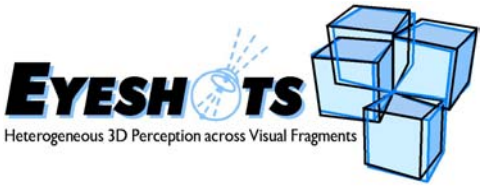


Project no.: FP7-ICT-217077
Project full title: Heterogeneous 3-D Perception across Visual Fragments
Project Acronym: EYESHOTS
Deliverable no: D1.2 (update)
Title of the deliverable: Non-visual depth cues and their influence on perception

Date of Delivery:	01 September 2010	
Organization name of lead contractor for this deliverable:	UG	
Author(s):	A. Canessa, M. Chessa, A. Gibaldi, F. Solari, S.P. Sabatini	
Participant(s):	UG	
Workpackage contributing to the deliverable:	WP1	
Nature:	Report	
Version:	3.1 (update)	
Total number of pages:	74	
Responsible person:	Silvio P. Sabatini	
Revised by:	N. Chumerin, M. Van Hulle	
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 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 are suggested.

Note: this deliverable is based on Deliverable D1.2 (dated 09 Sep 09), which has been updated with new results (see sections 8 and 9) and a more detailed description of the VR simulator used for the statistical analysis of the disparity patterns for a fixating observer (see Appendix B).
The work on the systematic analysis of the disparity patterns in natural scenarios (originally not planned to be part of deliverable D1.2) is still under development.

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 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 (see Appendix B) 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 [37][31][13] 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.

The major achievements can be summarized as follows:

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.

3. A statistical analysis of the disparity patterns for a fixating observer in a real-world environment highlighted differences between the different eye movement paradigms, and suggests the possibility of a mutual "calibration" of the vision and the oculomotor system to compensate disparity components due to the epipolar geometry.

2 Introduction

When we look around in a cluttered environment, or when we inspect a small object, the eyes coordinate themselves to make the lines of sight to intersect in the target dynamically, thus ensuring binocular fusion and accurate scanning of the object. In general, binocular coordinated movements of the eyes occur in a horizontal plane (azimuth), a vertical plane (elevation), or around the visual axes (torsion). Therefore, in addition to horizontal and vertical components of eyes rotations (responsible for version/vergence eye movements), we have extra torsional components, resulting in cycloversion/cyclovergence movements, which affect the two-dimensional (horizontal and vertical) disparity pattern in the peripheral part of the image of the fixated object, as well as in the background. Consequently, the strategy adopted to move the eyes can directly influence the perception of depth, the visual behavior, and eventually the 3D spatial awareness of the world around us. The functional role of binocular eye movements is even more puzzling if we consider that they obey motor constraints that specify the amount of the torsional angles for each gaze direction and vergence angle, according to the Listings law (**LL**) [39][16] and its binocular extension (**L2**) [26][32][25][23]. Since their first establishment and formulation, dating back to Donders and von Helmholtz, a role of the Listings laws in simplifying the motor control, facilitating stereopsis, or both has been advocated. Though, their perceptual consequences still remain an open question considering that, in principle, there are many other ways of reducing the degrees of freedom of eye rotation and solving the binocular correspondence problem, e.g., by adjusting the search zone on the retinae when the eyes move [5], but see [34].

In this report, starting from the seminal work of Tweed [38], we revisited the visuomotor theory of optimal binocular control by generalizing the visual constraint to include the coplanarity of the fixation planes. The resulting analysis confirmed the experimental evidences present in the literature for large and small vergence angles and suggests an extended **LL** to bridge the two extremes for far and near vision conditions, represented by **LL** and **L2**, respectively. Specifically, the effects of the different components of the cost functional are quantitatively evaluated with respect to (1) the relationships between the temporal rotations of the Listing's planes and the vergence and version angles, and (2) the amount of the resulting rotation eccentricities of the eyes. The major contributions of the report are: (1) the derivation of a general expression of the orientations of the eyes, which

is dependent on the coordinates of the fixation point only, and not on the adopted rotation system; (2) the solution of the visuo-motor optimization functional directly with respect to the rotation angles of Listing's planes, and (3) the derivation of a computational justification of the compromise between **LL** and **L2** that takes into account most of the experimental evidences on coordinated binocular eye movements. The rest of the document is organized as follows: in Section 3 the mathematical formalism for describing eye movements is introduced, and a new approach for describing cyclorotation is presented in Section 4. Section 5 describes the proposed visuo-motor optimization strategy. In Sections 6–8 the results are presented and discussed with respect to the existing literature, and to their implications for stereopsis and the design of anthropomorphic robot heads. The perceptual implications of binocular eye coordination on disparity estimation are discussed in Section 9. Specifically, the statistics of the patterns of binocular disparity for a fixating observer, obtained from the range data of real-world peripersonal scenes, have been analyzed. The use of the average disparity field as a *prior* model of the binocular correspondences is discussed with respect to the binocular energy-based models for disparity representation. Concluding remarks are presented in Section 10.

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. The reference position is usually defined as the one 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 cosines defines 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 a point ${}^{h}\boldsymbol{p}$ in the head-fixed system as:

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

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 [10]. 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, followed by a rotation by an angle H_{H} along the $\bar{\boldsymbol{h}}_{\mathrm{Y}}$ axis and finally by 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 azimuth, elevation and torsion 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. Indeed, 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 around 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 unit vector (i.e., versor) of the axis $\bar{\boldsymbol{n}}$, can be derived by introducing *quaternions* and *rotation vector* 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 and 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 product. 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 rotation axis of the quaternion resulting from the product $q \circ s$ becomes:

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

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:

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

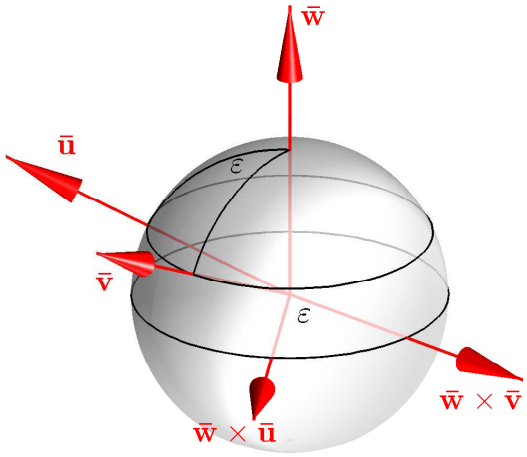


Figure 1: Determination of the eccentricity (ε) of a rotation around a generic axis $\bar{\boldsymbol{w}}$. The versor $\bar{\boldsymbol{u}}$ rotates into the versor $\bar{\boldsymbol{v}}$ while the versor $\bar{\boldsymbol{w}}$ remains fixed. Projecting $\bar{\boldsymbol{u}}$ and $\bar{\boldsymbol{v}}$ onto the plane whose normal is $\bar{\boldsymbol{w}}$ it is possible to derive the angle ε by which to rotate.

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:

$$\begin{aligned} q &= \sqrt{\left(\frac{1 + \frac{\bar{\boldsymbol{w}}\times\bar{\boldsymbol{u}}}{|\bar{\boldsymbol{w}}\times\bar{\boldsymbol{u}}|} \cdot \frac{\bar{\boldsymbol{w}}\times\bar{\boldsymbol{v}}}{|\bar{\boldsymbol{w}}\times\bar{\boldsymbol{v}}|}}{2}\right)} + \sqrt{\left(\frac{1 - \frac{\bar{\boldsymbol{w}}\times\bar{\boldsymbol{u}}}{|\bar{\boldsymbol{w}}\times\bar{\boldsymbol{u}}|} \cdot \frac{\bar{\boldsymbol{w}}\times\bar{\boldsymbol{v}}}{|\bar{\boldsymbol{w}}\times\bar{\boldsymbol{v}}|}}{2}\right)}\bar{\boldsymbol{w}} \\ &= \cos(\varepsilon/2) + \sin(\varepsilon/2)\bar{\boldsymbol{w}} \end{aligned} \tag{24}$$

or, equivalently, the rotation vector:

$$\boldsymbol{r} = \frac{\frac{\bar{\boldsymbol{w}}\times\bar{\boldsymbol{u}}}{|\bar{\boldsymbol{w}}\times\bar{\boldsymbol{u}}|} \times \frac{\bar{\boldsymbol{w}}\times\bar{\boldsymbol{v}}}{|\bar{\boldsymbol{w}}\times\bar{\boldsymbol{v}}|}}{1 + \frac{\bar{\boldsymbol{w}}\times\bar{\boldsymbol{u}}}{|\bar{\boldsymbol{w}}\times\bar{\boldsymbol{u}}|} \cdot \frac{\bar{\boldsymbol{w}}\times\bar{\boldsymbol{v}}}{|\bar{\boldsymbol{w}}\times\bar{\boldsymbol{v}}|}} = \tan(\varepsilon/2)\bar{\boldsymbol{w}} \tag{25}$$

3.2 Listing's Law

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: namely the azimuth and the elevation of the target. This implies that the eye could, in principle, assume an infinite number of torsional postures for any gaze direction. In other words, there are infinite ways to fixate any given target.

Donders [6] 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 asserted 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, independently 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 the 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 [39]. 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 the primary position constitutes the Listing's plane, see Figure 2. For now on we we will to Listing's Law by the abbreviation **LL**. Another way to see the **LL** is to consider the so called *Listing's half angle rule*, which is a generalization of the original **LL** that takes into account not only the primary position but also any other possible starting position. In this form, **LL** states that for any eye position, there is an associated velocity plane such that any position can be reached from that position by rotating about an axis that is confined to this particular plane. The orientation of velocity planes (and hence the rotational axes of the eye) depends on initial eye position: when the eye is in the primary position, the velocity plane is called Listing's plane, which is orthogonal to the gaze line. For any other eye position, the corresponding velocity plane is rotated half as far as the gaze line (half-angle rule), see Figure 3 [41].

3.3 Binocular extension of Listing's Law

LL 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 [2][25][26][28][36][38][32]. During convergence, the eyes' rotation axes still remain 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 4). These convergence-dependent changes of torsional positions (i.e., orientation of

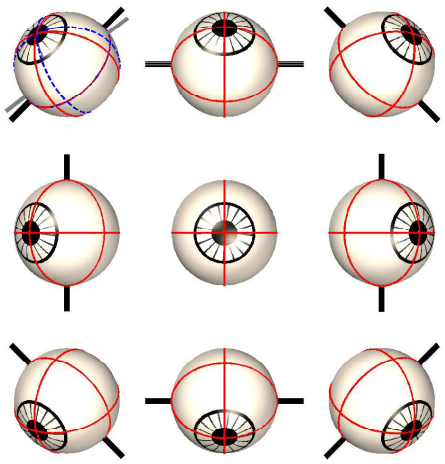


Figure 2: Graphical representation of the Listing's Law. The nine orientations drawn as solid lines correspond to the **LL**: they are obtained through rotation from the primary position about axes (thick solid lines) that lie on the Listing's plane (represented by the paper plane). The position drawn with dashed lines in the top left corner does not obey **LL** because the rotation to this position from primary position occurs about an axis (solid gray line) that is tilted out the paper plane.

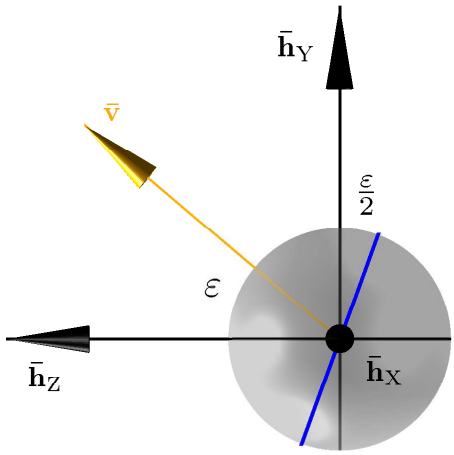


Figure 3: Graphical representation of the Listing's half angle rule. The line of sight when the eye is in the primary position coincides with the $\bar{\boldsymbol{h}}_z$ axis. In this position the Listing's plane orientation is vertical, orthogonal to the line of sight. When the eye starts to move from tertiary position $\bar{\boldsymbol{v}}$, characterized by an eccentricity ε, the orientation of the eye is determined by rotation about the axis that lies on a plane rotated in the same direction, but only half as much as the line of sight, that is $\varepsilon/2$ (blue line).

Listing's plane) have been referred to as the binocular extension of **LL** or, in brief, **L2** [41]. It is worth noting that **L2** is a generalization of the original, monocular, **LL**, 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. From the experimental data emerged a pro-

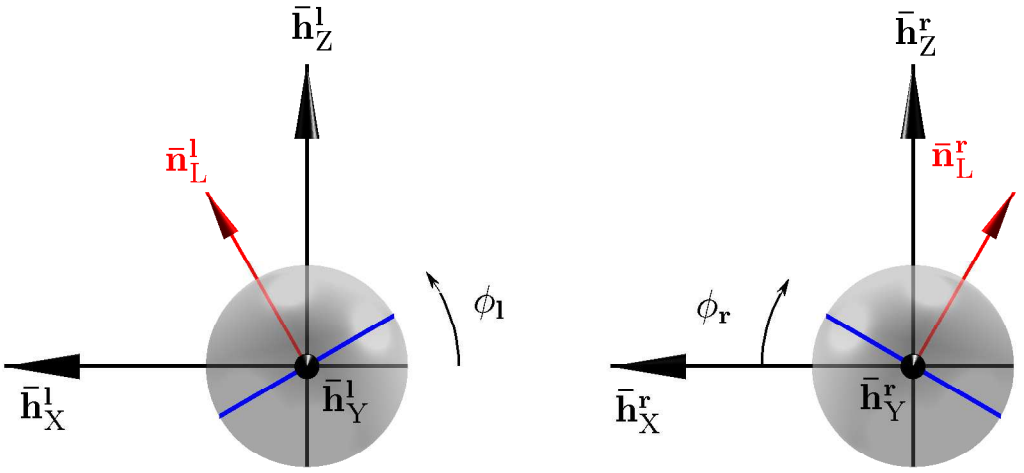


Figure 4: Graphical representation of the binocular extension of Listing's Law, or **L2**. During convergence, the Listing's plane is rotated temporally and symmetrically in each eye by an angle ϕ proportional to the vergence angle.

portionality between the ϕ angles and the vergence, so that in the literature it is well consolidated to express the ϕ's as a linear function of the vergence angle ν:

$$\phi_l = \mu_l \nu \tag{26}$$
$$\phi_r = -\mu_r \nu \tag{27}$$

where μ_l and μ_r are positive constants ranging between 0.20 and 0.41, values derived by fitting of experimental data [39][2][25][26][28][36][38][32][24]. Though, the values of those proportionality constants, and thus those of the rotation angles ϕ's are controversial.

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

$$\bar{\boldsymbol{n}}^l_L = [\sin\phi_l, 0, \cos\phi_l] \tag{28}$$
$$\bar{\boldsymbol{n}}^r_L = [\sin\phi_r, 0, \cos\phi_r]\,. \tag{29}$$

4 A new approach for describing binocular cyclo-rotations

On the basis of the experimental evidences [26] 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. 24 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 (i.e., primary 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 5, by specifying two rotation axes that bring $\bar{\boldsymbol{u}}$ into $\bar{\boldsymbol{v}}$. The first rotation axis is

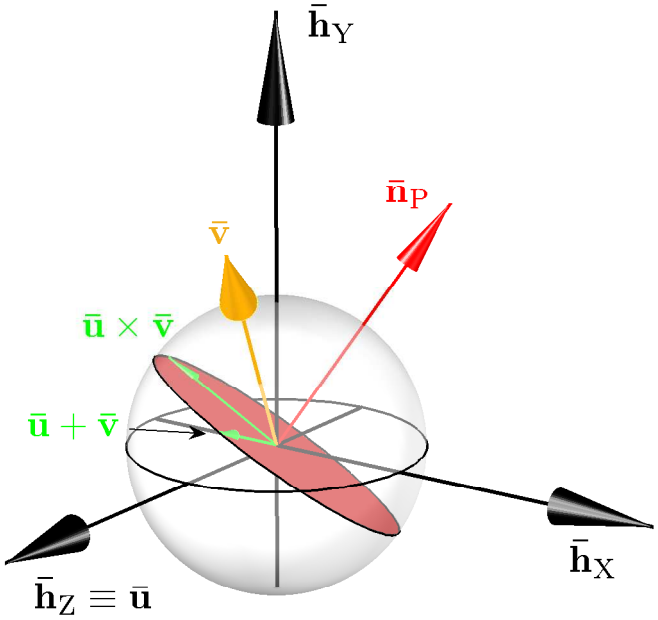


Figure 5: The plane colored in pink contains all the possible directions for rotation axes that bring the versor $\bar{\boldsymbol{u}}$, coincident with the $\bar{\boldsymbol{h}}_{\mathrm{Z}}$ axis, into the versor $\bar{\boldsymbol{v}}$. This plane is identified univocally by its normal versor $\bar{\boldsymbol{n}}_p$, obtained by taking the cross-product of the vector $(\bar{\boldsymbol{u}} \times \bar{\boldsymbol{v}})$ with the vector $(\bar{\boldsymbol{u}} + \bar{\boldsymbol{v}})$.

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:

$$\bar{\boldsymbol{n}}_p = (\bar{\boldsymbol{u}} \times \bar{\boldsymbol{v}}) \times (\bar{\boldsymbol{u}} + \bar{\boldsymbol{v}}) \equiv (\bar{\boldsymbol{v}} - \bar{\boldsymbol{u}}) \,. \tag{30}$$

The approach can be extended to the binocular case straightforwardly (see Figure 6), thus yielding to a pair of planes whose normal versors for the left and the right eye are given by:

$$\bar{\boldsymbol{n}}_p^l = (\bar{\boldsymbol{u}}^l \times \bar{\boldsymbol{v}}^l) \times (\bar{\boldsymbol{u}}^l + \bar{\boldsymbol{v}}^l) \equiv (\bar{\boldsymbol{v}}^l - \bar{\boldsymbol{u}}^l) \tag{31}$$

$$\bar{\boldsymbol{n}}_p^r = (\bar{\boldsymbol{u}}^r \times \bar{\boldsymbol{v}}^r) \times (\bar{\boldsymbol{u}}^r + \bar{\boldsymbol{v}}^r) \equiv (\bar{\boldsymbol{v}}^r - \bar{\boldsymbol{u}}^r) \,. \tag{32}$$

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 6).

Formally, by solving Eqs. (31)–(32) and Eqs. (28)–(29) we find the expressions of the rotation axes $\bar{\boldsymbol{w}}^l$ and $\bar{\boldsymbol{w}}^r$:

$$\bar{\boldsymbol{w}}^l(\phi_l) = \frac{\bar{\boldsymbol{n}}_L^l(\phi_l) \times \bar{\boldsymbol{n}}_p^l}{\left|\bar{\boldsymbol{n}}_L^l(\phi_l) \times \bar{\boldsymbol{n}}_p^l\right|} \tag{33}$$

$$\bar{\boldsymbol{w}}^r(\phi_r) = \frac{\bar{\boldsymbol{n}}_L^r(\phi_r) \times \bar{\boldsymbol{n}}_p^r}{\left|\bar{\boldsymbol{n}}_L^r(\phi_r) \times \bar{\boldsymbol{n}}_p^r\right|} \,. \tag{34}$$

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

$$q_V = \cos(V/2) + \sin(V/2)\bar{\boldsymbol{h}}_{\mathrm{X}} \,. \tag{35}$$

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

$$q_H = \cos(H/2) + \sin(H/2)\bar{\boldsymbol{h}}_{\mathrm{Y}} \,. \tag{36}$$

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

$$q_T = \cos(T/2) + \sin(T/2)\bar{\boldsymbol{h}}_{\mathrm{Z}} \,. \tag{37}$$

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

$$\begin{aligned} q &= q_V \circ q_H \circ q_T = \\ &= (c_{V/2}c_{H/2}c_{T/2} - s_{V/2}s_{H/2}s_{T/2}) + \\ &+ \bar{\boldsymbol{h}}_{\mathrm{X}}(c_{V/2}s_{H/2}s_{T/2} + s_{V/2}c_{H/2}c_{T/2}) + \\ &+ \bar{\boldsymbol{h}}_{\mathrm{Y}}(c_{V/2}s_{H/2}c_{T/2} - s_{V/2}c_{H/2}s_{T/2}) + \\ &+ \bar{\boldsymbol{h}}_{\mathrm{Z}}(c_{V/2}c_{H/2}s_{T/2} + s_{V/2}s_{H/2}c_{T/2}) \end{aligned} \tag{38}$$

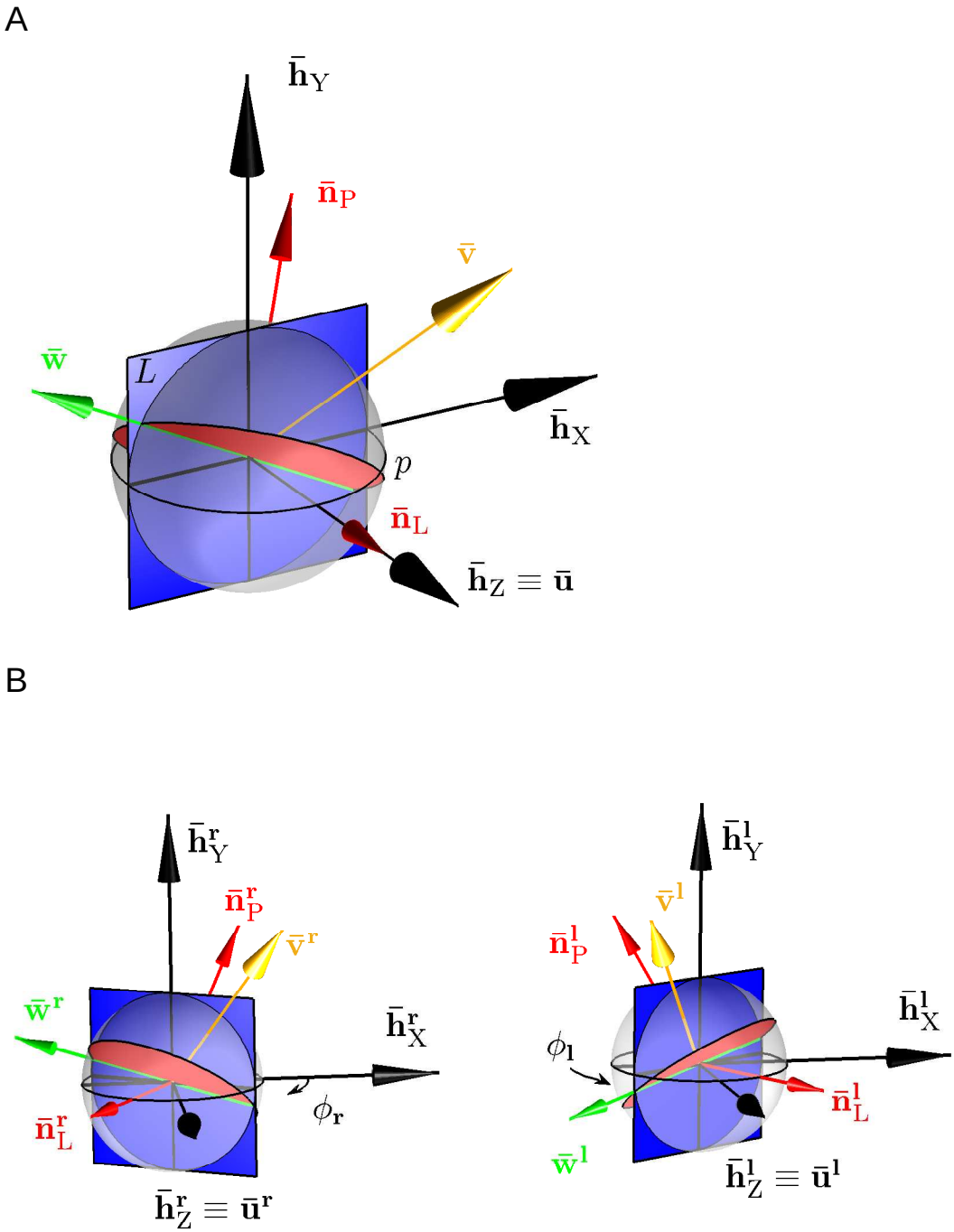


Figure 6: The intersection between the plane p and the Listing's plane L is the versor $\bar{\boldsymbol{w}}$ of the rotation axis that maps $\bar{\boldsymbol{u}}$ into $\bar{\boldsymbol{v}}$ respecting the **LL**. (A) The null vergence case. It is worth noting that L is not rotated with respect to the primary position, which corresponds to the $\bar{\boldsymbol{h}}_{\mathrm{z}}$ axis. (B) The rotation of 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 is rotated by an angle ϕ_r.

where $c_{V/2}$ and $s_{V/2}$ are the cosine and the sine of half the elevation angle V, and, similarly, the other terms denote the same quantities for the other rotation angles.

