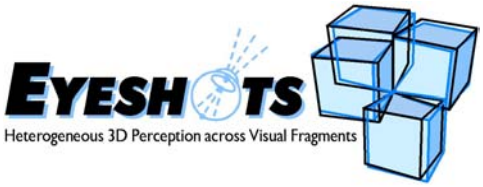


Project no.: FP7-ICT-217077
Project full title: Heterogeneous 3-D Perception across Visual Fragments
Project Acronym: EYESHOTS
Deliverable no: D1.2
Title of the deliverable: Non-visual depth cues and their influence on perception

Date of Delivery:	09 September 2009	
Organization name of lead contractor for this deliverable:	UG	
Author(s):	A. Canessa, M. Chessa, F. Solari, S.P. Sabatini	
Participant(s):	UG	
Workpackage contributing to the deliverable:	WP1	
Nature:	Report	
Version:	2.2	
Total number of pages:	44	
Responsible person:	Silvio P. Sabatini	
Revised by:	M. Lappe, R. Volcic, K. Havermann, S. Wulff	
Start date of project:	1 March 2008	**Duration:** 36 months

Project Co-funded by the European Commission within the Seventh Framework Programme		
Dissemination Level		
PU	Public	**X**
PP	Restricted to other program participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Abstract:

This deliverable reports on the first results of the analysis of the effects of complex binocular motor control strategies of the human eyes on the binocular visual correspondence problem. A general approach is proposed by which both vision and motor efficiency principles would guide proper eyes' postures, also taking into account the resources that motor and vision systems have at disposition. Possible strategies for embedding binocular fixation constraints in the neural mechanisms that underlie stereopsis (e.g., by gain modulations) are suggested.

Contents

1 Executive summary

One of the goals of the EYESHOTS project is to study the perceptual consequences of specific binocular eye coordination movements and their computational advantages on depth vision and interactive stereopsis. Relying on these specific motor behaviours, it is expected to improve the performances of the stereo vision modules of an active robot head, already at an early level of vision processing. Towards this goal, in WP1 (Task 1.2) we investigate how complex binocular motor control strategies of human eyes can help the brain to solve the binocular visual correspondence problem by interacting with the neural mechanisms that underlie stereopsis. This deliverable reports on the first results achieved in that direction, and specifically addresses the problem from a static (i.e. geometric) perspective by studying: (1) the perceptual reasons of the Listings law under the assumption of a global visuo-motor optimization strategy adopted by the oculomotor system; (2) the effects of the resulting torsional components on the perception of the visible surfaces on an observed object (i.e., local patches around the fixation point) for different gaze directions; (3) the different strategies we can adopt to embed fixation constraints posed by the oculomotor system into the binocular energy-based models of depth perception developed in WP2. In addition, we report on the realization of a VR simulator for binocular active vision systems, which can be used both for algorithmic and behavioural benchmarks for the whole duration of the project. The novelty of the approach is the use of VR as a tool to simulate, in closed perception/action loop, the behaviour of a binocular vision system that observes the scene, rather than just rendering the 3D perceptual illusion of the scene to a human observer.

Although the study of the perceptual consequences of Listings Law and its family of motor constraints has a long and rich history, dating back to Donders and von Helmholtz, their perceptual consequences still remain an open issue. We believe that the advantages of binocular visuo-motor strategies could be fully understood only if one jointly analyzes and models the problem of neural computation of stereo information, and if one takes into account the limited accuracy of the motor system. Unfortunately, models in this joint field are very seldom [45], [33], [17] and rarely address all the computational issues. In absence of such models, so far in robot vision, rectification techniques simply remove the problem by searching for correspondences along the epipolar lines or disregarding vertical disparities, but removing, in this way, any cognitive value related to active 3D eye movements in purposive vision. Hence, the computational principles pointed out by the analysis of the problem conducted by UG have been developed concurrently with the models defined in WP2 (cf. Task T2.1 and Task 2.2), in collaboration with K.U.Leuven.

We conclude that:

1. The results evidenced that the eyes should move both to maintain the coplanarity of the fixation planes (a property of a tilt-pan system) and to reduce the eccentricity of the rotation. Our approach confirms the experimental evidences present in the literature for large and small vergences, and proposes itself as a general model, forming a bridge between these two extremes (even for non-null version conditions). The resulting mean disparities pattern are strongly dependent on the current epipolar geometry of the system.

2. It is possible to introduce specific design strategies to modify the architectural parameters of the distributed representation of the disparity information. Predictable components of the disparity, which are related to the positions of the eyes in the orbits, have been profitably used to constraint the neural coding and decoding mechanisms of the population of binocular energy units.

The results reported in the presented study have been partially presented at the VISAPP'08, ICCNS'09, ECVP'09.

2 Introduction

The study of the perceptual consequences of Listings Law and its family of motor constraints has a long and rich history, dating back to Donders and von Helmholtz. Since the establishment of the validity of Listings Law in far vision, several studies on the binocular control of eye movements have focused on the deviations from Listing's law in near vision [28][34][26]. All these studies agree that eyes movements have a torsional component that varies with vergence and gaze elevation to reduce the cyclovergence and restricts the motion of the epipolar line, thus permitting stereo matching to work with smaller search zones [40]. Then, it has been demonstrated recently that the control of ocular torsion can be changed by a cyclodisparity stimulus. This suggests a view where ocular torsions are dynamically controlled to optimize binocular image alignment and simplify the perception of slanted surface [39][24][44], [27]. Notwithstanding the several lines of evidence of complex binocular motor control strategies of human eyes, their perceptual consequences still remain an open question considering that, in principle, the brain could solve the binocular visual correspondence problem by using 3D feedback signals for the orientation of both eyes [8]. From a geometrical point of view, eye rotations are usually decomposed into three sequential rotations about hierarchically nested axes (Helmholtz and Fick systems), but these description are strongly dependent on the order of these rotations. Even when quaternion algebra is adopted, Helmholtz or Fick angles are still commonly used to characterize the rotation vector components [18]. A new characterization of eye movements is developed (Section 3) that is dependent on the coordinates of the fixation point only, independently of the rotation system adopted. The experimental evidences (as Listing's Law and its binocular extension, L2) is taken into account. On the basis of this new characterization, in Section 4, we carry on a mathematical analysis to obtain optimal eye movements that maximize both motor efficiency and the perceptual advantages for stereo vision. The results evidence that the eyes should move both to maintain the coplanarity of the fixation planes (a property of a tilt-pan system) and to reduce the eccentricity of the rotation. Our approach confirms the experimental evidences presented in literature for large and small vergences, and proposes itself as a general model, forming a bridge between these two extremes (even for non-null version conditions). The perceptual implications of binocular eye coordination on disparity estimation are discussed in Section 5, where furthermore a Virtual Reality Tool for simulating the behaviour of an active vision system in real world situations is presented.

3 Geometric constraints in binocular eye coordination

3.1 Mathematics of 3D eye movements

The movement of the eye is the movement of a rigid body in a 3D space. By first approximation we can consider the eye as a center-fixed sphere, so its position is characterized only by a rotation around its center, since the translation can be neglected. Eye positions are usually classified in three groups: primary, secondary and tertiary positions. In primary position, the eye looks straight ahead and in this position the muscles exhibit the minimum force. From primary position any rotation about either the vertical or the horizontal axis brings the eye in a secondary position. In this case the eye looks to the left or to the right, or up or down. With a combination of rotation around both the horizontal and vertical axis the eye turns to a tertiary position. The target eye position is defined through the 3D rotation that, from a somewhat arbitrarily chosen reference position, brings the eye to that position. This reference position is usually defined as the position the eye assumes when the subject is looking straight ahead, while the head is kept upright. In order to describe the 3D eye position in space we need to define two coordinate frames: one head-fixed and one eye-fixed. Let $\langle h \rangle = \left\{ \bar{\boldsymbol{h}}_{\mathrm{X}}, \bar{\boldsymbol{h}}_{\mathrm{Y}}, \bar{\boldsymbol{h}}_{\mathrm{Z}} \right\}$ be the head-fixed and $\langle e \rangle = \{ \bar{\boldsymbol{e}}_{\mathrm{X}}, \bar{\boldsymbol{e}}_{\mathrm{Y}}, \bar{\boldsymbol{e}}_{\mathrm{Z}} \}$ be the eye-fixed, both right handed coordinate systems. $\bar{\boldsymbol{h}}_{\mathrm{Z}}$ points forward in the midplane close to the center of the oculomotor range, $\bar{\boldsymbol{h}}_{\mathrm{X}}$ points leftward through the inter-aural axis and $\bar{\boldsymbol{e}}_{\mathrm{Z}}$ points along the line of sight, and coincides with $\bar{\boldsymbol{h}}_{\mathrm{Z}}$ when the eye is in the reference position, i.e. when the eye looks straight ahead whilst the head is kept upright. These two systems have the same origin in the center of each eye. The configuration of the eye is completely determined if we know the position of the eye-fixed frame relative to the head-fixed one, i.e if we know the direction cosines between each couple of axes of the two systems. The 3×3 set of direction cosine define a transformation matrix $\mathbf{R}$ between the two systems. This matrix describes the mapping from a coordinate system to the other one and we can see it as an operator that transforms one reference frame into the other. Formally, we can write:

$$^{h}\boldsymbol{p} = {}^{\mathrm{h}}_{\mathrm{e}}\mathbf{R}\, {}^{e}\boldsymbol{p}, \tag{1}$$

in which ${}^{\mathrm{h}}_{\mathrm{e}}\mathbf{R}$ acts on the components of the vector ${}^{e}\boldsymbol{p}$ relative to the eye-fixed system, transforming it in the components of the same vector relative to the head-fixed system. Alternatively, the same operator ${}^{\mathrm{h}}_{\mathrm{e}}\mathbf{R}$ can be interpreted as an operator that acts on a vector ${}^{h}\boldsymbol{p}$ by rotating it into another vector ${}^{h}\boldsymbol{p'}$

$$^{h}\boldsymbol{p'} = {}^{\mathrm{h}}_{\mathrm{e}}\mathbf{R}\, {}^{h}\boldsymbol{p}\,, \tag{2}$$

both relative to the same head-fixed coordinate system. In the first case we are considering a rotation of the coordinate system, usually called passive rotation. On the contrary, in the second case we are considering a rotation of the single vector, usually called active rotation. In general, in this formulation we will consider active rotation with respect to the head-fixed system, if not differently indicated. Actually, it is possible to demonstrate

that only a set of three independent variables, function of the nine direction cosines, is sufficient to express the elements of the matrix $\mathbf{R}$.

In general the transformation of a coordinate system into another one can be obtained by three consecutive rotations in a well defined order about hierarchically nested axes. The angles associated with these rotations are three independent variables that can be chosen arbitrary according to different conventions used in every branch of mathematics, physics and engineering. Among them, the most commonly used in the eye movement research field are: the rotations specified by the Tait-Brian angles in the order defined by the Helmholtz sequence and the one specified by the Fick sequence. The Tait-Brian angles are also known as the Yaw, Pitch and Roll angle. Here we will refer to them as the azimuth H, elevation V and torsion T angles.

3.1.1 Helmholtz vs. Fick sequences

The Helmholtz sequence starts with a rotation by an angle T_{H} along the $\bar{\boldsymbol{h}}_{\mathrm{Z}}$ axis, follow by a rotation by an angle H_{H} along the $\bar{\boldsymbol{h}}_{\mathrm{Y}}$ axis and finally a rotation by an angle V_{H} along the $\bar{\boldsymbol{h}}_{\mathrm{X}}$ axis. The subscript H stands for Helmholtz. For the sake of compactness, for any angle A, CA and SA denote $\cos(A)$ and $\sin(A)$, respectively.

The first rotation is described through the matrix $\mathbf{R}_{T_{\mathrm{H}}}$:

$$\bar{\boldsymbol{e}}_k = \mathbf{R}_{T_{\mathrm{H}}} \bar{\boldsymbol{h}}_{\mathbf{k}}, \qquad \mathbf{k} = \mathrm{X}, \mathrm{Y}, \mathrm{Z} \tag{3}$$

where

$$\mathbf{R}_{T_{\mathrm{H}}} = \begin{pmatrix} \mathrm{C}T_{\mathrm{H}} & -\mathrm{S}T_{\mathrm{H}} & 0 \\ \mathrm{S}T_{\mathrm{H}} & \mathrm{C}T_{\mathrm{H}} & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{4}$$

The second rotation is described through the matrix $\mathbf{R}_{H_{\mathrm{H}}}$:

$$\bar{\boldsymbol{e}}_k = \mathbf{R}_{H_{\mathrm{H}}} \bar{\boldsymbol{h}}_{\mathbf{k}}, \qquad \mathbf{k} = \mathrm{X}, \mathrm{Y}, \mathrm{Z} \tag{5}$$

where

$$\mathbf{R}_{H_{\mathrm{H}}} = \begin{pmatrix} \mathrm{C}H_{\mathrm{H}} & 0 & -\mathrm{S}H_{\mathrm{H}} \\ 0 & 1 & 0 \\ \mathrm{S}H_{\mathrm{H}} & 0 & \mathrm{C}H_{\mathrm{H}} \end{pmatrix} \tag{6}$$

The third rotation is described through the matrix $\mathbf{R}_{V_{\mathrm{H}}}$:

$$\bar{\boldsymbol{e}}_k = \mathbf{R}_{V_{\mathrm{H}}} \bar{\boldsymbol{h}}_{\mathbf{k}}, \qquad \mathbf{k} = \mathrm{X}, \mathrm{Y}, \mathrm{Z} \tag{7}$$

where

$$\mathbf{R}_{V_{\mathrm{H}}} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & \mathrm{C}V_{\mathrm{H}} & -\mathrm{S}V_{\mathrm{H}} \\ 0 & \mathrm{S}V_{\mathrm{H}} & \mathrm{C}V_{\mathrm{H}} \end{pmatrix} \tag{8}$$

The complete transformation matrix $\mathbf{R}_{\mathrm{H}}$ is obtained by multiplying in cascade the matrices of the three single rotations:

$$\mathbf{R}_{\mathrm{H}} = \mathbf{R}_{V_{\mathrm{H}}} \mathbf{R}_{H_{\mathrm{H}}} \mathbf{R}_{T_{\mathrm{H}}} \tag{9}$$

thus obtaining:

$$\mathbf{R_H} = \begin{pmatrix} \mathrm{C}H_\mathrm{H}\mathrm{C}T_\mathrm{H} & -\mathrm{C}H_\mathrm{H}\mathrm{S}T_\mathrm{H} & \mathrm{S}H_\mathrm{H} \\ \mathrm{S}V_\mathrm{H}\mathrm{S}H_\mathrm{H}\mathrm{C}T_\mathrm{H} + \mathrm{C}V_\mathrm{H}\mathrm{S}T_\mathrm{H} & -\mathrm{S}V_\mathrm{H}\mathrm{S}H_\mathrm{H}\mathrm{S}T_\mathrm{H} + \mathrm{C}V_\mathrm{H}\mathrm{C}T_\mathrm{H} & -\mathrm{S}V_\mathrm{H}\mathrm{C}H_\mathrm{H} \\ -\mathrm{C}V_\mathrm{H}\mathrm{S}H_\mathrm{H}\mathrm{C}T_\mathrm{H} + \mathrm{S}V_\mathrm{H}\mathrm{S}T_\mathrm{H} & \mathrm{C}V_\mathrm{H}\mathrm{S}H_\mathrm{H}\mathrm{S}T_\mathrm{H} + \mathrm{S}V_\mathrm{H}\mathrm{C}T_\mathrm{H} & \mathrm{C}V_\mathrm{H}\mathrm{C}H_\mathrm{H} \end{pmatrix} \tag{10}$$

In the Fick sequence, instead, we have first a rotation by an angle T_F along the $\bar{\boldsymbol{h}}_\mathrm{Z}$ axis, followed by a rotation by an angle V_F along the $\bar{\boldsymbol{h}}_\mathrm{X}$ axis and finally a rotation by an angle H_F along the $\bar{\boldsymbol{h}}_\mathrm{Y}$ axis. The subscript F stands for Fick. The matrices of the single rotations are equal to those detailed above. The final transformation matrix $\mathbf{R}_\mathrm{F}$ is different:

$$\mathbf{R}_\mathrm{F} = \mathbf{R}_{H_\mathrm{F}}\mathbf{R}_{V_\mathrm{F}}\mathbf{R}_{T_\mathrm{F}} \tag{11}$$

$$\mathbf{R}_\mathrm{F} = \begin{pmatrix} \mathrm{C}H_\mathrm{F}\mathrm{C}T_\mathrm{F} + \mathrm{S}H_\mathrm{F}\mathrm{S}V_\mathrm{F}\mathrm{S}T_\mathrm{F} & -\mathrm{C}H_\mathrm{F}\mathrm{S}T_\mathrm{F} + \mathrm{S}H_\mathrm{F}\mathrm{S}V_\mathrm{F}\mathrm{C}T_\mathrm{F} & \mathrm{S}H_\mathrm{F}\mathrm{C}V_\mathrm{F} \\ \mathrm{C}V_\mathrm{F}\mathrm{S}T_\mathrm{F} & \mathrm{C}V_\mathrm{F}\mathrm{C}T_\mathrm{F} & -\mathrm{S}V_\mathrm{F} \\ -\mathrm{S}H_\mathrm{F}\mathrm{C}T_\mathrm{F} + \mathrm{C}H_\mathrm{F}\mathrm{S}T_\mathrm{F}\mathrm{S}V_\mathrm{F} & \mathrm{S}H_\mathrm{F}\mathrm{S}T_\mathrm{F} + \mathrm{C}H_\mathrm{F}\mathrm{C}T_\mathrm{F}\mathrm{S}V_\mathrm{F} & \mathrm{C}V_\mathrm{F}\mathrm{C}H_\mathrm{F} \end{pmatrix}. \tag{12}$$

It is worth noting that the same eye position is characterized by different values for the angles when described according to the Helmholtz or the Fick sequence. That is, different sequences of rotations lead to different azimuthal, elevational and torsional angle values for the same position of the eye. Considering a complete rotation as a composition of single standard rotations is not the only way. Instead, by a fundamental property of rigid body motion - the *Euler's Theorem* - for every two orientations of an object, the object can always move from the initial to the final position by a single rotation by an angle ε about a fixed axis $\bar{\boldsymbol{n}}$. An equivalent representation of the transformation of the coordinate system, as a function of the rotation angle ε and the versor of the axis $\bar{\boldsymbol{n}}$, can be derived by introducing *quaternions* and *rotation vectors* algebra.

