



Project no.: FP7-ICT-217077
Project full title: Heterogeneous 3-D Perception across Visual Fragments
Project Acronym: EYESHOTS
Deliverable no: D1.1
Title of the deliverable: Binocular eye coordination and its role in depth vision.

Date of Delivery:	07 March 2009	
Organization name of lead contractor for this deliverable:	UG	
Author(s):	G. Cannata, A. Canessa, M. Chessa, F. Solari, S.P. Sabatini	
Participant(s):	UG	
Workpackage contributing to the deliverable:	WP1	
Nature:	Report	
Version:	2.1	
Total number of pages:	48	
Responsible person:	Giorgio Cannata	
Revised by:	M. Lappe, R. Volcic	
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:

The problem of binocular eye coordination in stereovision is tackled from a computational point of view. Specifically, this deliverable focuses on the problem of controlling the vergence over the surface of an object assumed within the peri-personal workspace of a robot. Built upon this control module, a control scheme to make possible arbitrary motion of the vergence point is proposed. We conjecture that small and smooth sliding motions around local salient features might have a role to improve the perception of the 3D object properties. Biologically related constraints on torsional ocular postures are also considered as additional degrees of freedom on the eyes/cameras movements.

Contents

1	Executive Summary	3
2	Introduction	4
3	Notations	5
4	Vergence and Vergence Tracking Control	5
5	Vergence Control over a Surface \mathcal{S}	9
5.1	Specification of task errors	10
5.2	Vergence Control Signals	11
5.3	Simplified Control Strategies	12
5.3.1	Continuous Decoupled Control	13
5.3.2	Non-smooth Decoupled Control	14
5.4	Advanced Analysis	16
5.5	Vergence Control Simulations	19
6	Surface Tracking Under Vergence Conditions	21
6.1	Unconstrained Vergence Tracking	23
6.2	Constrained Vergence Control	25
6.3	Computing the vector \mathbf{k}_S	27
6.4	Tracking Control Simulations	30
7	Torsional ocular posture	30
7.1	The Listing's Law	33
7.2	Binocular Listing's Law	34
7.3	The meaning of Listing's Law	35
7.4	The meaning of binocular Listing's Law or L2	37
7.5	Implications on the disparity patterns	38
7.6	Open issues	40
A	Extraocular muscles and their actions	43
B	Mathematics of eye movements	44
C	Time derivative of unit geometric vectors	45
D	Simplified analysis of convergence properties of the decoupled control	47
E	Solving Equation (6.6)	47

1 Executive Summary

This document is a Technical Report which constitutes the Deliverable D1.1 entitled *Binocular Eye Coordination and its Role in Depth Vision* of the EU Project EYESHOTS. Deliverable D1.1 is part of the workpackage WP1: *Eye Movements for Exploration of the 3D Space*. In particular D1.1 is the *outcome* of the activities of the worktask **Task 1.1** *Eye Movements for Exploration of the 3D Space*. The goal of **Task 1.1** is the investigation of ocular motion strategies and the analysis of how they influence on perception and estimation of of 3D information (depth). This activity is mostly related to geometric and kinematics analysis of the vision system, while vision and image processing methods represent, depending on the case, *inputs* or *outputs* to/from the modules analyzed throughout this document.

The rationale adopted during **Task 1.1** has been the following. We have given for granted that large range ocular movements between *discrete* salient features (in a *quasi static environment*) obey to *high speed* transient ocular motion strategies (*saccades*) where motion dynamics strongly dominate, thus (probably) limiting the effects of visual feedback to the perception of depth. On the other hand, we have conjectured that between saccades *small range* and *low speed smooth* ocular motions might occur (or be implemented at least in the case of an artificial robotic vision system) with the goal of improving *locally* the depth perception.

To this aim, we have first analysed the ideal problem of *moving* the vergence point over a smooth surface in space (sections 5 and 6). This task has been tackled mostly from the kinematic control point of view to understand the requirements for its implementation (in particular in term of feedback requirements from vision modules). Furthermore, although the analysis refers to motions of *large* amplitude it should be interpreted as an idealization of small controlled motions enabling the enhancement of depth perception.

As a limit case, we have eventually considered the static conditions of the cameras when fixating an object within the *peri-personal* space¹ This investigation has led to a critic study of the relative posture of two cameras with respect to the perceived disparity information. In particular, the study aimed, on one hand to provide a perceptual interpretation of the *extended Listing's Law*, and on the other to analyze how different types of kinematic models (e.g. *pan-tilt* etc.) affect or limit its implementation.

The major results of this Report have strong connection with the remaining worktasks of WP1 as well as with **WP2**, which will provide at further

¹Roughly speaking the intersection of the working envelope of the human/robot arm with the field of view. Within the EYESHOTS project we always assume the robot head fixed, so that the *peri-personal* space is a stationary 3D volume.

stage of the project the main visual feedback information required for the actual control of the robot eyes planned in **Task 1.4**, which represents the final output of **WP1**.

2 Introduction

This reports will first address a class of co-ordinated motions of a couple of stereo cameras, and secondly, it will analyse the properties of a particular class of static geometric postures that are known in the literature as *Generalized Listing's Law* configurations.

The first goal is to specify a control model that allows us to coordinate the rotation of the cameras so that the vergence point is maintained *most of the time* (i.e. except for possible finite transient time intervals) over the surface of an object assumed within the peri-personal space of the robot.

Built upon this control module we shall investigate a control scheme that makes possible to *move* the vergence point so that it could slide over the object's surface. This kind of task is addressed within this report for arbitrary motions. We conjecture that small and smooth sliding motions around a *nominal* salient feature might have a role to improve the depth perception of the scene locally, and eventually provide the awareness of the local 3D object properties.

The second major goal of the report is instead related to the analysis of the effects of the relative *torsion* of the stereo cameras under the assumption of static configurations. Roughly speaking, for a given vergence point the cameras have in general two more degrees of freedom. It is known that in biological systems for short distance fixations a particular relative torsional component exists among the eyes, which is specified by a principle known as *Generalized Listing's Law*. We shall show which is the effect of torsion in terms of the perceived disparity and how different ocular kinematic models approximate *Generalized Listing's Law*. The structure of the report is the following. In section 4 the general statement of the vergence and vergence tracking control is given; then, in section 5 a vergence control law is proposed, including simulation results. The general model for vergence tracking is discussed in section 6, including simulation results. Finally in section 7 an interpretation of the *Generalized Listing's Law* is provided as long as its approximation by different ocular kinematic models.

3 Notations

In this document the following notations and conventions will be adopted:

- Scalar are described by lower case letters (with sub/super-scripts if needed), e.g. x , d .
- Angles are described by greek letters: e.g. θ , α .
- Points in 3D space are represented by capital letters (with sub/super-scripts if needed), e.g. P_0 , F_L
- 3D *geometric* vectors are represented by bold lower case variables (with sub/super-scripts if needed), e.g. \mathbf{v} , \mathbf{k}_L . Geometric vectors are not expressed with respect to any reference frame.
- 3D *algebraic* vectors are represented as bold lower case letters (with sub/super-scripts if needed) of the form: ${}^a\mathbf{x}$, where the superscript letter defines the reference frame used to represent the vector.
- Generic n -dimensional vectors are represented by capital letters: e.g. \mathbf{a} , \mathbf{b} . The dimension of the vector will be specified or self-evident from the context. *Algebraic* vectors of any dimension are always intended as *column vectors*.
- Vectors with a *bar* on top are normalized, e.g. $\bar{\mathbf{a}} = \mathbf{a}/|\mathbf{a}|$.
- Reference frames will be represented as letters within brackets, e.g. $\langle L \rangle$, $\langle R \rangle$.
- Matrices are represented by capital bold letters, e.g. \mathbf{Q} , \mathbf{A} . The dimensions of the matrix will be specified or self-evident from the context.
- Rotation matrices are expressed as capital bold letters as: ${}^a_b\mathbf{R}$ where $\langle b \rangle$ and $\langle a \rangle$ are the origin and target reference frames respectively.

4 Vergence and Vergence Tracking Control

The major goal of the next two sections is to analyze the mathematical conditions required to ensure the fulfillment of a particular visuo-motor task where vergence is achieved on a given surface \mathcal{S} , and vergence point is *moved* over the surface. To this aim consider a locally smooth surface in 3D space described by the implicit equation:

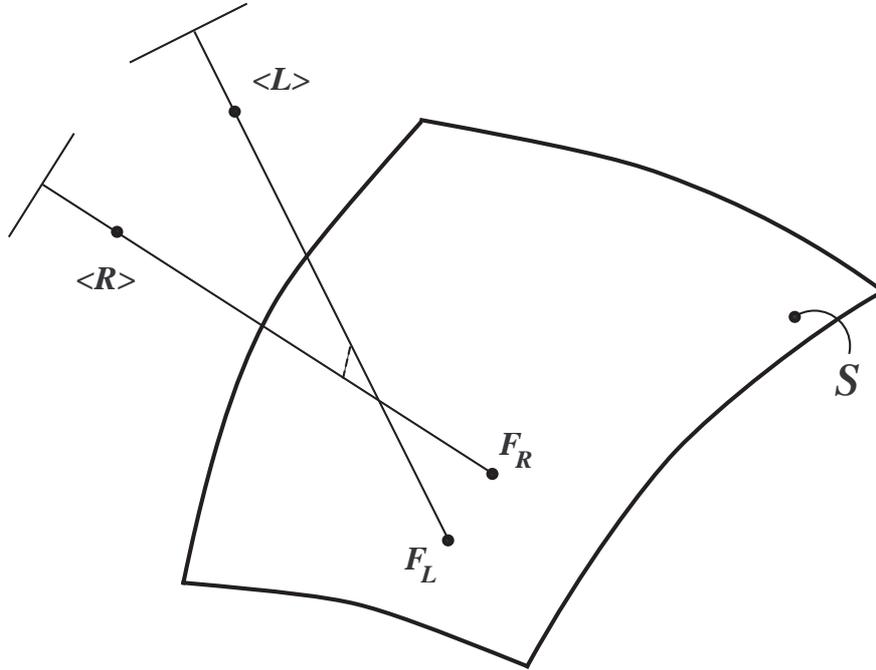


Figure 1: Transient phase: the vergence point (if any) $\boldsymbol{x}^* \notin \mathcal{S}$, rotation control of the SCS should make the fixation points to be coincident on \mathcal{S} .

$$\mathcal{S}(\boldsymbol{x}) = 0 \quad (4.1)$$

Let \boldsymbol{x}^* be the vergence point of a robot *Stereo Camera System* (SCS). The goal is to *move* the fixation points of both the cameras, by proper coordinated control of the rotation of the SCS, along the surface \mathcal{S} , so that they are coincident *most of the time*² which might lead to discontinuities in the 3D position of the fixation points. *Most of the time* in this sense refers to the requirement that transients required to regain vergence over \mathcal{S} are *short* with respect to the duration of the task.