Taking into account **L2**, each rotation axis for the left eye is perpendicular to the normal of the plane $\bar{\boldsymbol{n}}_L^l$ and each rotation axis for the right eye is perpendicular to the normal of the plane $\bar{\boldsymbol{n}}_L^r$ (see Figure 6B). If we define q_l and q_r the quaternion that represent the position for the left and the right eye, respectively, **L2** requires that [31]:

$$\boldsymbol{q}^l \cdot \bar{\boldsymbol{n}}_L^l = 0 \tag{39}$$

$$\boldsymbol{q}^r \cdot \bar{\boldsymbol{n}}_L^r = 0\,. \tag{40}$$

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

$$\begin{aligned} \tan\frac{T_l}{2} &= -\tan\frac{V_l}{2}\left[\frac{\tan\phi_l + \tan(H_l/2)}{1+\tan(H_l/2)\tan\phi_l}\right] \\ \tan\frac{T_r}{2} &= -\tan\frac{V_r}{2}\left[\frac{\tan\phi_r + \tan(H_r/2)}{1+\tan(H_r/2)\tan\phi_r}\right]\,. \end{aligned} \tag{41}$$

5 Binocular visuomotor coordination: optimization theories revisited

A justification of the **LL** with respect to both "motor" and "visual" efficiency criteria was first put forward by Helmholtz [40], primarily for monocular vision, and then generalized by Tweed [38], including implications for binocular vision. With the same spirit, our aim is to derive the pair of values for the ϕ's angles that allow 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:

$$\mathcal{F}(\phi_l, \phi_r) = (1-\alpha)\mathcal{M}(\phi_l, \phi_r) + \alpha\mathcal{V}(\phi_l, \phi_r) \tag{42}$$

where $\mathcal{M}$ is the motor constraint, $\mathcal{V}$ is the visual constraint (see Section 5.1 and Section 5.2), and α is a positive constant weighting factor ($0 \leq \alpha \leq 1$) that quantifies the relative importance of the two terms. Including both the terms, the eyes have to rotate around the visual axis for granting the visual efficiency be satisfied. It is worthwhile pointing out some differences with respect to Tweed's visuo-motor theory. In [38], 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 eye (cf., [38] p. 1943). 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 [30]): 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. This allows us to take into account the resources that motor and vision systems have at disposition.

5.1 Motor constraint

The motor term in Eq. (42) 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 [38], 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:

$$\mathcal{M}(\phi_l, \phi_r) = \varepsilon_l^2(\phi_l) + \varepsilon_r^2(\phi_r)\ . \tag{43}$$

where $\varepsilon_l(\phi_l)$ and $\varepsilon_r(\phi_r)$ are expressed, using Eq. 24, by:

$$\begin{aligned}
\varepsilon_l(\phi_l) &= 2\arccos\left(\sqrt{\frac{1+\frac{\bar{\boldsymbol{w}}^l(\phi_l)\times\bar{\boldsymbol{u}}^l}{|\bar{\boldsymbol{w}}^l(\phi_l)\times\bar{\boldsymbol{u}}^l|}\cdot\frac{\bar{\boldsymbol{w}}^l(\phi_l)\times\bar{\boldsymbol{v}}^l}{|\bar{\boldsymbol{w}}^l(\phi_l)\times\bar{\boldsymbol{v}}^l|}}{2}}\right) \\
\varepsilon_l(\phi_r) &= 2\arccos\left(\sqrt{\frac{1+\frac{\bar{\boldsymbol{w}}^r(\phi_r)\times\bar{\boldsymbol{u}}^r}{|\bar{\boldsymbol{w}}^r(\phi_r)\times\bar{\boldsymbol{u}}^r|}\cdot\frac{\bar{\boldsymbol{w}}^r(\phi_r)\times\bar{\boldsymbol{v}}^r}{|\bar{\boldsymbol{w}}^r(\phi_r)\times\bar{\boldsymbol{v}}^r|}}{2}}\right)\ .
\end{aligned} \tag{44}$$

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.

5.2 Visual constraints

The visual term in Eq. (42) is used to impose specific binocular correspondences between the stereo image pairs that particular reference surfaces project on the retinas. From this perspective, the visual constraint embraces two types of conditions:

$$\mathcal{V}(\phi_l, \phi_r) = \mathcal{V}_1(\phi_l, \phi_r) + \beta \mathcal{V}_2(\phi_l, \phi_r) \, . \tag{45}$$

With the first term we want to penalize a misalignment of the binocular projections on the horizontal and the vertical retinal meridians of a surface plane orthogonal to the gaze line [18][33][35]. The second term is an extension of the first one by which we impose the coplanarity of the fixation planes [19]. 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 [9][27], see also [21]. The alignment of the horizontal and vertical meridians brings to a situation like the one depicted in Figure 7.

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 7A, this condition can be expressed as:

$$\begin{aligned} \mathcal{V}_1(\phi_l, \phi_r) &= \left[1 - \left(\bar{\boldsymbol{n}}_s \times \bar{\boldsymbol{e}}^l_{\mathrm{Y}}(\phi_l)\right) \cdot \left(\bar{\boldsymbol{n}}_s \times \bar{\boldsymbol{e}}^r_{\mathrm{Y}}(\phi_r)\right)\right]^2 \\ &+ \left[1 - \left(\bar{\boldsymbol{n}}_s \times \bar{\boldsymbol{e}}^l_{\mathrm{X}}(\phi_l)\right) \cdot \left(\bar{\boldsymbol{n}}_s \times \bar{\boldsymbol{e}}^r_{\mathrm{X}}(\phi_r)\right)\right]^2 \end{aligned} \tag{46}$$

where $\bar{\boldsymbol{n}}_s$ is the normal to the reference surface, $\bar{\boldsymbol{e}}_{\mathrm{X}}$ and $\bar{\boldsymbol{e}}_{\mathrm{Y}}$, for both eyes, are directed along the horizontal and vertical retinal meridians and they depend by the values of the ϕ's through the following relations:

$$\bar{\boldsymbol{e}}_{\mathrm{X}} = q(\phi) \circ \bar{\boldsymbol{h}}_{\mathrm{X}} \circ q^{-1}(\phi) \qquad \bar{\boldsymbol{e}}_{\mathrm{Y}} = q(\phi) \circ \bar{\boldsymbol{h}}_{\mathrm{Y}} \circ q^{-1}(\phi) \, . \tag{47}$$

$$q(\phi) = \cos(\varepsilon(\phi)/2) + \sin(\varepsilon(\phi)/2)\bar{\boldsymbol{w}}(\phi) \tag{48}$$

To probe further the meaning of this visual constraint, it is worth noting that the orientation difference of the retinal projections of points in the 3D environment provides a visual cue for the orientation and inclination of a visible surface. **LL** generates a cyclovergence for both horizontal and vertical eye gaze changes, and this cyclovergence yields a

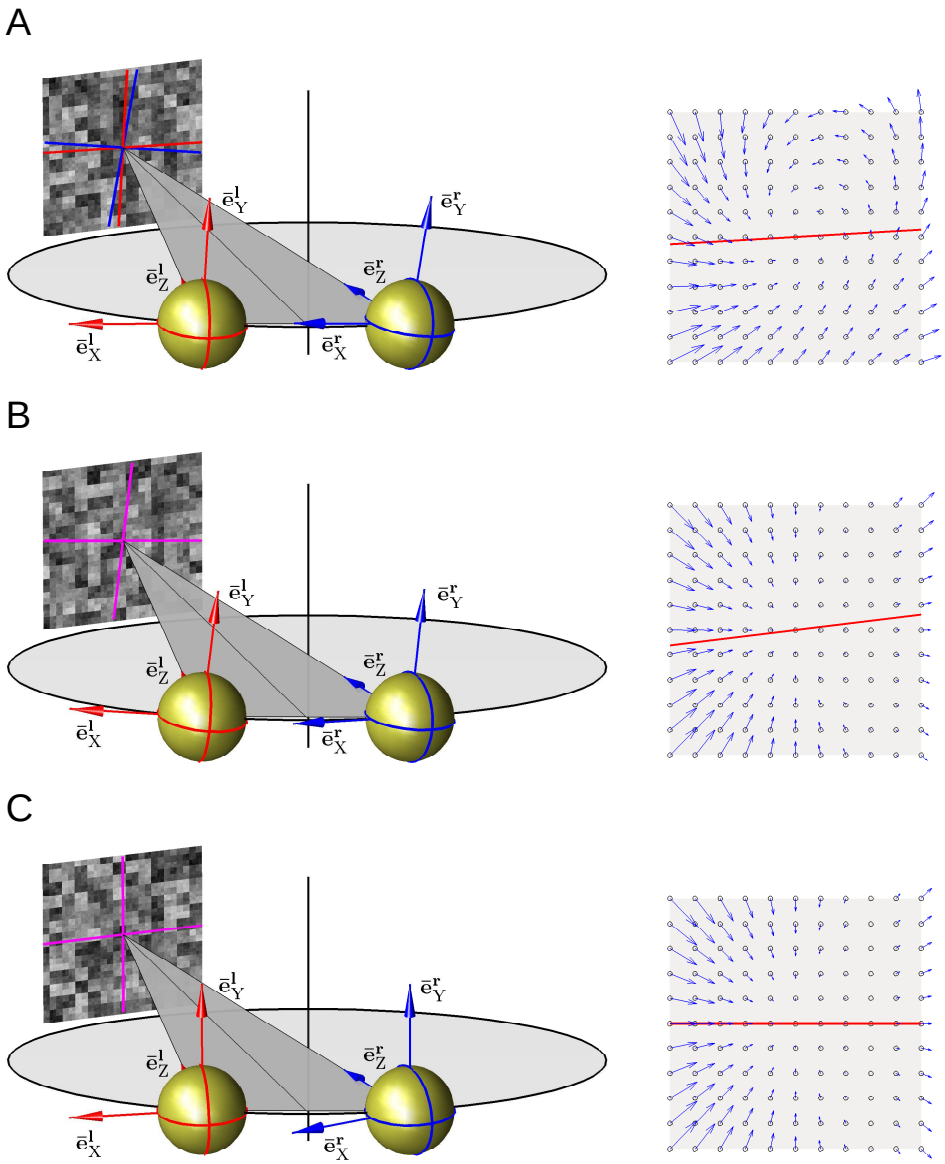


Figure 7: Graphical visualization of the binocular correspondences in the retinal planes imposed by the different constraints. For each fixation point, the reference surface is considered always orthogonal to the gaze line. In the three represented cases the eyes fixate the same point in the world but in different ways. (A) The eyes are characterized by a non optimal orientation. In this case the projections of the horizontal and vertical meridians for the left (blue lines) and the right (red lines) eye, respectively, are not aligned on the reference surface. (B) The backprojections of the horizontal and the vertical meridians of both eyes are aligned on the reference surface (purple lines). (C) In this case, besides the alignment of the meridians, the fixation planes of both eyes are coplanar, as it results from the parallelism of the versors $\bar{\boldsymbol{e}}_y^l$ and $\bar{\boldsymbol{e}}_y^r$. For each of the three represented situations the resulting disparity fields are depicted on the right. The red line is the projection of the visual plane on the retina of the left eye.

gradient of disparity along the vertical and the horizontal meridians. The torsional bias (cyclorotation) of the eyes influences the orientation of the lines on the retinal plane, and this variation generates ambiguities on the 3D information, altering a correct perception of the objects. The role of the visual constraint $\mathcal{V}_1$ is to maximize the visual efficiency aligning the retina (or better the retinal meridians) of both eyes, see Figure 7B. We will see that the visual constraint $\mathcal{V}_1$ yields equal torsions (expressed in Helmholtz coordinates) in the left and the right eye. Yet, in general, a redefinition of the visual constraint that models the desired "visual efficiency" might result in different values for torsions of both eyes.

Extended 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 [19], who claims that, with the head at rest, the eyes move without torsions 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 7C, to make the fixation planes of both eyes coplanar, we have to impose that $\bar{\boldsymbol{e}}^l_{\mathrm{Y}}$ and $\bar{\boldsymbol{e}}^r_{\mathrm{Y}}$, normal to the planes, be parallel. This defines the extension of the visual part of the cost function, namely the second visual criterion:

$$\mathcal{V}_2(\phi_l, \phi_r) = \left[1 - \bar{\boldsymbol{e}}^l_{\mathrm{Y}} \cdot \bar{\boldsymbol{e}}^r_{\mathrm{Y}}\right]^2 . \tag{49}$$

6 Results

We carried on a numerical minimization process in order to find the minimum of the functional $\mathcal{F}(\phi_l, \phi_r)$ with respect to the two variables ϕ_l and ϕ_r for different values of the weighting parameters α and β. The results are presented in the following two paragraphs: first analyzing the visual constraint $\mathcal{V}_1(\phi_l, \phi_r)$ only, and then including the second visual constraint $\mathcal{V}_2(\phi_l, \phi_r)$.

Alignment of retinal meridians ($\beta = 0$). In general, we should 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, we found 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 ϕ values that, besides maximizing the visual efficiency, they keep the eyes to move with

the minimum eccentricity. We have verified, in fact, that the mean worsening is about 0.05% with a peak value around 0.3%. A further discussion on this point can be found in Section 8.1. From the optimization we found that the values of the ϕ angles for both eyes are equal in magnitude but opposite in sign, see Figure 8. 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 ϕ 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. (41), we can obtain an analytical expression of the ϕ angles:

$$\phi_l = -\phi_r = \frac{1}{2} \arcsin\left(\frac{\sin(\nu/2)}{\cos(\gamma/2)}\right) \tag{50}$$

where $\nu = H_r - H_l$ and $\gamma = \frac{H_r + H_l}{2}$ are the vergence and the version angles, respectively. From Eq. 50 it is possible to see that the angles are not only a function of the 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 ϕ 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 ϕ 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 [34][35] always claimed the existence of two different and distinct strategies the eyes adopt to move and fixate: the **LL** for far fixation distances, which has motor advantages, and **L2**, which has visual advantages. In reality, the oculomotor system should strike a balance between them. We found that only one strategy could exist: a *generalized* Listing's Law (**LLx**). 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 the minimal eccentricity of the eye rotations. Yet, there might be the case that the binocular coordination of the eye rotations imposed by the **LL** have a different motivation, e.g., associated to a simplification or an increased robustness of the control strategy, as suggested by [29]. The oculomotor system faces indeed a serious problem in coordinate the eyes to hold their positions in three dimension thus maintaining a stable fixation. The eye is still only when the total torque which the globe is subjected to is zero, i.e., when the torque exerted by the six extraocular muscles balances the elastic torque of the orbital tissue. The oculomotor system has to cope with two levels of complexity: the high geometric non-linearities arising from the configuration of the muscles, and the high system redundancy. Porrill and colleagues [29] found that the muscle innervations associated with eye positions that are compatible with the **LL** are characterized by the property of being separable, thus facilitating the control of the eye plant. Accordingly, at least at first approximation, the innervations of the horizontal recti muscles corresponding to a given horizontal position are independent of the vertical eye position, while innervations of the four cyclovertical muscles, for a given vertical orientation, are unaffected by the horizontal eye position.

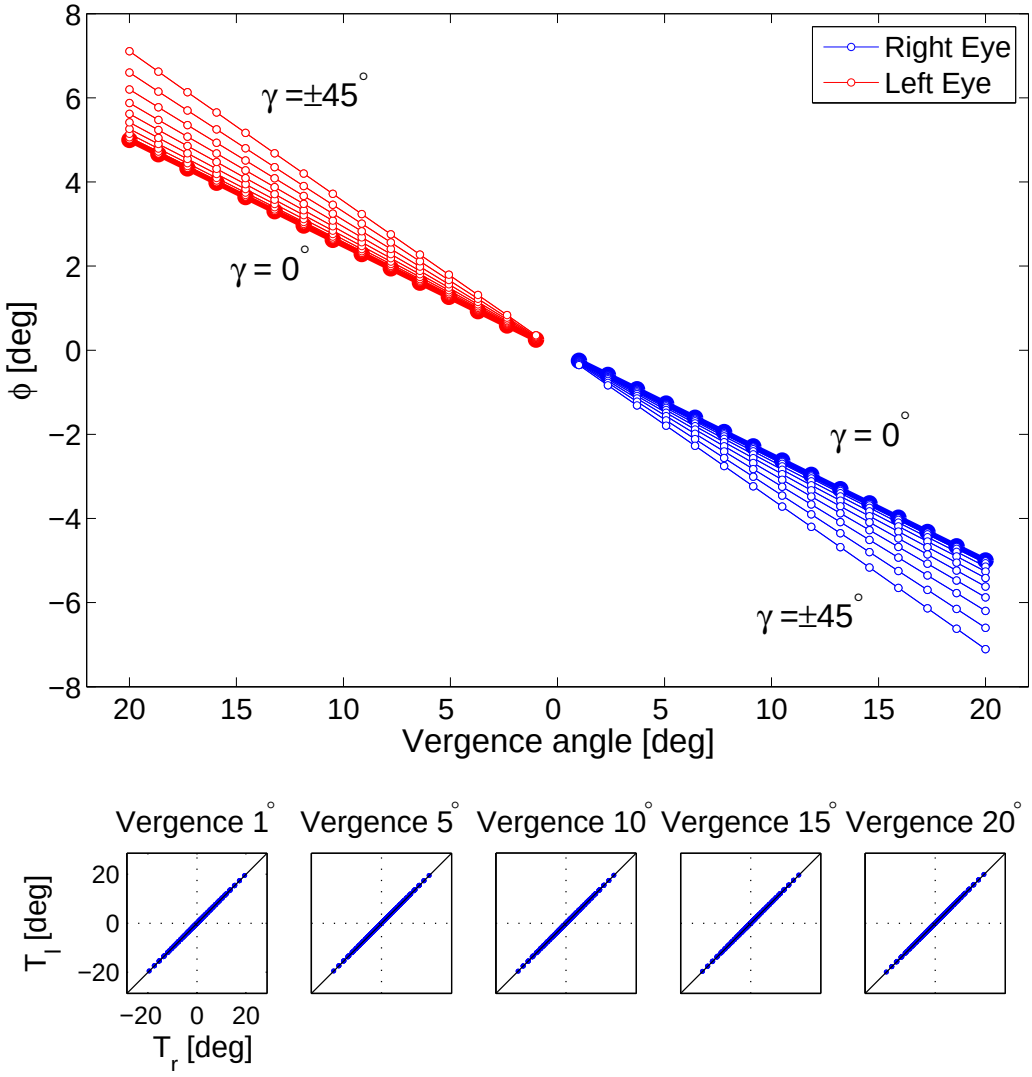


Figure 8: Rotation of the Listing's planes as a function of the vergence angle and the resulting eye torsion when only the visual constraint $\mathcal{V}_1$ is considered ($\beta = 0$). The values of the ϕ angles (ϕ_l in red and ϕ_r in blue) are depicted. It is worth noting that the ϕ angles are linear functions of the vergence, but the slope of the curves changes with the version γ. The thick line corresponds to the null-version case ($\gamma = 0°$). The insets at the bottom show the Helmholtz torsion T_r for the right eye plotted against the associated Helmholtz torsion T_l for the left eye. A strong identity relationship results between them. The range spanned by the torsions remains constant as the vergence varies.

Inclusion of the coplanarity of the fixation planes ($\beta \neq 0$). Differently from the previous case, now 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 a solution that allows us to nullify the cost of the visual component: this solution is represented by the classic tilt-pan system. Obviously, this type of solution is not biologically plausible, because it violates the **LL** for far fixation. In this condition, the ϕ angles should decrease with decreasing vergence, whereas in a tilt-pan system they remain confined in a wide range, see Figure 9.

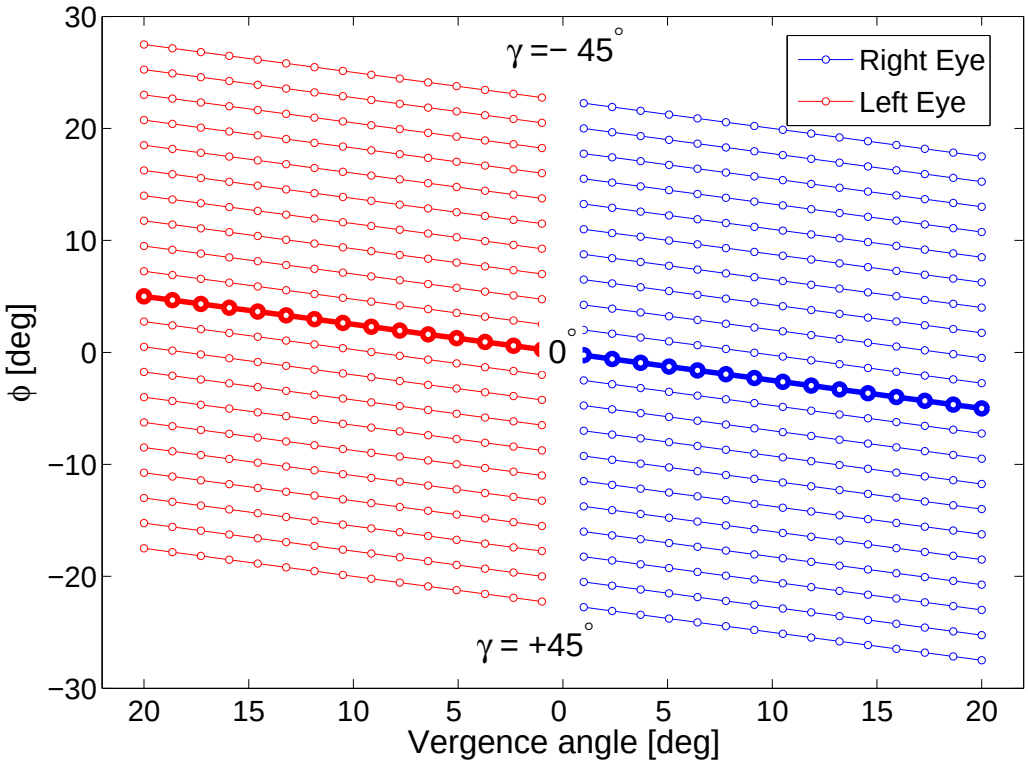


Figure 9: Rotation of the Listing's planes as a function of the vergence angle when the motor constraint $\mathcal{M}$ is not considered ($\alpha = 1$). The values of the ϕ angle (ϕ_l in red and ϕ_r in blue) are depicted. These results replicate the behavior of a classic tilt-pan robotic system. It is worth noting how the values of the ϕ angles violate the Listing's Law, since a non-null offset remains for small vergences when the version angle is not zero. The thick line corresponds to the null-version case ($\gamma = 0°$); non-null version values shifts up and down the curves, only, without changing their slope.

By a proper compromise between the motor and the visual components we found the value for the ϕ angles depicted in Figures 10 and 11.

Differently from the previous case, the ϕ's now are no more one the opposite of the other, so $\phi_l \neq -\phi_r$. However, they respect the **LL** for far fixation, approaching zero when the vergence decreases, and they behave like in a tilt-pan system when the vergence

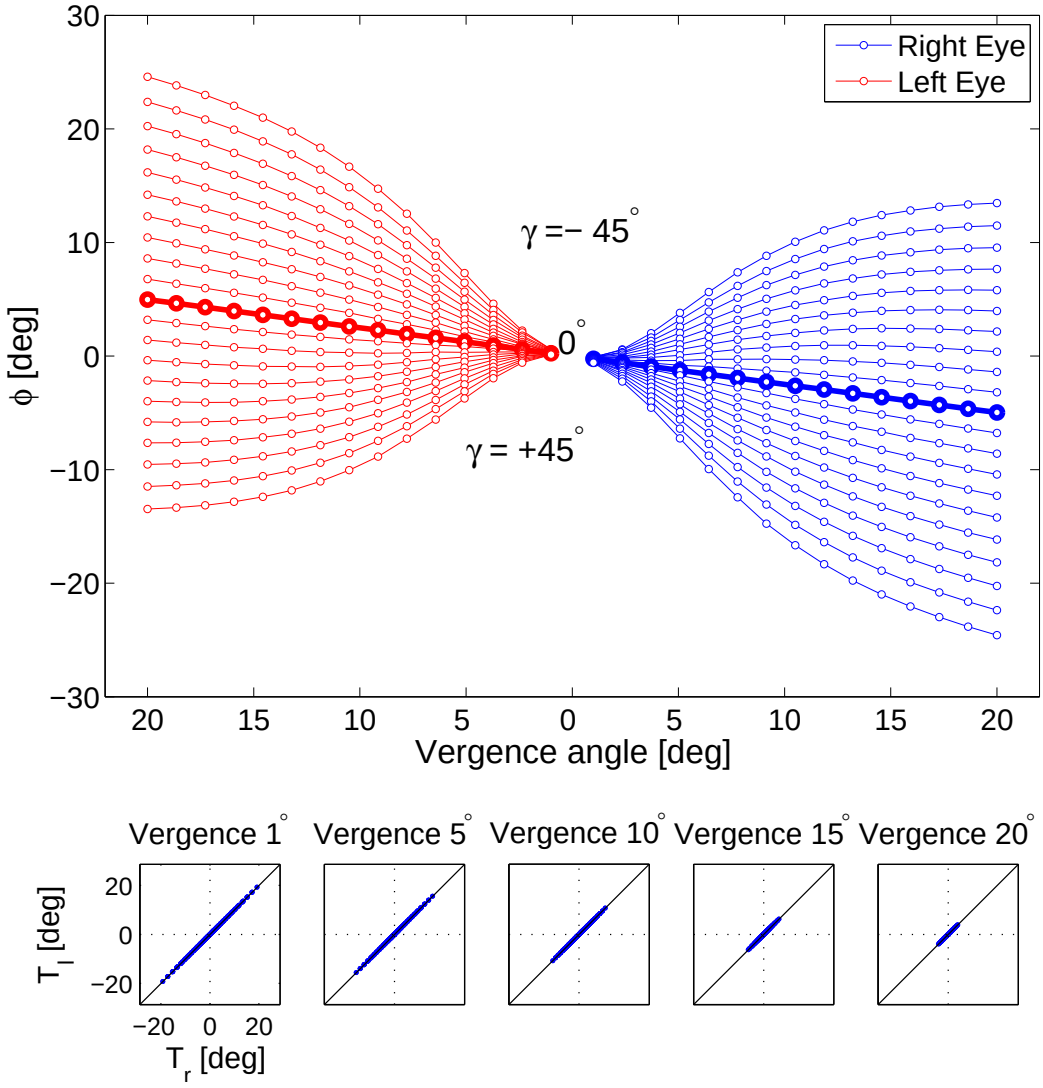


Figure 10: Rotation of the Listing's planes as a function of the vergence angle and the resulting eye torsion for different values of the optimization parameters. The values of the ϕ angle (ϕ_l in red and ϕ_r in blue) are depicted for $\alpha = 0.95$, $\beta = 3$. It is worth noting that, for small versions, the ϕ's are still linear functions of the vergence, with a slope that increases or decreases with the version. The behavior of the ϕ angles replicates what observed for a classical tilt-pan system for large vergences and tends to respect Listing's Law ($\phi \to 0$) for small vergence. The insets at the bottom of each figure show the associated Helmholtz torsion angles. Still, a strong identity relationship results between them, but the range spanned by the torsions decreases as the vergence increases.