3.1.2 Quaternions and rotation vectors

Quaternions provide a convenient mathematical notation for representing orientations and rotations of objects in three dimensions. We can think of a quaternion as a 3D vector augmented by a real number to make it a four element entity: this is usually called a hypercomplex number. Accordingly, quaternions are defined as the sum of four terms in the form:

$$q = 1 \cdot q_0 + i \cdot q_1 + j \cdot q_2 + k \cdot q_3 \tag{13}$$

where q_0, q_1, q_2 e q_3 are real numbers and i, j, k are symbolic elements with the following properties:

$$i^2 = j^2 = k^2 = ijk = -1 \tag{14}$$

The quaternion $q = 1 \cdot q_0 + i \cdot q_1 + j \cdot q_2 + k \cdot q_3$ can be interpreted as it would have a scalar component q_0 and a vectorial component ($\boldsymbol{q} = i \cdot q_1 + j \cdot q_2 + k \cdot q_3$), in which to the elements i, j, k it is possible to add a geometrical interpretation, considering them as the versors of $\langle h \rangle$. In quaternion notation, a rotation by an angle ε around an axis $\bar{\boldsymbol{n}}$ is represented by a quaternion

$$q = q_0 + \boldsymbol{q} \tag{15}$$

where

$$q_0 = \cos(\varepsilon/2) \text{ and } \boldsymbol{q} = |\boldsymbol{q}|\,\bar{\boldsymbol{n}} \tag{16}$$

with

$$|\boldsymbol{q}| = \sqrt{q_1^2 + q_2^2 + q_3^2} = \sin(\varepsilon/2) \tag{17}$$

For these reasons a quaternion is often represented in this form:

$$q = \cos(\varepsilon/2) + \sin(\varepsilon/2)\bar{\boldsymbol{n}} \tag{18}$$

The angle ε by which to rotate is usually called *rotation eccentricity*.

If $\boldsymbol{p}$ is a vector with three components, $\boldsymbol{p}' = q \circ \boldsymbol{p} \circ q^{-1}$ is the vector obtained after $\boldsymbol{p}$ is rotated by an angle ε about an axis parallel to $\boldsymbol{q}$, where q^{-1}, for unit norm quaternion, is expressed as

$$q^{-1} = q_0 - \boldsymbol{q} \tag{19}$$

while the product $\circ$ between two quaternions is defined as

$$q \circ s = (q_0 s_0 - \boldsymbol{q} \cdot \boldsymbol{s}) + (q_0 \boldsymbol{s} + s_0 \boldsymbol{q} + \boldsymbol{q} \times \boldsymbol{s}) \tag{20}$$

The rotation vector is only a different way to interpret a quaternion. In fact, since the scalar component q_0 and the norm of the vectorial component $|\boldsymbol{q}|$ contain the same information about the rotation, we can collapse them in a single term given by their quotient. The rotation vector $\boldsymbol{r}_q$ associated with a quaternion q that describes a rotation ε around an axis whose versor is $\bar{\boldsymbol{n}}$, is given by

$$\boldsymbol{r}_q = \frac{\boldsymbol{q}}{q_0} = \tan(\varepsilon/2)\frac{\boldsymbol{q}}{|\boldsymbol{q}|} = \tan(\varepsilon/2)\bar{\boldsymbol{n}} \tag{21}$$

In terms of rotation vectors the product $\circ$ becomes:

$$\boldsymbol{r}_q \circ \boldsymbol{r}_s = \frac{\boldsymbol{r}_q + \boldsymbol{r}_s + \boldsymbol{r}_q \times \boldsymbol{r}_s}{1 - \boldsymbol{r}_q \cdot \boldsymbol{r}_s}\,. \tag{22}$$

Within this framework, let us suppose we have two points $\bar{\boldsymbol{u}}$ and $\bar{\boldsymbol{v}}$ on the unit sphere, and that we want to rotate $\bar{\boldsymbol{u}}$ in such a way that it maps onto $\bar{\boldsymbol{v}}$, and a third point $\bar{\boldsymbol{w}}$ gets mapped onto itself. Which is the quaternion associated to this rotation? The rotation axis is given if we identify it with $\bar{\boldsymbol{w}}$. The angle ε remains undetermined. As shown in Figure 1, we can verify that the rotation angle is given by:

$$\cos\varepsilon = \frac{\bar{\boldsymbol{w}} \times \bar{\boldsymbol{u}}}{|\bar{\boldsymbol{w}} \times \bar{\boldsymbol{u}}|} \cdot \frac{\bar{\boldsymbol{w}} \times \bar{\boldsymbol{v}}}{|\bar{\boldsymbol{w}} \times \bar{\boldsymbol{v}}|} \tag{23}$$

Thus, by substituting Eq .23 into Eq .18 and by trigonometric manipulations, we can derive the expression of the quaternion that characterizes the desired rotation:

$$q = \sqrt{\left(\frac{1 + \frac{\bar{\boldsymbol{w}}\times\bar{\boldsymbol{u}}}{|\bar{\boldsymbol{w}}\times\bar{\boldsymbol{u}}|} \cdot \frac{\bar{\boldsymbol{w}}\times\bar{\boldsymbol{v}}}{|\bar{\boldsymbol{w}}\times\bar{\boldsymbol{v}}|}}{2}\right)} + \sqrt{\left(\frac{1 - \frac{\bar{\boldsymbol{w}}\times\bar{\boldsymbol{u}}}{|\bar{\boldsymbol{w}}\times\bar{\boldsymbol{u}}|} \cdot \frac{\bar{\boldsymbol{w}}\times\bar{\boldsymbol{v}}}{|\bar{\boldsymbol{w}}\times\bar{\boldsymbol{v}}|}}{2}\right)}\bar{\boldsymbol{w}} \tag{24}$$

$$= \cos(\varepsilon/2) + \sin(\varepsilon/2)\bar{\boldsymbol{w}} \tag{25}$$

or, equivalently, the rotation vector:

$$\boldsymbol{r} = \frac{\frac{\bar{\boldsymbol{w}}\times\bar{\boldsymbol{u}}}{|\bar{\boldsymbol{w}}\times\bar{\boldsymbol{u}}|} \times \frac{\bar{\boldsymbol{w}}\times\bar{\boldsymbol{v}}}{|\bar{\boldsymbol{w}}\times\bar{\boldsymbol{v}}|}}{1 + \frac{\bar{\boldsymbol{w}}\times\bar{\boldsymbol{u}}}{|\bar{\boldsymbol{w}}\times\bar{\boldsymbol{u}}|} \cdot \frac{\bar{\boldsymbol{w}}\times\bar{\boldsymbol{v}}}{|\bar{\boldsymbol{w}}\times\bar{\boldsymbol{v}}|}} = \tan(\varepsilon/2)\bar{\boldsymbol{w}} \tag{26}$$

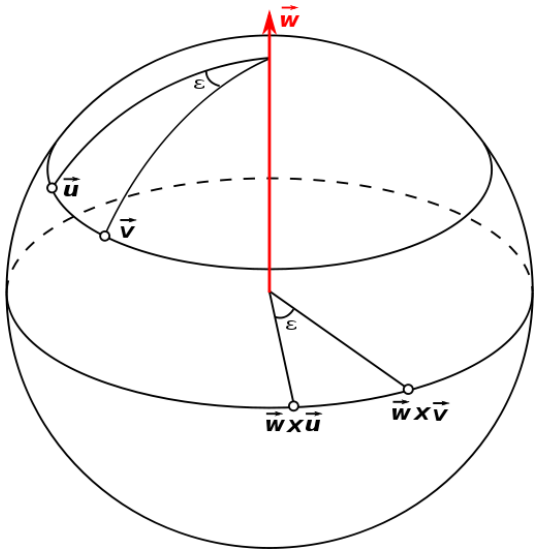


Figure 1: The versor $\bar{\boldsymbol{u}}$ rotates into the versor $\bar{\boldsymbol{v}}$ while the versor $\bar{\boldsymbol{w}}$ remains fixed. Projecting $\bar{\boldsymbol{u}}$ and $\bar{\boldsymbol{v}}$ onto the plane whose normal is $\bar{\boldsymbol{w}}$ it is possible to derive the angle ε by which to rotate.

As previously stated, the eye, like any rigid body, has three degrees of freedom. Though, only two angles are sufficient to determine the gaze direction for an eye: namely the azimuth and the elevation of the target. This implies that the eye could theoretically assume an infinite number of torsional posture for any gaze direction. In other words, there are infinite ways to fixate any given target.

3.2 Listing's Law [This section has been reported here from deliverable D1.1 for the sake of clarity]

Donders [9] discovered that, for a steady fixation condition with the head upright, the actual positions of the eye are restricted in such a way that there is only one eye position for every gaze direction. In other words, Donders found that the movement of the eye is restricted to a two-dimensional (2D) subspace of the whole three-dimensional (3D) space of all possible orientations. He observed that the torsional eye position is univocally related to the current pair of horizontal and vertical eye position, and postulated that the torsional position of the eye is always the same, independent of how the eye reaches a particular gaze direction. Listing's law goes one step further, by specifying the amount of such an ocular torsion. Listing's law states that, when the head is fixed, the eye assumes only those orientations that can be reached from primary position by a single rotation about an axis in a plane called Listing's plane. This plane is orthogonal to the line of

sight when the eye is in the primary position [53]. In other words, one can visualize any given eye movement as caused by rotation about an axis. The collection of these axes for all the rotations that start from primary position constitutes Listing's plane, see Figure 2.

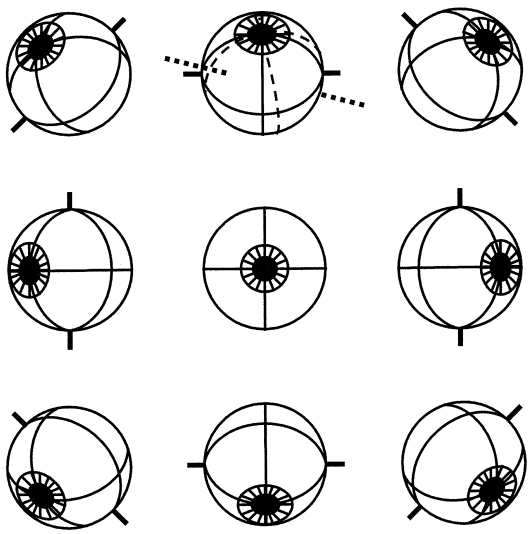


Figure 2: The nine orientations drawn as solid line correspond to the Listing's Law: they are obtained through rotation from the primary position onto these positions, about axes (thick solid lines) that lie on the Listing's plane (in this case represented by the paper plane). The position drawn in dashed line at the top center does not obey Listing's Law because the rotation to this position from primary position occurs about an axis (thick dotted line) that is tilted out the paper plane.

3.3 Binocular extension of Listing's Law (L2)

Listing's law applies when the eye fixates a target at optical infinity. However, the torsional posture of each eye changes when the eyes converge on a near object [4][26][28][30][42][52][34]. During convergence, the eyes' rotation axes still remains confined to planes for any vergence angle; however, as the eyes converge, these planes rotate temporally and roughly symmetrically by ϕ_l and ϕ_r angle, for the left and the right eye, respectively (see Figure 3). These convergence-dependent changes of torsional positions (i.e., orientation of Listing's plane) have been referred to as the binocular extension of Listing's law or L2. It is worth noting that L2 is a generalization of the original, monocular, Listing's Law, and reduces to it when the vergence angle is zero, as it occurs when the eye fixates a distant object.

In other words, as long as the vergence angle is fixed, there is still one and only one torsional position that the eye adopts for any gaze direction, but the torsion can vary when vergence changes. The more convergence exists, the more the plane rotates temporally, implying that during convergence, there is a relative excyclotorsion on upgaze, and a relative incyclotorsion on downgaze.

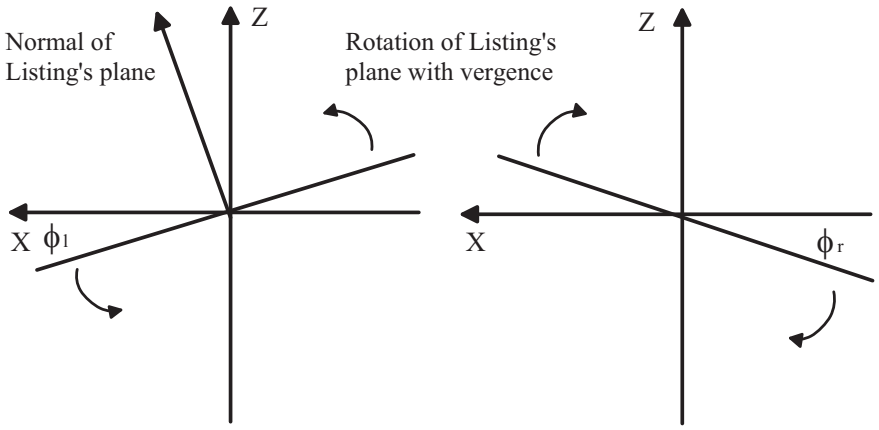


Figure 3: Binocular extension of Listing's Law. During convergence, the Listing's plane is rotated temporally and symmetrically in each eye by an angle ϕ proportional to the vergence angle.

From the experimental data emerged a proportionality between the ϕ angles and the vergence, so that in the literature it is well consolidated to express the ϕs as a linear function of the vergence angle ν:

$$\phi_l = \mu_l \nu \tag{27}$$
$$\phi_r = -\mu_r \nu \tag{28}$$

where μ_l and μ_r are positive constants ranging between 0.20 and 0.41, values derived by fitting of experimental data [53][4][26][28][30][42][52][34]. Though, the values of those proportionality constants, and thus those of the rotation angles ϕs are controversial. As a further generalization, we have to consider that the ϕs, as we said above, are related to the primary position, which seems to be temporally rotated in each eye by some angle λ_l and λ_r too, even in the case of a null vergence. In this case, the overall rotation of the Listing's plane becomes, for both eyes:

$$\psi_l = \lambda_l + \phi_l = \lambda_l + \mu_l \nu \tag{29}$$
$$\psi_r = \lambda_r + \phi_r = \lambda_r - \mu_r \nu \, . \tag{30}$$

Thus, we can write the normals of the two Listing's planes for the left and the right eye as:

$$\bar{\boldsymbol{n}}^l_{\mathscr{L}} = [\sin\psi_l, 0, \cos\psi_l] \tag{31}$$
$$\bar{\boldsymbol{n}}^r_{\mathscr{L}} = [\sin\psi_r, 0, \cos\psi_r] \tag{32}$$