Consider the figures 1 and 2, from a control point of view this task has two sub-goals:

- (i) taking \boldsymbol{x}^* on \mathcal{S} ;
- (ii) moving \boldsymbol{x}^* over \mathcal{S} ;

²We clearly expect that in practice, beyond the possible obvious control inaccuracies, the vergence conditions could not be ensured for any *visually bounded* a-priori unknown surface.

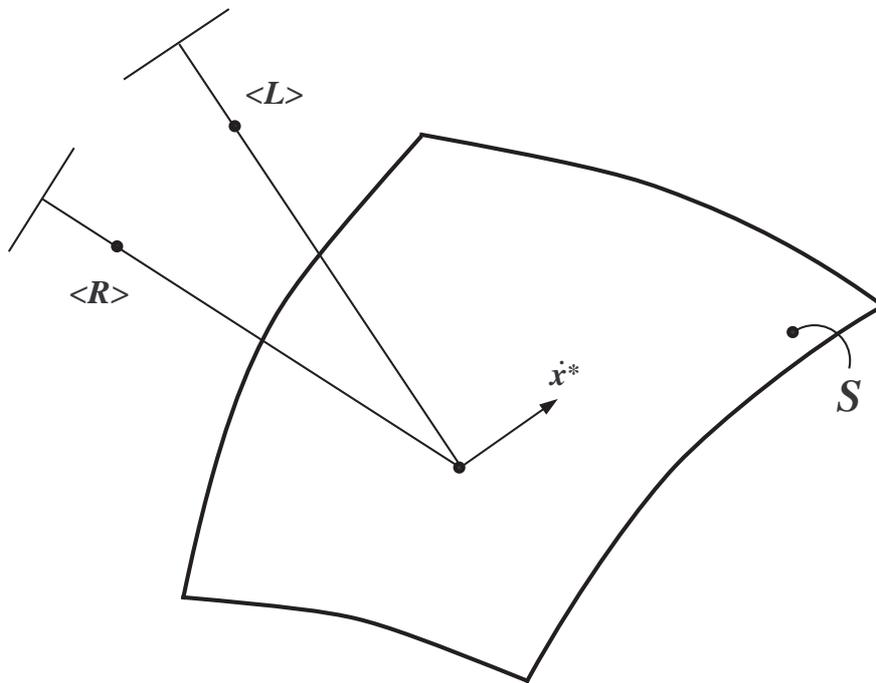


Figure 2: Tracking phase: co-ordinated rotation of the SCS makes x^* slide over \mathcal{S} .

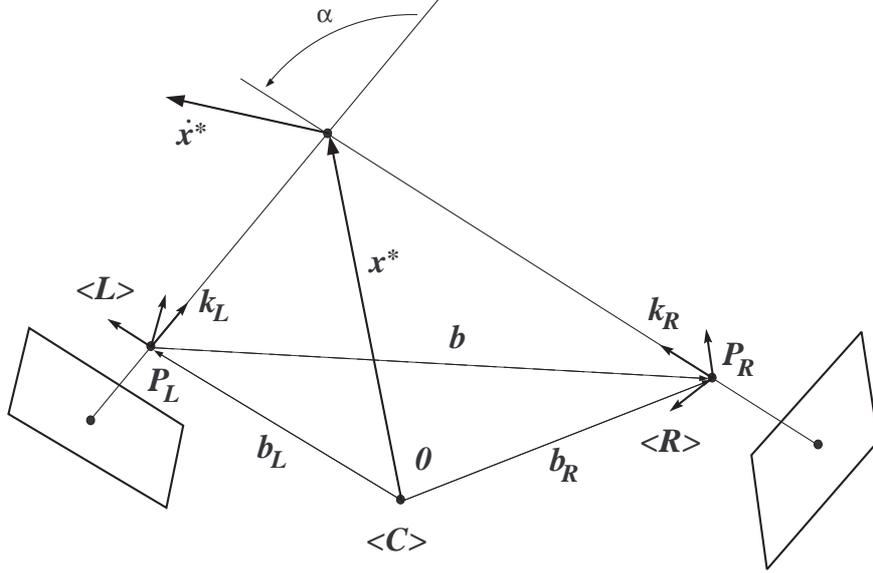


Figure 3: The general SCS configuration.

These two control tasks will in practice operate in parallel, assuming within this framework that vergence control has higher priority than tracking.

In the next two sections we shall first discuss how to address the vergence control problem, and then we will analyse the conditions required to move the vergence point over an a-priori unknown surface within the field of view of the SCS.

In the following we shall assume that both cameras rotate about axes passing through their *optic centers*, and also that the SCS is calibrated. In particular, with reference to figure 3, the geometry of the SCS is defined by the left and right camera reference frames $\langle L \rangle$ and $\langle R \rangle$ defined as:

$$\langle L \rangle = \{i_L, j_L, k_L\} \quad (4.2)$$

$$\langle R \rangle = \{i_R, j_R, k_R\} \quad (4.3)$$

where the vectors k_L and k_R are the fixation axes of the two cameras, while the other axes are specified in the most convenient way³ We shall also make reference to a third arbitrary *head fixed* reference frame $\langle C \rangle$ defined as:

$$\langle C \rangle = \{i_C, j_C, k_C\} \quad (4.4)$$

³For the sake of clarity, we assume that when the cameras have parallel optic axes then j_L and j_R are parallel and point upward, while i_L and i_R are coincident and point to *left*.

The frames $\langle L \rangle$ and $\langle R \rangle$ are located at points \mathbf{P}_L and \mathbf{P}_R corresponding to the rotation centers (and optic centers) of the cameras, while frame $\langle C \rangle$ is placed at a generic point $\mathbf{0}$. Then, the geometric vector \mathbf{b} defined as:

$$\mathbf{b} = (\mathbf{P}_R - \mathbf{P}_L) \quad (4.5)$$

defines the *baseline of the SCS*; furthermore, the displacement of the cameras with respect to $\langle C \rangle$ is expressed by the following vectors:

$$\mathbf{b}_L = (\mathbf{P}_L - \mathbf{0}) \quad (4.6)$$

$$\mathbf{b}_R = (\mathbf{P}_R - \mathbf{0}) \quad (4.7)$$

which are constant in $\langle C \rangle$, as well as \mathbf{b} . According to the above assumption all the points belonging to the optic axes of the cameras will be expressed in the following way:

$$\mathbf{x}_L = \mathbf{b}_L + z_L \mathbf{k}_L \quad (4.8)$$

$$\mathbf{x}_R = \mathbf{b}_R + z_R \mathbf{k}_R \quad (4.9)$$

where $z_L > 0$ and $z_R > 0$ are the distances of the fixation points from \mathbf{P}_L and \mathbf{P}_R .

5 Vergence Control over a Surface \mathcal{S}

Consider the situation shown in figure 4. The two cameras are fixating two distinct points belonging to a common smooth surface \mathcal{S} and their optical axes might not geometrically verge. The goal is to control the rotation of the two cameras so that they verge on an a-priori unspecified point $\mathbf{x}^* \in \mathcal{S}$. In this sense the control strategy should be *reactive* i.e. the control signal should be computed on the basis of visual and motor feedback, but without *high level* specification of the target points or features in the images.

In particular, \mathbf{F}_L and \mathbf{F}_R are the fixation points associated to $\langle L \rangle$ and $\langle R \rangle$ respectively, while \mathbf{u}_R is the projection of \mathbf{F}_L in $\langle R \rangle$ and \mathbf{u}_L is the projection of \mathbf{F}_R in $\langle L \rangle$.

Remark 1 *From the previous definitions \mathbf{u}_R is the disparity of \mathbf{F}_L and \mathbf{u}_L is the disparity of \mathbf{F}_R . Remind that we are idealizing as points small image areas. Therefore, \mathbf{u}_R and \mathbf{u}_L could be intended as average disparities (in the left and right cameras) of regions close to the fovea.*

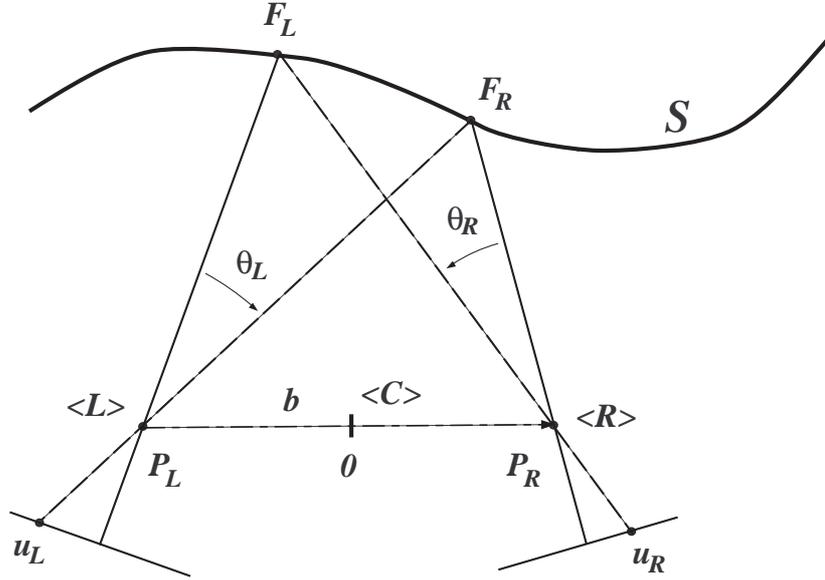


Figure 4: 2D sketch of the initial conditions of a SCS during a vergence task.

5.1 Specification of task errors

From a geometrical point of view the geometrical description of the task is to rotate $\langle L \rangle$ and $\langle R \rangle$ so that $\mathbf{F}_L \equiv \mathbf{F}_R$ so that also $\mathbf{u}_L \equiv \mathbf{u}_R$. Therefore, it is possible to define various *error* functions specifying the accomplishment of the vergence task. In fact from figure 4 we might consider:

1. $\mathbf{e} = (\mathbf{F}_R - \mathbf{F}_L)$

2. $\mathbf{e} = \begin{bmatrix} \mathbf{e}_L \\ \mathbf{e}_R \end{bmatrix}$ where

$$\mathbf{e}_L = (\mathbf{F}_L - \mathbf{P}_L) \times (\mathbf{F}_R - \mathbf{P}_L)$$

$$\mathbf{e}_R = (\mathbf{F}_R - \mathbf{P}_R) \times (\mathbf{F}_L - \mathbf{P}_R)$$

3. $\mathbf{e} = \begin{bmatrix} \mathbf{e}_L \\ \mathbf{e}_R \end{bmatrix}$ where

$$\mathbf{e}_L = \frac{(\mathbf{F}_L - \mathbf{P}_L)}{|\mathbf{F}_L - \mathbf{P}_L|} \times \frac{(\mathbf{F}_R - \mathbf{P}_L)}{|\mathbf{F}_R - \mathbf{P}_L|} \quad (5.1)$$

$$\mathbf{e}_R = \frac{(\mathbf{F}_R - \mathbf{P}_R)}{|\mathbf{F}_R - \mathbf{P}_R|} \times \frac{(\mathbf{F}_L - \mathbf{P}_R)}{|\mathbf{F}_L - \mathbf{P}_R|} \quad (5.2)$$

The first two proposed error functions are strictly related to *pure* 3D geometric information, while the third one can be directly expressed in term of image based data. In fact:

$$\mathbf{e}_L = -\mathbf{k}_L \times \frac{\mathbf{u}_L}{|\mathbf{u}_L|} = \sin \theta_L \quad (5.3)$$

$$\mathbf{e}_R = -\mathbf{k}_R \times \frac{\mathbf{u}_R}{|\mathbf{u}_R|} = \sin \theta_R \quad (5.4)$$

where the angles θ_L and θ_R are defined as

$$\theta_L = \widehat{\mathbf{F}_L \mathbf{P}_L \mathbf{F}_R} \quad (5.5)$$

$$\theta_R = \widehat{\mathbf{F}_R \mathbf{P}_R \mathbf{F}_L} \quad (5.6)$$

In the following we shall only consider the task errors specified by expressions (5.1) and (5.2) as they are the only computationally meaningful signals directly related to image coordinates.