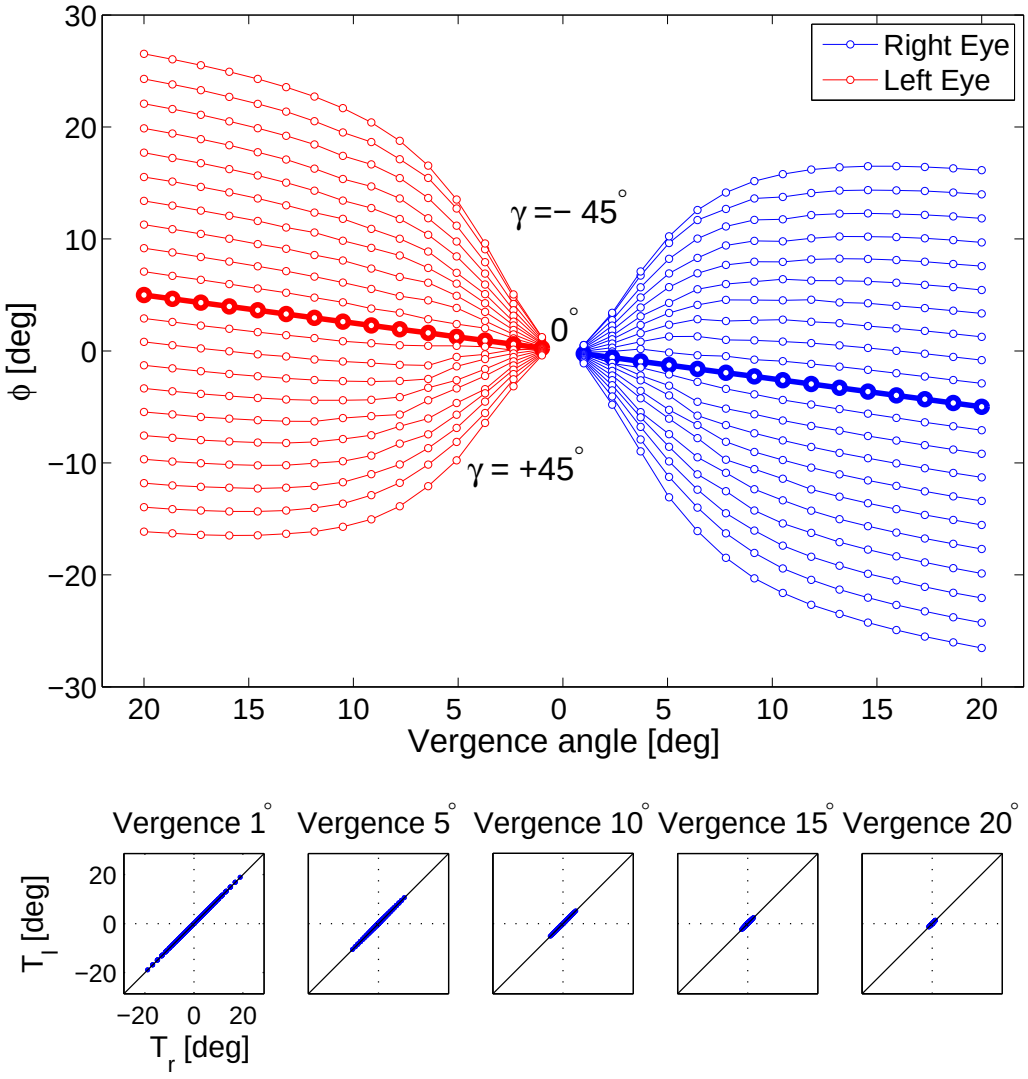


Figure 11: Same as in Figure 10 but for $\alpha = 0.95$, $\beta = 10$.

increases. It is worth noting that since a compromise exists, the eye rotations obtained with these ϕ angles are characterized by larger eccentricities, that penalize the motor efficiency. However, as in the previous case, we derived that the mean percentage error of eccentricity is in the range of $0.4\% \div 1.2\%$ with peak values in the range $4\% \div 6\%$. Also in this case we refer to Section 8.1 for deep information.

Moreover, for small versions, it is still respected the experimental evidences that both ϕ 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 visual constraint only, the ϕ angles span a wider range. 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.

7 Functional implications for depth vision

The visual constraints introduced in Section 4 are motivated by 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 favor of a broader perspective in which we take as a constraint 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. The epipolar lines are defined as the segments on the image plane of one eye on which all the possible matches of a given point on the retina of the other eye fall. Thus, they represent the loci of all the possible matching points 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 have to be found. From this perspective, having a general design strategy for the oculomotor system behavior, that minimizes the motion of the epipolar lines, would reduce the search zone and thus 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 as an optimization process in Section 5. Specifically, since the distances nearby to the fixation point tend to vary less than in the periphery, due to the local smoothness of the visible surfaces, we simplified the workspace around each fixation point, as delimited by a hyperboloid function (see Figure 12). The hyperboloid is always oriented along the gaze direction. For each retinal point for the left eye, we backproject a ray ${}^{e}\boldsymbol{v}^{l}$ and we find the intersection of this ray with the two sheets of the hyperboloid. We parameterize ${}^{e}\boldsymbol{v}^{l}$ through a vector ${}^{e}\boldsymbol{r}^{l}$, which represents the coordinates of any point in the image plane, and a multiplicative factor λ, so that:

$$ {}^{e}\boldsymbol{v}^{l} = \lambda {}^{e}\boldsymbol{r}^{l} \, . \tag{51} $$

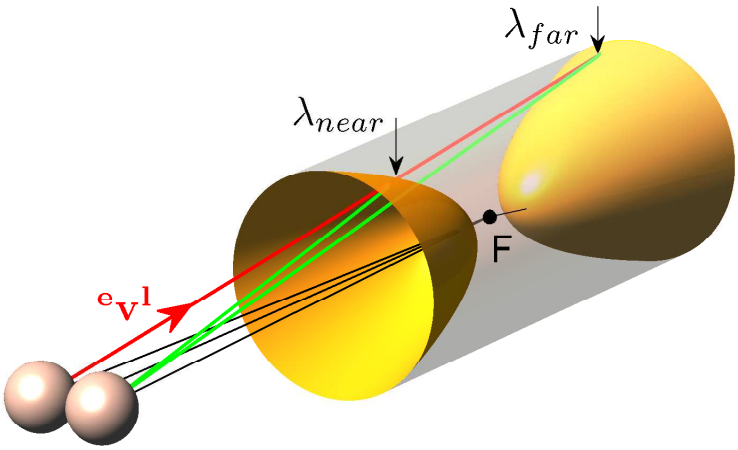


Figure 12: The three dimensional workspace used for analyzing the movement of the epipolar lines. Around each fixation point F, a volume space delimited by the two sheets of a hyperboloid function is considered. For each projecting ray ${}^{e}\boldsymbol{v}^{l}$ (red line) we calculate the intersection point with the two sheets and the corresponding values λ_{near} and λ_{far} of the parameter λ. By integrating between these two extremes we obtain the mean disparity for each projecting ray. The green lines are the backprojecting rays of the intersection points on the right retina.

Let define λ_{near} and λ_{far} the values for λ for which ${}^{e}\boldsymbol{v}^{l}$ intersects the two sheets of the hyperboloid. Now, exploiting this parameterization, it is possible to express the disparity in the following way:

$$\boldsymbol{d}({}^{e}\boldsymbol{r}^{l}, \lambda, \mathbf{R}^{\mathbf{r}}, \mathbf{R}^{\mathbf{l}}, \boldsymbol{p}^{\mathbf{l}}, \boldsymbol{p}^{\mathbf{r}}) = \mathbf{f}\frac{\left[\mathbf{R}^{\mathbf{r}}\right]^{\mathrm{T}}\left[\mathbf{R}^{\mathbf{l}}\lambda\,{}^{e}\boldsymbol{r}^{\mathbf{l}} + \boldsymbol{p}^{\mathbf{l}} - \boldsymbol{p}^{\mathbf{r}}\right]}{\left[\mathbf{R}^{\mathbf{r}}\right]^{\mathrm{T}}\bar{\boldsymbol{h}}_{\mathrm{Z}}\left[\mathbf{R}^{\mathbf{l}}\lambda\,{}^{e}\boldsymbol{r}^{\mathbf{l}} + \boldsymbol{p}^{\mathbf{l}} - \boldsymbol{p}^{\mathbf{r}}\right]}, \tag{52}$$

where f is the focal length of the visual system. If we fix the orientations of the eyes $\mathbf{R}^{\mathbf{l}}$ and $\mathbf{R}^{\mathbf{r}}$, and their position $\boldsymbol{p}^{l}$ and $\boldsymbol{p}^{r}$, the disparity $\boldsymbol{d}({}^{e}\boldsymbol{r}^{l}, \lambda)$ for each projecting ray becomes a function of the parameter λ, only. By integrating the Eq. 52 between λ_{near} and λ_{far} we can obtain the mean disparity:

$$\widehat{\boldsymbol{d}}({}^{e}\boldsymbol{r}^{l}) = \frac{1}{\lambda_{far} - \lambda_{near}}\int_{\lambda_{near}}^{\lambda_{far}} \boldsymbol{d}({}^{e}\boldsymbol{r}^{l}, \lambda)\, d\lambda\,. \tag{53}$$

Hence, we calculate the disparities of the intersection points $d_{far}({}^{e}\boldsymbol{r}^{l}, \lambda_{far})$ and $d_{near}({}^{e}\boldsymbol{r}^{l}, \lambda_{near})$. These three disparities, by construction, lie on the epipolar line of each retinal point. Hence, we have found the epipolar lines for different fixation points (varying the vergence in a range from 1° up to 20°, varying the version and the elevation in a range in between ±45°) and for the different visuo-motor strategies obtained by the minimization process for different values of the optimization parameters α and β. Figure 13 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

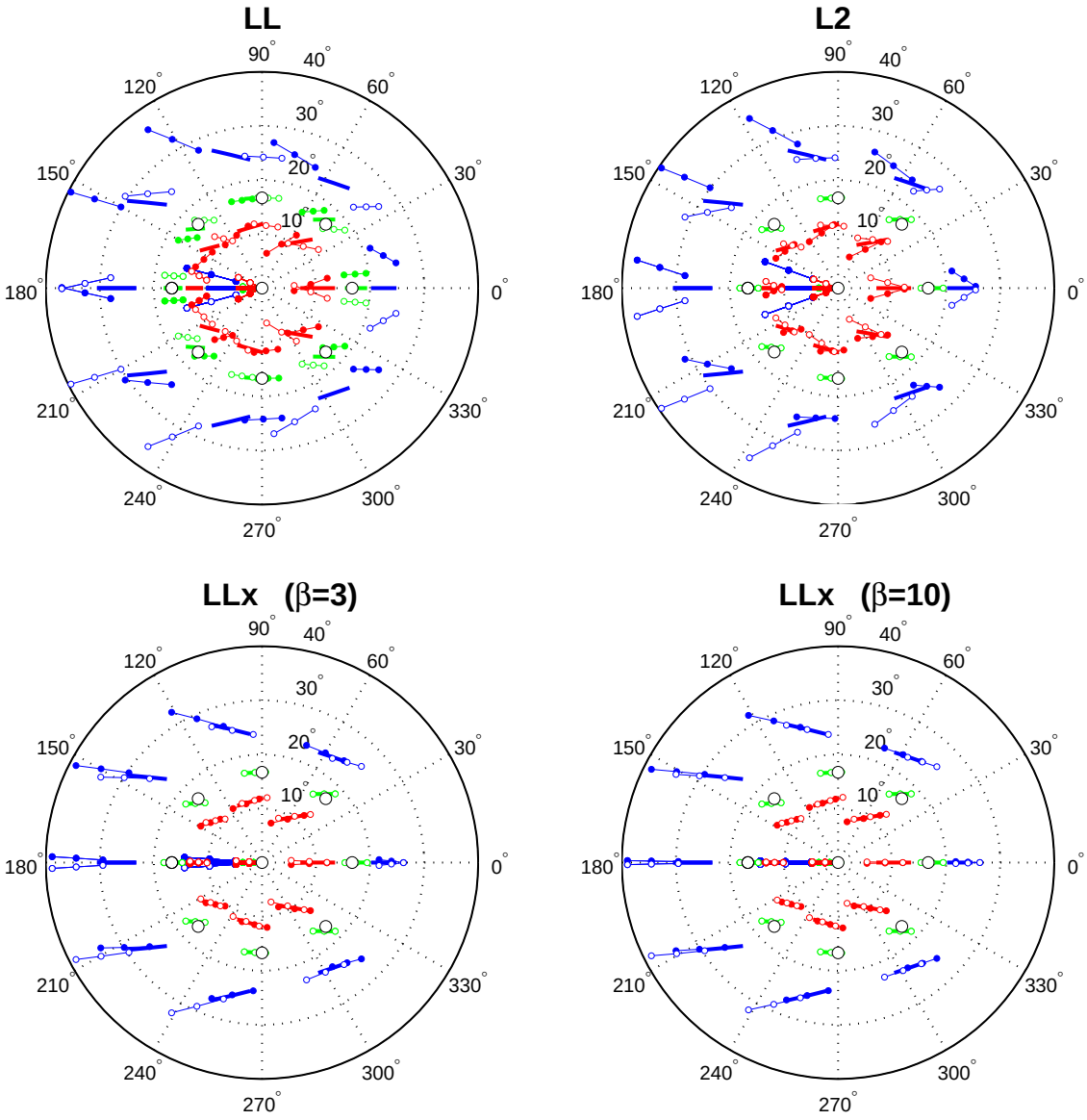


Figure 13: The resulting epipolar lines for the different strategies obtained from the optimization of visuomotor efficiency. The results are related to a fixed vergence of 20° and to different values of version and elevation. The blue, the green and the red families correspond to a version angle equal to $-45°$, $0°$ and $45°$, respectively. The filled circle, the no-circle and the open circle lines correspond to an elevation angle equal to $-45°$, $0°$ and $45°$, respectively. Listing is characterized by $\alpha = 0$, $\beta = 0$. $\mathcal{M} + \mathcal{V}_1$ is characterized by $\alpha = 0.95$, $\beta = 0$. $\mathcal{M} + \mathcal{V}_1 + \mathcal{V}_2$ ($\beta = 3$) is characterized by $\alpha = 0.95$, $\beta = 3$. $\mathcal{M} + \mathcal{V}_1 + \mathcal{V}_2$ ($\beta = 10$) is characterized by $\alpha = 0.95$, $\beta = 10$ and it is very close to the behavior of a classical tilt-pan system. The reference retinal points are indicated by large open circles.

justify the search for a "correct" (i.e., most convenient) strategy for the oculomotor system. There is a strong dependence of the disparities patterns on 3D gaze position, as it is evident from Figure 13 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.

8 Comparison with the literature

In the present study, we proposed a mathematical framework to investigate the functional implications of the rotations of the eyes described by the Listing's Laws. More precisely, we wanted to characterize the "sensory" or the "motor" nature of the **LL** and, in particular, of the **L2** under a visuo-motor optimization framework, by revisiting the original formulation proposed first by Tweed, Van Rijn, Van der Berg.

The results we obtained (by functional minimizations) are not in contrast with the experimental evidences and suggest a continuum of behaviors from far to near vision (i.e., from **LL** to **L2**) also for non-null version conditions.

In particular, we have found, as a result, (1) the identity of Helmholtz torsions ($T_l = T_r$), postulated by Tweed [38], for different instances of the visual constraint; and (2) the proportionality relationship, represented by the factor μ, between the rotation of the Listing's planes and the vergence angle. It is worth noting that, in our formulation, we directly expressed the rotation of the eyes as a function of the ϕ angles that characterize the temporal rotations of the Listing's planes, postulated by **L2** in near vision. Therefore, we did not look for the torsions that permit to meet the constraints posed by Listing, explicitly, rather the torsional components of the quaternions that describe the current positions of the eyes. Indeed, speaking in terms of torsions implies (1) the choice of the coordinate system to adopt to express the rotations, and (2) some assumptions on the relationships between T_l and T_r. Accordingly, if we would use the Fick coordinate system, instead of Helmholtz's, the relation between T_l and T_r must be different. Thus, working directly with the components of the rotation allows us reasoning in the most general way as possible.

LL has been known for more than a century. It was formulated by Listing in the first half of the 19^{th} century, but its significance was understood at all only once Helmholtz verified it by measurements of afterimage and included it in his treatise [40], spending more than 50 pages. This law states that for any position adopted by the eye, then there is a plane (velocity plane) such that the eye adopts only those orientations that can be reached from the current position by rotating about an axis that lies in this plane. Different positions have different planes, but there is one which is characterized by the fact that its plane is orthogonal to the gaze line: this position is therefore defined as the *primary position*, and its plane *Listing's plane*. This law is valid only when the the eyes are fixating at infinity, or when the vergence is null. Allen [1], in fact, first discovered

that **LL** is no more valid when the fixation comes near the subject. In 1992 Mok and colleagues [26], registering eyes postion during fixation on concave isovergence surfaces, observed an eyes' behavior different from that expected by the **LL**. In fact, in this case the Listing's planes tilted temporally by an angle proportional to the vergence, through a factor μ. This is generally known as the binocular extension of **LL**, or **L2**. From the mid-19^{th} many are the theories that have been proposed to explain why the eyes have to behave in these ways. The first, proposed by Fick and Wundt [40], is the so called *motor theory*. This states that **LL**, minimizing the eccentricity of eye rotation, maximizes the motor efficiency and optimizes the effort of the extraocular muscles. The problem is that this motor principle cannot explain **L2**. Helmholtz and Hering [40][17] went one step further. In their *visual theories*, subsequently taken up by Hepp [15], they hypothesized that **LL** brings some visual advantages, optimizing the retinal image flow. Also these theories unfortunately does not succeed in explaining the tilting of Listing's planes when vergence is different from zero. Finally, Tweed [38], exploiting the arguments of [32], that a consequence of **L2** was to make equal the Helmholtz torsional angles of both eyes, proposed his *visual-motor theory*. This theory is characterized by the combination of the motor theory of Fick and Wundt (i.e., the minimization of eccentricity) with the visual theory of Van Rijn and Van den Berg (i.e., the alignment of the retinal images of the visual plane). From the first time Tweed proposed the visual-motor theory, this has been taken as reference to explain **L2**. In reality, the same Tweed and many other researchers [38][34][35] was always of the opinion that eyes do not follow **L2** precisely, but they strike a compromise between the motor advantages of **LL** and the visual advantages of **L2**. Indeed, we have verified that it is not convenient to deal with a monocular **LL**, valid for null vergence, and a binocular version of it (or **L2**), valid for near fixation. We argue that only a "generalized" Listing's Law (**LLx**) can be considered, forming a continuum that ranges from large to small vergences, and that includes in itself all the observed experimental evidences. As a corollary result, we found is that it seems correct to think of the **LLx** as the result of an optimization process that takes into account both the maximization of a motor and a visual efficiency, as it is largely reported in the literature. We have also observed that the motor part might not play a crucial role in the minimization process if we consider only a visual constraint, which imposes that same points in the 3D world project on the horizontal and vertical meridians of both eyes. Hence, **L2** minimizes intrinsically the rotation eccentricity, without requiring compromises between motor and visual implications as Tweed, Schreiber and others suggested in their works. The addition of a second visual constraint, which includes the coplanarity of the fixation planes (cf. [19]), better represents this compromise, recollecting all the experimental evidences in a unique theory, the **LLx**.

8.1 Robotic models of human vision

For the sake of simplicity, the major part of the active binocular robot heads are characterized by a common motor for the elevation (tilt) and a pair of independent motors for the azimuth (pan), which are always fixed parallel to each other. Accordingly, binocular robot heads usually adhere to the Helmholtz coordinate model, as the natural configu-

ration to parameterize the rotational position of the eyes. Moreover, the common tilt implies a structural coplanarity of the fixation planes and prevents torsional rotations around the visual line. Hence, it is not surprising that in our argumentation the geometrical behavior of a standard tilt-pan system results as the extreme case of the extended visual constraint $(\mathcal{V}_1 + \mathcal{V}_2)$. Though, it is worth underlining that such a configuration links up with a Listing-compliant one when the value of the vergence angle decreases. The way by which one configuration turns into the other depends on the relative importance/weight of the motor and visual constraints, in line with the general principle of a Listings law as the oculomotor systems unique behavioral solution to its degrees of freedom problem [4] according to which the oculomotor system reduces its potentially redundant degrees of freedom when the behavioral situation demands or permits it. From this perspective, it might be functionally useful to dispose of a robotic platform capable of mimicking human-like binocular eye movements, since that would guarantee an optimal flexibility, being the *task* to condition the behavior of eyes movements. Despite their functional advantages, torsional eyes movements have been usually ignored by stereo research in Computer Vision and the implementation of robotics stereo head with 3 degrees of freedom for each camera are very seldom [22][11][8]. More specifically, mechanical implementation of Listings law (e.g., [3]), are in general more difficult to construct and control. As an alternative, the effect of torsion about the optical axis of each eye can be simulated by image processing. In this direction, Miles and Horaud [14] recently proposed a systematic formulation to synthesize the effect of Listings law on a set of calibrated images obtained from a standard (tilt-pan) robotic camera mounting. To this purpose, the torsion angle function, depending on the visual direction, has been derived and then the associated rotation matrices for a full rotation of the eyes. The principles and the relations derived in Section 3 and Section 4 for binocular near vision (e.g., cf. Eq. (41) and Eq. (50)), extend and complement the treatment of [14] by including the vergence angle, and can be directly used in their approach for further analysis of human-like binocular kinematics in robot heads. In this direction, we decided to systematically quantify the magnitude of Helmholtz torsions and the degree of failing in terms of motor efficiency for the different systems that result from the functional minimization. To conclude and to complete our treatment we decided to make the same analysis Hansard and Horaud [14] carried on. This for having an idea on which was the magnitude of Helmholtz torsions and the degree of worsening in term of motor efficiency among the different systems obtained by the minimization. Let us consider to locate the fixation points over spherical surfaces of different radii. The best way to describe this geometry is to use the spherical (or polar) coordinates: r, θ, ψ. Let θ be the azimuthal angle in the xy-plane from the x-axis with $0 \leq \theta < 360°$, ψ the polar angle (also known as the zenith) from the positive z-axis with $0 \leq \psi \leq 180°$, and r the distance (i.e., the radius) from a point located to the origin. Here, for the sake of uniformity with the previous analysis, instead of using the radial distance r, we have considered the corresponding vergence angle υ:

$$\upsilon = 2 \arctan\left(\frac{I}{2r}\right) , \tag{54}$$

where $I = 6$ cm is the interocular distance. Accordingly, we have defined a grid of fixation points with θ ranging from 0 and 360° by steps of 22.5°, ψ ranging from 0 to 75° by steps of 1°, and υ equal to $11°, 6°, 4°$ and 0°. For each fixation point it is possible to calculate the values of $T_l(\upsilon, \theta, \psi)$, $T_r(\upsilon, \theta, \psi)$, $\varepsilon_l(\upsilon, \theta, \psi)$ and $\varepsilon_r(\upsilon, \theta, \psi)$, i.e., the Helmholtz torsions and the rotation eccentricities for both eyes, measured for a fixation in the direction (θ, ψ) and at a vergence angle υ. Since, for both eyes, the torsions are equal and the eccentricity are symmetric we can remove the distinction between left and right in the following. To estimate which is the loss in term of motor efficiency we have defined an error function $\Delta\varepsilon_\alpha^\beta(\upsilon, \theta, \psi)$:

$$\Delta\varepsilon_\alpha^\beta(\upsilon, \theta, \psi) = \varepsilon_\alpha^\beta(\upsilon, \theta, \psi) - \varepsilon_0(\upsilon, \theta, \psi) \tag{55}$$

as the difference between the eccentricity ε_α^β, calculated for certain values of the constants α and β of the functional, and the minimal eccentricity ε_0, obtained for $\alpha = 0$. An average index of the degree of the error can be obtained by integrating $\Delta\varepsilon_\alpha^\beta(\upsilon, \theta, \psi)$ over a region of visual direction (θ, ψ). Exploiting the same approach used by [14] we have decided to integrate the absolute value of the error:

$$|\Delta\epsilon_\alpha^\beta(\upsilon)|_0^\varepsilon = \frac{1}{A(\varepsilon)} \int_0^{2\pi} \int_0^\varepsilon \sin(\psi) |\Delta\varepsilon_\alpha^\beta(\upsilon, \theta, \psi)| d\psi d\theta \tag{56}$$

where $\sin(\psi)$ is the scalar Jacobian, and the normalization term $A(\varepsilon)$ is the area of the spherical cap, over which the integration is performed:

$$A(\varepsilon) = 2\pi(1 - \cos\varepsilon) \, . \tag{57}$$

For what concerns the torsion, since the minimum value is 0, we integrated directly the absolute value of $T_\alpha^\beta(\upsilon, \theta, \psi)$:

$$|\tau_\alpha^\beta(\upsilon)|_0^\varepsilon = \frac{1}{A(\varepsilon)} \int_0^{2\pi} \int_0^\varepsilon \sin(\psi) |T_\alpha^\beta(\upsilon, \theta, \psi)| d\psi d\theta \, . \tag{58}$$

The integrals were evaluated for different values of the polar angle $\psi = 15°, 30°, 45°, 6°, 75°$, and for vergence distances equal to $\upsilon = 11°, 6°, 4°, 0°$. Note that the used values of vergence correspond to the typical range of the human peripersonal space (30-85 cm, for an interocular distance of 6 cm) for close object inspection and manipulation. The results obtained for torsions are shown in Table 1. The last row refers to the torsions predicted by the Listing's Law when the system fixates at infinity ($\upsilon = 0$), and can be directly compared to the ones derived in [14] and reported below the table in italics. As the vergence increases, the measured averaged torsions predicted by our generalized visuo-motor constraint diminish up to about 30%. It is worth noting that, on the contrary, for **LL** and **L2** the amount of torsion remains basically unaffected by vergence (values not shown, see the insets in Figure 8). The results obtained for the eccentricity are shown in Table 2, where a column relative to the Mean Absolute Percentage Error (MAPE) is added. The variation of eccentricity is extremely small for all the situations considered. The values in italics correspond to the average values of $|\Delta\epsilon_\alpha^\beta(\upsilon)|_0^\varepsilon$, over the different vergences, for **LLx**. For a direct comparison, the average values obtained for

L2 and a tilt-pan system are also reported. Considering the order of magnitude of the values, we can conclude that the motor efficiency, at least in terms of the minimal rotation eccentricity, is a negligible issue to explain the binocular eye movements predicted by **L2**.