3.4 A new approach for describing binocular cyclo-roations

On the basis of the experimental evidences [28] and by considering the mathematical formalism introduced in section 2, we express the orientation of the eyes through the

quaternions, in order to have no dependencies both on the particular rotations adopted, and on the particular sequence followed, as it occurs when one uses the rotation matrices. Eq .23 and Eq .25 allow us to express the quaternion that maps a vector $\bar{\boldsymbol{u}}$ onto a vector $\bar{\boldsymbol{v}}$, given the versor $\bar{\boldsymbol{w}}$ of the rotation axis. We denote with $\bar{\boldsymbol{v}}^L = [v^l_{\mathrm{X}}, v^l_{\mathrm{Y}}, v^l_{\mathrm{Z}}]$ and $\bar{\boldsymbol{v}}^R = [v^r_{\mathrm{X}}, v^r_{\mathrm{Y}}, v^r_{\mathrm{Z}}]$ the versors of a generic target $\boldsymbol{v}$ with respect to the two reference frames $\langle h \rangle^L$ and $\langle h \rangle^R$ of both eyes. Then, we identify $\bar{\boldsymbol{u}}^L = [u^l_{\mathrm{X}}, u^l_{\mathrm{Y}}, u^l_{\mathrm{Z}}]$ and $\bar{\boldsymbol{u}}^R = [u^r_{\mathrm{X}}, u^r_{\mathrm{Y}}, u^r_{\mathrm{Z}}]$ as directed along the lines of sight $\bar{\boldsymbol{e}}^L_{\mathrm{Z}}$ and $\bar{\boldsymbol{e}}^R_{\mathrm{Z}}$ of both eye when they are in their reference positions. These coincide with their primary positions and also with the versors $\bar{\boldsymbol{h}}^L_{\mathrm{Z}}$ and $\bar{\boldsymbol{h}}^R_{\mathrm{Z}}$. Let us first consider a single eye and the problem of aligning the gaze in the target's direction. We have to determine the versor $\bar{\boldsymbol{w}}$ of the axis around which to rotate the eye. The location of all the rotation axes that are instrumental to map the vector $\bar{\boldsymbol{u}}$ (line of sight) onto the position vector $\bar{\boldsymbol{v}}$ of the target is illustrated in Figure 4, by specifying two rotation axes that bring $\bar{\boldsymbol{u}}$ into $\bar{\boldsymbol{v}}$. The first rotation axis, obviously is given by the cross product $\bar{\boldsymbol{u}} \times \bar{\boldsymbol{v}}$. This axis is normal to the plane that contains $\bar{\boldsymbol{u}}$ and $\bar{\boldsymbol{v}}$ and it maps $\bar{\boldsymbol{u}}$ onto $\bar{\boldsymbol{v}}$ along the great circle. The second one is directed as the sum $\bar{\boldsymbol{u}} + \bar{\boldsymbol{v}}$. This axis lies on the plane containing $\bar{\boldsymbol{u}}$ and $\bar{\boldsymbol{v}}$, it bisects the angle between them, and with a rotation of π takes $\bar{\boldsymbol{u}}$ onto $\bar{\boldsymbol{v}}$. These two axes define a plane through the origin that represents the locus of all the possible rotation axes. The normal to this plane is given by:

$$\bar{\boldsymbol{n}}_p = (\bar{\boldsymbol{u}} \times \bar{\boldsymbol{v}}) \times (\bar{\boldsymbol{u}} + \bar{\boldsymbol{v}}) \equiv (\bar{\boldsymbol{v}} - \bar{\boldsymbol{u}}) \,. \tag{33}$$

The approach can be extended straightforwardly to the binocular case (see Figure 5b) thus yielding to a pair of planes whose normal versors for the left and the right eye are given by:

$$\bar{\boldsymbol{n}}^l_p = (\bar{\boldsymbol{u}}^l \times \bar{\boldsymbol{v}}^l) \times (\bar{\boldsymbol{u}}^l + \bar{\boldsymbol{v}}^l) \equiv (\bar{\boldsymbol{v}}^l - \bar{\boldsymbol{u}}^l) \tag{34}$$

$$\bar{\boldsymbol{n}}^r_p = (\bar{\boldsymbol{u}}^r \times \bar{\boldsymbol{v}}^r) \times (\bar{\boldsymbol{u}}^r + \bar{\boldsymbol{v}}^r) \equiv (\bar{\boldsymbol{v}}^r - \bar{\boldsymbol{u}}^r) \tag{35}$$

Among all the possible axes, we know, from experimental evidences, that the eyes adopt those orientation obtained by rotating along axes confined on the Listing plane, only. Now, for each eye, we have two planes: the first one contains all the axes that take $\bar{\boldsymbol{u}}$ into $\bar{\boldsymbol{v}}$, whereas the second one specifies a constraint for the possible orientation that they can assume. The intersection of these two planes defines the axis about which the eyes have to rotate (see Figure 5).

Formally, by solving Eqs .(34)-(35) and Eqs .(31)-(32) we find the expressions of the rotation axes $\bar{\boldsymbol{w}}^l$ and $\bar{\boldsymbol{w}}^r$:

$$\bar{\boldsymbol{w}}^l = \frac{\bar{\boldsymbol{n}}^l_{\mathscr{L}} \times \bar{\boldsymbol{n}}^l_p}{\left|\bar{\boldsymbol{n}}^l_{\mathscr{L}} \times \bar{\boldsymbol{n}}^l_p\right|} \tag{36}$$

$$\bar{\boldsymbol{w}}^r = \frac{\bar{\boldsymbol{n}}^r_{\mathscr{L}} \times \bar{\boldsymbol{n}}^r_p}{\left|\bar{\boldsymbol{n}}^r_{\mathscr{L}} \times \bar{\boldsymbol{n}}^r_p\right|} \,. \tag{37}$$

The constraints imposed by the Listing's Law, fix in such a sense the quaternion torsional components of the eye rotations; actually, the eyes' cyclo-rotations (or torsions) that are required physically, depend not only on the ϕ angles, but also on the 3D rotation

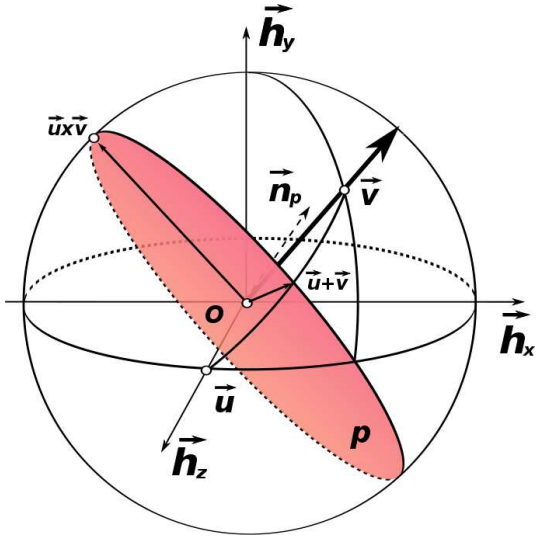


Figure 4: Plane containing all the possible directions for rotation axes which bring the versor $\bar{\boldsymbol{u}}$ into the versor $\bar{\boldsymbol{v}}$. This is the plane colored in pink. As usual, we represent this plane by its normal versor, which we denote by $\bar{\boldsymbol{n}}_p$, obtained by taking the cross-product of the vector $(\bar{\boldsymbol{u}} \times \bar{\boldsymbol{v}})$ with the vector $(\bar{\boldsymbol{u}} + \bar{\boldsymbol{v}})$.

coordinate system we use. This can be shown if we consider for instance an Helmholtz system. First of all, we have to write the quaternion associated to the three cardinal rotations. The rotation around $\bar{\boldsymbol{h}}_{\mathrm{X}}$ is represented by:

$$q_V = \cos(V/2) + \sin(V/2)\bar{\boldsymbol{h}}_{\mathrm{X}} \,. \tag{38}$$

The rotation around the $\bar{\boldsymbol{h}}_{\mathrm{Y}}$ is represented by:

$$q_{H^*} = \cos(H^*/2) + \sin(H^*/2)\bar{\boldsymbol{h}}_{\mathrm{Y}} \tag{39}$$

where $H^* = H - \lambda$ to account for the temporal rotations of the primary position. The rotation around the $\bar{\boldsymbol{h}}_{\mathrm{Z}}$ is represented by:

$$q_T = \cos(T/2) + \sin(T/2)\bar{\boldsymbol{h}}_{\mathrm{Z}} \,. \tag{40}$$

Hence, the overall quaternion is obtained by multiplying in cascade the cardinal quaternions in the order specified by the Helmholtz sequence:

$$\begin{aligned} q &= q_V \circ q_{H^*} \circ q_T = \\ &= (c_{V/2}c_{H^*/2}c_{T/2} - s_{V/2}s_{H^*/2}s_{T/2}) + \\ &+ \bar{\boldsymbol{h}}_{\mathrm{X}}(c_{V/2}s_{H^*/2}s_{T/2} + s_{V/2}c_{H^*/2}c_{T/2}) + \\ &+ \bar{\boldsymbol{h}}_{\mathrm{Y}}(c_{V/2}s_{H^*/2}c_{T/2} - s_{V/2}c_{H^*/2}s_{T/2}) + \\ &+ \bar{\boldsymbol{h}}_{\mathrm{Z}}(c_{V/2}c_{H^*/2}s_{T/2} + s_{V/2}s_{H^*/2}c_{T/2}) \end{aligned} \tag{41}$$

where $\mathrm{C}_{V/2}$ and $\mathrm{S}_{V/2}$ are the cosine and the sine of half the elevation angle V, etc.

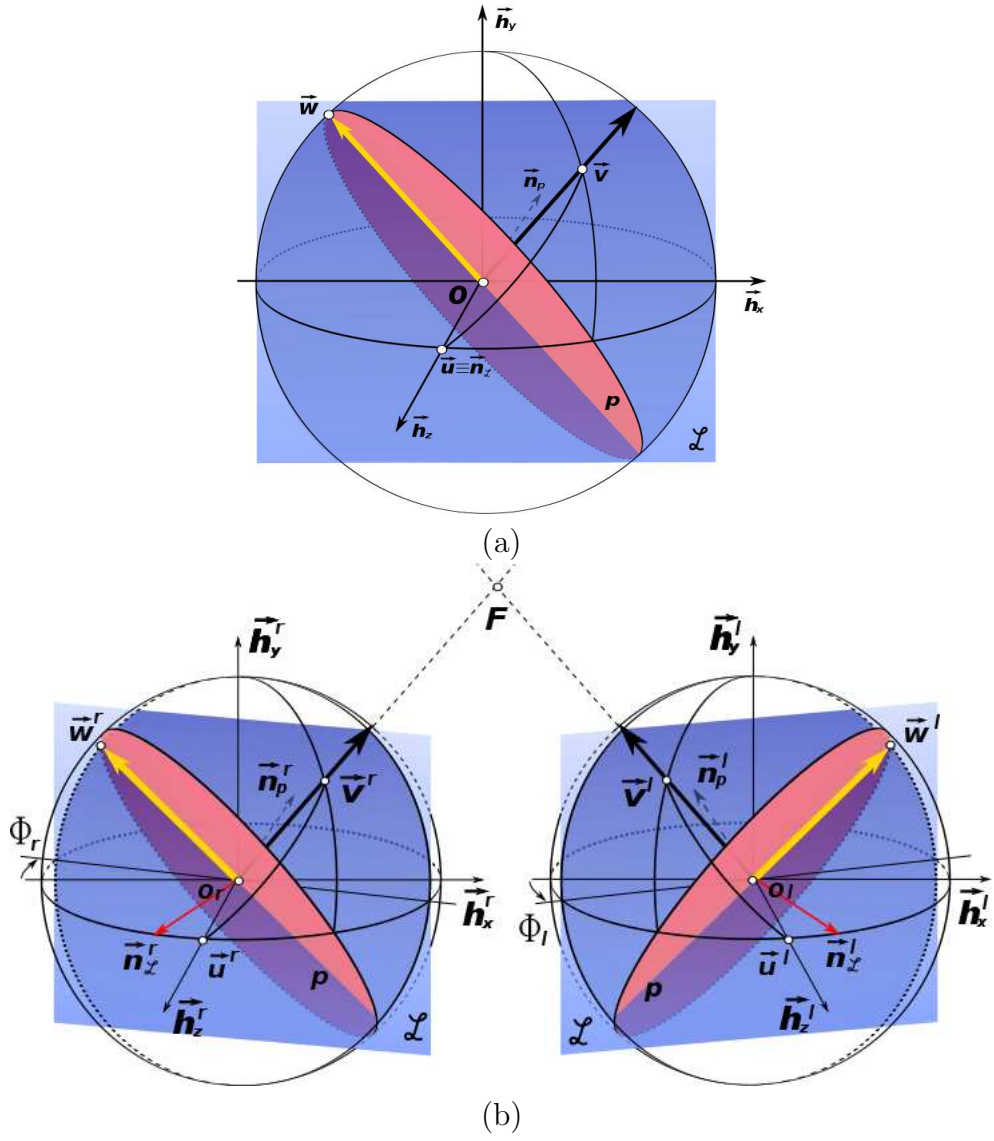


Figure 5: The intersection between the plane p and the Listing's plane $\mathscr{L}$ is the versor $\bar{\boldsymbol{w}}$ of the rotation axis that maps $\bar{\boldsymbol{u}}$ into $\bar{\boldsymbol{v}}$ respecting the Listing's Law. (a) The null vergence case. It is worth noting that $\mathscr{L}$ is not rotated with respect to the primary position, which correspond to the $\bar{\boldsymbol{h}}_z$ axis, since in the figure we have considered the value of λ set to zero. (b) The rotation of $\mathscr{L}$ for each eye in the binocular convergence case. The Listing's plane for the left eye is rotated by an angle ϕ_l while that for the right eye by an angle ϕ_r. Also in this case we have considered the values of λ_l and λ_r to be zero.

Taking into account L2, each rotation axis for the left eye is perpendicular to the normal of the plane $\bar{\boldsymbol{n}}^l_{\mathscr{L}}$ and each rotation axis for the right eye is perpendicular to the normal of the plane $\bar{\boldsymbol{n}}^r_{\mathscr{L}}$ (see Figure 5b). If we define q_l and q_r the quaternion representing the position for the left and right eye, respectively, L2 requires that [33]:

$$\boldsymbol{q}^l \cdot \bar{\boldsymbol{n}}^l_{\mathscr{L}} = 0 \tag{42}$$

$$\boldsymbol{q}^r \cdot \bar{\boldsymbol{n}}^r_{\mathscr{L}} = 0 \tag{43}$$

Solving these equations yields the following relationships that provide the Helmholtz torsion angles required by Listing:

$$\begin{aligned} \tan\frac{T_l}{2} &= -\tan\frac{V_l}{2}\left[\frac{\tan\psi_l + \tan(H_l^*/2)}{1+\tan(H_l^*/2)\tan\psi_l}\right] \\ \tan\frac{T_r}{2} &= -\tan\frac{V_r}{2}\left[\frac{\tan\psi_r + \tan(H_r^*/2)}{1+\tan(H_r^*/2)\tan\psi_r}\right] \end{aligned} \tag{44}$$

4 Visuomotor optimization constraints in binocular eye coordination

A justification of the Listing's Law with respect to both "motor"and "visual"efficiency criteria was first put forward by Helmholtz [54], primarily for monocular vision, and then generalized by Tweed [52], including implications for binocular vision. With the same spirit, our aim is to derive the pair of values for the ψ's angles that allows us to meet some visuo-motor optimality principle, which maximizes vision and motor efficiency. Similarly to Tweed's approach (cf. his visuo-motor theory), we define a cost function to be minimized which takes into account both the efficiency constraints:

$$\mathcal{F}(\psi_l, \psi_r) = (1-\alpha)\mathcal{M}(\psi_l, \psi_r) + \alpha\mathcal{V}(\psi_l, \psi_r) \tag{45}$$

where $\mathcal{M}$ is the motor constraint, $\mathcal{V}$ is the visual constraint, and α is a constant weighting factor that quantifies the relative importance of the two terms. Including both the terms, the eyes have to rotate around the visual axis for granting that the visual efficiency would be satisfied. It is worthwhile pointing out some differences with respect to Tweed's visuo-motor theory. In [52], Tweed defines the "visual constraint"as a condition on the eyes' postures, directly. Actually there is not a biunique correspondence between the alignment of the images of the visual plane and the condition of equi-cyclorotation of the eyes. Our major concern, here, is to define a new approach to the problem of the eye movements and their functional implications, which changes the perspective from which to face the problem (also with respect to [32]): contrary to starting from assumptions on the postures of the eyes and to analyzing their perceptual implications, we want to find general design criteria by which both vision and motor efficiency principles would guide proper eyes' postures, also taking into account the resources that motor and vision systems have at disposition.

4.1 Motor constraint

The motor term in Eq .(45) is introduced to characterize the primary position with the role of a "special"position for the oculomotor system, by which we want to move not too far. Accordingly, following [52], the motor efficiency is described in terms of the sum of the squared eccentricities of the rotations of the eyes, ε_l and ε_r:

$$\mathcal{M}(\psi_l, \psi_r) = \varepsilon_l^2 + \varepsilon_r^2 \tag{46}$$

Minimizing them we want to reduce the rotation amplitude of both eyes. In other words, we want the eyes not to drift too away from the primary position.

4.2 Visual constraint

The visual term in Eq .(45) imposes specific binocular correspondences between the stereo image pairs that particular reference surfaces project back on the retinas. As a starting point the visual constraint embraces two types of conditions:

$$\mathcal{V}(\psi_l, \psi_r) = \mathcal{V}_1(\psi_l, \psi_r) + \beta \mathcal{V}_2(\psi_l, \psi_r) \, . \tag{47}$$

With the first term we want to penalize a misalignment of the binocular projections on the horizontal and the vertical retinal meridian of a surface plane orthogonal to the gaze line. The second term is an extension of the first one by which we impose the coplanarity of the fixation planes. β is a positive constant that balances the relative importance of the two terms. Yet, the approach is generalizable to include different or additional viewing constraints to maximize the registration of the images of the local surface of a fixed object in dynamical situations.

First visual criterion. The reason behind this type of constraint relates to the fact that it gives rise to specific binocular correspondences in the retinal image planes, which we consider as "reference situations", invariant with respect to the gaze line. This consideration is dictated by the fact that it seems that the brain makes use of reference surfaces in order to judge the depth of the observed objects. The alignment of the horizontal and vertical meridia brings to a situation like the one depicted in Figure 6. When we observe the reference surface, we would like to have on the image plane only horizontal disparities along the horizontal meridian and only vertical disparities along the vertical meridian. With reference to Figure 6, this condition can be expressed as:

$$\mathcal{V}_1(\psi_l, \psi_r) = \left[1 - \left(\bar{\boldsymbol{n}}_s \times \bar{\boldsymbol{e}}_y^l\right) \cdot \left(\bar{\boldsymbol{n}}_s \times \bar{\boldsymbol{e}}_y^r\right)\right]^2 + \left[1 - \left(\bar{\boldsymbol{n}}_s \times \bar{\boldsymbol{e}}_x^l\right) \cdot \left(\bar{\boldsymbol{n}}_s \times \bar{\boldsymbol{e}}_x^r\right)\right]^2 \tag{48}$$

where $\bar{\boldsymbol{n}}_s$ is the normal to the reference surface, $\bar{\boldsymbol{e}}_x$ and $\bar{\boldsymbol{e}}_y$ are directed along the horizontal and vertical retinal meridians of both eyes.