5.2 Vergence Control Signals

In the following we will assume that the motion of the cameras could be controlled at kinematic level. Therefore, we will assume that the control signals correspond to the angular velocities assigned to the left $\langle L \rangle$ and right $\langle R \rangle$ frames, respectively ⁴. In order to compute the control signals we adopt a *Task Function* based approach [1]. To this aim let us define the following *Task Function*:

$$V = \frac{1}{2} |\mathbf{e}_L|^2 + \frac{1}{2} |\mathbf{e}_R|^2 = \frac{1}{2} \sin^2 \theta_L + \frac{1}{2} \sin^2 \theta_R \quad (5.7)$$

The time derivative of expression (5.7) can be expressed as:

$$\dot{V} = \mathbf{e}_L \cdot \dot{\mathbf{e}}_L + \mathbf{e}_R \cdot \dot{\mathbf{e}}_R \quad (5.8)$$

then, by using the definitions (5.1) and (5.2) \dot{V} can be rewritten as:

$$\dot{V} = \mathbf{e}_L \cdot \frac{d}{dt} \left[\mathbf{k}_L \times \frac{(\mathbf{F}_R - \mathbf{P}_L)}{|\mathbf{F}_R - \mathbf{P}_L|} \right] + \mathbf{e}_R \cdot \frac{d}{dt} \left[\mathbf{k}_R \times \frac{(\mathbf{F}_L - \mathbf{P}_R)}{|\mathbf{F}_L - \mathbf{P}_R|} \right] \quad (5.9)$$

By using the assumptions made on the kinematics of the SCS and taking into account the properties of the time derivatives of unit vectors (discussed in Appendix C), we can compute the time derivatives in expression (5.9) as:

⁴This assumption is quite common in the analysis of robot systems behavior and simplifies the analysis of the motion.

$$\begin{aligned}
& \frac{d}{dt} \left[\mathbf{k}_L \times \frac{(\mathbf{F}_R - \mathbf{P}_L)}{|\mathbf{F}_R - \mathbf{P}_L|} \right] = \\
& = -(\boldsymbol{\omega}_L \times \mathbf{k}_L) \times \bar{\mathbf{u}}_L + \frac{1}{|\mathbf{F}_R - \mathbf{P}_L|} \mathbf{k}_L \times \mathcal{P}_{\bar{\mathbf{u}}_L}^\perp \frac{d}{dt}(\mathbf{F}_R - \mathbf{P}_L) \quad (5.10)
\end{aligned}$$

$$\begin{aligned}
& \frac{d}{dt} \left[\mathbf{k}_R \times \frac{(\mathbf{F}_L - \mathbf{P}_R)}{|\mathbf{F}_L - \mathbf{P}_R|} \right] = \\
& = -(\boldsymbol{\omega}_R \times \mathbf{k}_R) \times \bar{\mathbf{u}}_R + \frac{1}{|\mathbf{F}_L - \mathbf{P}_R|} \mathbf{k}_R \times \mathcal{P}_{\bar{\mathbf{u}}_R}^\perp \frac{d}{dt}(\mathbf{F}_L - \mathbf{P}_R) \quad (5.11)
\end{aligned}$$

The control strategy required to achieve the condition of vergence onto the surface \mathcal{S} is to choose the control signals $\boldsymbol{\omega}_L$ and $\boldsymbol{\omega}_R$ such that $\dot{V} < 0$.

Remark 2 *The equalities (5.10) and (5.11) still involve time derivatives. As it will be shown in section 5.4, the time derivatives of $(\mathbf{F}_L - \mathbf{P}_R)$ and $(\mathbf{F}_R - \mathbf{P}_L)$ are related in involved form to $\boldsymbol{\omega}_R$ and $\boldsymbol{\omega}_L$, respectively. Which means that $\dot{\mathbf{e}}_L$ and $\dot{\mathbf{e}}_R$ are both functions of $\boldsymbol{\omega}_L$ and $\boldsymbol{\omega}_R$, i.e.*

$$\begin{aligned}
\dot{\mathbf{e}}_L &= -[(\boldsymbol{\omega}_L \times \mathbf{k}_L) \times \bar{\mathbf{u}}_L] + \mathbf{f}_L(\boldsymbol{\omega}_R) \\
\dot{\mathbf{e}}_R &= -[(\boldsymbol{\omega}_R \times \mathbf{k}_R) \times \bar{\mathbf{u}}_R] + \mathbf{f}_R(\boldsymbol{\omega}_L)
\end{aligned}$$

this means that the effect of the motion of eyes is coupled with respect to the relative orientation errors specified in (5.1) and (5.2).

5.3 Simplified Control Strategies

From the previous analysis we have found that \dot{V} could be also written as:

$$\begin{aligned}
\dot{V} &= -\dot{\mathbf{e}}_L \cdot [(\boldsymbol{\omega}_L \times \mathbf{k}_L) \times \bar{\mathbf{u}}_L] - \dot{\mathbf{e}}_R \cdot [(\boldsymbol{\omega}_R \times \mathbf{k}_R) \times \bar{\mathbf{u}}_R] \\
&\quad + \dot{\mathbf{e}}_L \cdot \mathbf{f}_L(\boldsymbol{\omega}_R) + \dot{\mathbf{e}}_R \cdot \mathbf{f}_R(\boldsymbol{\omega}_L) \quad (5.12)
\end{aligned}$$

Let us now assume that only the left camera is controlled to perform a rotation, while the right one is kept still. This is equivalent to say that we are only interested to control to zero only the *task error* \mathbf{e}_L . In particular consider a rotation of frame $\langle L \rangle$ such that the angle θ_L is decreasing. In this case \dot{V} could be expressed as:

$$\dot{V} = \mathbf{e}_L \cdot \dot{\mathbf{e}}_L = -\mathbf{e}_L \cdot [(\boldsymbol{\omega}_L \times \mathbf{k}_L) \times \bar{\mathbf{u}}_L] \quad (5.13)$$

in fact $\mathbf{f}_L = 0$ since we are assuming that the right camera is not moving, i.e. $\boldsymbol{\omega}_R = 0$. By exploiting the *cross products* in the expression above we obtain:

$$\dot{V} = -\mathbf{e}_L \cdot [(\boldsymbol{\omega}_L \times \mathbf{k}_L) \times \bar{\mathbf{u}}_L] = -(\boldsymbol{\omega}_L \times \mathbf{k}_L) \cdot [\bar{\mathbf{u}}_L \times \mathbf{e}_L] \quad (5.14)$$

By using the definition of *task error* defined in (5.3) the expression above can be expanded as:

$$\begin{aligned} \dot{V} &= (\boldsymbol{\omega}_L \times \mathbf{k}_L) \cdot [\bar{\mathbf{u}}_L \times (\mathbf{k}_L \times \bar{\mathbf{u}}_L)] = \\ &= (\boldsymbol{\omega}_L \times \mathbf{k}_L) \cdot [(\bar{\mathbf{u}}_L \cdot \bar{\mathbf{u}}_L)\mathbf{k}_L - (\bar{\mathbf{u}}_L \cdot \mathbf{k}_L)\bar{\mathbf{u}}_L] = \\ &= -(\bar{\mathbf{u}}_L \cdot \mathbf{k}_L) \bar{\mathbf{u}}_L \cdot (\boldsymbol{\omega}_L \times \mathbf{k}_L) = \\ &= -(\bar{\mathbf{u}}_L \cdot \mathbf{k}_L) (\mathbf{k}_L \times \bar{\mathbf{u}}_L) \cdot \boldsymbol{\omega}_L \end{aligned} \quad (5.15)$$

and eventually, again by (5.3) we obtain:

$$\dot{V} = (\bar{\mathbf{u}}_L \cdot \mathbf{k}_L) \mathbf{e}_L \cdot \boldsymbol{\omega}_L \quad (5.16)$$

In the above formula the coefficient $(\bar{\mathbf{u}}_L \cdot \mathbf{k}_L)$ is zero only in the *degenerate* case of $\mathbf{u}_L \rightarrow \infty$ (i.e. orthogonal to \mathbf{k}_L).

5.3.1 Continuous Decoupled Control

By choosing the control signal $\boldsymbol{\omega}_L$ as:

$$\boldsymbol{\omega}_L = -\gamma \frac{1}{(\bar{\mathbf{u}}_L \cdot \mathbf{k}_L)} \mathbf{e}_L \quad (5.17)$$

where γ is a positive constant feedback gain, then \dot{V} becomes:

$$\dot{V} = -\gamma \mathbf{e}_L \cdot \mathbf{e}_L < 0 \quad (5.18)$$

Then, control law (5.17) is a *continuous* feedback control signal which ensures the asymptotic convergence to zero of the *task error* \mathbf{e}_L .

Remark 3 *The control law (5.17) can also be reformulated as follows:*

$$\boldsymbol{\omega}_L = \gamma (\mathbf{k}_L \times \mathbf{u}_L) \quad (5.19)$$

In fact recall that $\bar{\mathbf{u}}_L = \mathbf{u}_L / |\mathbf{u}_L|$ and that $(\mathbf{u}_L \cdot \mathbf{k}_L) = 1$ (normalized focal length of the cameras).

An equivalent rationale could be applied assuming that the left camera is still while controlling the rotation of the right one. Therefore, we conjecture⁵ that two decoupled control signals of the form:

$$\boldsymbol{\omega}_L = \gamma_L (\mathbf{k}_L \times \mathbf{u}_L) \quad (5.20)$$

$$\boldsymbol{\omega}_R = \gamma_R (\mathbf{k}_R \times \mathbf{u}_R) \quad (5.21)$$

ensure the *asymptotic* convergence of the vergence point onto \mathcal{S} . Simulations reported in section 5.5 highlight the feature of this control algorithm.