LLx ($\beta = 10$)

$\lvert\tau^{10}_{.95}(\upsilon)\rvert^{\varepsilon}_{0}$	$\varepsilon = 15°$	$\varepsilon = 30°$	$\varepsilon = 45°$	$\varepsilon = 60°$	$\varepsilon = 75°$	
$\nu = 11°$	0.0698°	0.2858°	0.7813°	1.7941°	3.6709°	
$\nu = 6°$	0.1408°	0.6113°	1.7136°	3.5863°	6.4975°	
$\nu = 4°$	0.2530°	1.0057°	2.4480°	4.5405°	7.8998°	
$\nu = 0°$	0.3390°	1.2951°	2.9390°	5.4035°	8.9134°	
	0.314°	*1.280°*	*2.978°*	*5.596°*	*9.630°*	cf.[14]

Table 1: Absolute values of the torsion magnitude $\tau^{\beta}_{\alpha}(\theta, \psi)$ averaged over increasingly large regions of the viewing sphere and for different values of the distance vergence υ. The optimization parameters used are $\alpha = 0.95$, $\beta = 10$. The row in italics below the table shows the values obtained by [14], which can be directly compared to those we obtained in the null-vergence case ($\upsilon = 0$).

LLx ($\beta = 10$)

$\lvert\Delta\epsilon^{10}_{.95}(\upsilon)\rvert^{\varepsilon}_{0}$	$\varepsilon = 15°$	$\varepsilon = 30°$	$\varepsilon = 45°$	$\varepsilon = 60°$	$\varepsilon = 75°$	MAPE
$\nu = 11°$	0.0087°	0.0426°	0.1178°	0.2349°	0.3836°	0.4684%
$\nu = 6°$	0.0040°	0.0188°	0.0392°	0.0631°	0.0888°	0.1184%
$\nu = 4°$	0.0012°	0.0046°	0.0082°	0.0175°	0.0190°	0.0267%
	0.0046°	*0.0220°*	*0.0550°*	*0.1051°*	*0.1638°*	**LLx**
	0.0029°	*0.0058°*	*0.0087°*	*0.0117°*	*0.0149°*	**L2**
	0.0092°	*0.0610°*	*0.2046°*	*0.5123°*	*1.1100°*	**Tilt-Pan**

Table 2: Absolute values of the eccentricity error $\Delta\epsilon^{\beta}_{\alpha}(\theta, \psi)$ averaged over increasingly large regions of the viewing sphere and for different values of the distance vergence υ. The optimization parameters used are $\alpha = 0.95$, $\beta = 10$. In the last column the Mean Absolute Percentage Error is reported. The three rows in italics below the table show the corresponding averages (over the different vergences) for the **LLx**, together with those obtained for **L2** and a tilt-pan system.

9 Binocular disparity priors for a fixating observer

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. In this section, we describe how it is possible to exploit such oculomotor information in the specification of the architectural parameters of the distributed representation of disparity information developed in Workpart WP2. The predictable components of disparity 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 [20] [13], mapping the oculomotor constraints into the neural population coding and decoding strategies is still an open problem. As a starting point toward this goal, we have analyzed the influence of changes in the fixation position and of the 3D structure of the environment on the distribution of the disparity with respect to a reference scene acquired by a laser scanner.

9.1 Natural scene disparity patterns

Data acquisition - For the simulations shown in the following, we first captured 3D data from a real-world scene by using a 3D laser scanner (Konica Minolta Vivid 910, see Appendix A), with the optimal 3D measurement operating range from 0.6 m to 1.2 m, which is appropriate for analysing the disparity information experienced by an active observer in his/her peripersonal space. The system allows also capturing the color textures at a resolution of 640×480 pixels. Each scan cointained up to $307,200$ points within a variable field of view, which was adjusted with respect to the size of the object to be scanned. For this pilot experiment (the work will be further extended) we considered a cluttered desk with a collection of real-world objects (see Appendix B for image snapshots). The single objects as well as the whole scene were scanned, registered and merged together to obtain a full model of more than 13,000,000 of points. Off-line registrations of data guarantee an accuracy of about 0.1 mm. A full 360-degree view of the scene is acquired to minimize the occlusion problems that occur when one simulates changes in the vantage point of the virtual observer.

Virtual reality fixations - A Virtual Reality simulator has been implemented in C++, using OpenGL libraries and the Coin3D toolkit (www.coin3d.org). To obtain a stereoscopic visualization of the scene useful to mimic an active stereo vision system we have modified the SoCamera node of the Coin3D toolkit. In this way, we have obtained a fast tool, capable of handling the commonly used 3D modelling formats (e.g., VRML and OpenInventor) and the data acquired by the 3D laser scanner. The tool allows us to access the buffers used for the 3D rendering of the scenes: the 3D data and the textures were loaded in the virtual simulator, then the left and right projections, the horizontal and the vertical ground truth disparity maps, were obtained, *for each possible fixation point*. The developed tool is currently being used to create a database of real-world range data, and stereo image pairs for a variety of fixations. More details on the the simulator are reported in Appendix B.

Statistical analysis - For a given eye posture we computed the distribution of the horizontal and vertical disparities for all the objects whose images fall within an angle of

$\pm 22.5^(\circ)$ in both retinas. The other parameters used were: a resolution of 601×601 pixels, a focal length of 10 mm, and an interocular distance of 6 cm. We repeated the calculation for 100 different vantage points, corresponding to different positions and orientations of the cyclopean visual axis, and for a set of fixation points. The fixation points varied in the range of $0^\circ \div 360^\circ$ for the azimuth angle, and in the range of $0^\circ \div 31.82^\circ$ for the polar angle. More precisely, the fixation points were obtained by backprojecting a 11×11 grid of equally spaced points of the cyclopean retina on the closest visible surface of the scene. Under the same experimental conditions, the disparity patterns were calculated for two different eye movement paradigms (Listing and Tilt-Pan). Figures 14 and 15 demonstrate that large vertical disparities can occur in the peripheral field of view, especially for tertiary eye positions.

The mean vector disparity patterns, together with their standard ellipses (measuring the joint dispersion of the bivariate distribution) are shown in Figures 16 and 17, for fixations in the central position, for fixations in a tertiary position, and for the average over all the fixations we considered. It is worth noting that, as expected, the mean vector disparities are smaller for a Tilt-Pan system, though the statistical analysis revealed that their standard ellipses are smaller for Listing.

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

From the analysis of the simulation data, it is worth noting that the mean value of the disparities systematically change 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:

$$\vec{\delta} = \vec{\delta_s} + \vec{\delta_e}. \tag{59}$$

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 [7]. 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:

- 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 global mean pattern computed and to measure the residual 2D disparity ($\vec{\delta_s}$). It is worth noting that, by embedding the mean values computed with respect to the reference situation, the mean values of the disparities to be recovered become smaller and therefore more detectable by phase-shift mechanisms (see Fig. 18). Figures 19 and 20 show the estimation

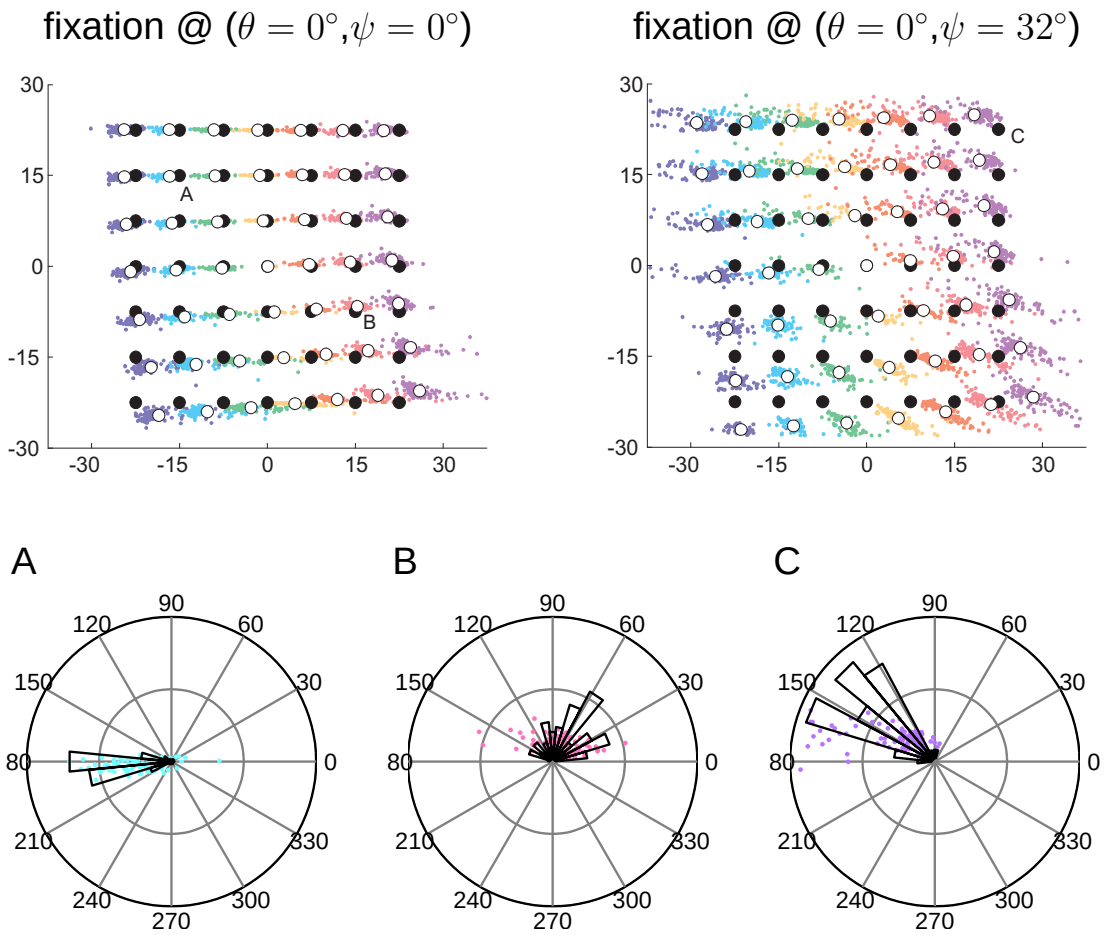


Figure 14: (Top) Patterns of binocular disparities computed for the primary position (left) and a tertiary position (right). The clouds of dots show the binocular disparities for the 100 analyzed scenes, and for a selection of 7×7 retinal reference positions (black circles) on the left retina. The white circles represent the corresponding vector mean disparities on the left retina. (Bottom) For the selected retinal locations (A,B,C), the circular histograms show the primarily anisotropic character of the disparity, by measuring the number of hits within a given orientation range. Note that the scales of the spokes on the histograms were arbitrarily adjusted with respect to the disparity values (colored points) for the sake of representation.

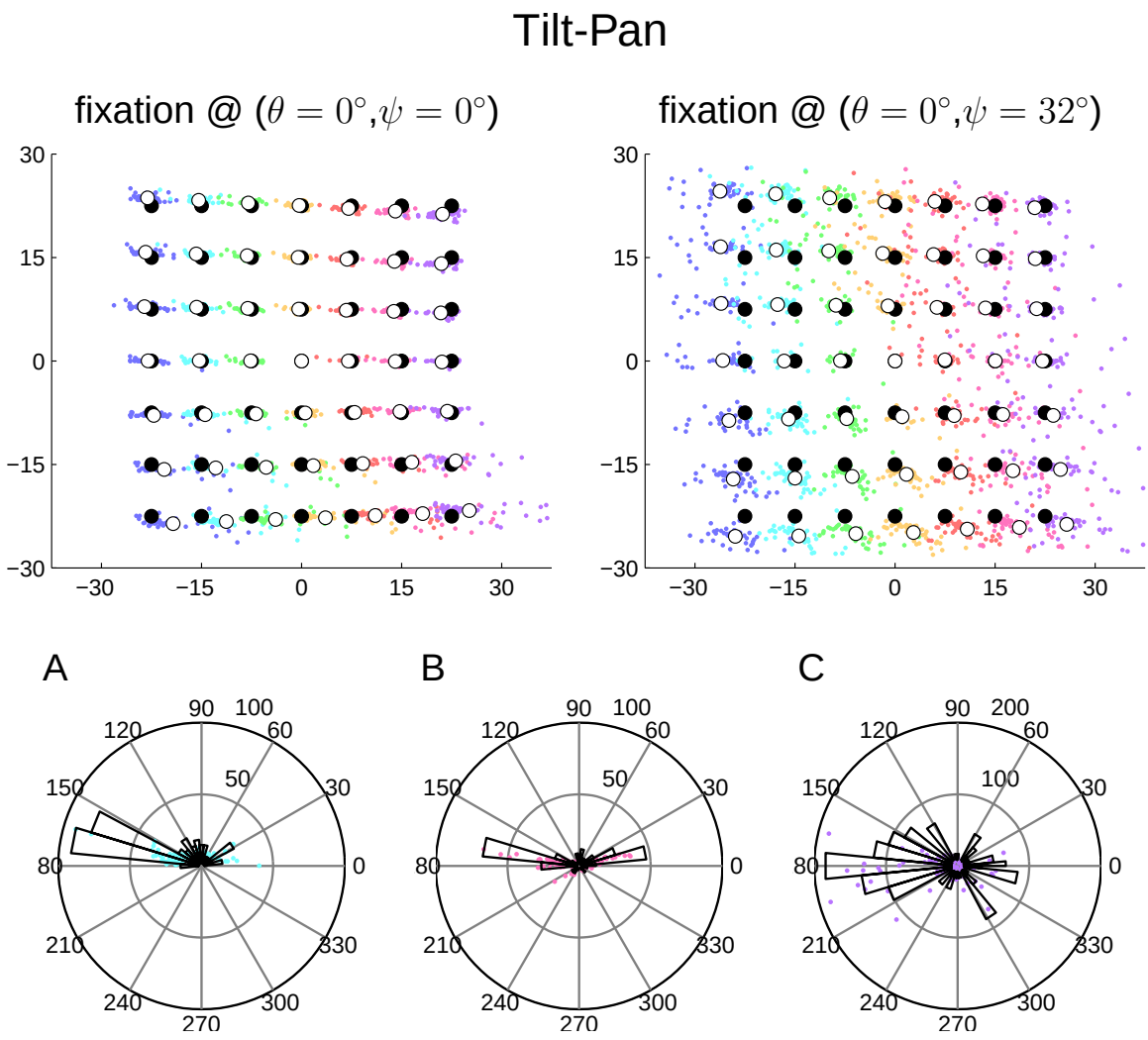


Figure 15: Same as for Figure 14 but for a Tilt-Pan system.

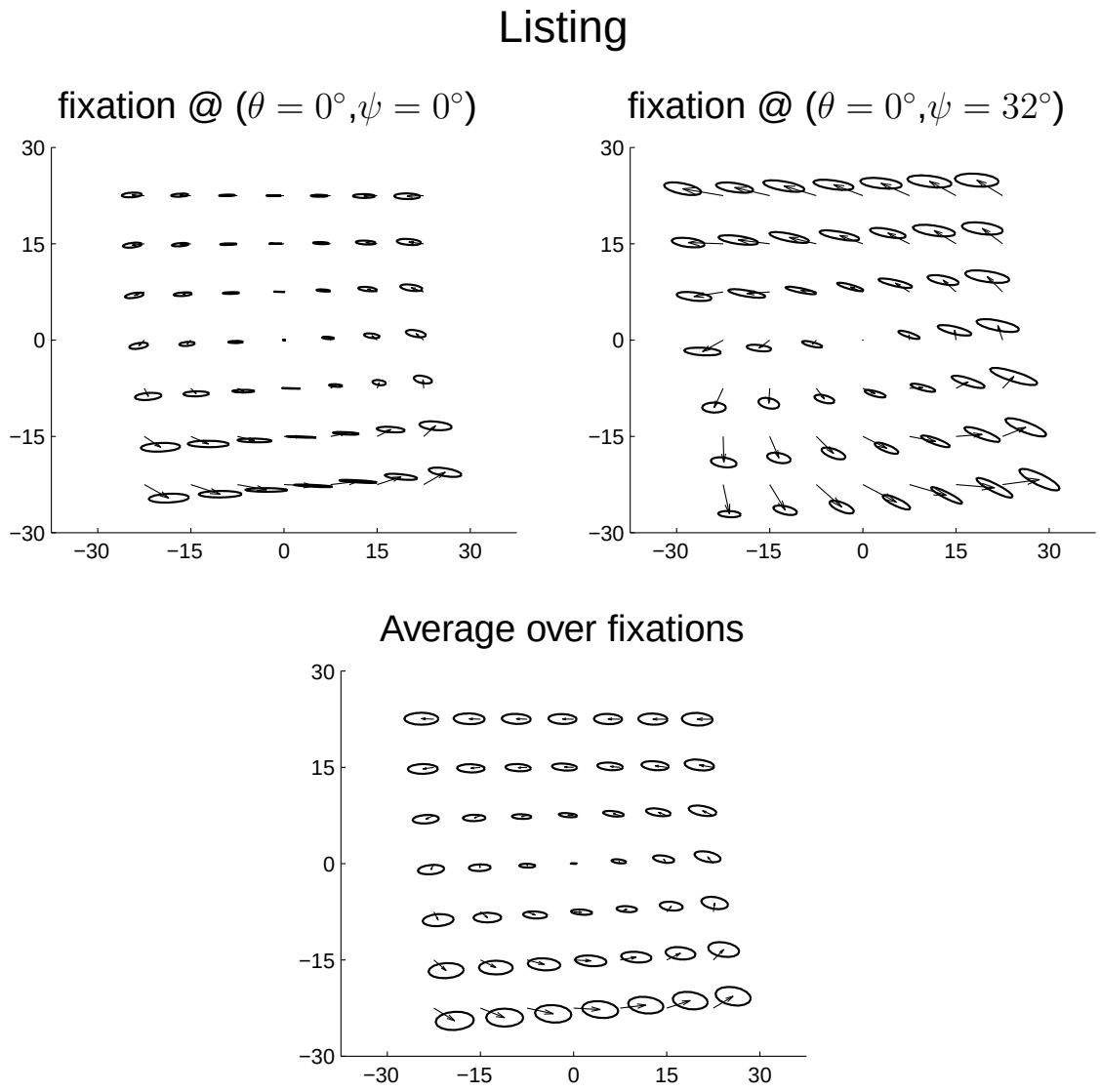


Figure 16: Mean vector disparity patterns and standard ellipses for two different fixations, and for the average over all the fixations.

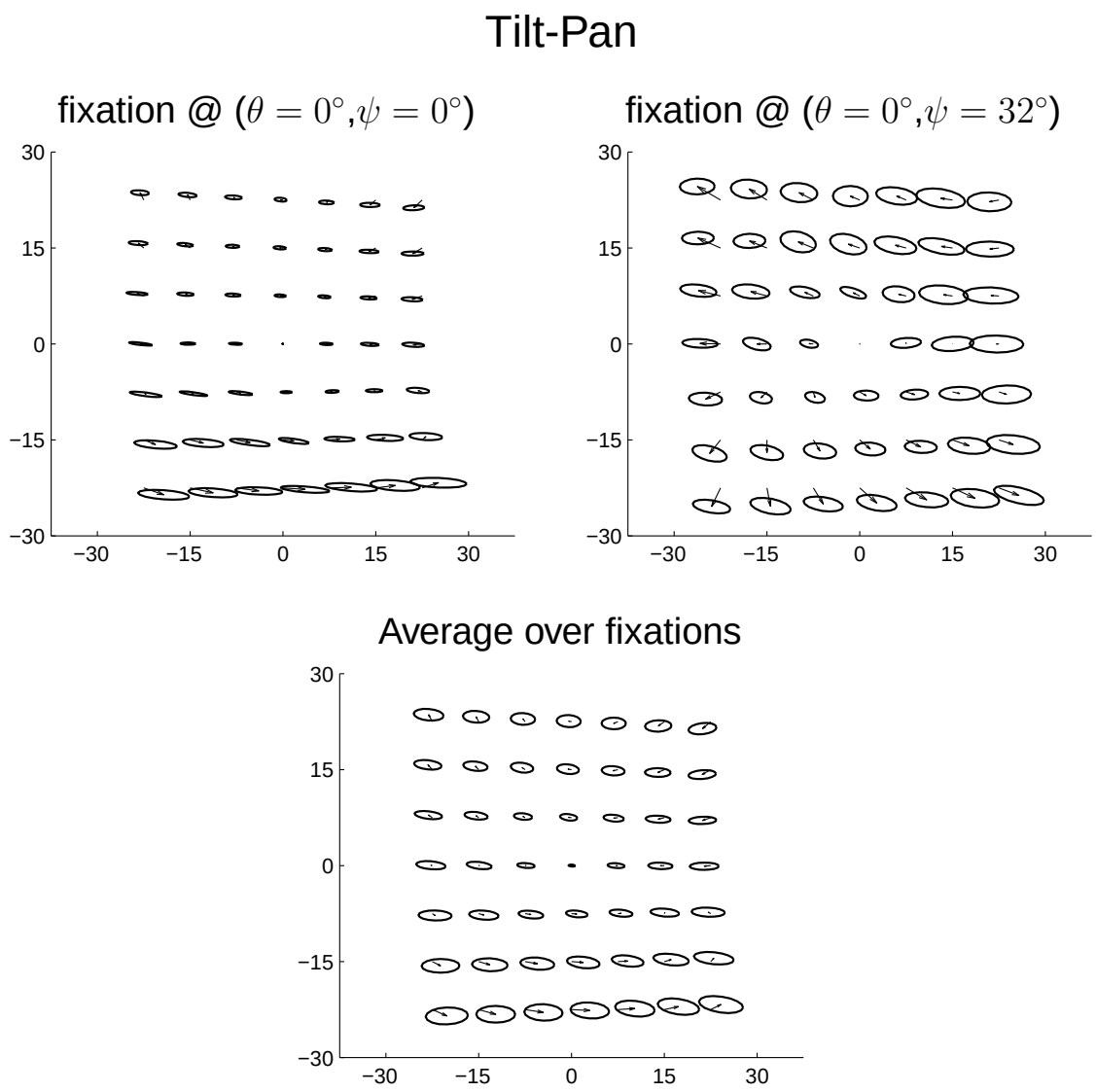


Figure 17: Same as for Figure 16 but for a Tilt-Pan system.

of the 2D disparity obtained from the cortical architecture 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, 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.