Second visual criterion. The second visual constraint imposes the coplanarity of the fixation planes. This second type of constraint is inspired by the works and the ideas of Jampel [20], who claims that, with the head at rest, the eyes move without torsions

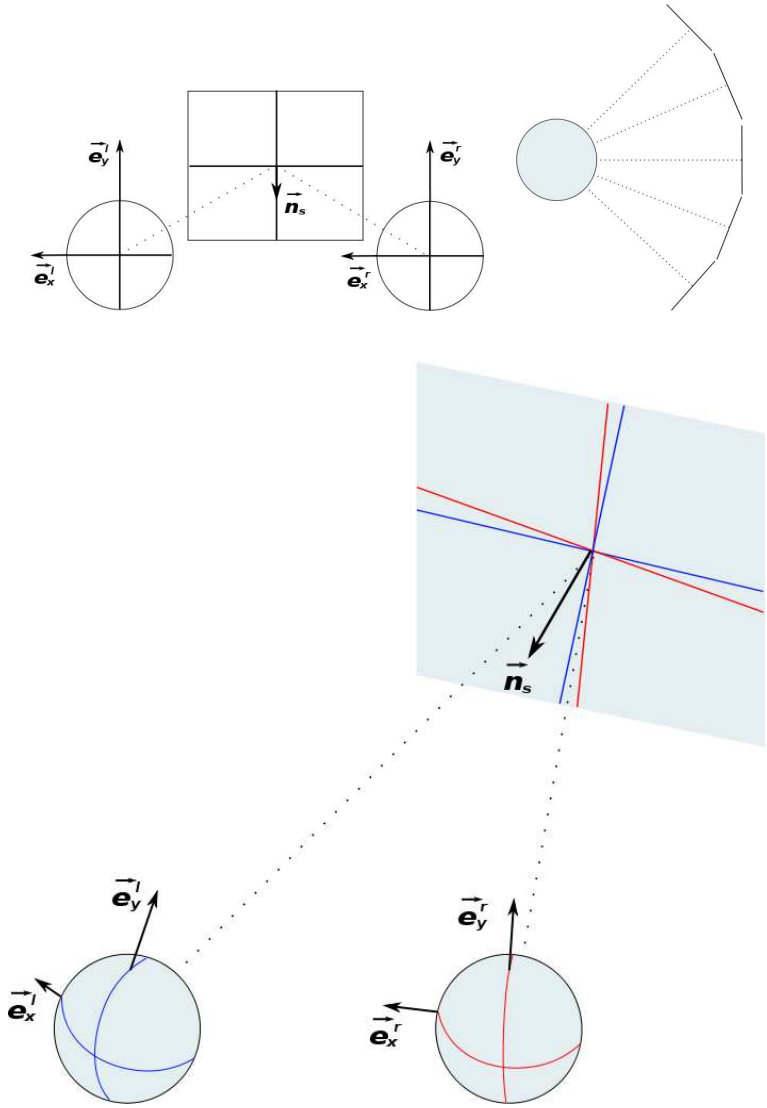


Figure 6: Top right: for every fixation point a surface (black line) orthogonal to the gaze line (dotted line) is always considered. Top left: the reference situation. The backprojections of the horizontal and the vertical meridia of both eyes are aligned on the reference surface, whose normal is $\bar{\boldsymbol{n}}_s$. In this case the fixation planes of both eyes are coplanar, as it results by the parallelism of the versors $\bar{\boldsymbol{e}}_y^l$ and $\bar{\boldsymbol{e}}_y^l$. Bottom: a situation in which the eyes are characterized by not optimal orientations. In this case the projections of the horizontal and vertical meridia for the left (blue line) and the right (red line) eye, respectively, are not aligned on the reference surface.

in any direction of gaze. The horizontal axes of both eyes are fixed in the head and collinear. In this way, the eyes move following the *Law of the fixation plane* by which the extraocular muscles, in both version and vergence movements, maintain the fixation planes coplanar. To reach a tertiary gaze position, the visual line rotates around the head-fixed horizontal axis, for elevating the gaze, and along an eye-fixed vertical axis to move the gaze laterally.

Each fixation plane is the plane through the fixation point and the nodal point of eye, that contains the horizontal retinal meridian. Accordingly, with reference to Figure 6 it is straightforward to describe the desired constraint: to make the fixation planes of both eyes coplanar, we have to impose that $\bar{\boldsymbol{e}}_y^l$ and $\bar{\boldsymbol{e}}_y^r$, normal to the planes, be parallel. Hence the extension of the visual part of the cost function is:

$$\mathcal{V}_2(\psi_l, \psi_r) = \left[1 - \bar{\boldsymbol{e}}_y^l \cdot \bar{\boldsymbol{e}}_y^r\right]^2 . \tag{49}$$

4.3 Results

The results are obtained by numerical minimization of the cost functionals with respect to the ψ angles.

$\beta = \mathbf{0}$ In general one has to consider the motor and the visual term equally important. Indeed, if we remove the motor efficiency constraint, the minimization of the functional yields infinite solutions that bring the retinal images to align, thus satisfying the visual constraint. By example, a tilt-pan system, in which torsions are intrinsically absent, is a solution. Surprisingly what we found is that the motor part of the cost function is not so important. In fact, even when the constant α is equal to 1, that is eliminating the contribution of $\mathcal{M}$ from the functional, we obtain for the ψ angles values which keep the eyes to move with the minimum eccentricity, than maximizing the visual efficiency. From the optimization we found that the values of the ψ angles for both eyes are equal in magnitude but opposite in sign, see Figure 7:

$$\psi_r = -\psi_l . \tag{50}$$

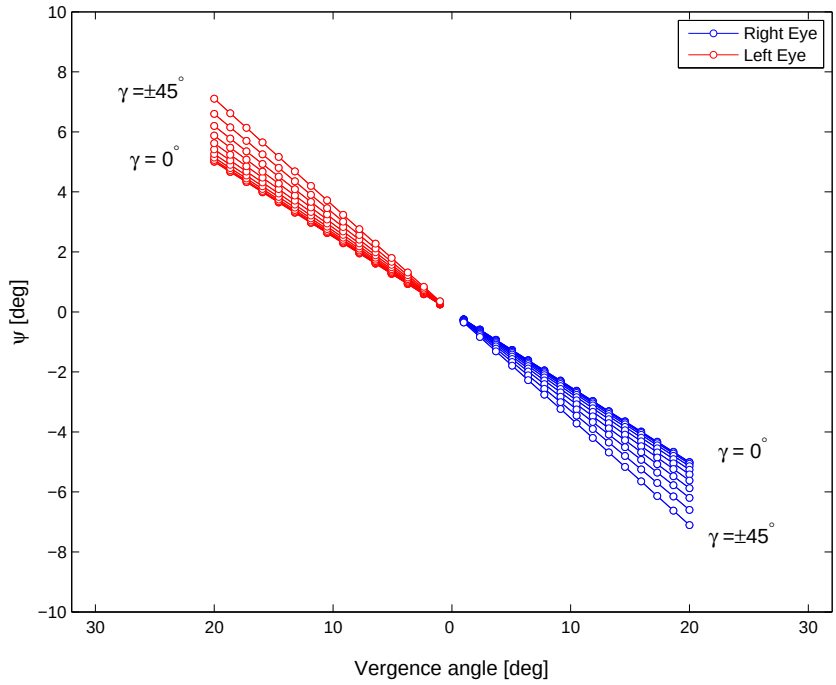
If we express the quaternions that describe the eyes' rotations, and finally derive the torsional angles in Helmholtz system, we find that the values of the ψ angles obtained by the optimization have the peculiarity to make T_l and T_r equal, see Figure 8.

In this way, we have obtained *as a result* what Tweed had yet imposed as the starting point of his minimization. From these considerations, using Eq .(44), we can obtain an analytical expression of the ψ angles:

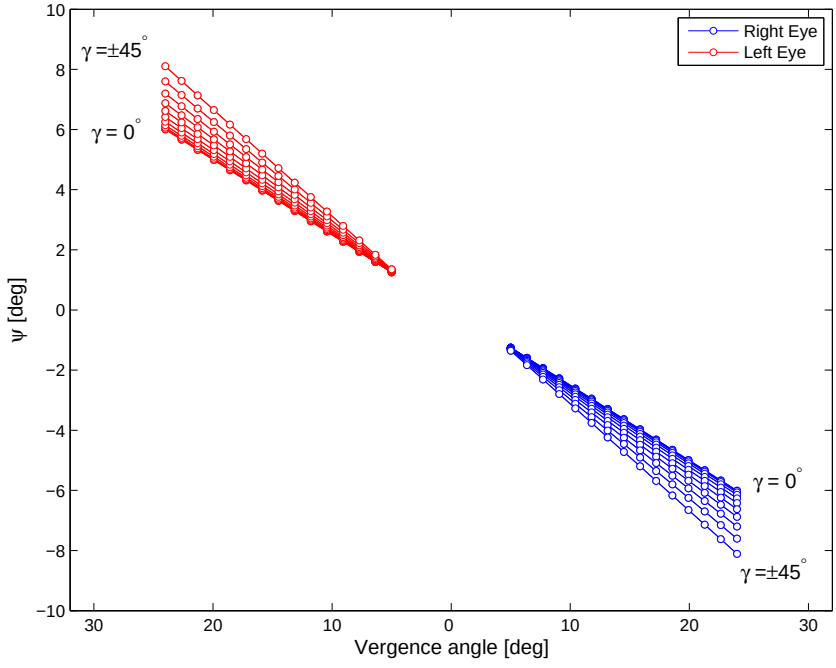
$$\psi_l = \frac{1}{2} \arcsin\left(\frac{\sin(\nu^*/2)}{\cos(\gamma^*/2)}\right) \tag{51}$$

$$\psi_r = -\frac{1}{2} \arcsin\left(\frac{\sin(\nu^*/2)}{\cos(\gamma^*/2)}\right) \tag{52}$$

where $\nu^* = H_r^* - H_l^*$ and $\gamma^* = \frac{H_r^* + H_l^*}{2}$ are the vergence and the version angles, respectively. From these equations it is possible to see that the angles are not only a function of the



(a)



(b)

Figure 7: Values of the ψ angles (ψ_l in red and ψ_r in blue) for (a) $\lambda_l = \lambda_r = 0°$ and (b) $\lambda_l = -\lambda_r = 2°$, in the case in which V_2 is not considered ($\beta = 0$). It is worth noting that the ψ angles are linear functions of the vergence, but the slope of the curves changes with the version γ. Changing the elevation does not influence the values of the ψ angles.

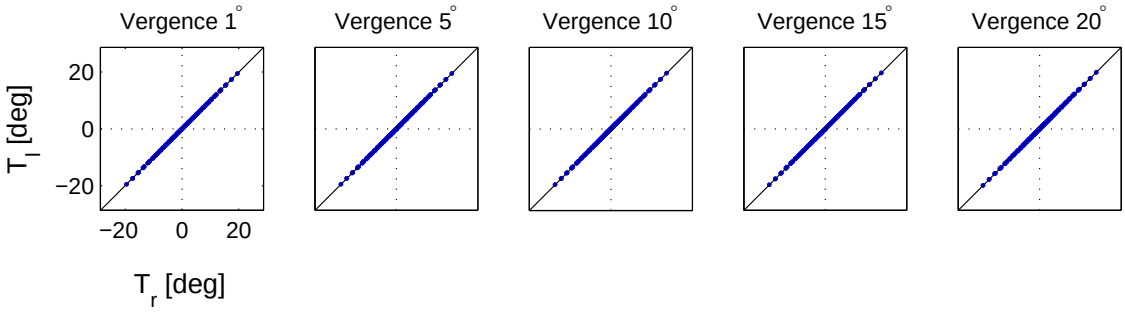
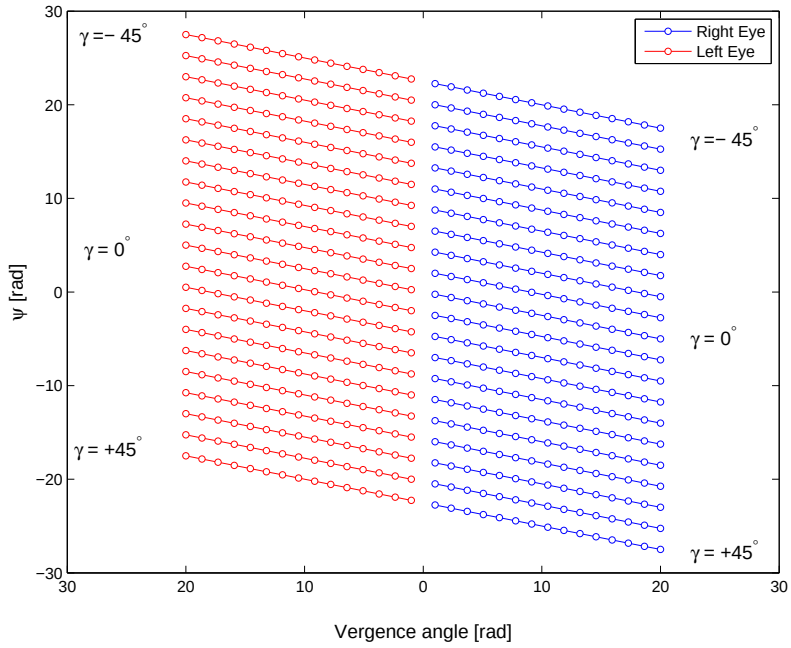


Figure 8: Helmholtz torsion T_r for the right eye plotted against the Helmholtz torsion T_l for the left eye. It is easy to observe the strong relationship that exist between them, $T_l = T_r$.

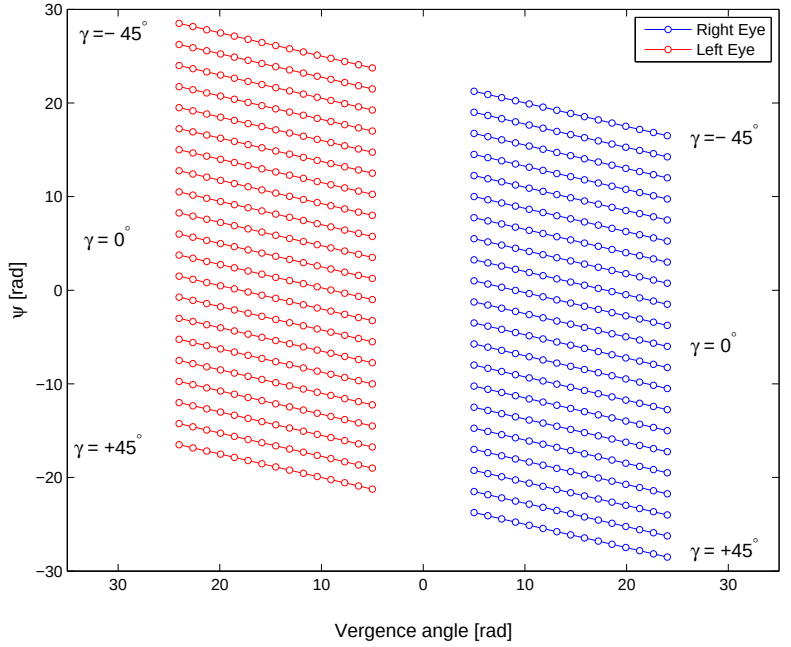
vergence ν^*, because also the version γ^* plays a role, whereas the elevation does not intervene at all. Moreover, when the version is null, it is always respected that the ψ angles would be a quarter of the vergence. Though, the multiplicative factor tends to increase to values greater than 0.25 whenever the gaze is not directed in the straight ahead direction. From the numerical results, we observed a wide range of values for the ψ angle for which there are no appreciable variations of the eccentricity. As a consequence, improving stereovision does not come at the price of a reduced motor efficiency. In the literature, several authors [40],[41] always claimed the existence of two different and distinct strategies the eyes adopt to move and fixate: the Listing' Law for far fixation distances, which has motor advantages, and its binocular extension (L2), which has visual advantages. Considering these two strategies as inconciliable, they always claimed that, in reality, the oculomotor system strikes a balance between them. Conversely, we found that only one strategy could exist: a *generalized* Listing's Law. Such a general strategy bridges the experimental data collected for very far and very near fixations, and it embraces both motor and visual constraints, without the need of a compromise, at least not in terms of minimal eccentricity of the eye rotations. Yet, there might be the case that the binocular coordination of the eye rotations imposed by the Listing's Law have a different motivation, e.g., associated to a simplification or robustness of the control strategy.

$\beta \neq \mathbf{0}$ Differently from the first functional, in this case the motor part, and thus the minimization of the eccentricity, plays a key role. In fact, if we do not consider the motor efficiency we obtain the solution that allows us to nullify the cost of the visual component: this solution is represented by the classical tilt-pan system. Obviously, this type of solution is not biologically plausible, because it violates the Listing's Law for far fixation. In this condition, the ψ angles should decrease with decreasing vergence, whereas in a tilt-pan system they remain confined in a wide range, see Figure 9.