Remark 4 *The proposed control scheme is equivalent to assume that each camera is rotated assuming the other is fixed. The speed of convergence and the final vergence point depend on the selection of the feedback gains γ_L and γ_R which establishes the relative speed rate of the two cameras. To achieve a symmetrical behavior of the system it is reasonable to assign the same feedback gain to both controllers.*

Remark 5 *The control law specified in (5.20) and (5.21) establishes the nominal angular speed of the cameras to accomplish the vergence task. In practice we must ensure that such a control law could be physically implementable. From (5.20) and (5.21) it is clear that the desired angular velocities must be parallel to the image planes of the cameras. This means that rotations of the cameras about their optic axes are specified. In fact, any rotation about \mathbf{k}_L or \mathbf{k}_R would affect the convergence properties⁶ of the algorithm. This means that the control law can be implemented by any robot eye system which could generate angular velocities which have components parallel to the image planes as defined by formulas (5.20) and (5.21). Therefore, any two degrees of freedom mechanism⁷ can implement the proposed control law (including pan-tilt cameras).*

5.3.2 Non-smooth Decoupled Control

Following the arguments of the previous section it is possible (in principle) to define another kind of feedback control algorithm. Assume again that only the left camera is rotated, from (5.16) we see that any control signal $\boldsymbol{\omega}_L$ such that $\mathbf{e}_L \cdot \boldsymbol{\omega}_L$ is negative for $\mathbf{e}_L \neq 0$ will ensure the convergence to zero of \mathbf{e}_L . In particular the largest is the magnitude of the control signals the highest is the convergence rate of the errors to zero.

⁵A formal proof of this property is not presently available. A simplified *geometric* argument is sketched in Appendix D.

⁶Under the assumption of motion of a single camera.

⁷Far from its singular configurations.

Let us assume that $|\boldsymbol{\omega}_L| \leq \Omega_{max}$ where Ω_{max} specifies the maximum admissible angular speed of the camera. Then \dot{V} can be made maximally negative by choosing $\boldsymbol{\omega}_L$ as:

$$\boldsymbol{\omega}_L = -\Omega_{max} \frac{\mathbf{e}_L}{|\mathbf{e}_L|} \text{sign}(\bar{\mathbf{u}}_L \cdot \mathbf{k}_L) \quad (5.22)$$

leading to

$$\dot{V} = -\Omega_{max} |\bar{\mathbf{u}}_L \cdot \mathbf{k}_L| |\mathbf{e}_L| \quad (5.23)$$

From (5.23) we have that $\dot{V} < 0$ for $|\mathbf{e}_L| \neq 0$ then $\sin \theta_L$ is decreasing so $|\bar{\mathbf{u}}_L \cdot \mathbf{k}_L| = \cos \theta_L$ is increasing. For any initial orientation of $\langle L \rangle$ and $\langle R \rangle$, assuming that at time $t = 0$ there is a *small* misalignment between \mathbf{k}_L and \mathbf{k}_R , we have:

$$\dot{V} \leq -\lambda \Omega_{max} |\mathbf{e}_L| \quad (5.24)$$

where $\lambda > 0$. The inequality above can be also rewritten as:

$$\dot{V} \leq -\lambda \Omega_{max} V^{1/2} \quad (5.25)$$

which can be integrated leading to:

$$|\mathbf{e}_L| < |\mathbf{e}_L(0)| - \frac{1}{2} \lambda \Omega_{max} t \quad (5.26)$$

Then \mathbf{e}_L will converge to 0 in finite time \bar{t} where:

$$\bar{t} < \frac{2 |\mathbf{e}_L(0)|}{\lambda \Omega_{max}} \quad (5.27)$$

By mimicking the reasoning from the previous section, we conjecture that this control law could be generalized to the case when both eye are controlled, in such a case we suggest the control laws

$$\boldsymbol{\omega}_L = -\Omega_{max} \frac{\mathbf{e}_L}{|\mathbf{e}_L|} \text{sign}(\bar{\mathbf{u}}_L \cdot \mathbf{k}_L) \quad (5.28)$$

$$\boldsymbol{\omega}_R = -\Omega_{max} \frac{\mathbf{e}_R}{|\mathbf{e}_R|} \text{sign}(\bar{\mathbf{u}}_R \cdot \mathbf{k}_R) \quad (5.29)$$

Remark 6 *Remarks 4 and 5 also apply to the control laws (5.28) and (5.29), which are decoupled as well as (5.20) and (5.21).*

5.4 Advanced Analysis

The control strategies proposed in the previous sections have been defined neglecting the two coupling terms $\mathbf{f}_L(\boldsymbol{\omega}_R)$ and $\mathbf{f}_R(\boldsymbol{\omega}_L)$ appearing in \dot{V} as shown in equation (5.12). For the sake of simplicity we shall refer in the following only to $\mathbf{f}_L(\boldsymbol{\omega}_R)$ and for symmetry we will generalize the results also to $\mathbf{f}_R(\boldsymbol{\omega}_L)$. From (5.10) $\mathbf{f}_L(\boldsymbol{\omega}_R)$ can be expressed as:

$$\mathbf{f}_L(\boldsymbol{\omega}_R) = \frac{1}{|\mathbf{F}_R - \mathbf{P}_L|} \mathbf{k}_L \times \mathcal{P}_{\mathbf{a}_L}^\perp \frac{d}{dt}(\mathbf{F}_R - \mathbf{P}_L) \quad (5.30)$$

and in order to exploit its structure we must compute the time derivative in the right hand side of the above equality. In particular, we have:

$$\begin{aligned} \frac{d}{dt}(\mathbf{F}_R - \mathbf{P}_L) &= \frac{d}{dt}(\mathbf{F}_R - \mathbf{P}_R + \mathbf{P}_R - \mathbf{P}_L) = \\ &= \frac{d}{dt}(\mathbf{F}_R - \mathbf{P}_R) + \frac{d}{dt}(\mathbf{P}_R - \mathbf{P}_L) = \frac{d}{dt}(z_R \mathbf{k}_R) + \frac{d}{dt} \mathbf{b} \end{aligned} \quad (5.31)$$

where z_R is the distance of the fixation point of the right camera from $\langle R \rangle$ and \mathbf{b} is the baseline. Then:

$$\frac{d}{dt}(\mathbf{F}_R - \mathbf{P}_L) = \dot{z}_R \mathbf{k}_R + z_R (\boldsymbol{\omega}_R \times \mathbf{k}_R). \quad (5.32)$$

We still have to compute \dot{z}_R which, as discussed below, depends on both $\boldsymbol{\omega}_R$ and on the shape of the surface \mathcal{S} . To this aim consider figures 5 and 6.

By assuming \mathcal{S} *piecewise smooth* then $d/dt (\mathbf{F}_R - \mathbf{P}_R)$ must be tangent to \mathcal{S} at \mathbf{F}_R , so it must be orthogonal to vector \mathbf{k}_s representing the normal to the surface at the fixation point; furthermore, from (5.32) it is a linear combination of vectors \mathbf{k}_R and $(\boldsymbol{\omega}_R \times \mathbf{k}_R)$. Therefore, $d/dt (\mathbf{F}_R - \mathbf{P}_R)$ must be orthogonal to both \mathbf{k}_s and $(\boldsymbol{\omega}_R \times \mathbf{k}_R) \times \mathbf{k}_R$. This can be also written as:

$$\frac{d}{dt}(\mathbf{F}_R - \mathbf{P}_R) = v \frac{\mathbf{k}_s \times [(\boldsymbol{\omega}_R \times \mathbf{k}_R) \times \mathbf{k}_R]}{|\mathbf{k}_s \times [(\boldsymbol{\omega}_R \times \mathbf{k}_R) \times \mathbf{k}_R]|} \quad (5.33)$$

where v takes into account the magnitude and direction of the derivative. By projecting the expressions (5.32) and (5.33) along the direction of $(\boldsymbol{\omega}_R \times \mathbf{k}_R)$ we obtain the following equality:

$$v \frac{\mathbf{k}_s \times [(\boldsymbol{\omega}_R \times \mathbf{k}_R) \times \mathbf{k}_R]}{|\mathbf{k}_s \times [(\boldsymbol{\omega}_R \times \mathbf{k}_R) \times \mathbf{k}_R]|} \cdot (\boldsymbol{\omega}_R \times \mathbf{k}_R) = z_R (\boldsymbol{\omega}_R \times \mathbf{k}_R) \cdot (\boldsymbol{\omega}_R \times \mathbf{k}_R) \quad (5.34)$$

The scalar product in the left hand side of the above equality can be simplified as follows:

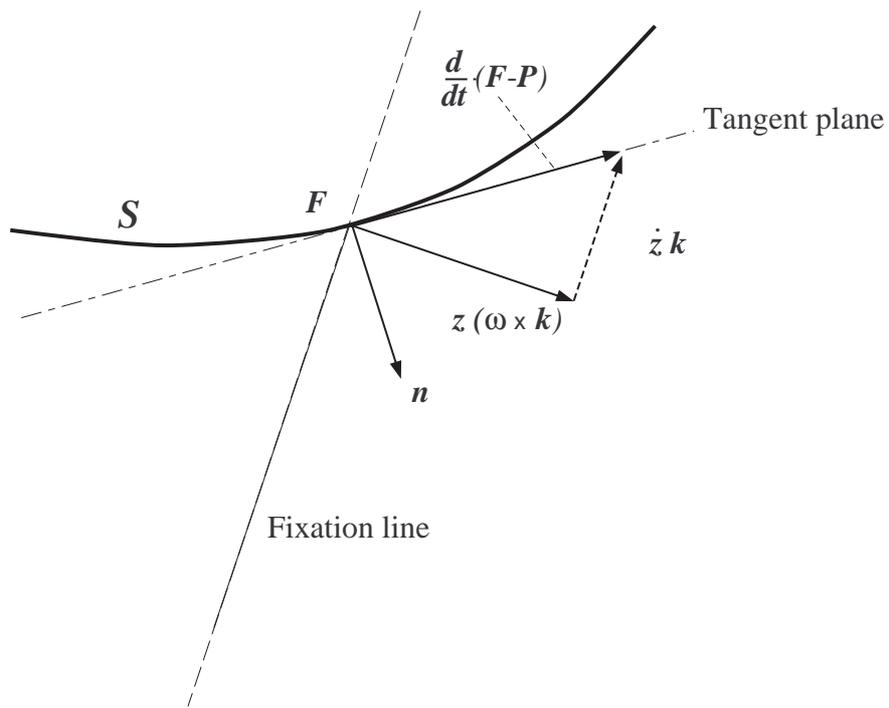
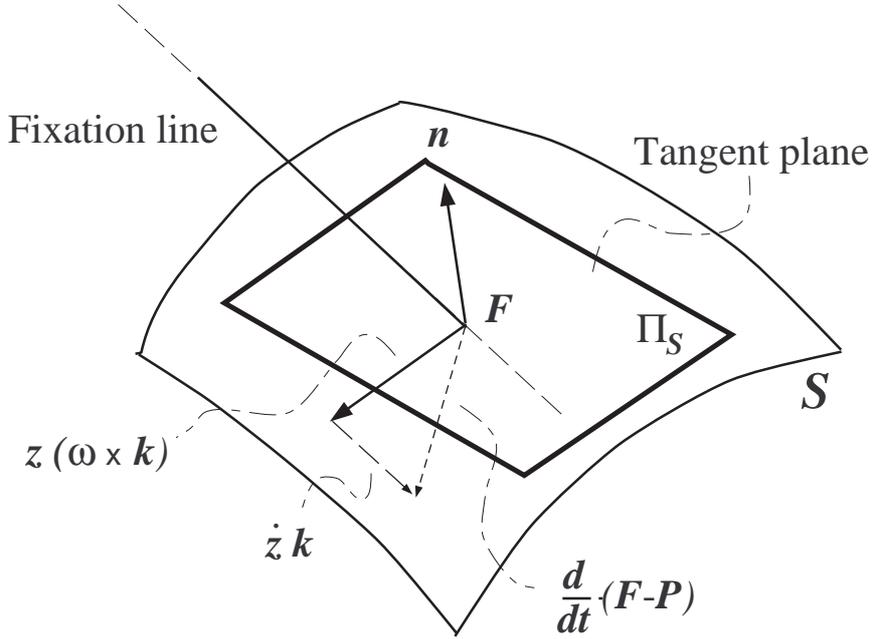


Figure 5: Geometric interpretation of \dot{z}_R : 2D sketch.