10 Conclusions

The functional implications of Listing's Law and its binocular extension are investigated on the basis of a new characterization of eye movements that depends on the coordinates of the fixation point only, but not on the rotation system adopted. On this ground, we carried out a mathematical analysis to derive the 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 standard binocular robot-heads) and to reduce the eccentricity of the rotation. Our approach confirms the experimental evidences presented in the literature for large and small vergence values, and proposes itself as a general model, forming a bridge between these two extreme cases, even for non-null version conditions. Finally, we have derived and compared the average spatial structure of the disparity fields for a series of binocular fixations in a natural environment. To the best of our knowledge, this is the first time the disparity patterns resulting from a close-range fixating observer in a natural environment have been collected and analyzed. A similar analysis was conducted by Hansard and Horaud in [12]. Though, in their work the authors warped a stereo image pair into a number of fixating views, instead of using a real 3D data set, thus being compelled to adopt several tricks to overcome the limitations in finding dense binocular correspondences in the warped stereo images. In our future work, we plan to extend the statistical analysis with additional data sets, to further characterize the variability of disparity information with respect to eye movements, and to relate such a variability with the variation of the tuning properties of disparity selective cells in (early) cortical areas.

Appendix A

Data sheets of the Laser Scanner used for 3D data collection.

Appendix B

Manuscript to be published as a chapter in the book "Virtual Reality", INTECH December 2010.

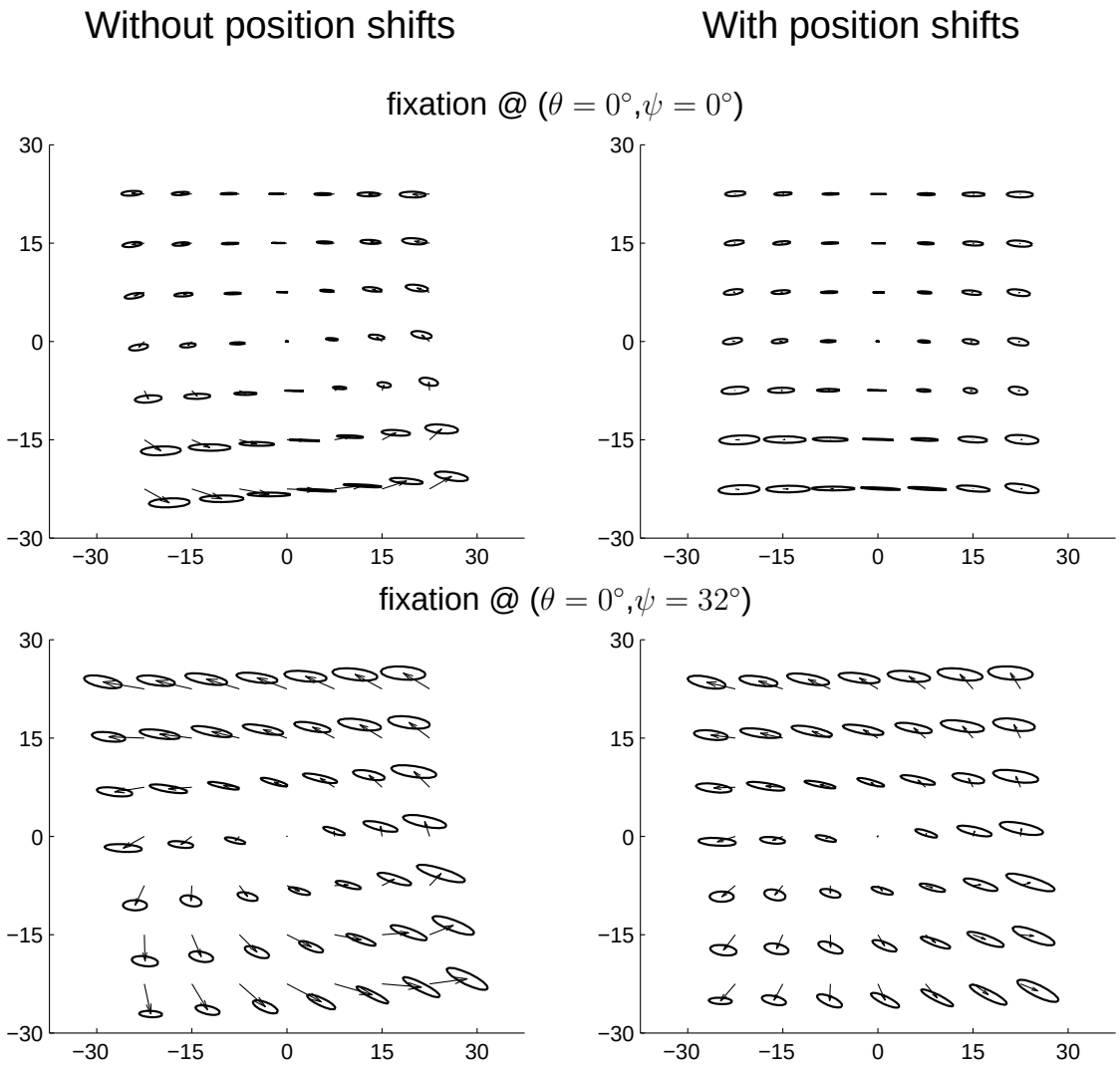


Figure 18: Mean vector disparity patterns and standard ellipses resulting after compensation with position shifts of the average disparity pattern due to the epipolar geometry (Listing). The patterns obtained without considering the position shifts are the same depicted in Fig. 16 and are reproduced here for the sake of comparison.

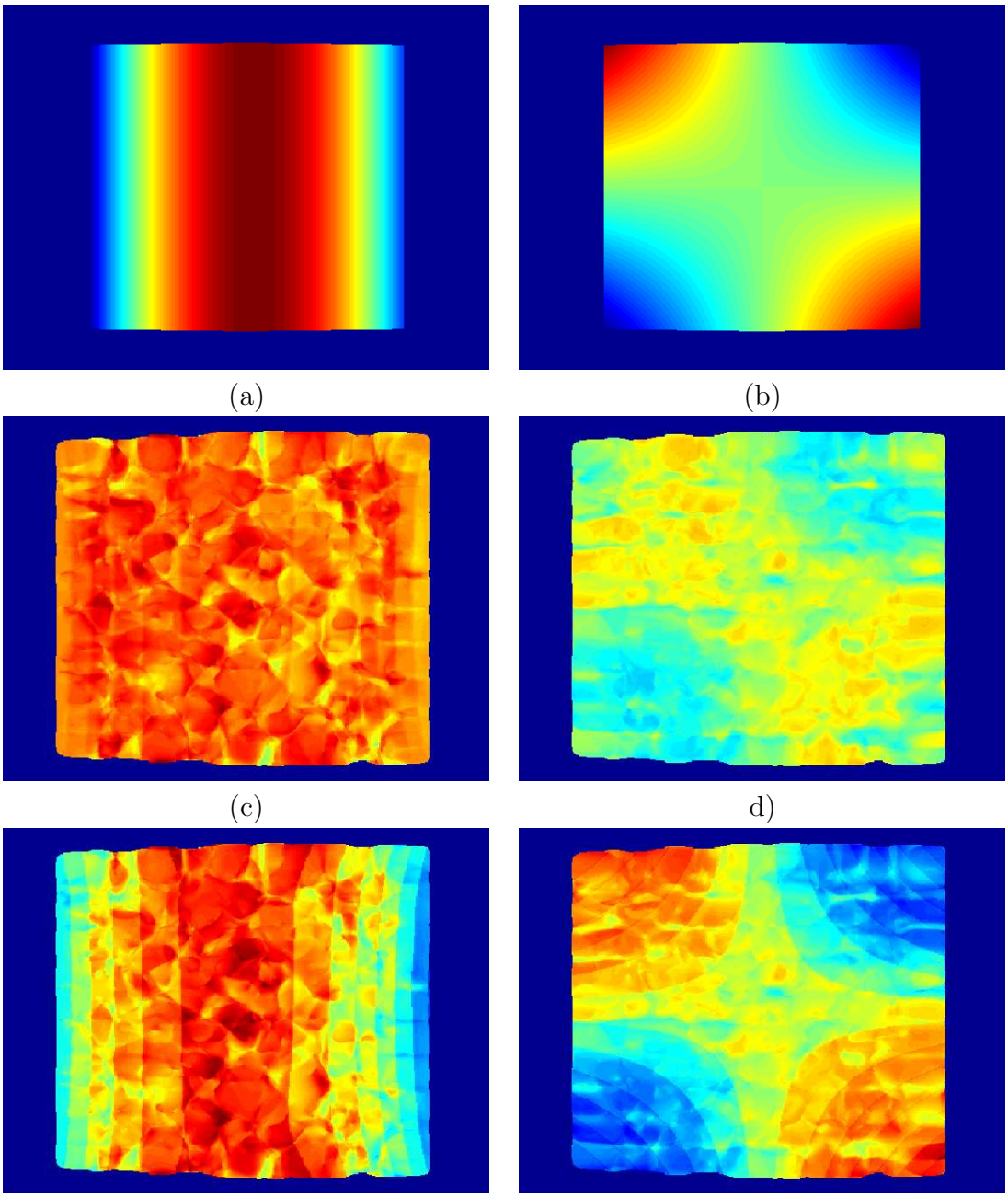


Figure 19: Disparity estimation by embedding fixation constraints into the binocular energy model for a stereo pair representing a fronto-parallel plane. (a)-(b) Ground truth horizontal and vertical disparity maps. (c)-(d) Estimation of the disparity by using the distributed architecture without embedding 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 65 cm 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.

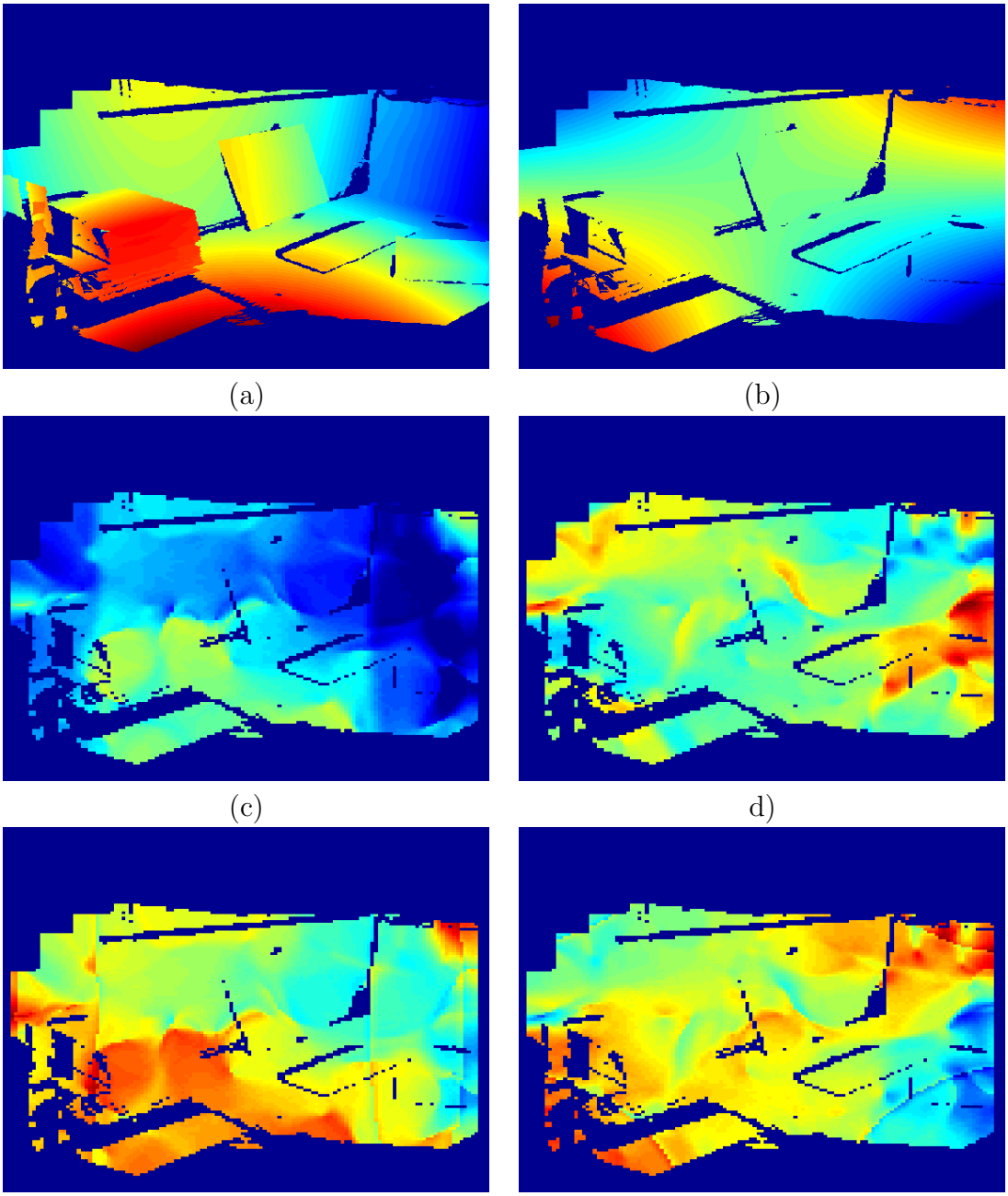


Figure 20: Disparity estimation by embedding fixation constraints into the binocular energy model for a stereo pair representing an indoor scenario acquired by using a laser scanner. (a)-(b) Ground truth horizontal and vertical disparity maps. (c)-(d) Estimation of the disparity by using the distributed 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.

References

[1] M.J. Allen. The dependence of cyclophoria on convergence, elevation and the system of axes. *American J. Optom.*, 31:297–307, 1954.

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

[3] G. Cannata and M. Maggiali. Models for the design of bioinspired robot eyes. *IEEE Trans. Robotics*, 24:27–44, 2008.

[4] J.D. Crawford. Visuomotor codes for three-dimensional saccades. In *Computational and Psychophysical Mechanism of Visual Coding*, pages 74–102. Cambridge University Press, New York, NY, USA, 1997.

[5] 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.

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

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

[8] Chun Him Fung and B.E. Shi. A biomimetic active stereo system with torsional control and saccade-like speed. pages 389 –392, 2006.

[9] A. Glennerster, S.P. McKee, and M.D. Birch. Evidence for surface-based processing of binocular disparity. *Current Biology*, 12:825 – 828, 2002.

[10] H. Goldstein, C. Poole, and J. Safko. *Classical Mechanics*. Addison Wesley, Cambridge, 2002.

[11] C. Gosselin, E. St-Pierre, and M. Gagné. On the development of the Agile Eye. *IEEE Robotics and Automation Society Magazine*, 3:29–37, 1996.

[12] M. Hansard and R. Horaud. Patterns of binocular disparity for a fixating observer. In *BVAI 2007, LNCS 4729*, pages 308–317. Springer-Verlag, Berlin, Heidelberg, 2007.

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

[14] M. Hansard and R. Horaud. Cyclorotation models for eyes and cameras. *IEEE Transactions on System, Man, and Cybernetics–Part B: Cybernetics*, 40:151–161, 2010.

[15] K. Hepp. Theoretical explanations of Listing's law and their implication for binocular vision. *Vis. Res.*, 35:3237–3241, 1995.

[16] K. Hepp. The eye of a mathematical physicist. *J. Stat. Phys.*, 134:1033–1057, 2009.

[17] E. Hering. *The theory of binocular vision.* New York: Plenum, 1868.

[18] I.Th.C. Hooge and A.V. van den Berg. Visually evoked cyclovergence and extended Listing's Law. *J. Neurophysiol.*, 83:2757–2775, 2000.

[19] 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.

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

[21] M.R.M. Jenkin, E.E. Milios, and J.K. Tsotsos. Cyclotorsion and the TRISH active stereo head. In *Int. Workshop on Stereoscopic and Three Dimensional Imaging*, pages 218–223, Santorini, Greece, September 6-8, 1995.

[22] M.R.M. Jenkin and J.K. Tsotsos. Active stereo vision and cyclotorsion. In *Proc. IEEE CVPR*, pages 806–811, 1994.

[23] J.S. Maxwell and C.M. Schor. The coordination of binocular eye movements: vertical and torsional alignment. *Vis. Res.*, 46:3537–3548, 2006.

[24] A.W.H. Minken, C.C.A.M. Gielen, and J.A.M. Van Gisbergen. An alternative three-dimensional interpretation of Hering's Equal-Innervation Law for version and vergence eye movements. *Vis. Res.*, 35:93–102, 1995.

[25] 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*, 101:331–345, 1994.

[26] D. Mok, A. Ro, W. Cadera, J.D. Crawford, and T. Vilis. Rotation of listing's plane during vergence. *Vis. Res.*, 32:2055–2064, 1992.

[27] Y. Petrov and A. Glennerster. Disparity with respect to a local reference plane as a dominant cue for stereoscopic depth relief. *Vision Research*, 46:4321 – 4332, 2006.

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

[29] J. Porrill, P.A. Warren, and P. Dean. A simple control law generates Listing's positions in a detailed model of the extraocular muscle system. *Vis. Res.*, 40:3743–3758, 2000.

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

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

[32] L.J. Van Rijn and A.V. Van den Berg. Binocular eye orientation during fixations: Listing's law extended to include eye vergence. *Vis. Res.*, 33:691–708, 1993.

[33] C.M. Schor, J.S. Maxwell, J. McCandless, and E. Graf. Adaptive control of vergence in humans. *Ann. N.Y. Acad. Sci.*, 956:297–305, 2002.

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

[35] 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.

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

[37] 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.

[38] D. Tweed. Visual-motor optimization in binocular control. *Vis. Res.*, 37:1939–1951, 1997.

[39] D. Tweed and T. Vilis. Geometric relations of eye position and velocity vectors during saccades. *Vis. Res.*, 30:111–127, 1990.

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

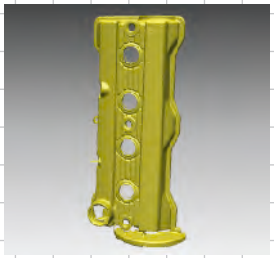
[41] A. Wong. Listing's law: clinical significance and implications for neural control. *Surv. Ophthalmol.*, 49:563–575, 2004.

KONICA MINOLTA

VIVID 910
NON-CONTACT 3-D DIGITIZER

3-D Digitizing - Breakthrough in Process Innovation

Konica Minolta's VIVID 910 the ideal 3D capture device for industrial applications in product design and manufacturing inspection.



VIVID 910

PET: Polygon editing software,
EAT: Easy alignment target-based registration

Compatible with all major 3D software for Modeling and CAD, CAM and CAT

The essentials of imaging

The Konica Minolta VIVID 910, Innovation in 3D Digitizing for both Product Design and Manufacturing.

The VIVID 910 is a non-contact 3-D digitizer, offering fast, precise capture of 3-D shapes. VIVID is ideal for applications in both product design and production. The designers find VIVID invaluable for "reverse engineering" or creating CAD data from physical models and design mock-ups. Production personnel use VIVID for Inspection and computer-aided dimensional testing (CAT). What's more, VIVID improves concurrent engineering by inexpensively making 3D data available throughout the enterprise.

Typical Applications of the VIVID 910

The VIVID 910 is employed in a variety of industries for the following applications :

Reverse Engineering (RE)/Rapid Prototyping (RP)

- Generation of design CAD data from physical modelsand data for detecting interference among mechanical parts from mock-ups.
- Generation of data of parts for which 3-D CAD data is unavailable.
- Verification and comparison of competitor's products with in-house products. Database creation.
- Generation and refinement of designs using actual models created through RP.
- Capture of data for finite element analysis.

Inspection (CAT)/CAE

- Alignment verification and dimensional inspection of components such as:

  metal castings & forgings,
  tooling dies and molds,
  plastic parts (pressure formed, rotational molds, injection),
  sheet metal stampings,
  wood products,
  composites and foam products.

Other Applications

- Food production
- Cultural Antiquities cataloging and publishing
- Dental & orthodontic appliances
- Cosmetic & Maxillofacial surgery
- Machine Vision

The Digitizer with camera like simplicity and refinement, Designed to excel in your Industrial Application VIVID 910

Your assurance of highly reliable data

The VIVID 910 offers the highest level of accuracy and reliability among non-contact digitizers. It excels at accurate and high-speed measurement of a variety of objects. In fact, as evidence of its accuracy, we offer a test report * (by special order) that measures its performance against artifacts traceable to national standards organizations. Konica Minolta is famous for our highly-reliable, measuring instruments that conform to ISO 9000 standards.

* VIVID 910 Certification of Performance is available by special order. KM offers a certification quantifying the VIVID's accuracy when measuring traceable artifacts. This service is of benefit to those who are implementing the ISO 9000 series of standards for quality assurance systems.

Measures objects of every size.

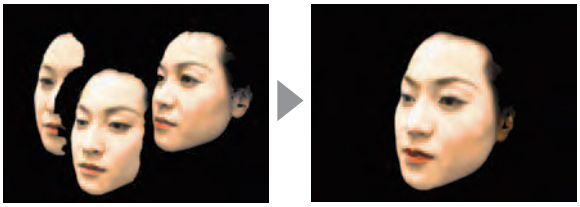
The VIVID 910 is provided with three interchangeable lenses that can accommodate measurement objects of various sizes and distances from the lens. A single scan is capable of capturing an angular field of view of approximately 10 square centimeters to 1 square meter.

Automatic configuration of detailed settings

The VIVID 910 incorporates the same automatic focus technology used in modern cameras. The optimal measurement distance is automatically detected through both passive and active AF (autofocus). In addition, the optimal laser intensity is obtained automatically through AE technology. The result is highly reliable measurements.

Provides 24-bit color images for outstanding texture mapping.

The CCD and RGB filter acquire rich, 24-bit full-color images. Since the acquired color images are on the same optical axis as the 3-D data, they can be used to create stunning, true-color models.



High-speed scanning capability

VIVID 910 is capable of capturing an object's shape and color in as little as 2.5 seconds. Our proprietary CCD readout technology measures up to 300,000 points at unsurpassed speed. When the subject is a moving object e.g. children, the human body and for other applications requiring higher speeds, an even faster mode is available that can complete a scan in a mere 0.3 seconds.

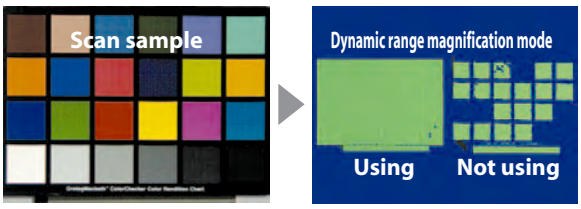
Fine Mode : 307,200 points/2.5 seconds
Fast Mode : 76,800 points/0.3 seconds

Designed to be portable and versatile

The VIVID 910 features a lightweight and compact body. It can operate without a host computer by recording data onto Compact Flash memory card. VIVID's integral LCD viewfinder can be used to set camera parameters and as a view-finder to frame the shot or review the data. As a result, the VIVID 910 offers convenience similar to that of a digital camera, so you can operate it wherever your subject may be located.

Dynamic range magnification mode

Objects with very dark to very bright regions are no longer a problem. The dynamic range magnification mode reduces the need for surface processing of objects with high-contrast surfaces (surfaces with both very light and very dark areas). This feature enables you to complete a measurement in only one operation.



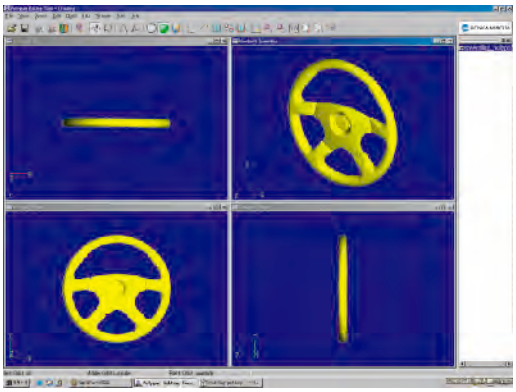
Benefit from the wide-ranging support provided by Konica Minolta, a leading maker of measuring instruments.

The VIVID 910 incorporates the services and expertise developed by Konica Minolta in the field of industrial measuring instruments such as colorimeters and measuring instruments for displays. We ensure your satisfaction by offering a wide range of optional support programs; that includes a periodical calibration service, a training by factory certified trainers and a network of consultants and systems integrators for custom installations.

Polygon Editing Tool (standard accessory)

Edit scanned data with complete freedom.

Our proprietary Polygon Editing Tool (PET) comes standard with the VIVID 910. PET enables you to control the VIVID 910 and easily scan, polygonize, edit, and convert the scanned data into any of several common data formats. Multiple scans can be easily registered and merged into a single watertight polygonal model. Editing functions include: fill holes; filter irregular polygons and noise; and perform smoothing. PET exports data in industry-standard formats including: DXF, STL etc. for accurate transfer to a variety of Modeling, Inspection CAD, CAM and CAT 3-D applications. In addition, a SDK (software development kit) is included to enable you to drive the VIVID 910 from your own software application.