By a proper compromise between the motor and the visual components we found the value for the ψ angles depicted in Figure 10 and Figure 11. They respect the Listing's



(a)



(b)

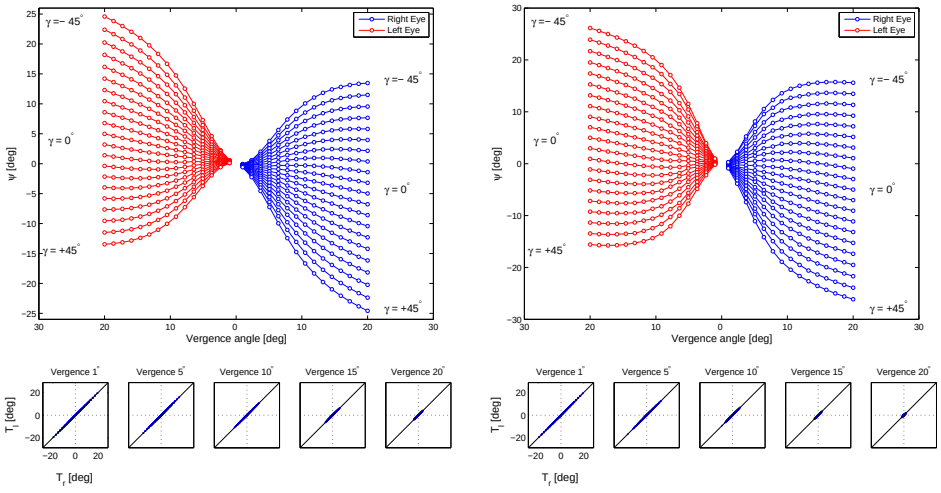
Figure 9: Values of the ψ angle (ψ_l in red and ψ_r in blue) for (a) $\lambda_l = \lambda_r = 0°$ and (b) $\lambda_l = -\lambda_r = 2°$, in the case in which $\mathcal{M}$ is not considered ($\alpha = 1$). These results replicate the behaviour of a classical tilt-pan robotic system. It is worth noting how the values of the ψ angles violate the Listing's Law, since they do not go to zero for small vergences. In this case the version γ only shifts up and down the curves without changing their slope.

Law for far fixation, approaching zero when the vergence decreases, and they behave like in a tilt-pan system when the vergence increases. It is worth noting that since a compromise exists, the eye rotations obtained with these ψ angles are characterized by larger eccentricities, that penalize the motor efficiency. However, we derived that the mean worsening is in the range of $0.5 \div 1\%$ with a peak value around $5 \div 6\%$.

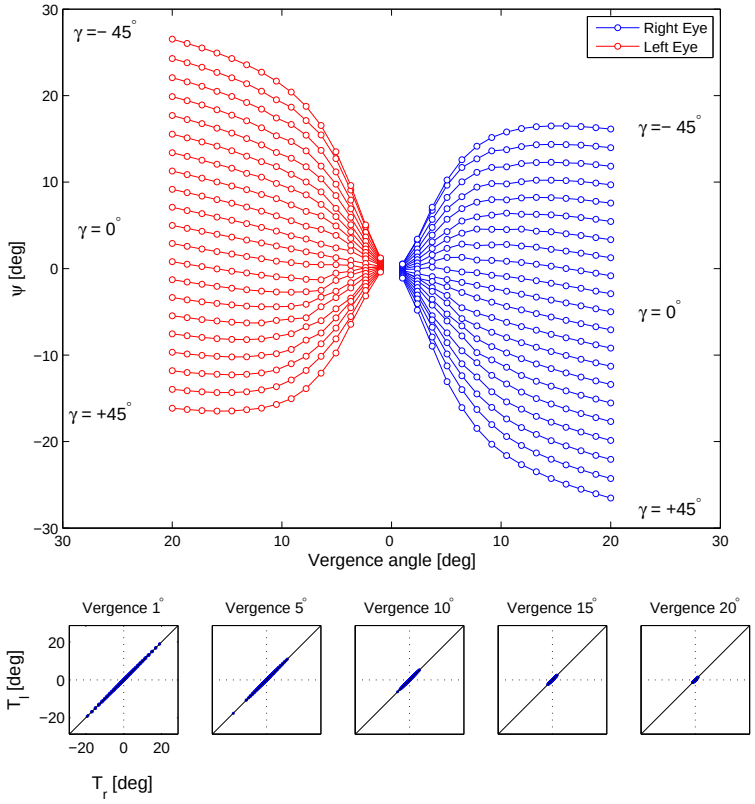
Moreover, for small versions, it is still respected the experimental evidences that both ψ are linear functions of the vergence, with a slope close to 0.25. Finally, for a fixed vergence, on the contrary to what we obtained with the first cost function, the ψ angles span a wider range then the one observed for the first cost function. Maybe this could be the explanation of the controversy about the value of the multiplicative constant μ, that links the rotation of the Listing's plane with the vergence, and also of the temporal rotation bias λ of the primary position.

5 Functional implications for depth vision

The visual constraints described in the previous Section have been introduced because of their computational advantages for stereopsis, at least in particular situations, which we take as reference. In general, stereo vision efficiency is always related to the properties of the disparity patterns in the retinal plane. Therefore, it would be more interesting breaking away from specific instances, like "reference surfaces"and their associated disparity patterns, in favour of a broader perspective in which we take as constraints all the disparities that could fall on the retinas. In this direction, perhaps the most characterizing/descriptive elements that we have at our disposal are the epipolar lines. As it is well known, the epipolar lines are defined as the segments on the image plane of one eye on which all the possible matching points of a given point on the retina of the other eye fall. Thus, they represent the loci of all the possible matches for every retinal location. When we look straight ahead at infinity (i.e., with parallel optical axes) all the epipolar lines are horizontal. Conversely, whenever the gaze changes and the vergence increases, the epipolar lines move and become more and more tilted. This movement causes an increase of the observed disparities, and, as a consequence, the vision system has to cope with larger search zones within the stereo correspondences are to find. From this perspective, having a general design strategy for the oculomotor system behavior, that minimizes the movements of the epipolar lines, would reduce the search zone and thus reduce the computational cost of finding visual correspondences. As a preliminary step in this direction, we have measured, *a posteriori*, the epipolar lines for the different systems we have discussed and characterized by the optimization process in the previous Section. Specifically, we consider around each fixation point a workspace delimited by an hyperboloid function, to simulate the fact that near the fixation point the distances tend to vary less than in periphery. The hyperboloid is always oriented along the gaze direction. For each retinal point for the left eye, we backproject a ray and we find the intersection of this ray with the two sheets of the hyperboloid. Then, we calculated analytically the mean disparity, integrating between these two extremes, and the disparities of the intersection points. These three disparities, by construction, lie on the epipolar line of each retinal point. Hence, we have found all the epipolar lines for different fixation

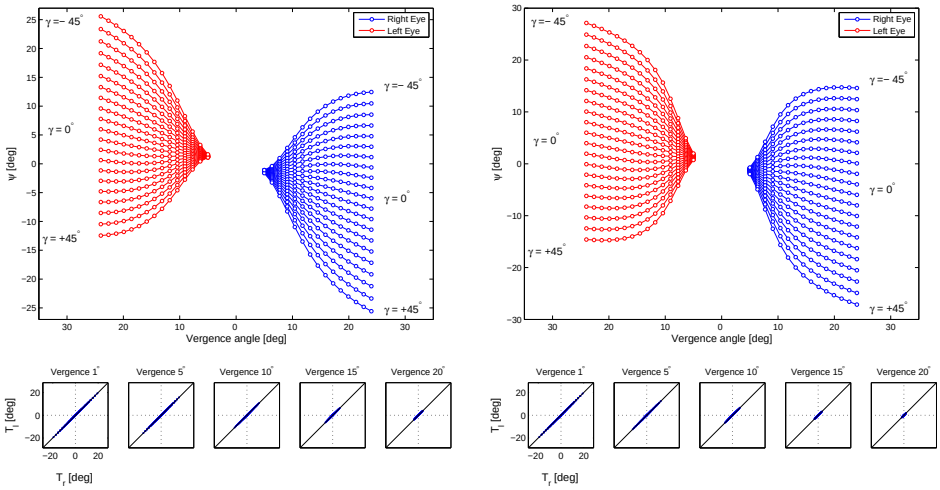


(a) (b)

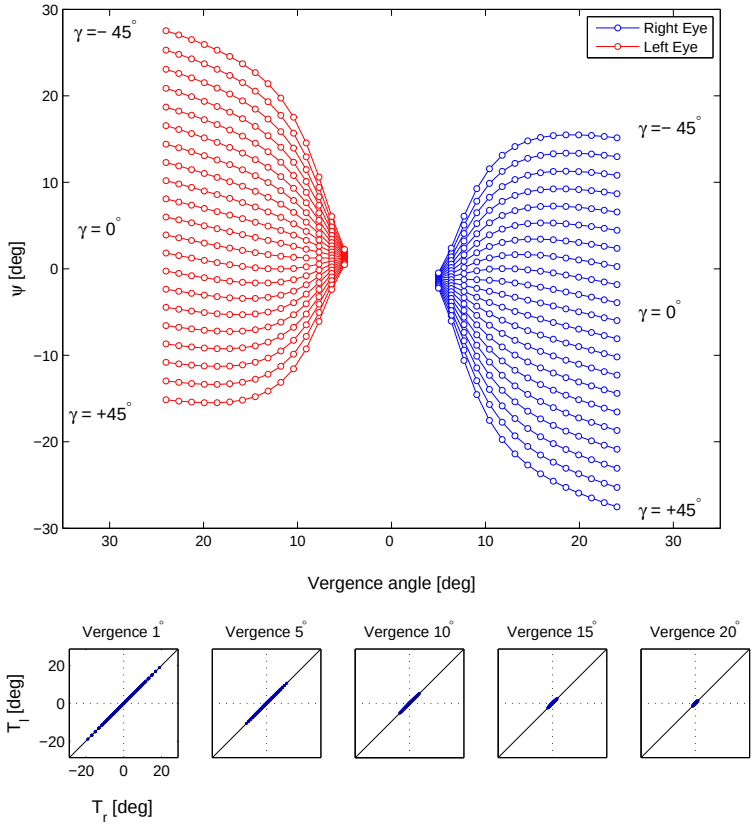


(c)

Figure 10: Values of the ψ angle (ψ_l in red and ψ_r in blue) for $\lambda_l = \lambda_r = 0°$ for different values of the optimization parameters. (a) $\alpha = 0.95$, $\beta = 3.15$. (b) $\alpha = 0.95$, $\beta = 7.36$. (c) $\alpha = 0.95$, $\beta = 10.52$. It is worth noting how for small versions the ψ's are still linear functions of the vergence, with a slope that increases or decreases with the version. The behaviour of the ψ angles replicates what observed for a classical tilt-pan system for large vergences and tends to respect Listing's Law (approaching zero) for small vergences. The insets at the bottom of each figures show the associated Helmholtz torsion angles.



(a) (b)



(c)

Figure 11: Values of the ψ angle (ψ_l in red and ψ_r in blue) and of the Helmholtz torsion for $\lambda_l = \lambda_r = 2°$. For explanation of the figure see Figure 10.

points (varying the version, the vergence and the elevation) and for different visuo-motor strategy obtained by the minimization process for different values of the optimization parameters α and β. Figure 12 shows the results.

We can observe how the movements of the eyes affect the geometry of the epipolar lines, and how different the search zone are. These results justify the search for a "correct"(i.e., most convenient) strategy for the oculomotor system. There is a strong dependence of 3D gaze position on the disparities patterns, and it is easy to see this fact by looking at the mean disparities, located at the center of each epipolar segment. These mean values are characterized by an offset with respect to the reference point, indicated by the large open circles. Moreover, the resulting mean disparity patterns are strongly dependent on the current epipolar geometry of the system. The mean disparities for a given vergence angle tend to move more for a change of the version than for a change of the elevation. By changing the vergence, the global behavior remains unchanged, only the magnitudes of the disparities vary proportionally.

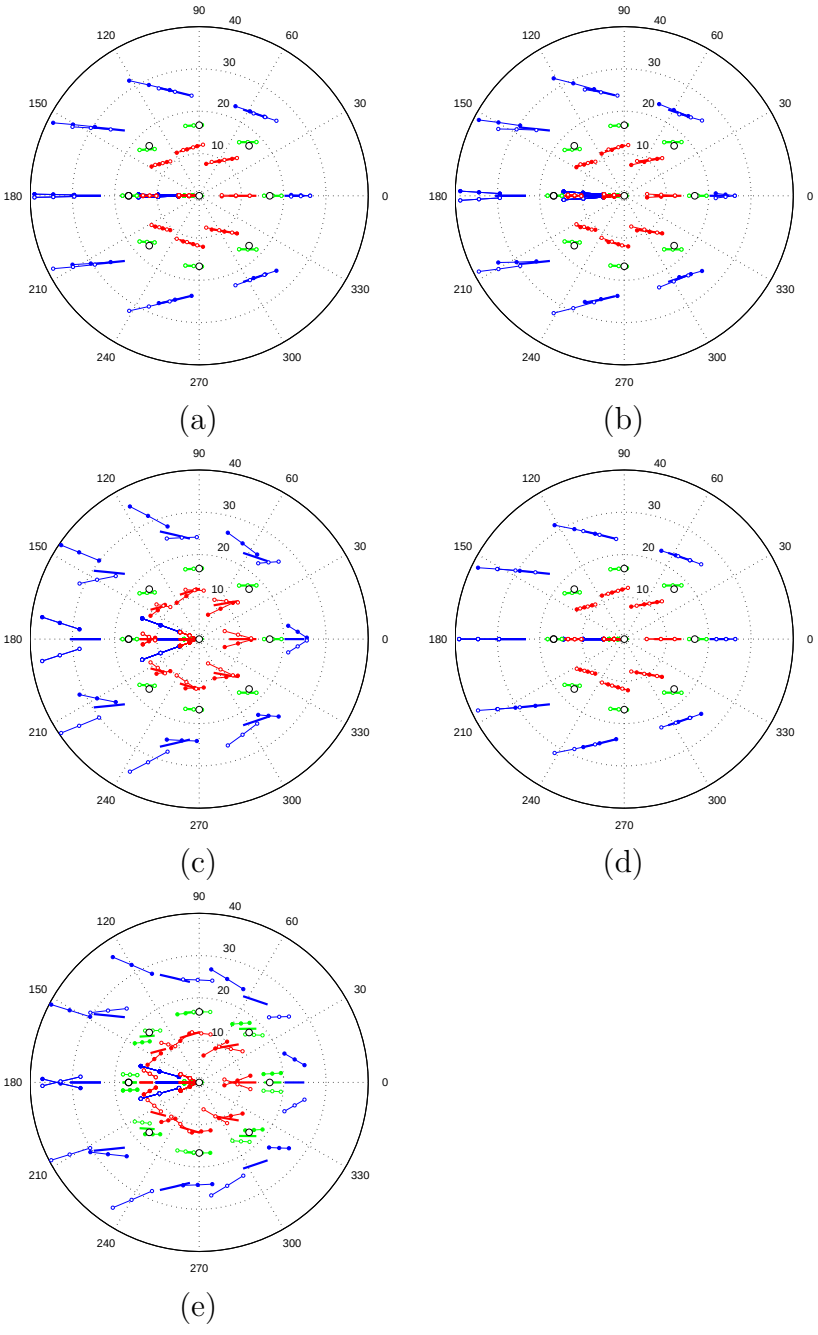


Figure 12: The resulting epipolar lines for the different strategies obtained from the minimization. The results are relative to a fixed vergence of 20° and to different values of the version and the elevation. The blue, the green and the red families are for a version angle equal to $-45°$, $0°$ and $45°$, respectively. The filled circle, the no-circle and the open circle lines are for an elevation angle equal to $-45°$, $0°$ and $45°$, respectively. (a) $\alpha_1 = 0.95$, $\alpha_2 = 10$. (b) $\alpha_1 = 0.95$, $\alpha_2 = 3$. (c) $\alpha_1 = 0.95$, $\alpha_2 = 0$. (d) $\alpha_1 = 1$, $\alpha_2 = 1$ (equal to a tilt-pan system). (e) $\alpha_1 = 0$, $\alpha_2 = 0$. The reference retinal points are indicate by large open circles.

5.1 How to embed fixational constraints into binocular energy-based models of depth perception

In this section we describe how it is possible to introduce specific design strategies and to modify the architectural parameters of the distributed representation of disparity information, described in more detail in Deliverable D2.1.

5.1.1 Design strategies and architectural parameters

In natural viewing conditions, the disparity distributions (horizontal and vertical) critically depend on the orientation of the eyes. Over relatively large visual angles, the retinal disparity patterns experienced by a binocular vergent system engaged in natural viewing present predictable components related to the positions of the eyes in the orbits. The predictable components may be used as priors to optimally allocate the computational resources to ease the recovery of the unpredictable components of disparity, which are dependent on the structure of the scene, only. Although, from a conceptual point of view, the oculomotor parametrization of active stereopsis is a well-established issue [21] [17], mapping the oculomotor constraints into the neural population coding and decoding strategies is still an open problem.