Figure 6: Geometric interpretation of z_R : 3D sketch.

$$\begin{aligned}
 & \{ \mathbf{k}_s \times [(\boldsymbol{\omega}_R \times \mathbf{k}_R) \times \mathbf{k}_R] \} \cdot (\boldsymbol{\omega}_R \times \mathbf{k}_R) = \\
 & = \{ [(\boldsymbol{\omega}_R \times \mathbf{k}_R) \times \mathbf{k}_R] \times (\boldsymbol{\omega}_R \times \mathbf{k}_R) \} \cdot \mathbf{k}_s = \\
 & = [(\boldsymbol{\omega}_R \times \mathbf{k}_R) \cdot (\boldsymbol{\omega}_R \times \mathbf{k}_R)] (\mathbf{k}_R \cdot \mathbf{k}_s)
 \end{aligned} \tag{5.35}$$

By using the right hand side of (5.35), the equality (5.33) can be solved with respect to v as follows:

$$v = z_R \frac{|\mathbf{k}_s \times [(\boldsymbol{\omega}_R \times \mathbf{k}_R) \times \mathbf{k}_R]|}{(\mathbf{k}_R \cdot \mathbf{k}_s)}. \tag{5.36}$$

From the formula above we have that always $v \geq 0$ since at any *visible* fixation point F_R the normal to the surface \mathcal{S} must have a component pointing toward the camera. Then, by using (5.32), (5.33) and (5.36) we can write:

$$\begin{aligned}
\dot{z}_R &= \mathbf{k}_R \cdot \frac{d}{dt}(\mathbf{F}_R - \mathbf{P}_R) = \\
&= v \frac{\mathbf{k}_R \cdot \{\mathbf{k}_s \times [(\boldsymbol{\omega}_R \times \mathbf{k}_R) \times \mathbf{k}_R]\}}{|\mathbf{k}_s \times [(\boldsymbol{\omega}_R \times \mathbf{k}_R) \times \mathbf{k}_R]|} = \\
&= v \frac{\mathbf{k}_s \cdot \{[(\boldsymbol{\omega}_R \times \mathbf{k}_R) \times \mathbf{k}_R] \times \mathbf{k}_R\}}{|\mathbf{k}_s \times [(\boldsymbol{\omega}_R \times \mathbf{k}_R) \times \mathbf{k}_R]|} = \\
&= v \frac{\mathbf{k}_s \cdot \{[(\boldsymbol{\omega}_R \cdot \mathbf{k}_R)\mathbf{k}_R - (\mathbf{k}_R \cdot \mathbf{k}_R)\boldsymbol{\omega}_R] \times \mathbf{k}_R\}}{|\mathbf{k}_s \times [(\boldsymbol{\omega}_R \times \mathbf{k}_R) \times \mathbf{k}_R]|} = \\
&= v \frac{\mathbf{k}_s \cdot (\boldsymbol{\omega}_R \times \mathbf{k}_R)}{|\mathbf{k}_s \times [(\boldsymbol{\omega}_R \times \mathbf{k}_R) \times \mathbf{k}_R]|} \tag{5.37}
\end{aligned}$$

By substituting expression (5.36) into the above formula we obtain:

$$\dot{z}_R = -z_R \frac{(\mathbf{k}_R \times \mathbf{k}_s) \cdot \boldsymbol{\omega}_R}{(\mathbf{k}_R \cdot \mathbf{k}_s)} \tag{5.38}$$

which relates the rate of change of z_R to the control signals and to the local shape of the surface \mathcal{S} .

5.5 Vergence Control Simulations

In this section we present a few simulation results which show the effectiveness of the proposed control strategy based on (5.20) and (5.21). In all the following tests we have assumed an ideal independent *pan-tilt* kinematic model for both cameras⁸

Test 1 The cameras are looking at a vertical plane passing through a point at a distance of 0.35 m and rotated about the y -axis by 45 deg. The cameras have a common feedback gain $\gamma = 5$ and initial conditions $(q_L^1, q_L^2) = (q_R^1, q_R^2) = (0, 0)$ (where q_j^1 is the tilt angle, and q_j^2 is the pan angle) such that the optic axes are parallel. Figures 7 and 8 show the transient response for the left and right cameras. Please note that since the cameras have the same elevation at time $t = 0$ the task is a pure vergence control.

Test 2 The cameras are looking at the same plane as in Test 1 under the same operating conditions except for the initial conditions $(q_L^1, q_L^2) =$

⁸All the simulations proposed in this section and in section 6.4 have been performed using a kinematic simulator implemented using MATLAB and SIMULINK. Short video clips showing the simulated behavior of the SCS are available in the EYESHOTS web-site along with this report.

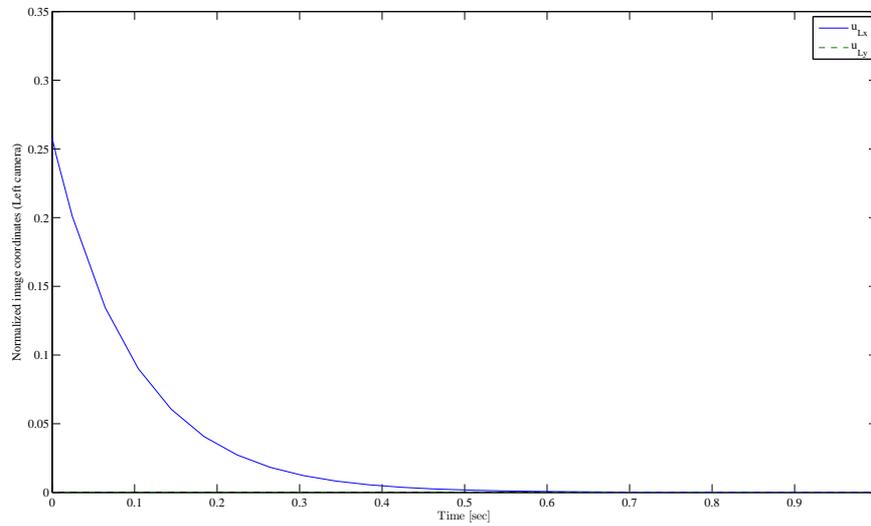


Figure 7: Transient response of the smooth control for the left camera (Test 1)

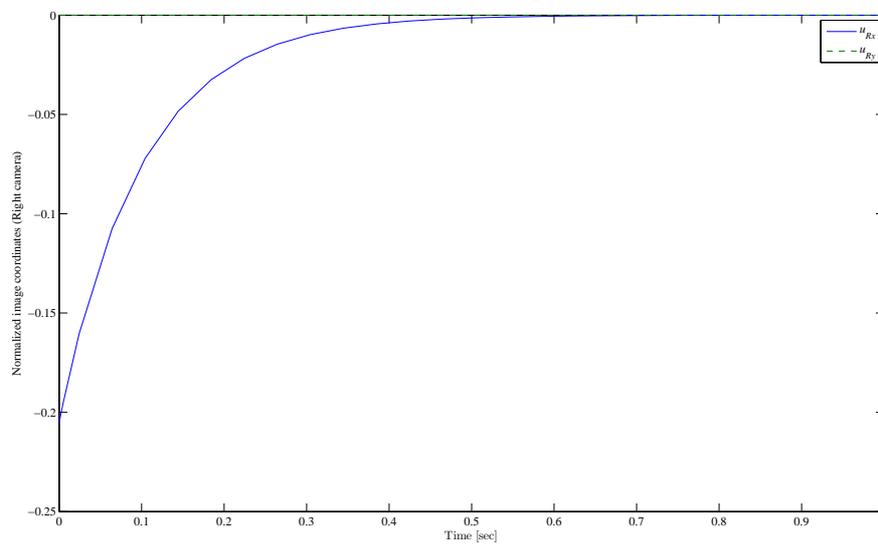


Figure 8: Transient response of the smooth control for the right camera (Test 1)

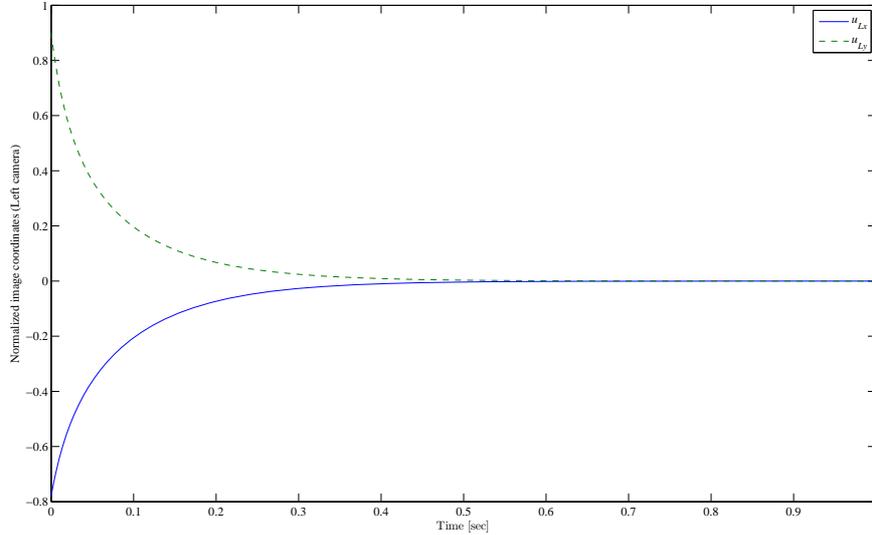


Figure 9: Transient response of the smooth control for the left camera (Test 2)

$(-\pi/32, -\pi/6)$ and $(q_R^1, q_R^2) = (\pi/6, \pi/12)$. The transient response is shown in figures 9 and 10. In this case it is clear that an initial *vertical disparity* is present; therefore, the task is not a pure vergence control as a correction of the elevation is also performed.