Features

Data read	Proprietary formats: CAM, CDM, VVD, SCN General format: STL
Data conversion	Converts from proprietary format to various common formats. Polygon: DXF, Wavefront, Softimage, VRML 2.0, STL, MGF Point group: ASCII
Functions	Automatic data registration, data merging, smoothing, sub-sampling and curvature-based decimation, polygon checking, texture blending, and other functions
Editing	Rotation, transfer, elimination of point groups, and hole filling with data interpolation
Remote camera operation	Image capture, reference depth of field setting, dynamic range magnification mode, laser power setting, readout of camera data
Display	Wireframe, shading, texture mapping

Data Registration and Merging

Computer Requirements

PC/AT-compatible workstation capable of running, Windows® 2000 or Windows® XP

Operating system	Windows® 2000 Professional SP4, Windows® XP Professional SP2 (x64 Edition not supported)
CPU	Pentium 4 or higher
Memory	1024 MB minimum (2048 MB recommended)
Display	1024 x 768 minimum 1280 x 1024 or higher is recommended when using Easy Align Tool for automatic marker registration.
Graphics	OpenGL-compatible video card(Contact us for details.)
SCSI	Adaptec SCSI interface card Note: Contact us for details of tested models.
Drive	CD-ROM drive

Software

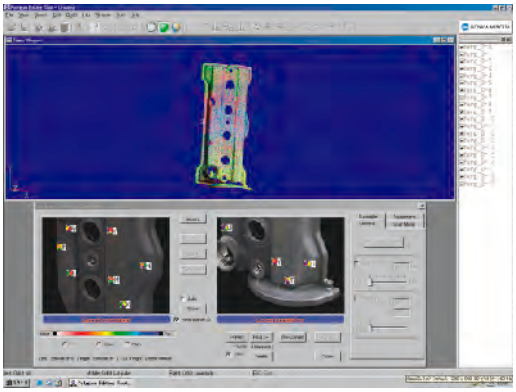
Easy Align Tool automatic target-based registration software (optional)

The automatic data registration tool that's simple and user-friendly. Reduces registration time by 66%!

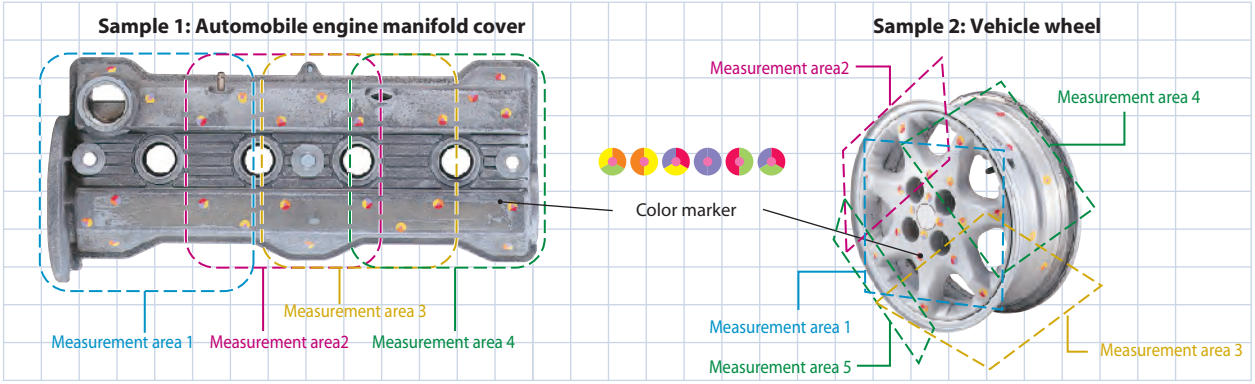
Automatic Alignment:

Alignment of individual scans has been a challenging task for some. But no longer, Easy Alignment Tool has changed all that. Simply place one of Konica Minolta's proprietary color markers on or near the object to be measured. Now, scan the same object from a different perspective with enough overlap so that at least three of the same markers are included in the second scan.

The new data is automatically registered (with coordinates aligned) with the previously scan. You can see what has been scanned and what has been missed, greatly reducing the time required for capture and post processing. EAT's takes the work out of the measurement of objects (Sample 1) that cannot be placed on a Rotary turn table, or objects (Sample 2) that require multiple measurements from varying points of view.



• This software is an optional add-on to the Polygon Editing Tool. Requires Polygon Editing Tool ver. 1.2 or higher.



Process Innovation with the VIVID 910

Schematic Design

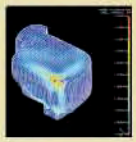
Reverse Engineering

1 Perspective drawings → 2 Mock-ups → 3 3-D measurement → 4 Schematic CAD data (Surface models) → 5 Revision of mock-ups → Reproduction for Rapid Prototyping (replicas)

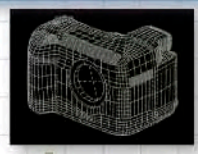
Package examination

Structural Design

1 Structural design → 2 Metal mold requirements → 3 CAE analysis



3-D CAD data for mass production



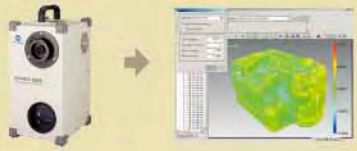
Products

Approval of Metal Molds

1 Design of metal molds → 2 Test shots → 3 3-D measurement → 4 Comparison with CAD data and inspections (CAT) → 5 Approval of metal molds

Inspection of Mass-produced Parts

1 3-D measurement (Sampling or 100%) → 2 Comparison with CAD data and inspections (CAT)



Applications and Data Flow for the VIVID 910

Measuring with the VIVID 910

Further simplify operations with optional accessories – the Rotary Stage Set, Benchtop Frame Set, and Easy Align Tool (automatic registration software).

Anthropometrical measurement

CG

Cataloging of cultural artifacts

CT data → Various Reverse Modeler

Various CG software (optional)

Editing of Measurement Data

Polygon Editing Tool (standard accessory)

Rapid Prototyping

Reverse Engineering

Various Reverse Modeler

Various Reverse Modeler

• CAT
• Inspection

IGES

Importing of CAD data

Various 3-D CAD/CAE → Various rapid prototyping systems

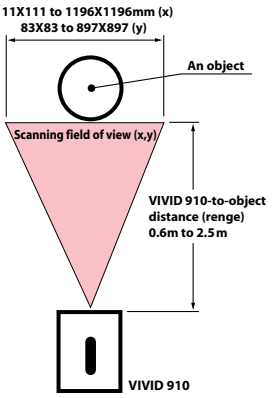
Report output

Various 3-D CAD

Theory of Operation

Basic Principle

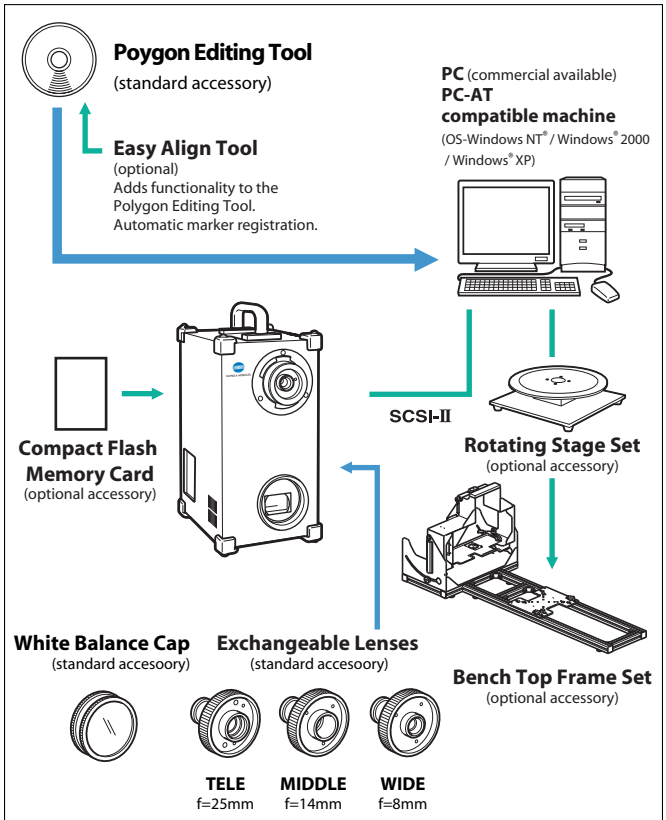
The VIVID 910 uses LASER triangulation. The object is scanned by a plane of laser light coming from the VIVID's source aperture. The plane of light is swept across the field of view by a mirror, rotated by a precise galvanometer. This LASER light is reflected from the surface of the scanned object. Each scan line is observed by a single frame, captured by the CCD camera. The contour of the surface is derived from the shape of the image of each reflected scan line. The entire area is captured in 2.5 seconds (0.3 seconds in FAST mode), and the surface shape is converted to a lattice of over 300,000 vertices (connected points). VIVID gives you more than a point cloud; a polygonal-mesh is created with all connectivity information retained, thereby eliminating geometric ambiguities and improving detail capture. A brilliant (24-bit) color image is captured at the same time by the same CCD. Unlike other scanners, the VIVID has no parallax error, its "spot - on"!



High Accuracy Measurement

A high-accuracy scanner and a high-accuracy Calibration facility unit to be used for calculation of 3-D data have been developed for the VIVID 910.
The 3-D reference chart traceable to the national standards has also been established to utilize the technology and algorithm that enable higher accuracy measurement.

System Block Diagram



■ Optional accessories

- Rotating Stage Set
- Bench Top Frame Set
- Compact Flash Memory Card
- PC Card Adapter
- Tripod Set
- Easy Align Tool

Specifications

Type	Non-contact 3D digitizer VIVID 910
Measuring method	Triangulation light block method
Auto Focus method	Image surface AF (contrast method), active AF
Light-Receiving Lens (Exchangeable)	TELE: Focal distance f=25mm MIDDLE: Focal distance f=14mm WIDE: Focal distance f=8mm
Scan Range (Depth of field)	0.6 to 2.5m (2m for WIDE)
Optimal 3D measurement Range	0.6 to 1.2m
Laser class	Class 2 (IEC 60825-1), Class 1 (FDA)
Laser Scan Method	Galvanometer-driven rotating mirror
X Direction Input Range (Varies with the distance)	111 to 463mm (TELE), 198 to 823mm (MIDDLE), 359 to 1196mm (WIDE)
Y Direction Input Range (Varies with the distance)	83 to 347mm (TELE), 148 to 618mm (MIDDLE), 269 to 897mm (WIDE)
Z Direction Input Range (Varies with the distance)	40 to 500mm (TELE), 70 to 800mm (MIDDLE), 110 to 750mm (WIDE/FINE mode)
Accuracy	TELE X: ±0.22mm, Y: ±0.16mm, Z: ±0.10mm to the Z reference plane (Conditions:TELE/FINEmode,Konica Minolta's standard)
Input Time	0.3 sec (FAST mode), 2.5 sec (FINE mode), 0.5 sec (COLOR)
Transfer Time to Host Computer	Approx. 1 sec (FAST mode), 1.5 sec (FINE mode)
Ambient Lighting Condition	Office Environment, 500 lx or less
Imaging Element	3-D data:1/3-inch frame transfer CCD (340,000 pixels) Color data:3-D data is shared (color separation by rotary filter).
Number of Output Pixels	3-D data : 307,000 (for FINE mode), 76,800 (for FAST mode) Color data : 640 x 480 x 24 bits color depth
Output Format	3-D data : Konica Minolta format, & (STL, DXF, OBJ, ASCII points, VRML) (Converted to 3-D data by the Polygon Editing Software/ standard accessory) Color data : RGB 24-bit raster scan data
Recording Medium	Compact Flash memory card (128MB)
Data File Size	Total 3-D and color data capacity: 1.6MB per data (for FAST mode), 3.6MB per data (for FINE mode)
Viewfinder	5.7-inch LCD (320 x 240 pixels)*1
Output Interface	SCSI II (DMA synchronous transfer)
Power	Commercial AC power 100 to 240V (50 to 60Hz), rated current 0.6A (when 100Vac is input)
Dimensions (WxHxD)	213 x 413 x 271 mm (8-3/8 x 16-1/4 x 10-11/16 in.)
Weight	Approx.11kg (25 lbs)
Operating temperature/ humidity range*2	10 to 40°C, relative humidity 65% or less with no condensation
Storage temperature/ humidity range	-10 to 50°C, relative humidity 85% or less (at 35°C) with no condensation

*1 Contains Mercury in the backlighting of LCD used for display, Dispose According to Local, State or Federal Laws.

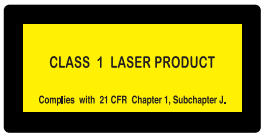
*2 Operating temperature/humidity range of products for North America : 10 to 40°C, relative humidity 50% or less (at 40°C) with no condensation.

- Specifications are subject to change without notice.
- Product names in this brochure are trademarks of their respective companies.

SAFETY PRECAUTIONS

Read all safety and operating instructions before operating the VIVID 910.

- Use only a power source of the specified rating. Improper connection may cause a fire or electric shock..
- Do not stare into the laser beam. (MAX. 30mW 690nm / CLASS 1 (FDA), CLASS 2 (IEC) LASER PRODUCT)



Certificate No : YKA 0937154
Registration Date : March 3, 1995

Certificate No : JQA-E-80027
Registration Date : March 12, 1997

KONICA MINOLTA SENSING, INC. **3-91, Daisennishimachi, Sakaiku, Sakai, Osaka 590-8551, Japan**
EMail : 3dsales@konicaminolta.jp Web : http://konicaminolta.jp/pr/se_3d

Konica Minolta Sensing Americas, Inc. 101 Williams Drive, Ramsey, New Jersey 07446, U.S.A. Phone: 888-473-2656 (in USA), 201-236-4300 (outside USA) FAX: 201-785-2480
EMail : vivid3d@se.konicaminolta.us Web : http://se.konicaminolta.us/3d

Konica Minolta (CHINA) Investment Ltd. SE Sales Division Rm.29A,K Cross Region Plaza, No.899 Lingling Rd., Shanghai, China Phone: +86-021-5489 0202 FAX: +86-021-5489 0005

Konica Minolta Sensing Singapore Pte Ltd. 10, Teban Gardens Crescent, Singapore 608923 Phone: +65 6563-5533 FAX: +65 6560-9721

Addresses and telephone/fax numbers are subject to change without notice. For the latest contact information, please refer to the KONICA MINOLTA SENSING Worldwide Offices web page (link below). **http://konicaminolta.com/about/se/contact.html**

9242-4880-12 AGMFPK⑨ Printed in Japan

Appendix B

Virtual reality to simulate visual tasks for robotic systems

Manuela Chessa, Fabio Solari, Silvio P. Sabatini
Department of Biophysical and Electronic Engineering, University of Genoa
Via all'Opera Pia 11/A - 16145 Genova,Italy
{*manuela.chessa, fabio.solari, silvio.sabatini*}*@unige.it, www.pspc.dibe.unige.it*

1. Introduction

Virtual reality (VR) can be used as a tool to analyze the interactions between the visual system of a robotic agent and the environment, with the aim of designing the algorithms to solve the visual tasks necessary to properly behave into the 3D world. The novelty of our approach lies in the use of the VR as a tool to simulate the behavior of vision systems. The visual system of a robot (e.g., an autonomous vehicle, an active vision system, or a driving assistance system) and its interplay with the environment can be modeled through the geometrical relationships between the virtual stereo cameras and the virtual 3D world. Differently from conventional applications, where VR is used for the perceptual rendering of the visual information to a human observer, in the proposed approach, a virtual world is rendered to simulate the actual projections on the cameras of a robotic system. In this way, machine vision algorithms can be quantitatively validated by using the ground truth data provided by the knowledge of both the structure of the environment and the vision system.

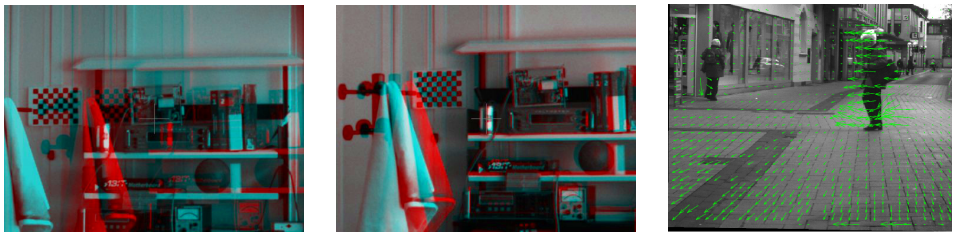
In computer vision (Trucco & Verri, 1998; Forsyth & Ponce, 2002), in particular for motion analysis and depth reconstruction, it is important to quantitatively assess the progress in the field, but too often the researchers reported only qualitative results on the performance of their algorithms due to the lack of calibrated image database. To overcome this problem, recent works in the literature describe test beds for a quantitative evaluation of the vision algorithms by providing both sequences of images and ground truth disparity and optic flow maps (Scharstein & Szeliski, 2002; Baker et al., 2007). A different approach is to generate image sequences and stereo pairs by using a database of range images collected by a laser range-finder (Yang & Purves, 2003; Liu et al., 2008).

In general, the major drawback of the calibrated data sets is the lack of interactivity: it is not possible to change the scene and the camera point of view. In order to face the limits of these approaches, several authors proposed robot simulators equipped with visual sensors and capable to act in virtual environments. Nevertheless, such software tools are capable of accurately simulating the physics of robots, rather than their visual systems. In many works, the stereo vision is intended for future developments (Jørgensen & Petersen, 2008; Awaad et al., 2008), whereas other robot simulators in the literature have a binocular vision

system (Okada et al., 2002; Ulusoy et al., 2004), but they work on stereo image pairs where parallel axis cameras are used. More recently, a commercial application (Michel, 2004) and an open source project for cognitive robotics research (Tikhanoff et al., 2008) have been developed both capable to fixate a target, nevertheless the ground truth data are not provided.

2. The visual system simulator

Figure 1a-b shows the real-world images gathered by a binocular robotic head, that is fixating a specific point in the scene through vergence movements. It is worth noting that both horizontal and vertical disparities have quite large values in the periphery, while disparities are zero in the fixation point. Analogously, if we look at the motion field generated by an agent moving in the environment (see Fig. 1c), where both still and moving objects are present the resulting optic flow is composed both by ego-motion components, due to motion of the observer, and by the independent movements of the objects in the scene.



(a) (b) (c)

Fig. 1. Binocular snapshots obtained by real-world vision systems. (a)-(b): The stereo image pairs are acquired by a binocular active vision system (`http://www.searise.eu/`) for different stereo configurations: the visual axes of the cameras are (a) kept parallel, (b) convergent for fixating an object in the scene (the small tin). The anaglyphs are obtained with the left image on the red channel and the right image on the green and blue channels. The interocular distance is 30 cm and the camera resolution is 1392×1236 pixels with a focal length of 7.3 mm. The distance between the cameras and the objects is between 4 m and 6 m. It is worth noting that both horizontal and vertical disparities are present. (c): Optic flow superimposed on a snapshot of the relative image sequence, obtained by a car, equipped with a pair of stereo cameras with parallel visual axes , moving in a complex real environment. The resolution of the cameras is 1392×1040 pixels with a focal length of 6.5 mm, and the baseline is 33 cm (`http://pspc.dibe.unige.it/drivsco/`). Different situations are represented: ego-motion (due to the motion of the car) and a translating independent movement of a pedestrian (only the left frame is shown).

The aim of the work described in this chapter is to simulate the active vision system of a robot acting and moving in an environment rather than the mechanical movements of the robot itself. In particular, we aim to precisely simulate the movements (e.g. vergence and version) of the two cameras and of the robot in order to provide the binocular views and the related ground truth data (horizontal and vertical disparities and binocular motion field). Thus, our VR tool can be used for two different purposes (see Fig. 2):

1. to obtain binocular image sequences with related ground truth, to quantitatively assess the performances of computer vision algorithms;

2. to simulate the closed loop interaction between visual perception and action of the robot.

The binocular image sequences provided by the VR engine could be processed by computer vision algorithms in order to obtain the visual features necessary to the control strategy of the robot movements. These control signals act as an input to the VR engine, thus simulating the robot movements in the virtual environment, then the updated binocular views are obtained. In the following, a detailed description of the model of a robotic visual system is presented.

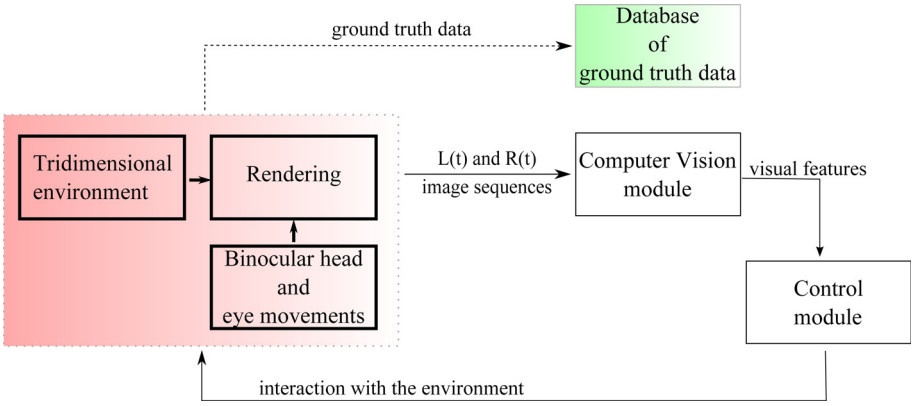


Fig. 2. The proposed active vision system simulator. Mutual interactions between a robot and the environment can be emulated to validate the visual processing modules in a closed perception-action loop and to obtain calibrated ground truth data.

2.1 Tridimensional environment

The 3D scene is described by using the VRML format. Together with its successor X3D, VRML has been accepted as an international standard for specifying vertices and edges for 3D polygons, along with the surface color, UV mapped textures, shininess and transparency. Though a large number of VRML models are available, e.g. on the web, they usually have not photo-realistic textures and they are often characterized by simple 3D structures. To overcome this problem, a dataset of 3D scenes, acquired in controlled but cluttered laboratory conditions, has been created by using a scanner laser. The results presented in Section 6 are obtained by using the dataset obtained in our laboratory.

It is worth noting that the complex 3D VRML models can be easily replaced by simple geometric figures (cubes, cones, planes) with or without textures at any time, in order to use the simulator as an agile testing platform for the development of complex computer vision algorithms.

2.2 Rendering

The scene is rendered in an on-screen OpenGL context. Moreover, the `SoOffScreenRenderer` class is used for rendering scenes in off-screen buffers and to save to disk the sequence of stereo pairs. The renderer can produce stereo images of different resolution and acquired by cameras with different field of views. In particular, one can set the following parameters :

- resolution of the cameras (the maximum possible resolution depends on the resolution of the textures and on the number of points of the 3D model);
- horizontal and vertical field of view (HFOV and VFOV, respectively);
- distance from camera position to the near clipping plane in the camera's view volume, also referred to as a viewing frustum, (`nearDistance`);
- distance from camera position to the far clipping plane in the camera's view volume (`farDistance`);
- distance from camera position to the point of focus (`focalDistance`).

2.3 Binocular head and eye movements

The visual system, presented in this Section, is able to generate the sequence of stereo image pairs of a binocular head moving in the 3D space and fixating a 3D point $(X^F, Y^F, Z^F,)$. The geometry of the system and the parameters that can be set are shown in Figure 3.

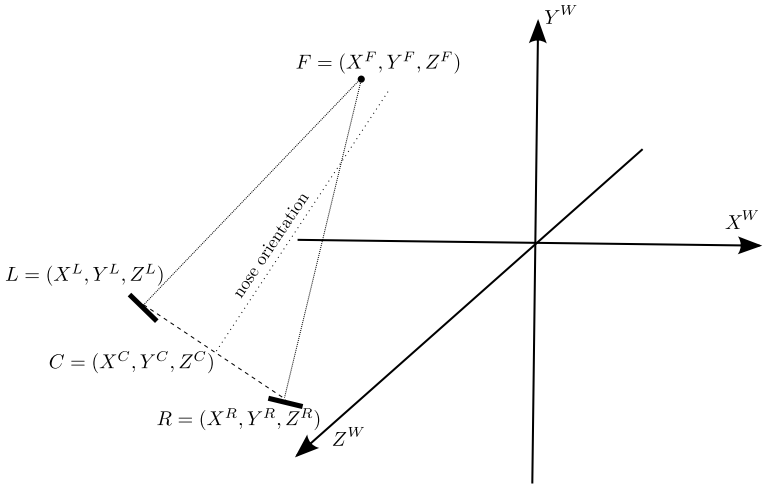


Fig. 3. Schematic representation of the geometry of the binocular active vision system.

The head is characterized by the following parameters (each expressed with respect to the world reference frame (X^W, Y^W, Z^W)):

- cyclopic position $\boldsymbol{C} = (X^C, Y^C, Z^C,)$;
- nose orientation;
- fixation point $\boldsymbol{F} = (X^F, Y^F, Z^F,)$

Once the initial position of the head is fixed, then different behaviours are possible:

- to move the eyes by keeping the head (position and orientation) fixed;
- to change the orientation of the head, thus mimicking the movements of the neck;
- to change both the orientation and the position of the head, thus generating more complex motion patterns.