As a starting point, we have analyzed the influence of changes in the fixation location and of the 3D structure of the environment on the distribution of the disparity. To achieve this goal we have computed the statistics of the disparity distribution for different fixation points and for different values of position and orientation in the 3D environment of a flat surface. In particular, we have considered 125 different fixation points (X_f, Y_f, Z_f), that correspond to version and elevation angles uniformly distributed between $-\pi/4$ and $\pi/4$, and to an Euclidean distance from the eyes that ranges from 30cm to 100cm. For each fixation point we have considered 1000 different scenes, composed of planes centered in $(X_f \pm 20, Y_f \pm 20, Z_f \pm 50)$, and randomly oriented around the X and the Y axes. Figure 13 shows the histograms, representing the distribution of the vertical and the horizontal disparities, for different retinal locations and for two different fixation points (zero elevation and zero version, and elevation and version equals to $\pi/8$, respectively). It is worth noting that the mean value of the disparities changes with the fixation point, thus it is possible to divide the disparity in two parts: the first ($\vec{\delta_s}$), unpredictable, due to the structure of the 3D scene and the second ($\vec{\delta_e}$), more predictable, due to the geometry of the binocular system. Hence we can write:

$$\vec{\delta} = \vec{\delta_s} + \vec{\delta_e}. \tag{53}$$

The component of the disparity due to the epipolar geometry of the system ($\vec{\delta_e}$) can be embedded in the distributed representation of disparity information with the position shifts mechanisms [10]. The position-shift model assumes that the left and right receptive fields of a simple cell are always identical in shape but can be centered at different spatial locations (see D2.1 for further details). To embed the position shifts in the distributed representation we can consider two different situations:

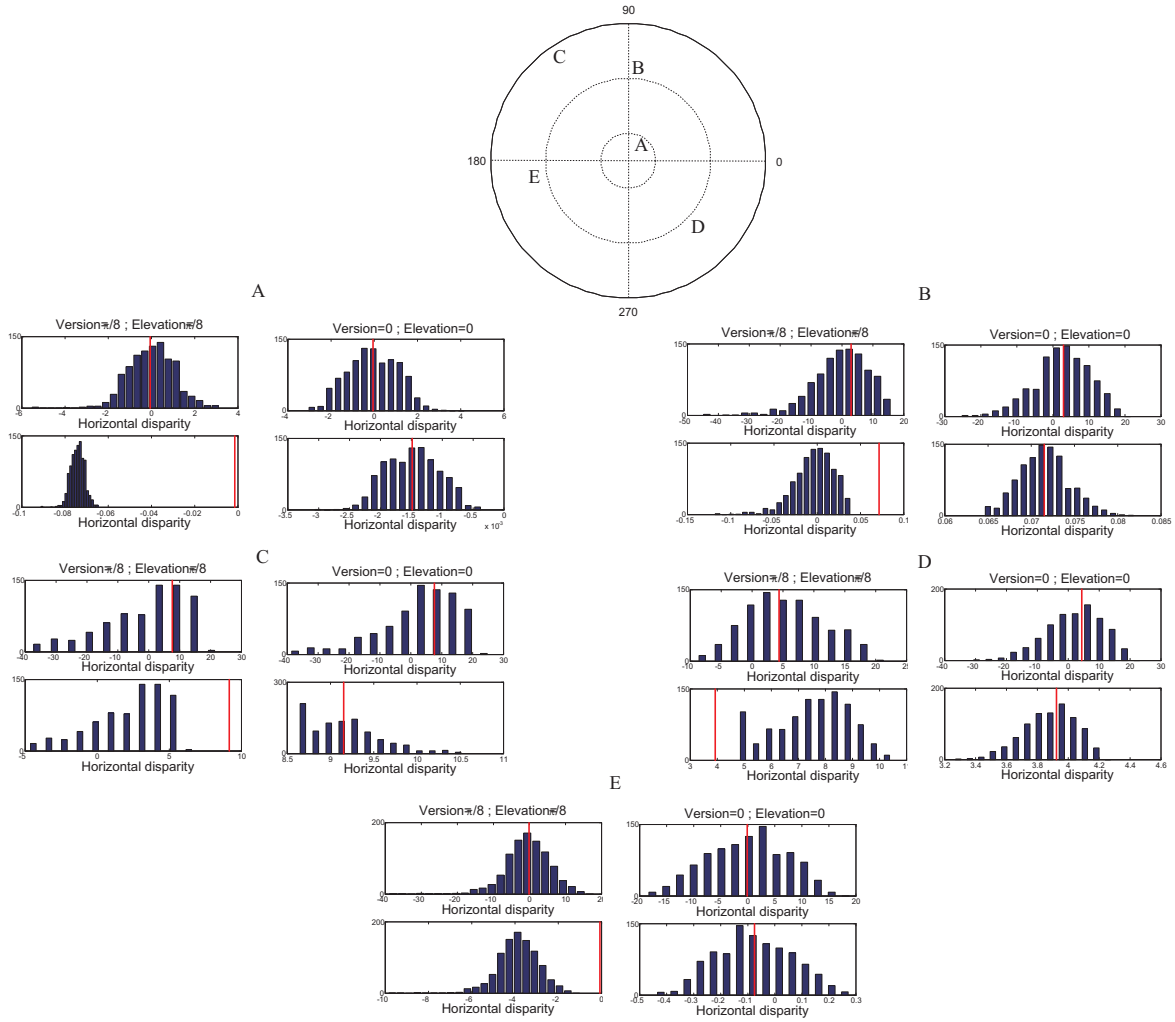


Figure 13: Distribution of the horizontal and vertical disparities. For each retinal point (here we have indicated 5 different point with capital letters A-E), we have computed the distribution of the horizontal (first row) and vertical (second row) disparities, for two different values of the gaze. Values corresponding to elevation $\pi/8$ and version $\pi/8$ are shown in the first column, whereas values corresponding to zero elevation and version are shown in the second column. The red lines correspond to the mean values computed with respect to the zero elevation and version situation.

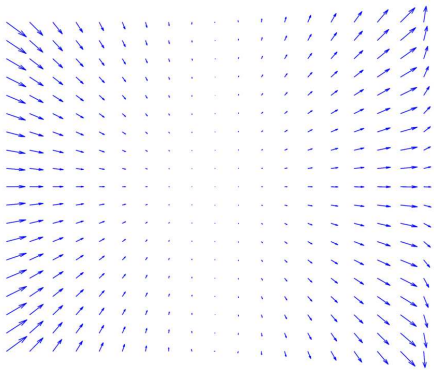
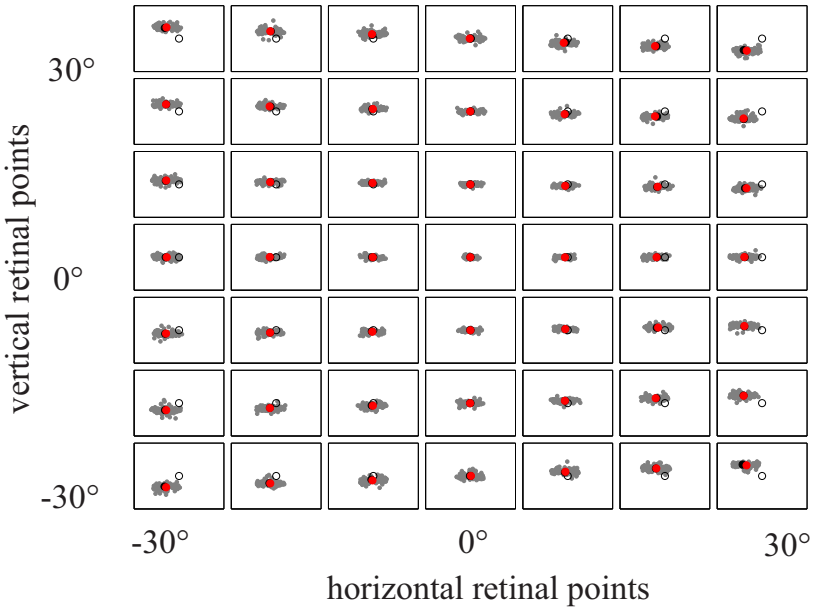


Figure 14: Mean pattern of horizontal and vertical disparities, computed in a reference situation with zero elevation and version angles, with a fixation point at a distance of 65cm.

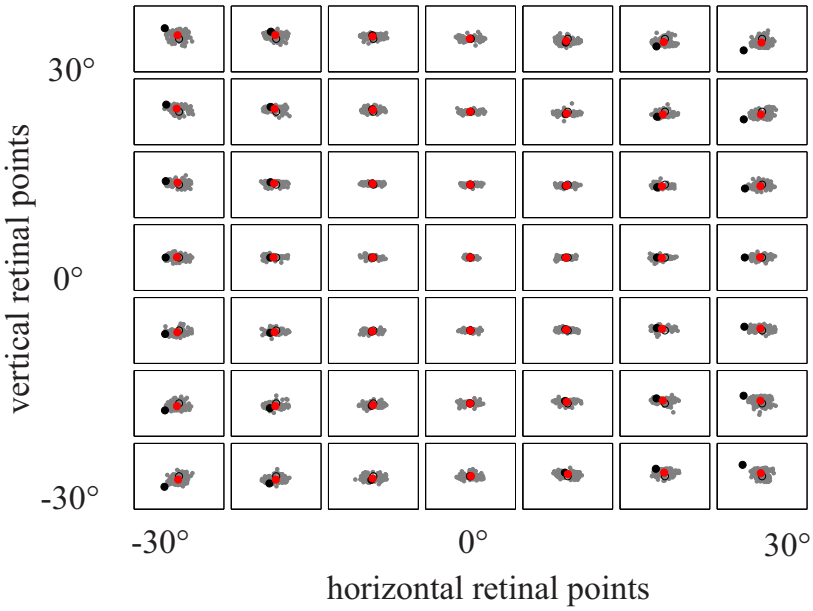
- We can take into account the mean value of the horizontal and vertical disparities in a reference situation corresponding to a fixation point characterized by null elevation and version angle and at a distance of 65 cm;
- We can continuously adapt the position shifts with respect to the fixation point.

We have decided to apply the position shifts mechanism with respect to the mean pattern computed in the reference situation (see Figure 14) and to recover the residual 2D disparity ($\vec{\delta_s}$) by using the phase-shift mechanism. Figure 15 shows the distribution of the estimated disparities between the left and the right retinas without and with a compensation of the predictable components of the disparity pattern. For different points of the left retina (open circles) we estimated the corresponding point for the right retina (gray dots) by a population of disparity detectors. The disparity relates to point image projections of randomly positioned objects in the peripersonal space. The red dots represent the mean of the estimated disparities, whereas the black dots represent the true mean disparity. It is worth noting that, by embedding the mean values computed with respect to the reference situation, the mean values of the estimated disparities (red dots) become closer to the true values of disparities (black dots). Figures 16 and 17 show the estimation of the 2D disparity without global components compensation and by embedding these components into the model, for a frontoparallel plane and for a scene obtained with the VR simulator (see Section 5.2), respectively. It is worth noting that the reliability of the disparity representation is improved, by embedding the component due to the epipolar geometry of the system.

The position shifts mechanism can be seen as a pre-wired design strategy that takes into account an initial adaptation of the system with respect to a "typical" viewing condition. Moreover, the activity of the population of neurons in the distributed representation can be adapted, in a dynamic way, by changing the distribution of the units, and by this minimize the necessary resources while preserving reliable estimates. To this aim, we have tested if a prior knowledge of a particular feature (e.g the sign of the disparity value of the disparity, or the range of the values) can be used to redistribute the sensitivity



(a)



(b)

Figure 15: The subplots represent a grid of 7x7 retinal points for a value of the gaze corresponding to zero elevation and zero version angles. For different points of the left retina (open circles) we estimated the corresponding point for the right retina (Gray dots) by a population of disparity detectors. The disparity relates to point image projections of randomly positioned objects in the peripersonal space, when the fixation point is in the primary position. The red dots represent the mean of the disparities, whereas the black dots represent the true mean disparity. Distribution of the estimated disparities between the left and the right retina are depicted (a) without and (b) with the compensation of the predictable components of the disparity pattern.

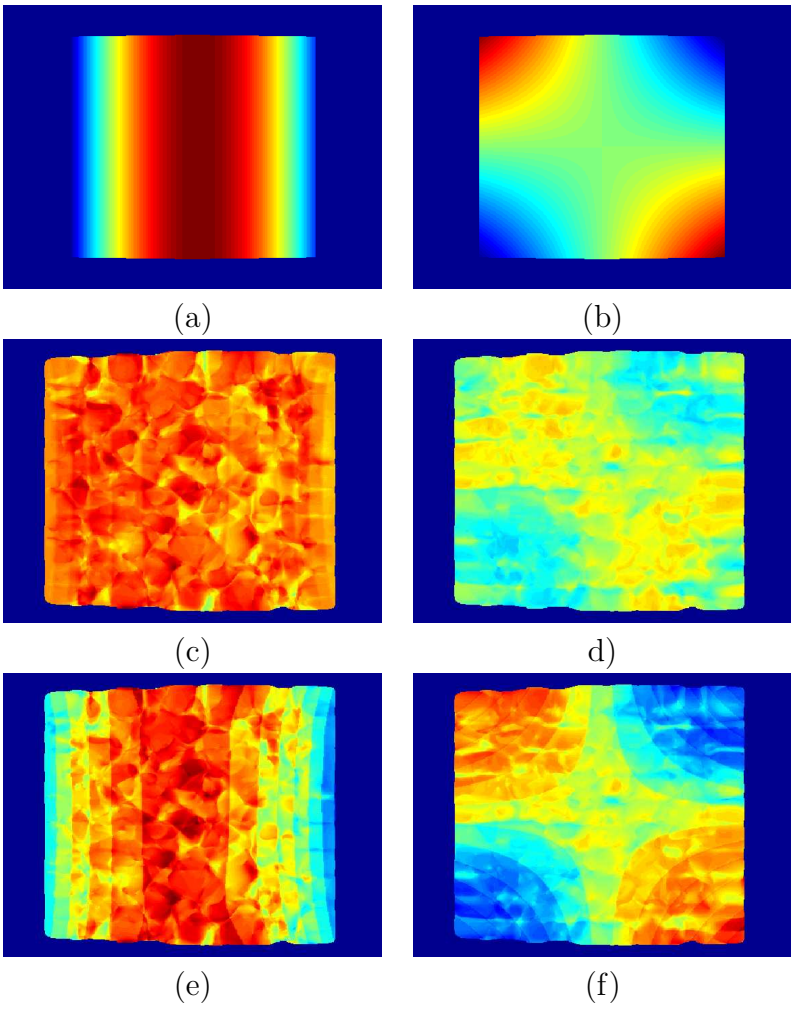


Figure 16: Disparity estimation by embedding fixation constraints into the binocular energy model for a stereo pair representing a fronto-parallel plane. (a)-(b) Ground truth horizontal and vertical disparity maps. (c)-(d) Estimation of the disparity by using the distributed architecture without embedding any fixation constraint. (e)-(f) Estimation of the disparity by using the distributed architecture by embedding the fixation constraints: a position shift derived from the mean values of disparities, accordingly to the statistics previously described, when the fixation point is at 65cm from the observer, with zero elevation and version angles. The results are obtained by using 43×43 pixels receptive fields, tuned to a disparity range from -8 to 8 pixels.

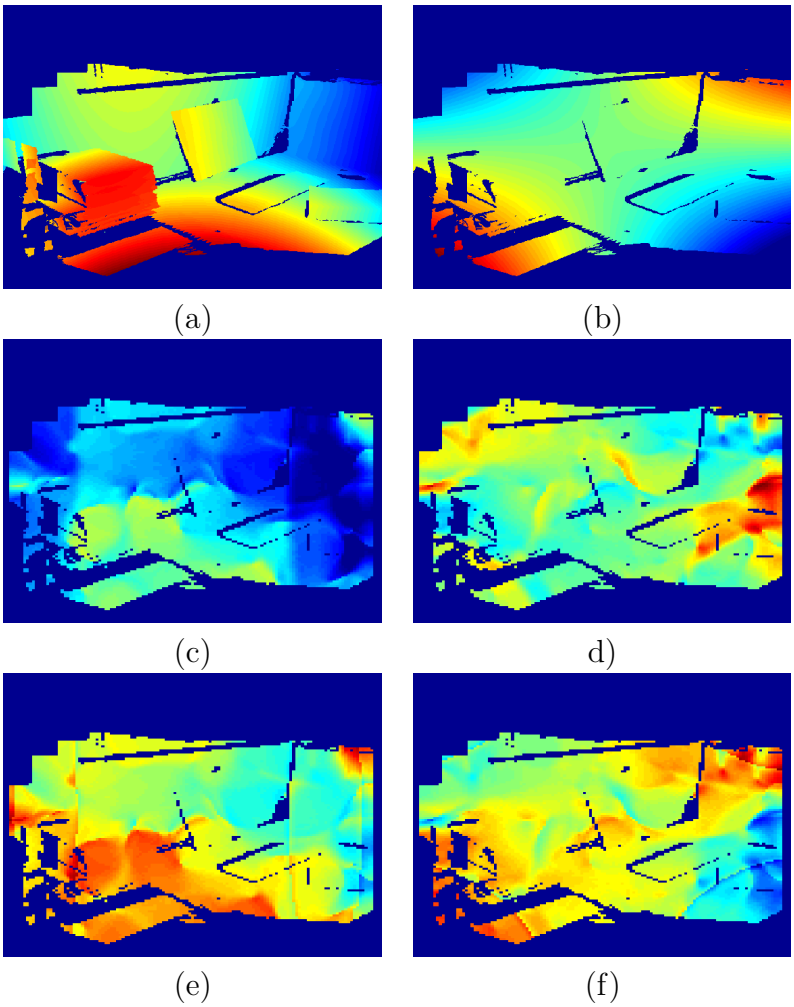


Figure 17: Disparity estimation by embedding fixation constraints into the binocular energy model for a stereo pair representing an indoor scenario acquired by using a laser scanner. (a)-(b) Ground truth horizontal and vertical disparity maps. (c)-(d) Estimation of the disparity by using the distributed architecture without embedding any fixation constraint. (e)-(f) Estimation of the disparity by using the distributed architecture by embedding the fixation constraints: a position shift derived from the mean values of disparities, accordingly to the statistics previously described, when the fixation point is at 65cm from the observer, with zero elevation and version angles. The results are obtained by using 43×43 pixels receptive fields, tuned to a disparity range from -8 to 8 pixels.