Test 3 Same condition as in the previous test except that now the cameras are looking at a sphere of radius 0.25 m and centered at the point ${}^C\mathbf{x}_0 = (-0.1, -0.1, 0.5)^T$. The transient response is shown in figures 11 and 12. Also, in this case the the controller performs simultaneously a vergence control and an elevation adjustment.

6 Surface Tracking Under Vergence Conditions

In this section we will analyse the conditions required to *move* the vergence point over a *locally smooth* surface \mathcal{S} . The rationale is as follows. We first assume that the SCS is verging on a generic point \mathbf{x}^* (expressed with respect to frame $\langle C \rangle$) and study the conditions required to *move* \mathbf{x}^* in 3D space⁹.

⁹This preliminary part of the analysis is clearly not meaningful from the point of view of the active vision or visual servoing paradigms, however it is technically useful to address

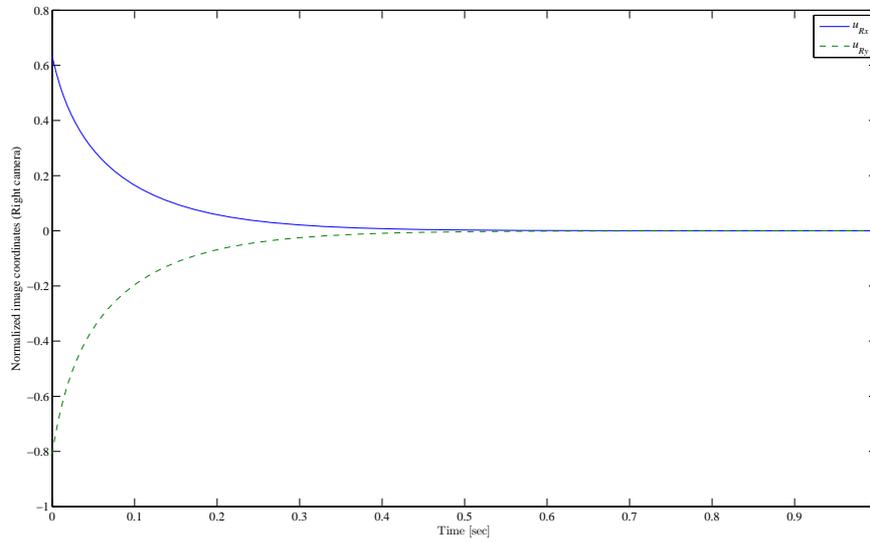


Figure 10: Transient response of the smooth control for the right camera (Test 2)

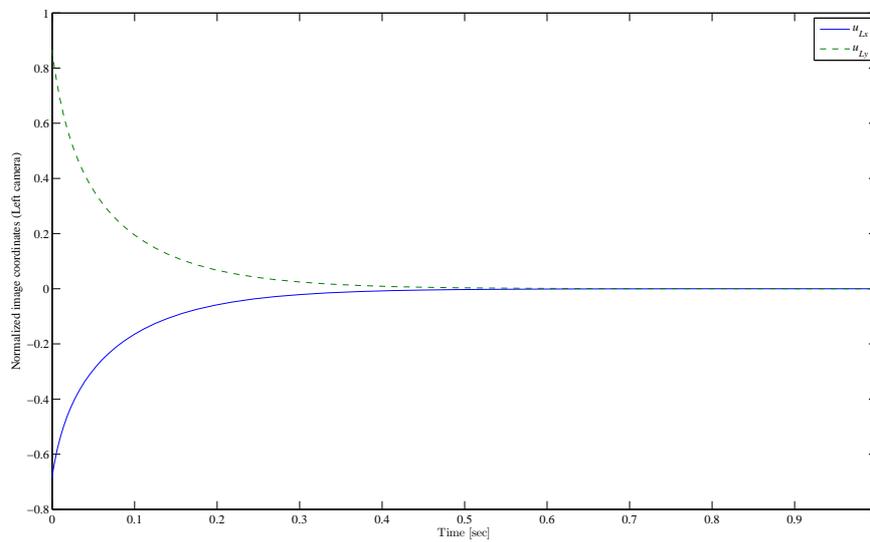


Figure 11: Transient response of the smooth control for the left camera (Test 3)

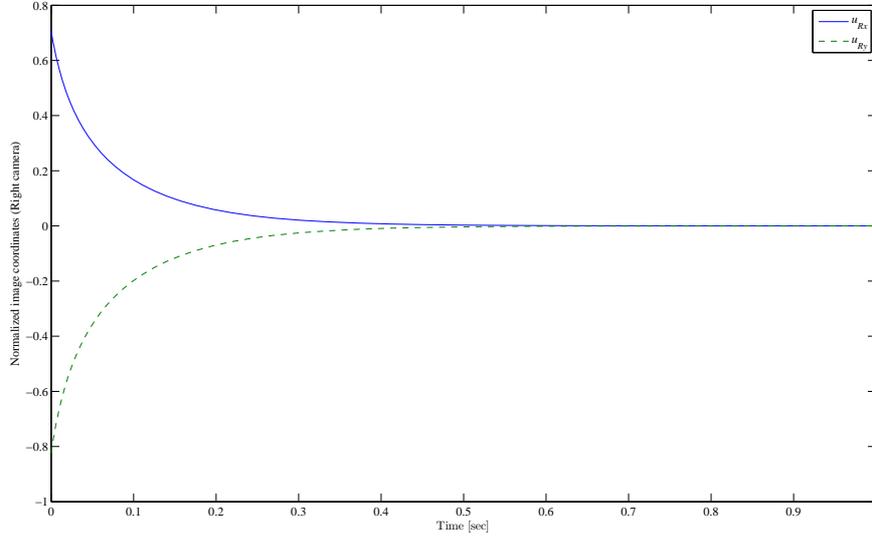


Figure 12: Transient response of the smooth control for the right camera (Test 3)

Then, we find a particular solution to the general problem leading to a *nominal* trajectory of \mathbf{x}^* along \mathcal{S} . Finally, we shall discuss some possible methods to exploit from image data the quantities required to implement the proposed control scheme¹⁰.

6.1 Unconstrained Vergence Tracking

This section specifies the nominal motion conditions required to keep the vergence on a point \mathbf{x}^* freely moving in 3D space.

We assume that at time $t = 0$ the SCS is verging on some point \mathbf{x}^* , so that the fixation points must coincide, i.e. $(\mathbf{F}_L - \mathbf{0}) = (\mathbf{P}_R - \mathbf{0}) = \mathbf{x}^*$. Then, according to expressions (4.8) and (4.9) the following equalities must hold for some values of $z_L > 0$ and $z_R > 0$:

$$\mathbf{x}^* = \mathbf{b}_L + z_L/\mathbf{k}_L \quad (6.1)$$

$$\mathbf{x}^* = \mathbf{b}_R + z_R/\mathbf{k}_R \quad (6.2)$$

the actual tracking control problem.

¹⁰The validation of these procedures using *real* vision systems is not within the scope of this deliverable. The goal here is to evaluate the technical feasibility of the method (also using simulative tests).

The above system of equations can be also rewritten as:

$$\mathbf{0} = (\mathbf{b}_R - \mathbf{b}_L) + [z_R \mathbf{k}_R - z_L \mathbf{k}_L] \quad (6.3)$$

$$2\mathbf{x}^* = (\mathbf{b}_R + \mathbf{b}_L) + [z_R \mathbf{k}_R + z_L \mathbf{k}_L] \quad (6.4)$$

Let assume for the sake of simplicity that:

$$(\mathbf{b}_R + \mathbf{b}_L) = \mathbf{0} \quad (6.5)$$

i.e. $\mathbf{0}$ is assumed to be the mid-point of the baseline. Then, by (4.5) and (6.5) the equations (6.3) and (6.4) can be rewritten as:

$$z_R \mathbf{k}_R - z_L \mathbf{k}_L = \mathbf{b} \quad (6.6)$$

$$\mathbf{x}^* = \frac{1}{2} [z_R \mathbf{k}_R + z_L \mathbf{k}_L] \quad (6.7)$$

Equation (6.6) is necessary and sufficient for the vergence condition, while (6.7) specifies the actual vergence point in 3D space.

The system (6.6) can be solved for z_L and z_R , if and only if the vectors \mathbf{k}_L , \mathbf{k}_R and \mathbf{b} are co-planar, i.e.:

$$(\mathbf{k}_R \times \mathbf{k}_L) \cdot \mathbf{b} = 0 \quad (6.8)$$

and $(\mathbf{k}_R \times \mathbf{k}_L) \neq \mathbf{0}$. Under these conditions the vectors:

$$\left\{ \mathbf{k}_R, \mathbf{k}_L, \frac{\mathbf{k}_R \times \mathbf{k}_L}{|\mathbf{k}_R \times \mathbf{k}_L|} \right\} \triangleq \{ \mathbf{k}_R, \mathbf{k}_L, \mathbf{n} \} \quad (6.9)$$

form a *moving* reference frame common to both cameras. Further details related to the solution of system (6.6) are discussed in Appendix E.

Remark 7 *It is worth noting that $|\mathbf{k}_R \times \mathbf{k}_L| = |\sin \alpha|$ being α the vergence angle of the SCS.*

Let us now compute the time derivative in frame $\langle C \rangle$ of equations (6.3) and (6.4), then we obtain:

$$\dot{\mathbf{x}}^* = \dot{z}_L \mathbf{k}_L + z_L (\boldsymbol{\omega}_L \times \mathbf{k}_L) \quad (6.10)$$

$$\dot{\mathbf{x}}^* = \dot{z}_R \mathbf{k}_R + z_R (\boldsymbol{\omega}_R \times \mathbf{k}_R) \quad (6.11)$$

Consider now the following local camera frames:

$$\langle L' \rangle = \{ \mathbf{k}_L, \mathbf{n}, (\mathbf{k}_L \times \mathbf{n}) \} \quad (6.12)$$

$$\langle R' \rangle = \{ \mathbf{k}_R, \mathbf{n}, (\mathbf{k}_R \times \mathbf{n}) \} \quad (6.13)$$

By projecting equations (6.10) onto frame $\langle L' \rangle$ and (6.11) onto $\langle R' \rangle$ we obtain:

$$\begin{aligned} \mathbf{k}_L \cdot \dot{\mathbf{x}}^* &= \dot{z}_L \\ \mathbf{n} \cdot \dot{\mathbf{x}}^* &= z_L (\mathbf{k}_L \times \mathbf{n}) \cdot \boldsymbol{\omega}_L \\ (\mathbf{k}_L \times \mathbf{n}) \cdot \dot{\mathbf{x}}^* &= z_L [\mathbf{k}_L \times (\mathbf{k}_L \times \mathbf{n})] \cdot \boldsymbol{\omega}_L \end{aligned} \quad (6.14)$$

and

$$\begin{aligned} \mathbf{k}_R \cdot \dot{\mathbf{x}}^* &= \dot{z}_R \\ \mathbf{n} \cdot \dot{\mathbf{x}}^* &= z_R (\mathbf{k}_R \times \mathbf{n}) \cdot \boldsymbol{\omega}_R \\ (\mathbf{k}_R \times \mathbf{n}) \cdot \dot{\mathbf{x}}^* &= z_R [\mathbf{k}_R \times (\mathbf{k}_R \times \mathbf{n})] \cdot \boldsymbol{\omega}_R \end{aligned} \quad (6.15)$$