These situations imply the study of different perceptual problems, from scene exploration to navigation with ego-motion. Thus, in the following (see Section 6), we will present the results obtained in different situations.
For the sake of clarity and simplicity, in the following we will consider the position $\boldsymbol{C} = (X^C, Y^C, Z^C,)$ and the orientation of the head fixed, thus only the ocular movements will be considered. In Section 3.3.1 different stereo systems will be described (e.g. pan-tilt, tilt-pan, etc.), the simulator can switch through all these different behaviours. The results presented in the following consider a situation in which the eyes can rotate around an arbitrary axis, chosen in order to obtain the minimum rotation to make the ocular axis rotate from the initial position to the target position (see Section 3.3.1).

2.4 Database of ground truth data

In the literature several database of ground truth data can be found, to quantitatively assess optic flow and disparity measures.
One of the best known and widely used is the *Yosemite sequence*, that has been used extensively for experimentation and quantitative evaluation of the performances of optical flow computation techniques, camera motion estimation, and structure from motion algorithms. The data was originally generated by Lynn Quam at SRI and David Heeger (Heeger, 1987) was the first to use it for optical flow experimentation. The sequence is generated by taking an aerial image of Yosemite valley and texture mapping it onto a depth map of the valley. A synthetic sequence is generated by flying through the valley.
Other simple, but widely used, image sequences with associated ground truth data are the *Translating tree* and the *Diverging tree* by (Fleet & Jepson, 1990). Moreover, it is possible to find the *Marbled-Block sequence*, recorded and first evaluated by (Otte & Nagel, 1995), a polyhedral scene with a moving marbled block and moving camera.
A large number of algorithms for the estimation of optic flow have been benchmarked, by using these sequences. Unfortunately, it is difficult to know how relevant these results are to real 3D imagery, with all its associated complexities (for example motion discontinuities, complex 3D surfaces, camera noise, specular highlights, shadows, atmospherics, transparency). To this aim (McCane et al., 2001) have used two methods to generate more complex sequences with ground-truth data: a ray-tracer which generates optical flow, and a Tcl/Tk tool which allows them to generate ground truth optical flow from simple (i.e. polygonal) real sequences with a little help from the user.
Nevertheless, these sequences are too simple and the needs of providing more complex situation leads to the creation of databases that include much more complex real and synthetic scenes, with non-rigid motions (Baker et al., 2007). The authors rather than collecting a single benchmark dataset (with its inherent limitations), they collect four different sets, each satisfying a different subset of desirable properties. A proper combination of these datasets could be sufficient to allow a rigorous evaluation of optical flow algorithms.
Analogously, for the estimation of binocular disparity, synthetic images have been used extensively for quantitative comparisons of stereo methods, but they are often restricted to simple geometries and textures (e.g., random-dot stereograms). Furthermore, problems arising with real cameras are seldom modeled, e.g., aliasing, slight misalignment, noise, lens aberrations, and fluctuations in gain and bias. Some well known stereo pairs, with ground truth, are used by researcher to benchmark their algorithms: the *Tsukuba* stereo pair (Nakamura et al., 1996), and *Sawtooth* and *Venus* created by (Scharstein & Szeliski, 2002). Though these sequence are widely used also in recent papers, in the last years the progress in the performances of stereo

algorithms is quickly outfacing the ability of these stereo data sets to discriminate among the best-performing algorithms, thus motivating the need for more challenging scenes with accurate ground truth information. To this end, (Scharstein & Szeliski, 2003) describe a method for acquiring high-complexity stereo image pairs with pixel-accurate correspondence information using structured light.

Nevertheless, databases for the evaluation of the performances of *active* stereo systems are still missing. The stereo geometry of the existing database is fixed, and characterized by parallel axis cameras. By using the software environment we developed, it is possible to collect a large number of data in different situations: e.g. vergent stereo cameras with different fixation points and orientation of the eyes, optic flow maps obtained for different ego-motion velocities, or different gaze orientation. The true disparity and optic flow maps can be stored together with the 3D data from which they have been generated and the corresponding image sequences. These data can be used for future algorithm benchmarking also by other researchers in the Computer Vision community. A tool capable of continuously generating ground truth data can be used online together with the visual processing algorithms to have a continuous assessment of their reliability. Moreover, the use of textured 3D models, acquired in real-world conditions, can solve the lack of realism that affects many datasets in the literature.

2.5 Computer Vision module

Visual features (e.g. edges, disparity, optic flow) are extracted by the sequence of binocular images by the Computer Vision module. It can implement any kind of computer vision algorithm. The faithful detection of the motion and of the distance of the objects in the visual scene is a desirable feature of any artificial vision system designed to operate in unknown environments characterized by conditions variable in time in an often unpredictable way. In the context of an ongoing research project (EYESHOTS, 2008) aimed to investigate the potential role of motor information in the early stages of human binocular vision, the computation of disparity and optic flow has been implemented in the simulator by a distributed neuromorphic architecture, described in (Chessa et al., 2009). In such distributed representations, or population codes, the information is encoded by the activity pattern of a network of simple and complex neurons, that are selective for elemental vision attributes: oriented edges, direction of motion, color, texture, and binocular disparity (Adelson & Bergen, 1991). In this way, it is possible to use the simulator to study adaptation mechanisms of the responses of the neural units on the basis of the relative orientation of the eyes.

2.6 Control module

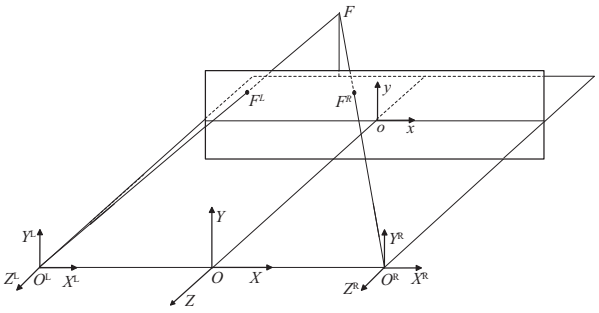
This module generates the control signal that is responsible for the camera/eye movements, in particular for version and vergence, and for the movement of the neck (rotation and position). By considering the neck fixed, and thus focusing on eye movements only, the simulator has been exploited to study a model of vergence control based on a dual-mode paradigm (Gibaldi et al., 2010; Hung et al., 1986). The goal of the vergence control module is to produce the control signals for the eyes to bring and keep the fixation point on the surface of the object of interest without changing the gaze direction. Since the task is to nullify the disparity in fovea, the vergence control module receives inputs from the same disparity detector population response described in Section 2.5 and converts it into the speed rotation of each eye. Other control models can be easily adopted to replace the existing one, in order to achieve different behaviours or to compare different algorithms and approaches.

3. Geometry of the stereo vision

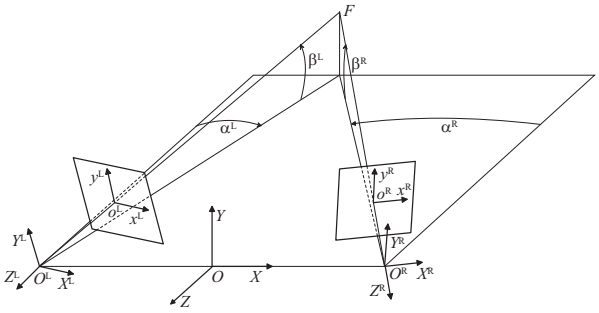
In the literature the most frequently used methods to render stereo image pairs are (Bourke & Morse, 2007; Grinberg et al., 1994): (1) the off-axis technique, usually used to create a perception of depth for a human observer and (2) the toe-in technique that can simulate the actual intensity patterns impinging on the cameras of a robotic head.

3.1 Off-axis technique

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 4a: **F** represents the location of the virtual point perceived when looking at the stereo pair composed by $\mathbf{F}^L$ and $\mathbf{F}^R$.



(a)



(b)

Fig. 4. (a) 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 **F**. The baseline b is denoted by $O^L O^R$. (b) 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 **F**. The baseline b is denoted by $O^L O^R$, the pan angles by α^L and α^R, and the tilt angles by β^L and β^R.

To produce a perception of depth for a human observer, it is necessary to pay attention to some specific geometrical parameters of the stereo acquisition setup (both actual and virtual) (Grinberg et al., 1994):

- the image planes have to be parallel;
- the optical points should be offset relative to the center of the image;
- the distance between the two optical centers have to be equal to the interpupillary distance;
- the field of view of the cameras must be equal to the angle subtended by the display screen;
- the ratio between the focal length of the cameras and the viewing distance of the screen should be equal to the ratio between the width of the screen and of the image plane.

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 (Southard, 1992).
However, it is difficult to perceptually render a large interval of 3D space without a visual stress, since the eye of the observer have to maintain accommodation on the display screen (at a fixed distance), thus lacking the natural relationship between accommodation and vergence eye movements, and the distance of the objects (Wann et al., 1995). Moreover, the visual discomfort is also due to spatial imperfections of the stereo image pair (Kooi & Toet, 2004). The main factors yielding visual discomfort are: vertical disparity; crosstalk, that is a transparent overlay of the left image over the right image and vice versa; blur, that is different resolutions of the stereo image pair.

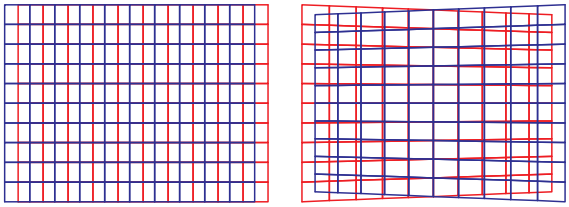
3.2 Toe-in technique

Since our aim is to simulate the actual images acquired by the vergent pan-tilt cameras of a robotic head, the correct way to create the stereo pairs is the toe-in method: each camera is pointed at a single target 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 model is shown in Figure 4b.
It is worth noting that, for specific application fields, the toe-in technique is also used for the perceptual rendering of the stereo image pair to a human observer. In the field of the telerobotic applications (Ferre et al., 2008; Bernardino et al., 2007), it is important to perceive veridical distances in the remote environment, and the toe-in technique allows choosing where the stereo images are properly fused and the optimal remote working area. However, the parallel axes configuration is again effective when a large workspace is necessary, e.g. for exploration vehicles. The toe-in method is also helpful in the field of stereoscopic television (Yamanoue, 2006), since the perception of the 3D scene is more easily manipulated, and the objects can be seen between the observer and the display screen, i.e. it is possible to render the crossed, zero, and uncrossed disparity.

The disparity patterns produced by the off-axis and toe-in techniques are shown in Figure 5a and Figure 5b, respectively.

3.3 Mathematics of the toe-in technique

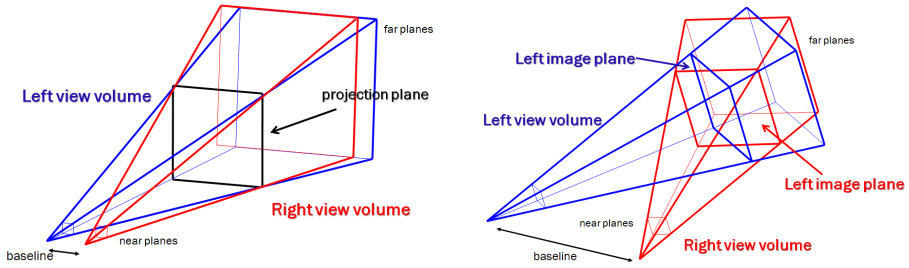
Our aim is to formally describe the toe-in technique in order to generate stereo image pairs like in a pan-tilt robotic head. To this purpose, the skewed frustum (see Fig. 6a) (necessary



(a) (b)

Fig. 5. 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.

to obtain the off-axis stereo technique) is no longer necessary. Accordingly, we introduced the possibility of pointing the left and the right optical axes at a single 3D target point, by rotating two symmetric frustums (see Fig. 6b), in order to obtain the left and the right views both fixating a point **F**.



(a) (b)

Fig. 6. (a) The two skewed frustums for the off-axis technique. (b) The two view volumes of the stereo cameras for the toe-in technique.

In general, the two camera frames $\mathbf{X}^L$ and $\mathbf{X}^R$ are related by a rigid-body transformation in the following way:

$$\mathbf{X}^R = \mathcal{R}\mathbf{X}^L + \mathcal{T} \tag{1}$$

where $\mathcal{R}$ and $\mathcal{T}$ denote the rotation matrix and the translation, respectively. The coordinate transformation described by Eq. 1 can be converted to a linear transformation by using homogeneous coordinates (Ma et al., 2004). In the following, we use the homogeneous coordinates to describe the coordinate transformation that brings the cameras from a parallel axes configuration to a convergent one.
The translation for the left and the right view volume can be obtained by applying the following translation matrix:

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

Then the azimuthal rotation (α^L and α^R) and the elevation (β^L and β^R) are obtained with the following rotation matrices:

$$\mathbf{R}_{\alpha}^{L/R} = \begin{bmatrix} \cos\alpha^{L/R} & 0 & \sin\alpha^{L/R} & 0 \\ 0 & 1 & 0 & 0 \\ -\sin\alpha^{L/R} & 0 & \cos\alpha^{L/R} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{3}$$

$$\mathbf{R}_{\beta}^{L/R} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\beta^{L/R} & -\sin\beta^{L/R} & 0 \\ 0 & \sin\beta^{L/R} & \cos\beta^{L/R} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{4}$$

The complete roto-translation of the view-volumes is:

$$\begin{bmatrix} \mathbf{O}^{L/R} \\ 1 \end{bmatrix} = \mathbf{R}_{\beta}^{L/R}\mathbf{R}_{\alpha}^{L/R}\mathbf{T}^{L/R}\begin{bmatrix} \mathbf{O} \\ 1 \end{bmatrix} \tag{5}$$

Thus, the projection direction is set to the target point **F**, then the left and the right views project onto two different planes, as it can be seen in Figure 4b.
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 axes and to decide the location of the fixation point. This emulates the behavior of a couple of verging pan-tilt cameras.

3.3.1 Camera rotations

In general, the frame transformation can be described by consecutive rotations (and translations), the specific rotation described by Eq. 5 is the Helmholtz sequence (neglecting the torsion of the camera, i.e. the rotation around the visual axis). This rotation sequence is related to the gimbal system of the actual camera (we are simulating). In particular, the horizontal axis is fixed to the robotic head, and the vertical axis rotates gimbal fashion around the horizontal axis (Haslwanter, 1995). That is, first we rotate through $\beta^{L/R}$ around the horizontal axis, then we rotate through $\alpha^{L/R}$ around the new updated vertical axis.
We can simulate a different gimbal system by using the Fick sequence (i.e., the vertical axis is fixed to the robotic head), described by:

$$\begin{bmatrix} \mathbf{O}^{L/R} \\ 1 \end{bmatrix} = \mathbf{R}_{\alpha}^{L/R}\mathbf{R}_{\beta}^{L/R}\mathbf{T}^{L/R}\begin{bmatrix} \mathbf{O} \\ 1 \end{bmatrix} \tag{6}$$

It is worth noting that the fixation point is described by different values of the angles.
For non conventional cameras, e.g. (Cannata & Maggiali, 2008), it is also possible to describe the camera rotation movements from the initial position to the final one through a single rotation by a given angle γ around a fixed axis $\mathbf{a}_{\gamma}$:

$$\begin{bmatrix} \mathbf{O}^{L/R} \\ 1 \end{bmatrix} = \mathbf{R}^{L/R}(\gamma, \mathbf{a}_{\gamma})\mathbf{T}^{L/R}\begin{bmatrix} \mathbf{O} \\ 1 \end{bmatrix} \tag{7}$$

In this way, we can study how the additive degrees of freedom of bio-inspired systems may have effects on the computational processing of visual features.

3.3.2 General camera model

A simple and widely used model for the cameras is characterized by the following assumptions: the vertical and horizontal axes of rotation are orthogonal, through the nodal points, and aligned with the image planes. However, the commercial cameras without a careful engineering can violate the previous assumptions, and also the cameras equipped with a zoom, since the position of the nodal point changes with respect to the position of the image plane as a function of focal length. A general camera model (Davis and Chen; 2003; Jain et al.; 2006; Horaud et al.; 2006) takes into account that the pan and tilt can have arbitrary axes, and the image plane are rigid objects that rotate around such axes. The actual camera geometry is described by:

$$\begin{bmatrix} \mathbf{O}^{L/R} \\ 1 \end{bmatrix} = \mathbf{T}_{pan}\mathbf{R}_{pan}\mathbf{T}_{pan}^{-1}\mathbf{T}_{tilt}\mathbf{R}_{tilt}\mathbf{T}_{tilt}^{-1}\begin{bmatrix} \mathbf{O} \\ 1 \end{bmatrix} \tag{8}$$

where $\mathbf{R}$ denotes a rotation around the tilt/pan axis, and $\mathbf{T}$ denotes a translation from the origin to each axis. In particular, the following steps are performed: first a translation $\mathbf{T}$ to the center of rotation, then a rotation $\mathbf{R}$ around the respective axis, and eventually a back translation for allowing the projection.

Figure 7 and 8 show the horizontal and vertical disparity maps for the different gimbal systems and for the general camera model. The stereo virtual cameras are fixating nine targets on a fronto-parallel plane, the central target is straight ahead and the other eight targets are symmetrically displaced at $\pm 14^{\circ}$. The baseline of the cameras is 6.5 cm with a field of view of 21°, and the plane is at 65 cm from the cameras. For the general camera model, we simulated a displacement of the nodal points of 0.6 cm, and a misalignment of the tilt and pan axes with respect to the image plane of 3°.

4. Geometry of the motion flow

In many robotic applications it is important to know how the coordinates of a point and its velocity change as the camera moves. The camera frame is the reference frame and we describe both the camera motion and the objects in the environment relative to it. The coordinates of a point $\mathbf{X}_0$ (at time $t = 0$) are described as a function of time t by the following relationship (Ma et al.; 2004):

$$\mathbf{X}(t) = \mathcal{R}(t)\mathbf{X}_0 + \mathcal{T}(t) \tag{9}$$

where $\mathcal{R}(t)$ and $\mathcal{T}(t)$ denote a trajectory that describes a continuous rotational and translational motion.

From the transformation of coordinates described by Eq. 9, the velocity of the point of coordinates $\mathbf{X}(t)$ relative to the camera frame (see Fig. 9) can be derived (Longuet-Higgins and Prazdny; 1980):

$$\dot{\mathbf{X}}(t) = \boldsymbol{\omega}(t) \times \mathbf{X}(t) + \mathbf{v}(t) \tag{10}$$

where $\times$ denotes the cross product, $\boldsymbol{\omega}(t)$ and $\mathbf{v}(t)$ denote the angular velocity and the translational velocity of the camera, respectively.

Figure 10 shows the motion fields for different kinds of camera movements. For the sake of simplicity, the visual axes are kept parallel and only the left frame is shown. The virtual set-up is the same of Fig. 7 and 8.

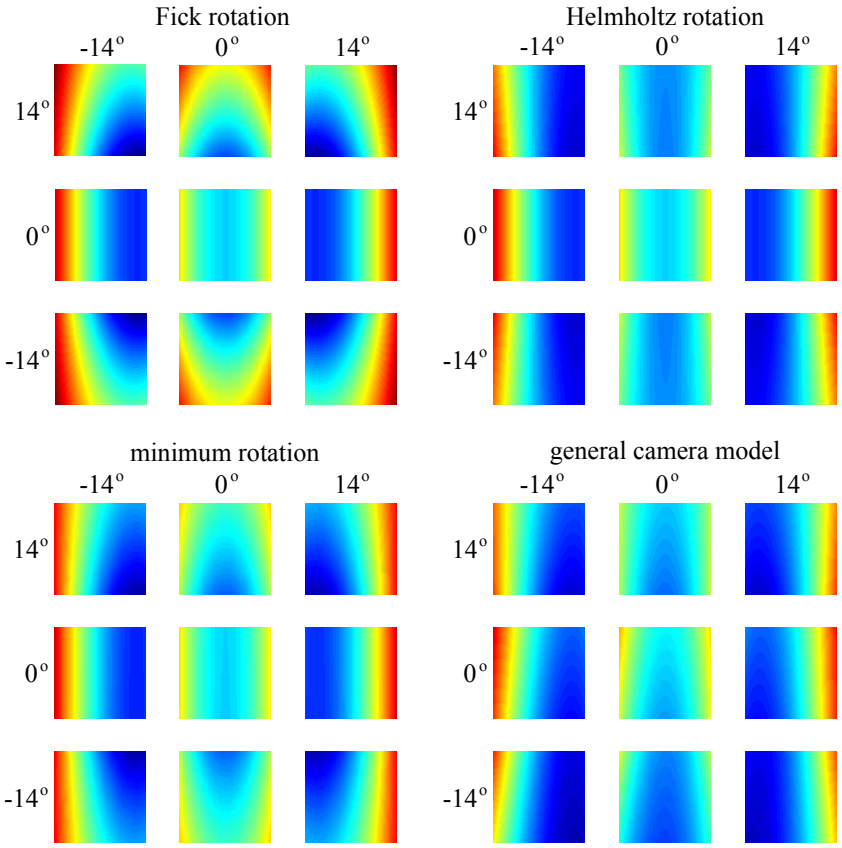


Fig. 7. Horizontal disparity patterns for different kinds of rotations and camera models. For each panel nine different gaze directions are shown. The disparity values are coded from red (uncrossed disparity) to blue (crossed disparity).

5. Software implementation

The virtual reality tool we propose is based on a C++ / OpenGL architecture and on the Coin3D graphic toolkit (`www.coin3D.org`). Coin3D is a high level 3D graphic toolkit for developing cross-platform real time 3D visualization and visual simulation software. It is portable over a wide range of platforms, it is built on OpenGL and uses scene graph data structures to render 3D graphics in real time. Coin3D is fully compatible with SGI Open Inventor 2.1, the de-facto standard for 3D visualization in the scientific and engineering communities. Both OpenGL and Coin3D code co-exist in our application.

In order to obtain a stereoscopic visualization of the scene useful to mimic an active stereo system, rather than to make a human perceive stereoscopy (see Section 3 for further details), we have not used the stereo rendering of the `SoCamera` node in the library, since it adopts the off-axis geometry. We have created our own `Cameras` class, that contains a pointer to a

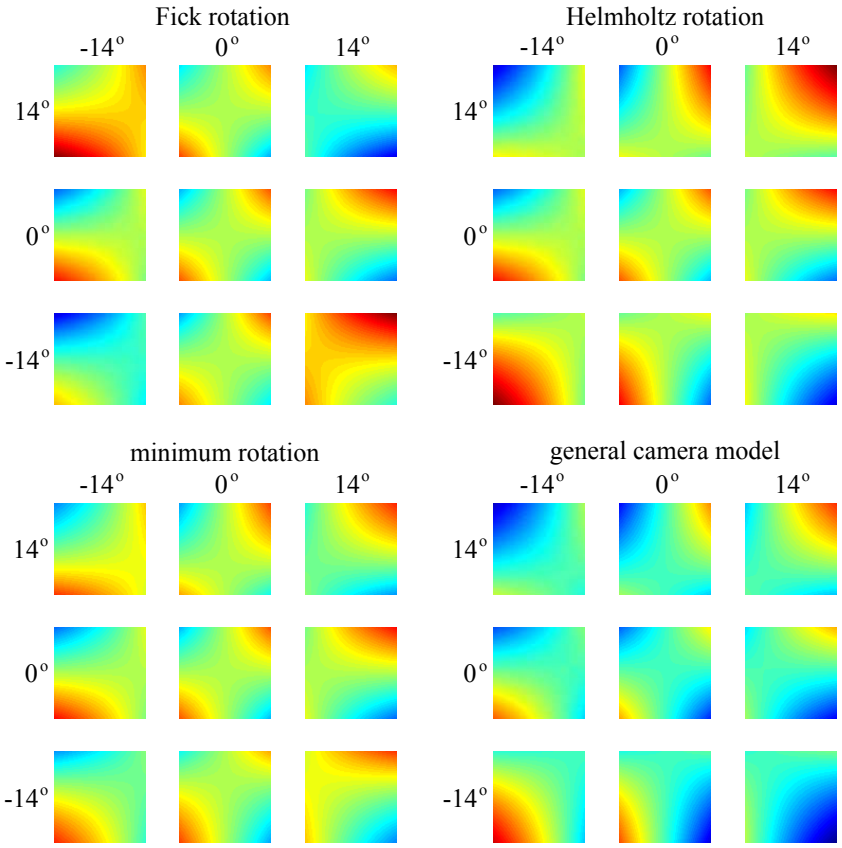


Fig. 8. Vertical disparity patterns for different kind of rotations and camera model. Same notation of Fig.‘7

`SoPerspectiveCamera`, which can be moved in the left, right and cyclopic position. The class stores the status of the head:

- 3D position of the neck;
- projection direction of the cyclopic view;
- direction of the baseline, computed as the cross product between the projection direction and the up vector;

and the status of each view:

- 3D position and rotation (R^R and R^L) computed with respect to the $(0,0,-1)$ axis;
- values of the depth buffer with respect to the actual position of the camera.

The left and the right views are continuously updated after having computed the rotation R^R and R^L necessary to fixate the target. Also the position of the neck, the projection direction, and the direction of the baseline can be updated if the neck is moving.