	# cells	AVG	STD	%
Venus stereo pair				
without phase shifts redistribution	33	0.84	0.63	91
with phase shifts redistribution	17	0.72	0.56	91
Tsukuba stereo pair				
without phase shifts redistribution	33	0.36	0.37	56
with phase shifts redistribution	17	0.28	0.20	91

Table 1: Average error (AVG), standard deviation of the error (STD) and density (expressed as the percentage of estimated values with respect to the total number of pixels in the image) for the Venus and the Tsukuba stereo pairs, without and with the redistribution of the cells. The redistribution has been performed by taking into account the known sign of the true disparity values. The first column shows the number of cells (for each spatial orientation) necessary to obtain the estimation of the disparity.

coverage of the cells' population and its density, by properly choosing the phase-shifts, while keeping fixed the other parameters. In this Section we show the compared results, obtained by using two stereo image pairs (Venus and Tsukuba) for which the ground truth is available [38]. We have redistributed the cells of the population, accordingly to the (known) sign of the disparity and we have compared the results with the ground truth disparity maps. Table 1 shows how the same (and in certain cases better) reliability is obtained by halving the units of the population. These preliminary results show how, in perspective, the activity of the distributed representation can be modulated by using the gaze information. In Figure 13 it is evident how the distribution of the disparities depends on the gaze position, this information can be thus used as a prior to reallocate the resources, e.g. by redistributing the cells of the population, or by modulating their responses (see Section 5.1.2).

5.1.2 Gain fields

Gain modulation, which is present in many cortical areas, is a change in the response amplitude of a neuron that is not accompanied by a modification of response selectivity. Just as population coding, in which ensembles of neurons encode information collectively, is a ubiquitous form of neuronal representation, gain modulation appears to be a widespread mechanism of neuronal computation. In particular, it allows information from different sensory and cognitive modalities to be combined. Modulation of the neural activity (or gain fields) by eye position have been described in many sites of the dorsal cortical pathway implicated in oculomotor outputs and involved in representation of space. In particular, modulations have been described in V3a [11], in MT [23], in MST [29][35][46], in VIP [5], in LIP [1][15][16], in PO (V6, V6a) [12][13][14] and 7a [2][1]. Moreover, Connors observed [6], [7] that also attention can give raise to gain modulation. He found that the visual response of many V4 neurons are modulated by a multiplicative gain factor that depends on where attention is directed. Several authors tried to give a justification to such a widespread mechanism. Gain modulation appears in a wide range of contexts, from the gaze-direction dependence of visual neurons in posterior parietal cortex [2] to the effects of attention [25][47]. It has been used to explain how the brain makes the remapping from a retinocentric to an egocentric frame of reference and the sensorimotor transformation useful for reaching and grasping tasks [36][31]. Moreover, with this mechanism it is possible to obtain translation and scale invariances useful for recognizing objects independently of the location and the size of their images on the retina, (cf. neurons of the area IT [37]). Obviously one can object that the cortical areas in which this mechanism is present are all high cortical regions of the brain, associated with motor and superior cognitive functions, but this is not true. In fact, gain fields has also been discovered in the primary visual cortex V1, where the activity of many disparity selective cells seems to be modulated by the vergence and the gaze direction signals [49][50][48]. In the same way as in the models cited above, used to explain sensorimotor transformations and attentional invariance, we expect to use the gain modulation in our population of binocular energy cells in order to modify their activity with the fixation distance and the direction of gaze. Thus, it allows us to reallocate the phase which characterize the selectivity of the cells and to shift, like in a position-shift model, their receptive field accordingly with the epipolar geometry of the system.

5.2 A virtual reality tool for the simulation of active vision systems

In 3D computer vision [51] and in particular for the stereoscopic vision, it is important to assess quantitatively the progress in the field. To solve this problem, we can find tools in the literature that provide test beds for a quantitative evaluation of the stereo algorithms (e.g the Middlebury database, `http://vision.middlebury.edu/stereo`). In general, these calibrated data sets have two major drawbacks: they only consider parallel axes stereo geometry, and they are not interactive, thus it is not possible to change the scene and the camera point of view. On the contrary, the interaction between the visual scene and the vision system is the main characteristic of an active vision system, thus we have designed a virtual reality tool that implements the requirements imposed by an active vision system and allows the changing of the geometry of the virtual stereo cameras as a function of the visual input to the active system. Such a tool, exploiting the ground truth available from the virtual world and the related projected stereo images, provides a way to validate the behavior of an active vision system in a controlled and realistic scenario. In particular, the simulator aims to precisely simulate the vergence movements of the two cameras in order to provide the stereo views and the related ground truth data (horizontal and vertical disparities). Thus, the virtual system can be used for two different purposes: (a) to produce visual behaviors, in a closed loop with a control strategy of the vergence movements guided by a vision-based information; (b) to obtain stereo sequences with related ground truth, to quantitatively assess the performances of the binocular energy-based models of depth perception.

5.2.1 The computation of a stereo image pair

In the virtual reality literature the main methods to render stereo image pairs are [3]: (1) the off-axis technique, usually used to create a perception of depth for a human observer (e.g. in applications like 3D cinema or immersive videogames) and (2) the toe-in technique that can simulate the actual intensity patterns impinging on the cameras of a robotic head.

In the off-axis technique, the stereo images are generated by projecting the objects in the scene onto the display plane for each camera; such projection plane has the same position and orientation for both camera projections. The model of the virtual setup is shown in Figure 18: $\mathbf{F}$ represents the location of the virtual point perceived when looking at the stereo pair composed by $\mathbf{F}^L$ and $\mathbf{F}^R$. This is the correct way to create stereo pairs that are displayed on stereoscopic devices for human observers. This technique introduces no vertical disparity, thus it does not cause discomfort for the users [43].

Since our aim is to simulate the actual images acquired by the cameras of a verging robotic head, the correct way to create the stereo pairs is the toe-in method: each camera is pointed at a single focal point (the fixation point) through a proper rotation. The geometrical sketch of the optical setup of an active stereo system and of the related toe-in technique is shown in Figure 19. The relation between the 3D world coordinates $\mathbf{X} = (X, Y, Z)$ and the homogeneous image coordinates $\mathbf{x}^L = (x^L, y^L, 1)$ and $\mathbf{x}^R = (x^R, y^R, 1)$ for the toe-in technique is described by a general perspective projection model. A generic

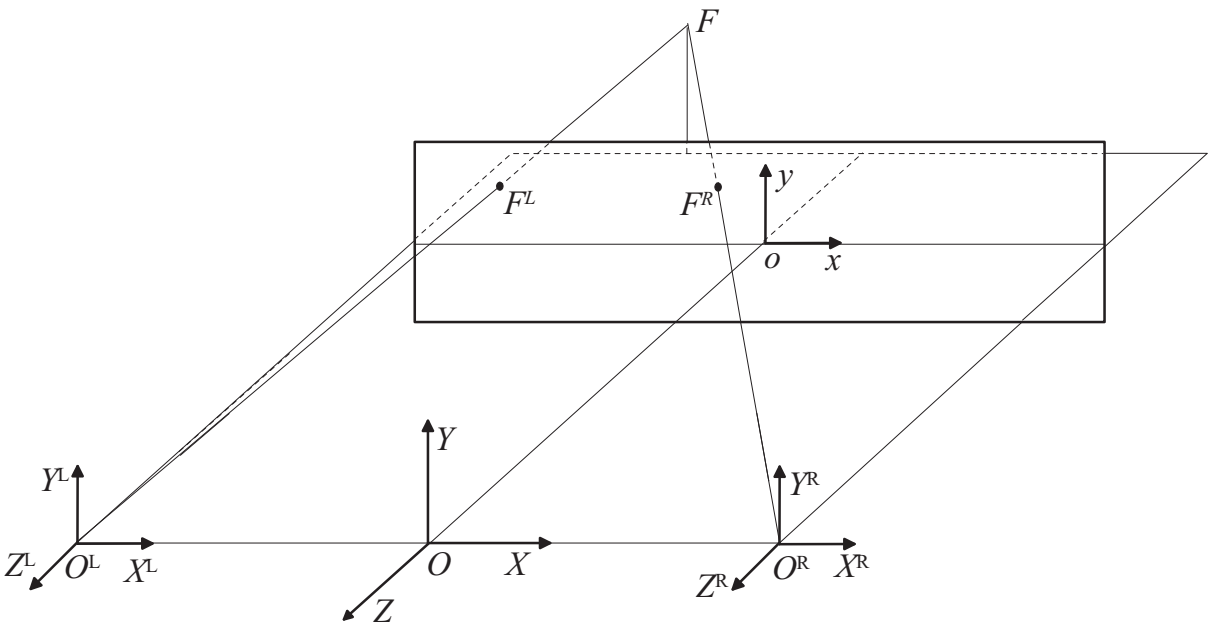


Figure 18: Geometrical sketch of the off-axis technique. The left and right camera frames: (X^L, Y^L, Z^L) and (X^R, Y^R, Z^R). The image plane (x, o, y) and the focal length Oo. The image points $\mathbf{F}^L$ and $\mathbf{F}^R$ are the stereo projection of the virtual point $\mathbf{F}$. The baseline b is denoted by $O^L O^R$.

point $\mathbf{X}$ in the world coordinates is mapped onto image plane points $\mathbf{x}^L$ and $\mathbf{x}^R$ on the left and right cameras, respectively.

The disparity patterns produced by the off-axis and toe-in techniques are shown in Figure 20a and Figure 20b, respectively.

5.2.2 Implementation

The virtual reality tool we have developed is based on a C++ / OpenGL architecture and on the Coin3D graphic toolkit (`www.coin3D.org`). Coin3D is built on OpenGL and uses scene graph data structures to render 3D graphics in real time. To implement the stereo geometry that we have previously described we modified the `SoCamera` node in Coin3D distribution. The `SoCamera` class is the abstract base class for camera definition nodes and it can be used to obtain a stereoscopic visualization of the scene. The stereoscopic technique usually implemented is the off-axis technique, but our aim was to add the toe-in technique, to generate stereo pairs like in a vergent stereoscopic robotic head. The implementation, described in this section of the Deliverable, follows the Helmholtz sequence for the rotation and refers to a tilt-pan stereo system, thus the torsional components are not considered. It is worth noting that the followed approach is general and any movements can be implemented into the simulator

Accordingly, we introduced the possibility of pointing the left and the right views at a single focal point, keeping fixed and symmetric the two view volumes and rotating them. To obtain the left and the right views both fixating a point $\mathbf{F}$, a symmetric view volume

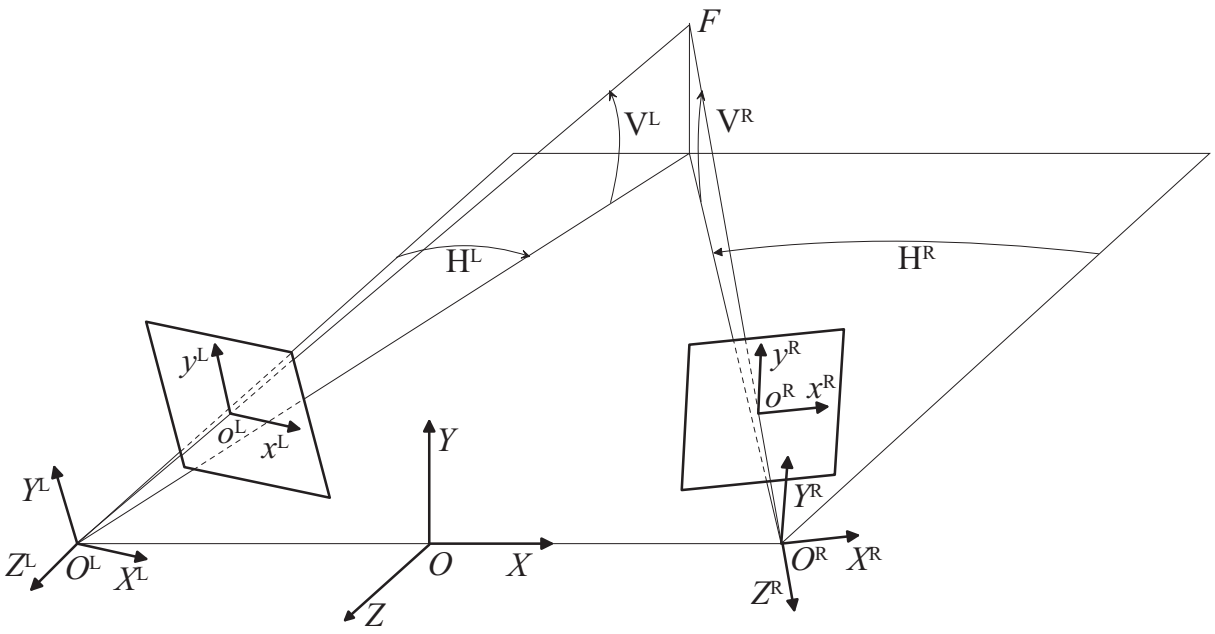


Figure 19: Geometrical sketch of the toe-in technique. The left and right camera frames: (X^L, Y^L, Z^L) and (X^R, Y^R, Z^R). The left and right image planes: (x^L, o^L, y^L) and (x^R, o^R, y^R). The left and right focal lengths: $O^L o^L = O^R o^R$, named f_0. The camera optical axes $O^L F$ and $O^R F$ are adjusted to fixation point $\mathbf{F}$. The baseline b is denoted by $O^L O^R$, the azimuth angles by H^L and H^R, and the elevation angles by V^L and V^R.

is created, centered in the position $\mathbf{O} = (X, Y, Z)$ (see Fig. 19). The skewed frustum (necessary to obtain the off-axis stereo technique) is no longer necessary. The view volume is then translated to the positions $\mathbf{O}^L = (X^L, Y^L, Z^L)$ and $\mathbf{O}^R = (X^R, Y^R, Z^R)$ in order to obtain the stereo separation b. The translation for the left and the right view volume can be obtained by applying the following translation matrix:

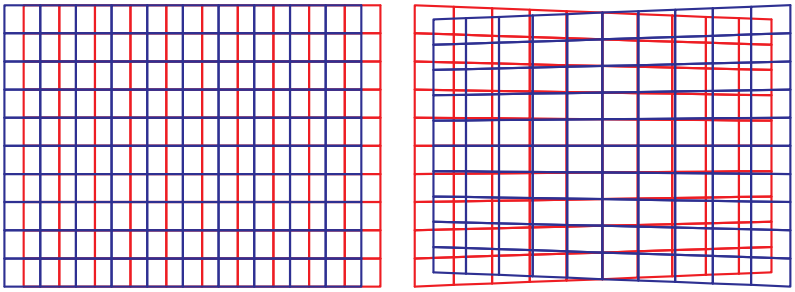
$$\mathbf{T}^{L/R} = \begin{bmatrix} 1 & 0 & \pm\frac{b}{2} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{54}$$

Then the azimuthal rotation (H^L and H^R) and the elevation (V^L and V^R) are obtained by applying the rotation matrices described by Eqs. 6 and 8. The complete roto-translation of the view-volumes (by considering homogeneous coordinates) is:

$$\begin{bmatrix} \mathbf{O}^{L/R} \\ 1 \end{bmatrix} = \mathbf{R}_V^{L/R} \mathbf{R}_H^{L/R} \mathbf{T}^{L/R} \begin{bmatrix} \mathbf{O} \\ 1 \end{bmatrix} \tag{55}$$

Thus, the projection direction is set to the target point $\mathbf{F}$, then the left and the right views project onto two different planes, as it can be seen in Figure 19.

In this way, it is possible to insert a camera in the scene (e.g. a perspective camera), to obtain a stereoscopic representation with convergent axis and to decide the location of the fixation point, in order to emulate the behavior of a stereo system.



(a) (b)

Figure 20: The projections of a fronto-parallel square onto the image planes, drawn in red for the left image and blue for the right. The texture applied to the square is a regular grid. (a) The projection obtained with the off-axis technique: only horizontal disparity is introduced. (b) The projection obtained with the toe-in technique: both vertical and horizontal disparities are introduced.