From the definitions of $\langle L' \rangle$ and $\langle R' \rangle$ we have that $[\mathbf{k}_L \times (\mathbf{k}_L \times \mathbf{n})] = [\mathbf{k}_R \times (\mathbf{k}_R \times \mathbf{n})] = -\mathbf{n}$. Therefore, we can compute the *feedforward* control angular velocities required to achieve (nominally) the specified tracking task:

$${}^{L'}\boldsymbol{\omega}_L = \frac{1}{z_L} \begin{bmatrix} 0 \\ (\mathbf{n} \times \mathbf{k}_L) \cdot \dot{\mathbf{x}}^* \\ \mathbf{n} \cdot \dot{\mathbf{x}}^* \end{bmatrix} \quad (6.16)$$

$${}^{R'}\boldsymbol{\omega}_R = \frac{1}{z_R} \begin{bmatrix} 0 \\ (\mathbf{n} \times \mathbf{k}_R) \cdot \dot{\mathbf{x}}^* \\ \mathbf{n} \cdot \dot{\mathbf{x}}^* \end{bmatrix} \quad (6.17)$$

Remark 8 Notice that any rotation of the cameras about the corresponding optic axes does not affect the motion of \mathbf{x}^* . Therefore, $\boldsymbol{\omega}_L$ and $\boldsymbol{\omega}_R$ can have any component along the axes \mathbf{k}_L and \mathbf{k}_R . For the sake of simplicity in the next formulas we shall set to zero these components. However, in practice the rotation kinematics of the cameras could introduce torsional¹¹ components which, as stated above, do not modify the properties of the computed angular velocities.

6.2 Constrained Vergence Control

On the basis of the previous discussion the conditions which keep the point \mathbf{x}^* sliding over a *locally smooth* surface \mathcal{S} can be now determined. In this case $\dot{\mathbf{x}}^*$ is no more arbitrary, but is constrained to be tangent to \mathcal{S} at any time, see figure 13.

¹¹By *torsional* component we intend here the component of the angular velocity about the fixation axis. This definition is by no way well posed in the literature. As a matter of fact the torsional component is referred to the axis \mathbf{k}_C in the statement of *Listing's Law*.

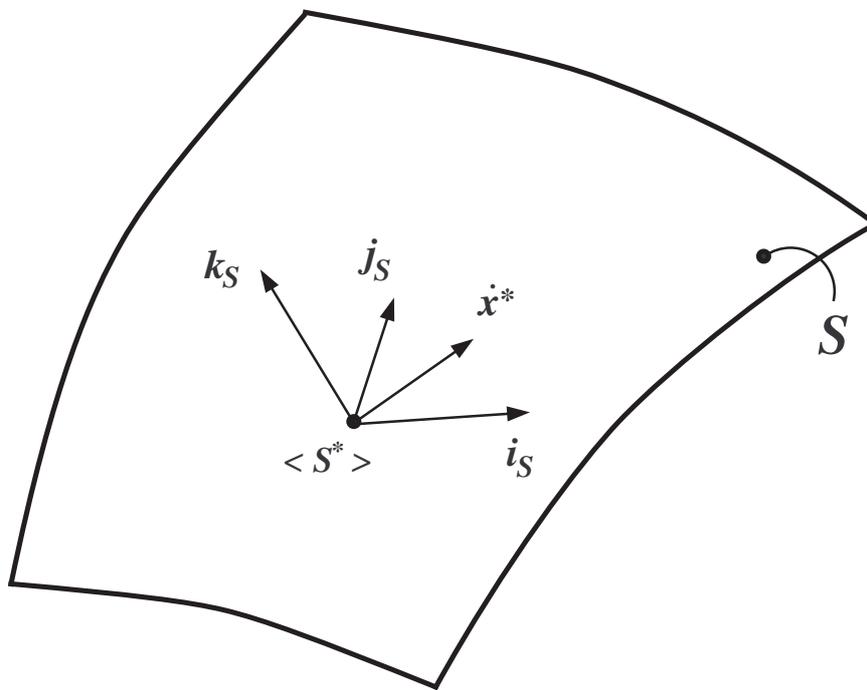


Figure 13: Constrained motion of \boldsymbol{x}^* along the surface \mathcal{S} .

The frame $\langle S^* \rangle = \{\mathbf{i}_S, \mathbf{j}_S, \mathbf{k}_S\}$ defines a local coordinate system moving with $\mathbf{x}^* \in \mathcal{S}$. The vector \mathbf{k}_S was already introduced in section 5.4

In fact if $\mathbf{x}^* \in \mathcal{S}$, from (4.1) the following equality must hold:

$$\dot{S}(\mathbf{x}^*) = 0 \quad \iff \quad \left. \frac{\partial S}{\partial \mathbf{x}} \right|_{\mathbf{x}=\mathbf{x}^*} \cdot \dot{\mathbf{x}}^* = 0 \quad (6.18)$$

then

$$\mathbf{k}_S = \frac{\frac{\partial S}{\partial \mathbf{x}}}{\left| \frac{\partial S}{\partial \mathbf{x}} \right|} \Bigg|_{\mathbf{x}=\mathbf{x}^*} \quad (6.19)$$

and therefore $\mathbf{x}^* \in \text{span}(\mathbf{i}_S, \mathbf{j}_S)$.

Let us now assume we are given a generic *velocity command* in 3D, and call it $\dot{\mathbf{x}}_0^*$, then we may compute an admissible reference velocity command as follows (see Appendix E):

$$\dot{\mathbf{x}}^* = \mathcal{P}_{\mathbf{k}_S}^\perp \dot{\mathbf{x}}_0^* \quad (6.20)$$

By assuming that all the vectors are expressed with respect to a known, common reference frame, e.g. $\langle C \rangle$, the above formula can be also expressed as

$${}^C \dot{\mathbf{x}}^* = [\mathbf{I} - {}^C \mathbf{k}_S {}^C \mathbf{k}_S^T] {}^C \dot{\mathbf{x}}_0^* \quad (6.21)$$

6.3 Computing the vector \mathbf{k}_S

In order to compute \mathbf{k}_S the following procedure is proposed. Consider the scenario sketched in figure 14.

Let \mathbf{P}'_L and \mathbf{P}'_R be the projections of the the fixation points respectively of the left camera and right cameras on the corresponding image planes Π_L and Π_R . Consider an *image segment* on Π_L and call it γ_L , then γ_L is the projection of a segment of a curve $\gamma_S \in \mathcal{S}$. By construction γ_S is (at least locally) a regular planar curve and its osculatory plane Π_{γ_L} is defined by the orientation of $\gamma_L \in \Pi_L$ and \mathbf{k}_L .

Consider now the projection of γ_S in the camera frame $\langle R \rangle$. This is a segment of a curve γ_R , which, under vergence conditions over \mathcal{S} , passes through \mathbf{P}'_R .

Let \mathbf{t}_S be the tangent vector to γ_S at \mathbf{x}^* ; then, $\mathbf{t}_S \in T$, where T is the tangent plane to \mathcal{S} at \mathbf{x}^* . The line passing through \mathbf{x}^* and directed as \mathbf{t}_S is projected in $\langle L \rangle$ and $\langle R \rangle$ with two lines parallel to the vectors \mathbf{t}_L and \mathbf{t}_R passing through \mathbf{P}'_L and \mathbf{P}'_R , respectively.

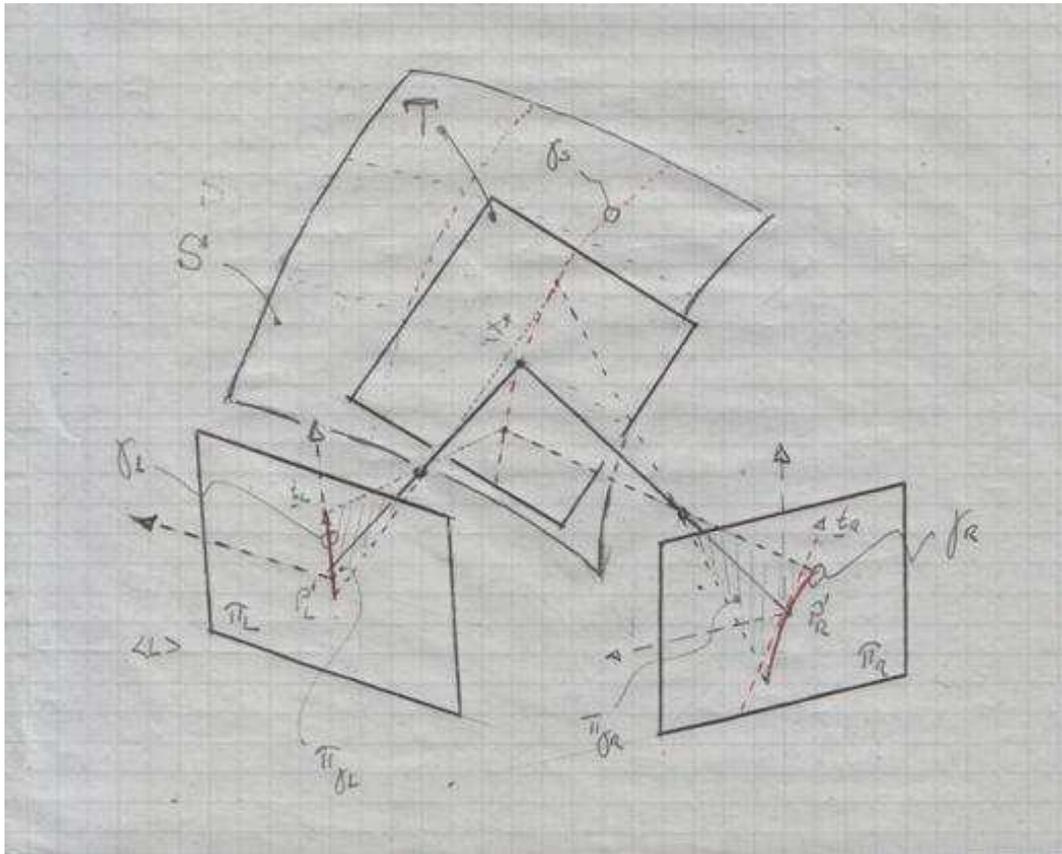


Figure 14: Correspondence of curves passing through $x^* \in \mathcal{S}$.

