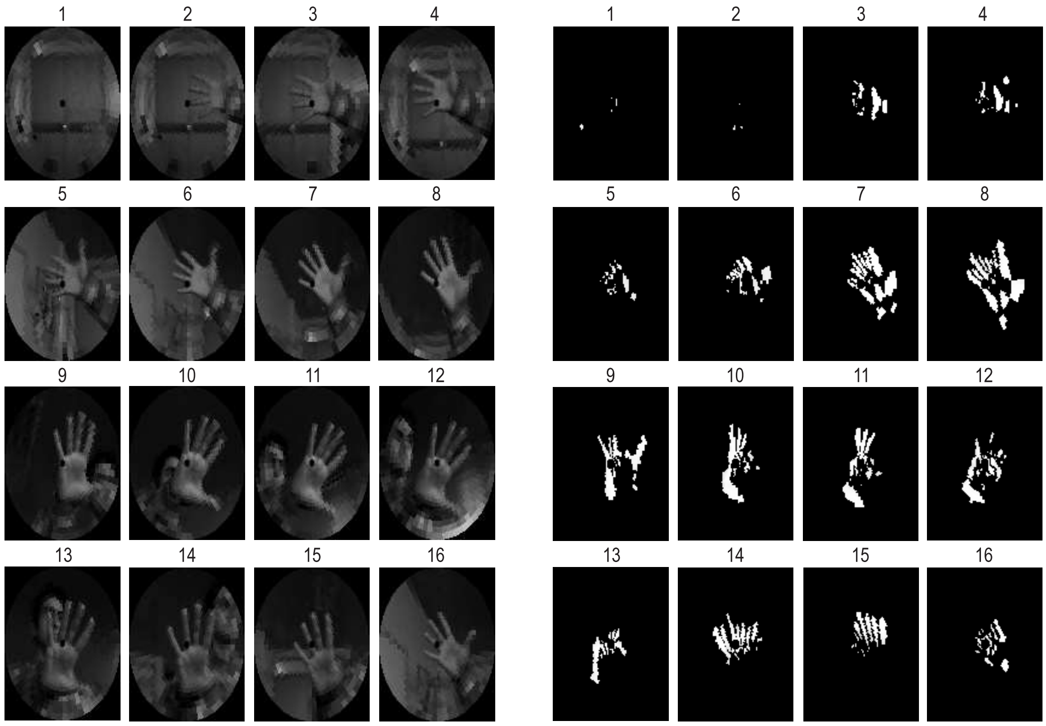

Fig. 10. Hand tracking. Gray-level images (left) and segmentation output (right).

Their close external coupling permits a robust cooperation in a great diversity of situations and conditions – each behaviour enhances the perceptual requirements of the other.

References

[1] R. Andersson. Dynamic sensing in a ping-pong playing robot. *IEEE J. Robotics and Automation*, 5(6):728–739, December 1989.

[2] A. Bernardino and J. Santos-Victor. Sensor geom. for dynamic vergence. In *Workshop on Perform. Charact. of Vision Algorithms*, Cambridge, UK, April 1996.

[3] A. Bernardino and J. Santos Victor. Vergence control for robotic heads using log-polar images. In *Proc. of the IEEE/RJS IROS*, pages 1264–1271, Osaka, Japan, November 1996.

[4] K. Brunnström, J. Eklundh, and T. Uhlin. Active fixation for scene exploration. *IJCV*, 17(2):137–162, February 1996.

[5] C. Capurro, F. Panerai, and G. Sandini. Dynamic vergence using log-polar images. *IJCV*, 24(1):79–94, August 1997.

[6] R. Carpenter. *Movements of the eyes.* Pion, London, 1988.

[7] D. Coombs and C. Brown. Real-time binocular smooth pursuit. *IJCV*, 11(2):147–164, October 1993.

[8] B. Espiau, F. Chaumette, and P. Rives. A new approach to visual servoing in robotics. *IEEE J. Robotics and Automation*, 8(3):313–326, June 1992.

[9] C. Fermüller and Y. Aloimonos. The role of fixation in visual motion analysis. *IJCV*, 11(2):165–186, October 1993.

[10] D. Fleet, A. Jepson, and M. Jenkin. Phase-based disparity measurement. *CVGIP*, 53(2):198–210, March 1991.

[11] G. Franklin, J. Powell, and M. Workman. *Digital Control of Dynamic Systems.* Addison-Wesley, 1990.

[12] D. Gennery. Visual tracking of known three-dimensional objects. *IJCV*, 7(3):243–270, April 1992.

[13] N. Griswald, J. Lee, and C. Weiman. Binocular fusion revisited utilizing a log-polar tessellation. *CVIP*, pages 421–457, 1992.

[14] B. Horn and B. Shunk. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[15] D. Koller, K. Daniilidis, and H. Nagel. Model-based object tracking in monocular image sequences of road traffic scenes. *IJCV*, 10(3):257–281, June 1993.

[16] D. Lowe. Robust model-based motion tracking through the integration of search and estimation. *IJCV*, 8(2):113–122, August 1992.

[17] K. Ludwig, H. Neumann, and B. Neumann. Robust estimation of local stereoscopic depth. In *Robust Computer Vision.* Wichmann, 1992.

[18] M. Pobuda and C. Erkelens. The relation between absolute disparity and ocular vergence. *Biological Cybernetics*, 68:221–228, 1993.

[19] D. Robinson. The oculomotor control system: A review. *Proc. of the IEEE*, 56(6), June 1968.

[20] Santos-Victor, Sandini, Curotto, and Garibaldi. Divergent stereo in autonomous navigation : From bees to robots. *IJCV*, 14(2):159–177, 1995.

[21] J. Santos-Victor, F. van Trigt, and J. Sentieiro. Medusa - a stereo head for active vision. In *Proc. of the Int. Symposium on Intelligent Robotic Systems*, Grenoble, France, July 1994.

[22] E. Schwartz. Spatial mapping in the primate sensory projection : Analytic structure and relevance to perception. *Biological Cybernetics*, 25:181–194, 1977.

[23] J. Semmlow, G. Hung, J. Horng, and K. Ciuffreda. Disparity vergence eye movements exhibit preprogrammed motor control. *Vision Research*, 34(10):1335–1343, 1994.

[24] R. Wallace, P. Ong, B. Bederson, and E. Schwartz. Space variant image processing. *IJCV*, 13(1):71–90, September 1995.