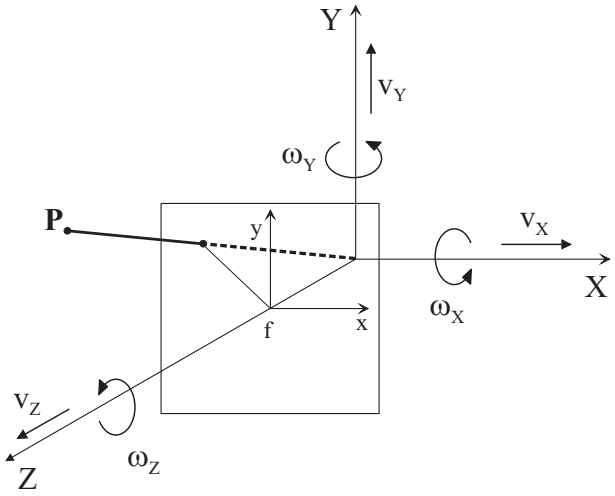


Fig. 9. Viewer-centered coordinate frame. The relative motion between an observer and the scene can be described at each instant t as a rigid-body motion, by means of two vectors (i.e., kinetic characteristics): the translational velocity $\mathbf{v} = (v_X, v_Y, v_Z)^T$, and the angular velocity $\boldsymbol{\omega} = (\omega_X, \omega_Y, \omega_Z)^T$.

The scene from the point of view of the two stereo cameras is then rendered both in the on-screen OpenGL context and in the off-screen buffer. At the same time the depth buffer is read and stored. It is worth noting that, since Coin3D library does not easily allow the users to access and store the depth buffer, the `SoOffscreenRender` class has been modified in order to add this feature. After such a modification it is possible to access both the color buffer and the depth buffer.
The ground truth maps can be then generated and stored.

5.1 Ground truth data generation

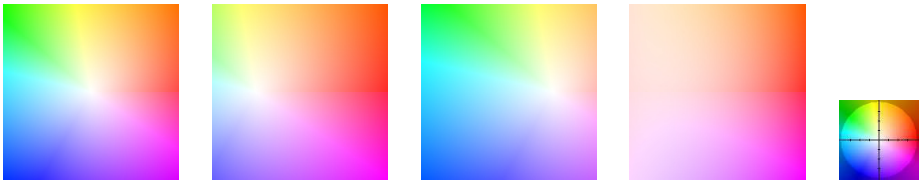
To compute the ground truth data it is necessary to exploit the resources available from the graphics engine by combining them through the computer vision relationships that describe a 3D moving scene and the geometry of two views, typically used to obtain a 3D reconstruction.

5.1.1 Stereo cameras

Formally, by considering two static views, the two camera reference frames are related by a rigid body transformation described by the rotation matrix $\mathcal{R}$ and the translation $\mathcal{T}$ (see Eq. 1), thus the two projections (left and right) are related in the following way (Ma et al., 2004):

$$\lambda^R \mathbf{x}^R = \mathcal{R}\lambda^L \mathbf{x}^L + \mathcal{T} \tag{11}$$

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.
In order to define the disparity, we explicitly write the projection equations for Eq. 5 (Helmholtz sequence). The relation between the 3D world coordinates $\mathbf{X} = (X, Y, Z)$ and the



(a) (b) (c) (d) (e)

Fig. 10. Motion fields for different camera movements: (a) the v_Z camera velocity produces an expansion pattern with the focus of expansion in the center; the superposition of a v_X velocity moves the focus of expansion on the left (b) or on the right (c) as a function of its sign; (d) the cameras move with v_Z and a rotational velocity ω_Y. (e) The color-coding scheme used for the representation: the hue represents the velocity direction, while its magnitude is represented by the saturation.

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 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. It is worth noting that the fixation point $\mathbf{F}$ in Figure 4b is projected onto the origins of the left and right image planes, since the vergence movement makes the optical axes of the two cameras to intersect in $\mathbf{F}$. For identical left and right focal lengths f_0, the left image coordinates are (Volpel & Theimer, 1995):

$$
\begin{aligned}
x^L &= f_0 \frac{X_+ \cos\alpha^L + Z \sin\alpha^L}{X_+ \sin\alpha^L \cos\beta^L - Y \sin\beta^L - Z \cos\alpha^L \cos\beta^L} \\
y^L &= f_0 \frac{X_+ \sin\alpha^L \sin\beta^L + Y \cos\beta^L - Z \cos\alpha^L \sin\beta^L}{X_+ \sin\alpha^L \cos\beta^L - Y \sin\beta^L - Z \cos\alpha^L \cos\beta^L}
\end{aligned}
\tag{12}
$$

where $X_+ = X + b/2$. Similarly, the right image coordinates are obtained by replacing α^L, β^L and X_+ in the previous equations with α^R, β^R and $X_- = X - b/2$, respectively. We can define the horizontal disparity $d_x = x^R - x^L$ and the vertical disparity $d_y = y^R - y^L$, that establish the relationship between a world point $\mathbf{X}$ and its associated disparity vector $\mathbf{d}$.

5.1.2 A moving camera

Considering the similarities between the stereo and motion problems, as they both look for correspondences between different frames or between left and right views, the generalizations of the two static views approach to a moving camera is in principle straightforward. Though, the description of the stereoscopically displaced cameras and of the moving camera are equivalent only if the spatial and temporal differences between frame are small enough, since the motion field is a differential concept, but not the stereo disparity. In particular, the following conditions must be satisfied: small rotations, small field of view, and v_Z small with respect to the distance of the objects from the camera. These assumptions are related to the analysis of video streams, where the camera motion is slow with respect to the frame rate (sampling frequency) of the acquisition device. Thus, we can treat the motion of the camera as continuous (Ma et al., 2004; Trucco & Verri, 1998; Adiv, 1985).

The relationship that relates the image velocity (motion field) $\dot{\mathbf{x}}$ of the image point $\mathbf{x}$ to the angular ($\boldsymbol{\omega}$) and the linear ($\mathbf{v}$) velocities of the camera and to the depth values is described by

the following equation (see also Eq. 10):

$$\dot{\mathbf{x}} = \boldsymbol{\omega} \times \mathbf{x} + \frac{1}{\lambda}\mathbf{v} - \frac{\dot{\lambda}}{\lambda}\mathbf{x} \tag{13}$$

where λ and $\dot{\lambda}$ are the depth and its temporal derivative, respectively.
For planar perspective projection, i.e. $\lambda = Z$, we have that the image motion field $\dot{\mathbf{x}}$ is expressible as a function of image position $\mathbf{x} = (x, y)$ and surface depth $Z = Z(x, y)$ (i.e., the depth of the object projecting in (x, y) at current time):

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \frac{1}{Z}\begin{bmatrix} f_0 & 0 & -x \\ 0 & f_0 & -y \end{bmatrix}\mathbf{v} + \begin{bmatrix} -xy/f_0 & (f_0 + x^2/f_0) & -y \\ -(f_0 + y^2/f_0) & xy/f_0 & x \end{bmatrix}\boldsymbol{\omega} \tag{14}$$

To apply the relationship described by Eqs. 11 and 13 we first read the z-buffer (w) of the camera through the method added in the `SoOffScreenRenderer` class, then we obtain the depth values with respect to the reference frame of the camera in the following way:

$$\lambda = \frac{f\,n}{w(f-n) - f} \tag{15}$$

where f and n represent the values of the far and the near planes of the virtual camera, respectively. Finally, from Eq. 11 it is possible to compute the ground truth disparity maps $\mathbf{d}$, and from Eq. 13 it is possible to obtain the ground truth motion field $\dot{\mathbf{x}}$.

6. Results for different visual tasks

The proposed VR tool can be used to simulate any interaction between the observer and the scene. In particular, in the following two different situations will be considered and analyzed:

1. Scene exploration, where both the head and the scene are fixed, and only ocular movements are considered.
2. Robotic navigation, by considering monocular vision, only.

6.1 Active vision - Scene exploration

By keeping fixed the position and the orientation of the head, 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) to use the system to obtain sequences where the fixation points are chosen on the surfaces of the objects in the scene; (2) to use the system in cooperation with an algorithm that implements a vergence/version strategy. In the first case, it is not possible to fixate beyond 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 visual input for selecting a new target point.
To compute the fixation point in 3D coordinates, starting from its 2D projection, the `SoRayPickAction` class has been used. It contains the methods for setting up a ray from the near plane to the far plane, from the 2D point in the projection plane. Then the first hit, the one that corresponds to the first visible surface, is taken as the 3D coordinates, the system should fixate.

Figure 11 shows the active exploration of an indoor scene, representing a desktop and different objects at various distances, acquired by using a laser scanner. The simulator aims to mimic the behavior of a human-like robotic system acting in the peripersonal space. Accordingly, the interocular distance between the two cameras is set to 6 cm and the distance between the cameras and the center of the scene is about 80 cm. The fixation points have been chosen arbitrary, thus simulating an active exploration of the scene, and in their proximity the disparity between the left and the right projections is zero, while getting far from the fixation point both horizontal and vertical disparities emerge, as it can be seen in the ground truth data.
Instead of directly computing the 3D coordinates of the fixation points, it is possible to analyze the sequence of images and the corresponding disparity maps, while performing vergence movements. In Figure 12 it is possible to see the red-cyan anaglyph of the stereo pairs and the corresponding disparity maps, before having reached the fixation point (upper part of the figure) and when fixation is achieved onto a target (bottom). The plots on the right show the variation of the disparity in the center of the image at each time step (upper part) and the variation of the actual depth of the fixation point with respect to the desired value (bottom).

6.2 Robotic navigation

In this Section, the simulator is used to obtain sequences acquired by a moving observer. The position and the orientation of the head can be changed, in order to mimic the navigation in the virtual environment. For the sake of simplicity, the ocular movements are not considered and the visual axes are kept parallel. It is worth noting that the eye movements necessary to actively explore the scene, considered in the previous Section, could be embedded if necessary. Figure 13 shows the sequence of images and the related ground truth optic flow fields for the different movements of the observer.

7. Conclusion

In conclusion, a tool that uses the VR to simulate the actual projections impinging the cameras of an active visual system rather than to render the 3D visual information for a stereoscopic display, has been developed. The simulator works on the 3D data that can be synthetically generated or acquired by a laser scanner and performs both cameras and robot movements following the strategies adopted by different active stereo vision systems, including bio-mimetic ones.
The virtual reality tool is capable of generating pairs of stereo images like the ones that can be obtained by a verging pan-tilt robotic head and the related ground truth data, disparity maps and motion field.
To obtain such a behavior the toe-in stereoscopic technique is preferred to the off-axis technique. By proper roto-translations of the view volumes, we can create benchmark sequences for vision systems with convergent axis. Moreover, by using the precise 3D position of the objects these vision systems can interact with the scene in a proper way. A data set of stereo image pairs and the related ground truth disparities and motion fields are available for the Robotics and Computer Vision community at the web site `www.pspc.dibe.unige.it/Research/vr.html`.
Although the main purpose of this work is to obtain sufficiently complex scenarios for benchmarking an active vision system, complex photo-realistic scenes can be easily obtained by using the 3D data and textures acquired by laser scanners, which capture detailed, highly accurate, and full color objects to build 3D virtual models at an affordable computational cost.

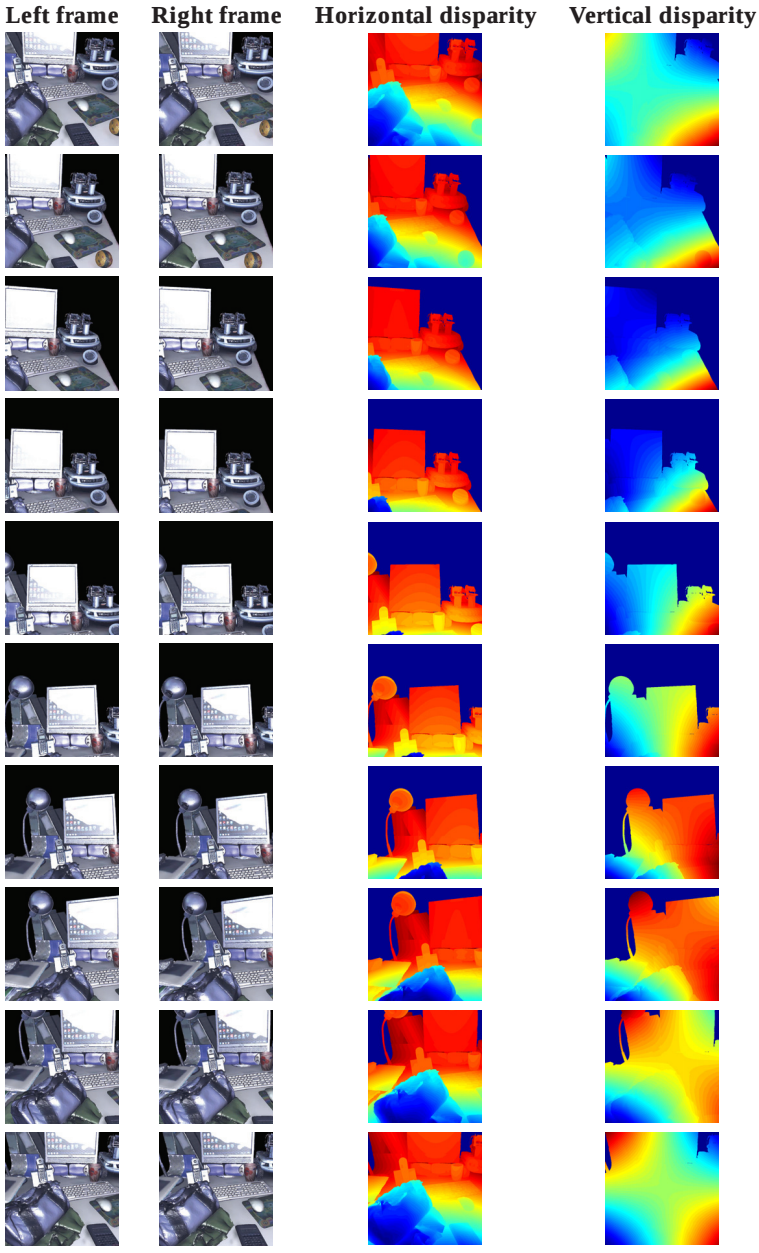


Fig. 11. Active exploration of a scene. The position and the orientation of the head is fixed, whereas the eyes are moving in order to explore the scene. The scenario mimics a typical indoor scene. The disparity values are coded from red (uncrossed disparity) to blue (crossed disparity).

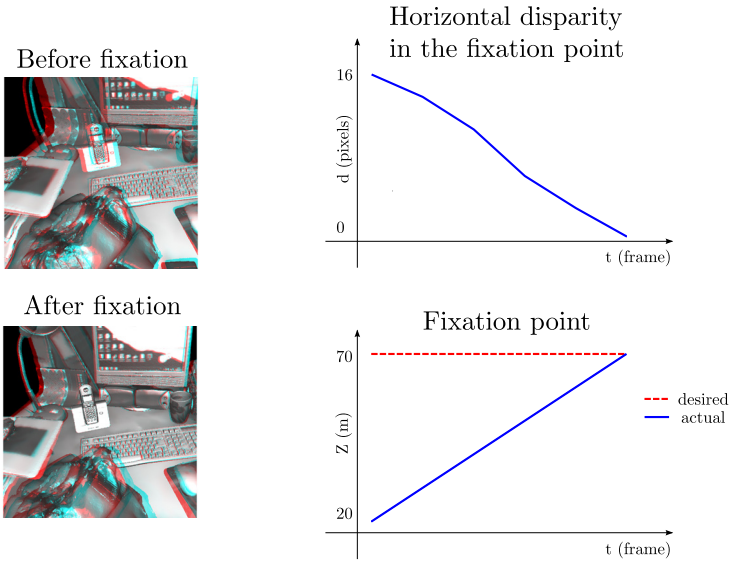


Fig. 12. (left) The anaglyph images before and after a vergent movement of the cameras in order to correctly fuse the left and right images of the target object. (right) The plots show the variation of the disparity in the center of the stereo images and the related depth of actual and desired fixation point. It is worth noting that the vergence movements are synthetically generated (i.e., not driven by the visual information).

In this way improving the photo-realistic quality of the 3D scene does not endanger the definition of a realistic model of the interactions between the vision system and the observed scene. As part of a future work, we plan to modify the standard pan-tilt behaviour by including more biologically plausible constraints on the camera movements Schreiber et al. (2001); Van Rijn & Van den Berg (1993) and to integrate vergence / version strategies in the system in order to have a fully active tool that interacts with the virtual environments.

Acknowledgements

We wish to thank Luca Spallarossa for the helpful comments, and Andrea Canessa and Agostino Gibaldi for the acquisition, registration and post-processing of the 3D data.
This work has been partially supported by EU Projects FP7-ICT 217077 “EYESHOTS” and FP7-ICT 215866 “SEARISE”.

8. References

Adelson, E. & Bergen, J. (1991). The plenoptic and the elements of early vision, *in* M. Landy & J. Movshon (eds), *Computational Models of Visual Processing*, MIT Press, pp. 3–20.

Adiv, G. (1985). Determining three-dimensional motion and structure from optical flow generated by several moving objects, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **7**: 384–401.

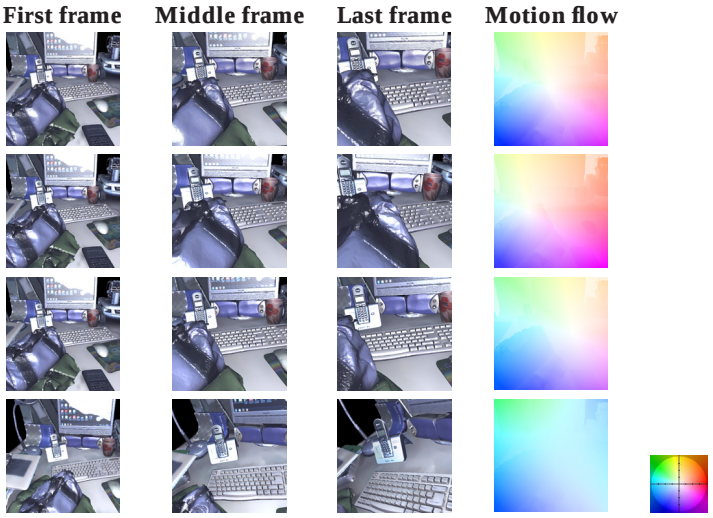


Fig. 13. Robotic navigation in an indoor scenario. Different situation are taken into account. First row: the robot has v_Z velocity, only. Thus the focus of expansion is in the center. Second and third row: positive and negative v_X are introduced, thus the focus of expansion move to the left and to the right, respectively. Fourth row: a rotation around the Y axis is combined with a translation aking v_Z.

Awaad, I., Hartanto, R., León, B. & Plöger, P. (2008). A software system for robotic learning by experimentation, *Workshop on robot simulators (IROS08)*.

Baker, S., Scharstein, D., Lewis, J., Michael, S., Black, J. & Szeliski, R. (2007). A database and evaluation methodology for optical flow, *ICCV*, pp. 1–8.

Bernardino, A., Santos-Victor, J., Ferre, M. & Sanchez-Urn, M. (2007). Stereoscopic image visualization for telerobotics. experiments with active binocular cameras, *Advances in Telerobotics*, pp. 77–90.

Bourke, P. & Morse, P. (2007). Stereoscopy: Theory and practice, *Workshop at 13th International Conference on Virtual Systems and Multimedia* .

Cannata, G. & Maggiali, M. (2008). Models for the design of bioinspired robot eyes, *IEEE Transactions on Robotics* **24**: 27–44.

Chessa, M., Sabatini, S. & Solari, F. (2009). A fast joint bioinspired algorithm for optic flow and two-dimensional disparity estimation, *in* M. Fritz, B. Schiele & J. Piater (eds), *Computer Vision Systems*, Vol. 5815 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, pp. 184–193.

Davis, J. & Chen, X. (2003). Calibrating pan-tilt cameras in wide-area surveillance networks, *Proc. of IEEE International Conference on Computer Vision*, IEEE Computer Society, pp. 144–150.

EYESHOTS (2008). *FP7-ICT 217077 "EYESHOTS"*, `http://www.eyeshots.it/`.

Ferre, P., Aracil, R. & Sanchez-Uran, M. (2008). Stereoscopic human interfaces, *IEEE Robotics & Automation Magazine* **15**(4): 50–57.

Fleet, D. & Jepson, A. (1990). Computation of component image velocity from local phase information, *International Journal of Computer Vision* **5**(1): 77–104.

Forsyth, D. & Ponce, J. (2002). *Computer Vision: A Modern Approach*, Prentice Hall.

Gibaldi, A., Chessa, M., Canessa, A., Sabatini, S. & Solari, F. (2010). A cortical model for binocular vergence control without explicit calculation of disparity, *Neurocomputing* **73**(7-9): 1065 – 1073.

Grinberg, V., Podnar, G. & Siegel, M. (1994). Geometry of binocular imaging, *Proc. of the IS&T/SPIE Symp. on Electronic Imaging, Stereoscopic Displays and applications*, Vol. 2177, pp. 56–65.

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

Heeger, D. (1987). Model for the extraction of image flow, *Journal of the Optical Society of America A* **4**(1): 1445–1471.

Horaud, R., Knossow, D. & Michaelis, M. (2006). Camera cooperation for achieving visual attention, *Machine Vision and Applications* **16**: 331–342.

Hung, G., Semmlow, J. & Ciufferda, K. (1986). A dual-mode dynamic model of the vergence eye movement system, *Biomedical Engineering, IEEE Transactions on* **BME-33**(11): 1021–1028.

Jain, A., Kopell, D., Kakligian, K. & Wang, Y. (2006). Using stationarydynamic camera assemblies for wide-area video surveillance and selective attention, *In IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 1, pp. 537–544.

Jørgensen, J. & Petersen, H. (2008). Usage of simulations to plan stable grasping of unknown objects with a 3-fingered schunk hand, *Workshop on robot simulators (IROS08)*.

Kooi, F. & Toet, A. (2004). Visual comfort of binocular and 3d displays, *Displays* **25**(2-3): 99–108.

Liu, Y., Bovik, A. & Cormack, L. (2008). Disparity statistics in natural scenes, *Journal of Vision* **8**(11): 1–14.

Longuet-Higgins, H. & Prazdny, K. (1980). The interpretation of a moving retinal image, *Phil. Trans. R. Soc. Lond. B* **208**: 385–397.

Ma, Y., Soatto, S., Kosecka, J. & Sastry, S. (2004). *An Invitation to 3D Vision. From Images to Geometric Models*, Springer-Verlag.

McCane, B., Novins, K., Crannitch, D. & Galvin, B. (2001). On benchmarking optical flow, *Computer Vision and Image Understanding* **84**(1): 126–143.

Michel, O. (2004). Webots: Professional mobile robot simulation, *International Journal of Advanced Robotic Systems* **1**(1): 39–42.

Nakamura, Y., Matsuura, T., Satoh, K. & Ohta, Y. (1996). Occlusion detectable stereo-occlusion patterns in camera matrix, *Computer Vision and Pattern Recognition, 1996. Proceedings CVPR '96, 1996 IEEE Computer Society Conference on*, pp. 371 –378.

Okada, K., Kino, Y. & Kanehiro, F. (2002). Rapid development system for humanoid vision-based behaviors with real-virtual common interface, *IEEE/RSJ IROS*.

Otte, M. & Nagel, H. (1995). Estimation of optical flow based on higher-order spatiotemporal derivatives in interlaced and non-interlaced image sequences, *Artif. Intell.* **78**(1-2): 5–43.

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

Scharstein, D. & Szeliski, R. (2003). High-accuracy stereo depth maps using structured light, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 1, pp. 195–202.

Schreiber, K. M., Crawford, J. D., Fetter, M. & Tweed, D. B. (2001). The motor side of depth vision, *Nature* **410**: 819–822.

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

Tikhanoff, V., Cangelosi, A., Fitzpatrick, P., Metta, G., Natale, L. & Nori, F. (2008). An open-source simulator for cognitive robotics research: The prototype of the iCub humanoid robot simulator, *Workshop on Performance Metrics for Intelligent Systems Workshop*.

Trucco, E. & Verri, A. (1998). *Introductory Techniques for 3-D Computer Vision*, Prentice Hall.

Ulusoy, I., Halici, U. & Leblebicioglu, K. (2004). 3D cognitive map construction by active stereo vision in a virtual world, *Lecture notes in Computer Science* **3280**: 400–409.

Van Rijn, L. & Van den Berg, A. (1993). Binocular eye orientation during fixations: Listing's law extended to include eye vergence, *Vision Research* **33**: 691–708.

Volpel, B. & Theimer, W. (1995). Localization uncertainty in area-based stereo algorithms, *IEEE Transactions on Systems, Man and Cybernetics* **25**(12): 1628–1634.

Wann, J. P., Rushton, S. & Mon-Williams, M. (1995). Natural problems for stereoscopic depth perception in virtual environments., *Vision research* **35**(19): 2731–2736.

Yamanoue, H. (2006). The differences between toed-in camera configurations and parallel camera configurations in shooting stereoscopic images, *Multimedia and Expo, IEEE International Conference on* pp. 1701–1704.

Yang, Z. & Purves, D. (2003). Image/source statistics of surfaces in natural scenes, *Network: Comput. Neural Syst.* **14**: 371–390.