5.2.3 Active vision implementation

The described tool is active in the sense that the fixation point **F** of the stereo cameras varies to explore the scene. We can distinguish two possible scenarios: (1) we can use the system to obtain sequences where the fixation points are chosen on the surfaces of the objects in the scene, to obtain a "perfect" vergence; (2) we can use the system in cooperation with an algorithm that implements a vergence/version strategy. In the first case, it is not possible to fixate beside or in front of the objects. In the second case, the vergence/version algorithm gives us an estimate of the fixation point, the system adapts itself looking at this point and the snapshots of the scene are then used as a new input for selecting a new target point.

If we focused on the first issue, and we want the system to fixate points laying on the objects' surfaces, it is necessary to derive the 3D coordinates of all the visible surfaces. This information can be obtained from the z-buffer with the `glReadPixels` function, from which we obtain the 3D window coordinates, that are mapped into the object coordinates, through the function `gluUnproject`, by using the transformations defined by the `ModelView` matrix, the `Projection` matrix and the `Viewport` [19, 55].

5.2.4 Ground truth data generation

To compute the ground truth data for the horizontal and vertical disparities of the stereo image pairs, given the projection of a 3D virtual point in one image plane, we have to look for the correspondent projection in the other image plane. Formally, the two camera reference frames are related by a rigid body transformation described by the rotation matrix $\mathcal{R}$ and the translation $\mathcal{T}$. The left and right projections are related by the same transformation in the following way [22]:

$$\lambda^R \mathbf{x}^R = \mathcal{R}\lambda^L \mathbf{x}^L + \mathcal{T} \tag{56}$$

where $\mathbf{x}^L$ and $\mathbf{x}^R$ are the homogeneous coordinates in the two image planes, and λ^L and λ^R are the depth values.

To apply the relationship described by Eq. 56 we first read the z-buffer (w) of the two stereo views through the `glReadPixels` function, then we obtain the depth values with respect to the reference frames of the two cameras in the following way:

$$\lambda^{L/R} = \frac{f\ n}{w^{L/R}(f-n) - f} \tag{57}$$

where f and n represent the values of the far and the near planes of the virtual camera. Starting from the image coordinate $\mathbf{x}^L$ of the left image and the depth values $\lambda^{L/R}$ obtained by Eq. 57, we obtain the image coordinate $\mathbf{x}^R$ of the right view by combining the roto-translation described in Eq. 55 and Eq. 56 in the following way:

$$\lambda^R \mathbf{x}^R = \mathbf{R}^R \mathbf{T}^R (\mathbf{T}^L)^{-1} (\mathbf{R}^L)^{-1} \lambda^L \mathbf{x}^L \tag{58}$$

where $\mathbf{R}^{L/R} = \mathbf{R}_V^{L/R} \mathbf{R}_H^{L/R}$. Finally the horizontal disparity $d_x = x^R - x^L$ and the vertical disparity $d_y = y^R - y^L$ are computed.

In this way, we have obtained a fast tool, capable of handling the commonly used 3D modeling formats (e.g., VRML and OpenInventor) and the data acquired by a 3D laser scanner (Konica Minolta Vivid 910), specifically purchased by UG for benchmarking the active vision strategies developed in EYESHOTS with real-world conditions with known ground truth.

Figures 21a,e show the VRML models of two indoor scenes, acquired by the 3D laser scanner. The 3D data and the textures have been loaded in the virtual simulator, then the left and right projections (see the anaglyph of Figures 21b,f), the horizontal and the vertical ground truth disparity maps (see Figures 21c,d,g,h), can be obtained, for each possible fixation point.

The developed tool is currently being used to create a database of stereo image pairs with data about the vergence points and the ground truth disparities (see, `www.pspc.dibe.unige.it/Research/vr.html`), to be used in WP2 and WP3 for learning adaptive vergence strategies, as well as 3D visual descriptors for object identity. In general, the tool can be used both for algorithmic and behavioural benchmarks for the whole duration of the project. The simulator and the general approach followed have been presented at the VISAPP09 Conference and at the ECVP'09 Conference.

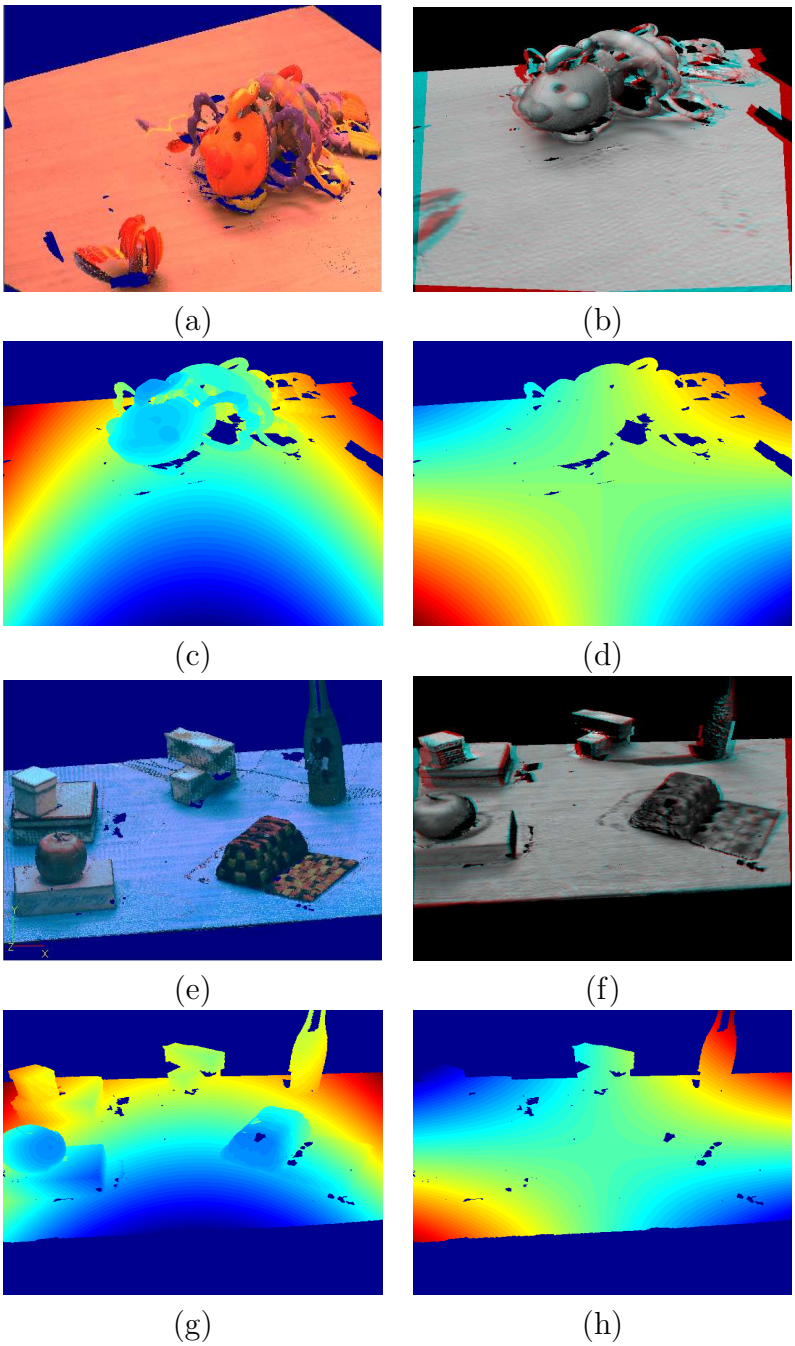


Figure 21: (a)-(e) VRML models of two indoor scenes, acquired by the 3D laser scanner. (b)-(f) anaglyph of the left and right view when the system is verging in the middle of the table in both situations. (c)-(d)-(g)-(h) horizontal and vertical ground truth disparity map for the two stereo pairs.

References

[1] R.A Andersen, R.M. Bracewell, S. Barash, J.W. Gnadt, and L. Fogassi. Eye position effects on visual, memory and saccade-related aactivity in areas LIP and 7a of macaque. *J. Neurosci.*, 10:1176–1196, 1990.

[2] R.A. Andersen and V.B. Mountcastle. The influence of the angle of gaze upon the excitability of the light-sensitive neurons of the posterior parietal cortex. *J. Neurosci.*, 3:532–548, 1983.

[3] P. Bourke and P. Morse. Stereoscopy: Theory and practice. *Workshop at 13th International Conference on Virtual Systems and Multimedia*, September 2007.

[4] P. Bruno and A.V. Van den Berg. Relative orientation of primary position of the two eyes. *Vision Res.*, 37(7):935–947, 1997.

[5] C.L. Colby, J.R. Duhamel, and M.E. Goldberg. Ventral intraparietal area of the macaque: anatomic location and visual response properties. *J. Neurophysiol.*, 69:902–914, 1993.

[6] C.E. Connors, D.C. Preddie, J.L. Gallant, and D.C. Van Essen. Response in area V4 depend on the spatial relationship between stimulus and attention. *J. Neurophysiol.*, 75:1306–1308, 1996.

[7] C.E. Connors, D.C. Preddie, J.L. Gallant, and D.C. Van Essen. Spatial attention effects in macaque area V4. *J. Neurosci.*, 17:3201–3214, 1997.

[8] J.D. Crawford, J.C. Martinez-Trujillo, and E.M. Klier. Neural control of three-dimensional eye and head movements. *Curr. Opinion in Neurobiology*, 13:655–662, 2003.

[9] F.C. Donders. Beitrag zur Lehre von den Bewegungen des menschlichen Auges. *Holland Beit Anatom Physiolog Wiss,*, 1:105–145, 1848.

[10] D.J. Fleet, H. Wagner, and D.J. Heeger. Neural encoding of binocular disparity: Energy models, position shifts and phase shifts. *Vision Res.*, 36(12):1839–1857, 1996.

[11] C. Galletti and P.P. Bataglini. Gaze-dependent visual neurons in area V3a of monkey prestriate cortex. *J. Neurosci.*, 9:1112–1125, 1989.

[12] C. Galletti, P.P. Bataglini, and P. Fattori. Functional properties of neurons in the anterior bank of the parieto-occipital sulcus of the macaque monkey. *Eur. J. Neurosci.*, 3:452–461, 1991.

[13] C. Galletti, P.P. Bataglini, and P. Fattori. Eye position influence on the parieto-occipital area PO (V6) of the macaque monkey. *Eur. J. Neurosci.*, 7:2486–2501, 1995.

[14] C. Galletti, P. Fattori, P.P. Bataglini, S. Shipp, and S. Zeki. Functional demarcation of a border between areas V6 and V6a in the superior parietal gyrus of the macaque monkey. *Eur. J. Neurosci.*, 8:30–52, 1996.

[15] J.W. Gnadt and L.E. Mays. Depth-tuning in area LIP by disparity and accomodative cues. *Soc. Neurosci. Abstr.*, 17:1113, 1991.

[16] J.W. Gnadt and L.E. Mays. Neurons in monkey parietal area LIP are tuned for eye-movement parameters in three dimensional space. *J. Neurophysiol.*, 73:280–297, 1995.

[17] M. Hansard and R. Horaud. Cyclopean geometry of binocular vision. *JOSA A*, 25(9):2357–2369, 2008.

[18] T. Haslwanter. Mathematics of three-dimensional eye rotations. *Vision Res.*, 35(12):1727–1739, 1995.

[19] D. Hearn and M. P. Baker. *Computer Graphics, C Version, 2/E.* Prentice Hall, 1997.

[20] R.S. Jampel. The function of the extraocular muscles, the theory of the coplanarity of the fixation planes. *Journal of Neurological Sciences,*, 280:1–9, 2009.

[21] M.R.M. Jenkin. Stereopsis near the horoptor. *Proc. 4th ICARCV*, 1996.

[22] Y. Ma, S. Soatto, and S. Kosecka, J.and Sastry. *An Invitation to 3D Vision. From Images to Geometric Models.* Springer-Verlag, 2004.

[23] J.H.R. Maunsell and D.C. Van Essen. Functional properties of neurons in middle temporal area of the macaque monkey. II binocular interactions and sensitivity to binocular disparity. *J. Neurophysiol.*, 49:1148–1167, 1983.

[24] J.S. Maxwell, E.W. Graf, and C.M. Schor. Adaptation of torsional eye alignment in relation to smooth pursuit and saccades. *Vision Res,*, 41:3735 3749, 2001.

[25] C.J. McAdams and J.H.R Maunsell. Effects of attention on orientation-tuning function of single neurons in macaque cortical area V4. *J. Neurosci.*, 19:431–441, 1999.

[26] A.W.H. Minken and J.A.M. Van Gisbergen. A three dimensional analysis of vergence movements at various level of elevation. *Experimental Brain Research*, 101(2):331–345, 1994.

[27] H. Misslisch, D. Tweed, and B.J. Hess. Stereopsis outweighs gravity in the control of the eyes. *J. Neurosci.*, 21(3):126–130, 2002.

[28] D. Mok, A. Ro, W. Cadera, J.D. Crawford, and T. Vilis. Rotation of listing's plane during vergence. *Vision Res.*, 32:2055–2064, 1990.

[29] W.T. Newsome, R.H. Wurtz, and H. Komatsu. Relation of cortical areas MT and MST to pursuit eye movements. II differentiation of retinal from extraretinal inputs. *J. Neurophysiol.*, 60:604–644, 1988.

[30] J. Porrill, J.P. Ivins, and J.P Frisby. The variation of torsion with vergence and elevation. *Vision Res.*, 39:3934–3950, 1999.

[31] A. Pouget and T.J. Sejnowki. Spatial transformation in the parietal cortex using basis function. *J. Cogn. Neurosci.*, 9:222–237, 1997.

[32] J.C.A. Read and B.G. Cumming. Understanding the cortical specialization for horizontal disparity. *Neural Computation,*, 16:19832020, 2004.

[33] J.C.A. Read and B.G. Cumming. Does depth perception require vertical-disparity detectors? *Journal of Vision,*, 6(12):1323–1355, 2006.

[34] L.J. Van Rijn and A.V. Van den Berg. Binocular eye orientation during fixations: Listing's law extended to include eye vergence. *Vision Res.*, 33(5/6):691–708, 1993.

[35] J.P. Roy, H. Komatsu, and R.H. Wurtz. Disparity sensitivity of neurons in monkey extrastriate area MST. *J. Neurosci.*, 12:2478–2492, 1992.

[36] E. Salinas and L.F. Abbott. Transfer of coded information from sensory to motor networks. *J. Neurosci.*, 15(10):6461–6474, 1995.

[37] E. Salinas and L.F. Abbott. Invariant visual responses from attentional gain fields. *J. Neurophsiol.*, 77:3267–3272, 1997.

[38] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1/2/3):7–42, 2002.

[39] C.M. Schor, J.S. Maxwell, and E.W. Graf. Plasticity of convergence-dependent variations of cyclovergence with vertical gaze. *Vision Res,*, 41:3353 3369, 2001.

[40] K. Schreiber, J.D. Crawford, M. Fetter, and D. Tweed. The motor side of depth vision. *Nature*, 410:819–822, 2001.

[41] K.M. Schreiber, D.B. Tweed, and C.M. Schor. The extended horopter: Quantifying retinal correspondence across changes of 3d eye position. *Journal of Vision*, 6:64–74, 2006.

[42] R.A.B. Somani, J.F.X. Desouza, D. Tweed, and T. Vilis. Visual test of Listing's Law during vergence. *Vision Res.*, 38(6):911–923, 1998.

[43] D.A. Southard. Transformations for stereoscopic visual simulation. *Computers & Graphics*, 16(4):401–410, 1992.

[44] H. Steffen, M. Walker, and D.S. Zee. Changes in Listings plane after sustained vertical fusion. *Invest. Ophthalmol. Vis. Sci.*, 4:668672, 2002.

[45] W.M Theimer and H.A. Mallot. Phase-based binocular vergence control and depth reconstruction using active vision. *CVGIP: Image Understanding*, 60(3):343–358, 1994.

[46] P. Thier and R.G. Erickson. Responses of visual-tracking neurons from cortical area MST-I to visual, eye and head motion. *Vision Res.*, 39:3934–3950, 1999.

[47] S. Treue and J.C. Martinez-Trujillo. Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, 399:575–579, 1999.

[48] Y. Trotter and S. Celebrini. Gaze direction controls response gain in primary visual-cortex neurons. *Nature*, 38:239–242, 1999.

[49] Y. Trotter, S. Celebrini, B. Stricanne, S. Thorpe, and M. Imbert. Modulation of neural stereoscopic processing in primate area V1 by the viewing distance. *Science*, 257(5074):1279–1281, 1992.

[50] Y. Trotter, S. Celebrini, B. Stricanne, S. Thorpe, and M. Imbert. Neural processing of stereopsis as a function of viewing distance in primate visual cortical area V1. *J. Neurophysiol.*, 76(5):2872–2285, 1996.

[51] E. Trucco and A. Verri. *Introductory Techniques for 3-D Computer Vision.* Prentice Hall, 1998.

[52] D. Tweed. Visual-motor optimization in binocular control. *Vision Res.*, 37(14):1939–1951, 1997.

[53] D. Tweed and T. Vilis. Geometric relations of eye position and velocity vectors during saccades. *Vision Res.*, 30(1):111–127, 1990.

[54] H. von Helmholtz. *Handbuch der Physiologischen Optik.*, volume 3. Voss, Hamburg, 1867.

[55] R.S. Wrigh, B. Lipchack, and N. Haemel. *OpenGL superbible, Fourth Edition, Comprehensive Tutorial and Reference.* Addison-Wesley, 2007.