Following the previous rationale it is clear that \mathbf{t}_L specifies the direction of the line γ_L and is known, since γ_L is an arbitrary selected line in Π_L . On the other hand, the vector \mathbf{t}_R is unknown and should be computed by estimating the tangent of γ_R at point \mathbf{P}'_R .

Remark 9 *Computing γ_R is a specific image processing task (while γ_L is given).*

Once \mathbf{t}_L and \mathbf{t}_R are given, \mathbf{t}_S can be computed as the intersection of planes Π_{γ_L} and Π_{γ_R} formally defined as:

$$\mathbf{n}_L \cdot (\mathbf{P} - \mathbf{P}_L) = 0 \quad (6.22)$$

$$\mathbf{n}_R \cdot (\mathbf{P} - \mathbf{P}_R) = 0 \quad (6.23)$$

where

$$\mathbf{n}_L \triangleq (\mathbf{k}_L \times \mathbf{t}_L) \quad (6.24)$$

$$\mathbf{n}_R \triangleq (\mathbf{k}_R \times \mathbf{t}_R) \quad (6.25)$$

Then

$$\mathbf{t}_S = \frac{\mathbf{n}_R \times \mathbf{n}_L}{|\mathbf{n}_R \times \mathbf{n}_L|} \quad (6.26)$$

The procedure described above is *numerical* and should be repeated for different selections of orientations of \mathbf{t}_L ¹². In this way it is possible to compute a set of vectors $\{\mathbf{t}_s^1 \mathbf{t}_s^2 \dots \mathbf{t}_s^m\}$ which (at least nominally) must span the plane T .

Therefore, in order to compute \mathbf{k}_S we must find the *common normal* to all the vectors \mathbf{t}_s^i . A possible method is to find \mathbf{k}_S as the vector (of unit norm) which minimizes the following cost function:

$$J = \sum_{i=1}^m (\mathbf{k}_S \cdot \mathbf{t}_s^i)^2 \quad (6.27)$$

By projecting all the vectors \mathbf{t}_s^i on a common reference frame, e.g. $\langle C \rangle$, we can rewrite the cost function (6.27) as:

¹²The same rationale can be applied by selecting straight segments \mathbf{t}_R passing through the origin of the right camera, and repeating the procedure starting from $\langle R \rangle$.

$$\begin{aligned}
J &= \sum_{i=1}^m {}^C \mathbf{k}_S^T {}^C \mathbf{t}_S^i ({}^C \mathbf{t}_S^i)^T {}^C \mathbf{k}_S = \\
&= {}^C \mathbf{k}_S^T \left[\sum_{i=1}^m {}^C \mathbf{t}_S^i ({}^C \mathbf{t}_S^i)^T \right] {}^C \mathbf{k}_S = \\
&= {}^C \mathbf{k}_S^T {}^C \mathbf{T}_S {}^C \mathbf{k}_S
\end{aligned} \tag{6.28}$$

In the above formula the matrix ${}^C \mathbf{T}_S$ is a symmetric semi-positive definite matrix; hence, \mathbf{k}_S can be computed as the eigenvector of ${}^C \mathbf{T}_S$ associated with its least eigenvalue¹³.

6.4 Tracking Control Simulations

In this section we present a few simulation results which show the effectiveness of the proposed feedforward control strategy (6.16) and (6.17) with and without projection of the reference velocity as shown in equation (6.21). Where not otherwise specified, the simulations have been performed under the conditions discussed in section 5.5.

Test 4 The cameras are fixating the point reached at the end of the task described in Test 1. The initial conditions of the cameras joint angles are $(q_L^1, q_L^2) = (0, -1.282 \times 10^{-1})$ and $(q_R^1, q_R^2) = (0, 1.023 \times 10^{-1})$, while $\gamma = 1$. The reference velocity command \mathbf{x}_0^* has a magnitude of 0.25 *m/sec*, components only the $x - y$ plane and it generates an 8-shaped trajectory. The response errors of the system using the projection algorithm discussed in section 6.2 and 6.3 are shown in figures 15 and 16. The magnitude of the tracking errors is virtually 0 being well below the machine accuracy limits.

Test 5 The cameras perform the same task discussed above, but the projection algorithm of section 6.2 is not applied in this case. The response errors of the system are shown in figures 17 and 18. In this case the tracking errors are significant and related to the slantness of the target plane where the eyes try to keep the vergence during the tracking.

7 Torsional ocular posture

The control model described in the previous sections does not involve any torsional action of the cameras. Here, we consider the possible effects of

¹³Standard SVD decomposition methods are appropriate for this task.

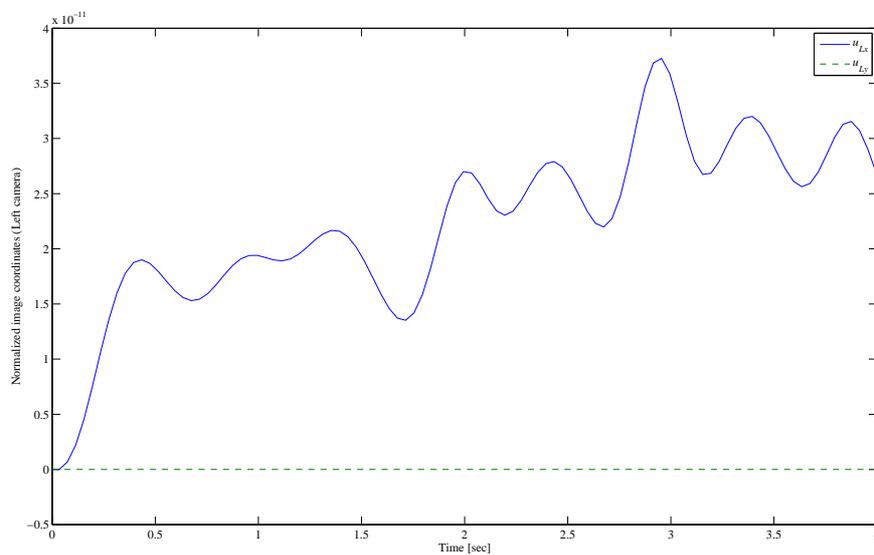


Figure 15: Transient response of the smooth control for the left camera (Test 4)

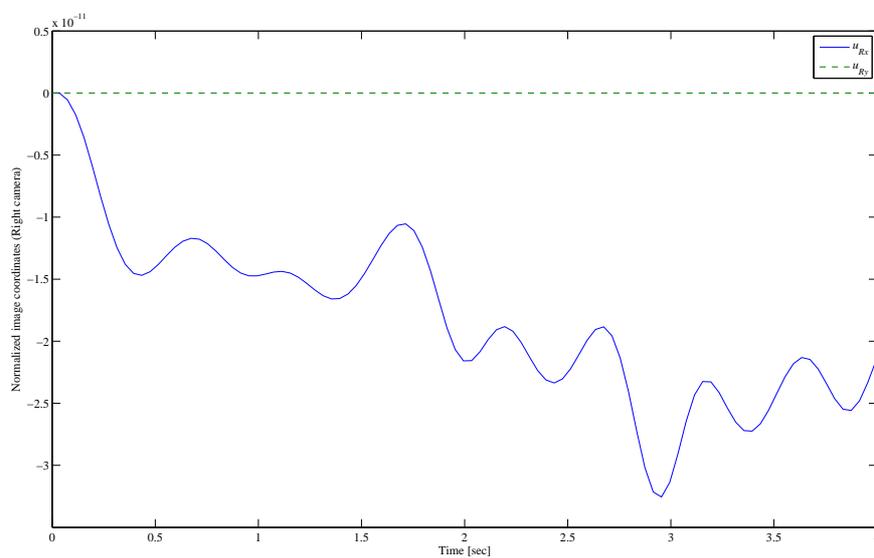


Figure 16: Transient response of the smooth control for the right camera (Test 4)

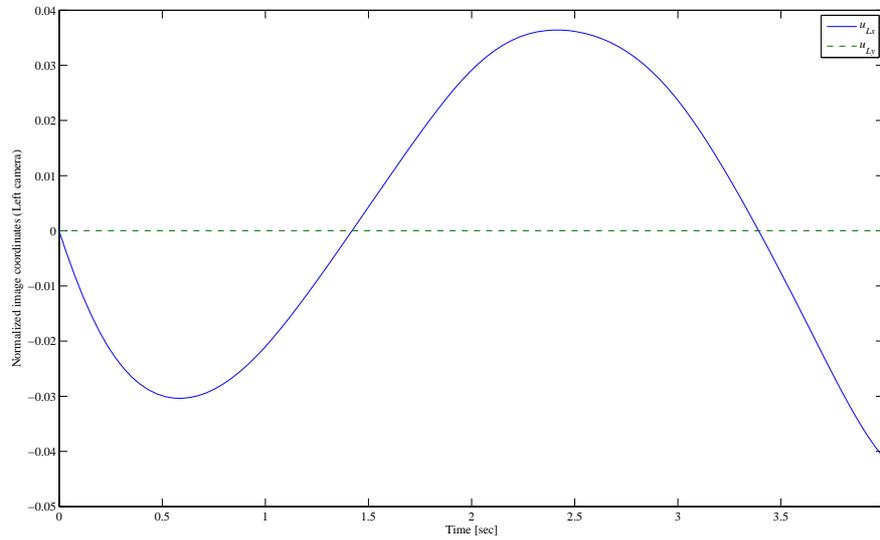


Figure 17: Transient response of the smooth control for the left camera (Test 5)

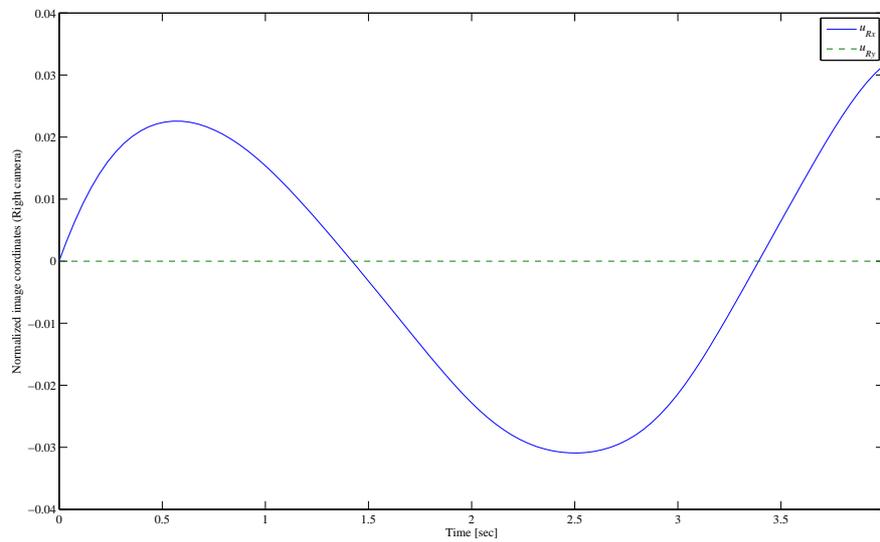


Figure 18: Transient response of the smooth control for the right camera (Test 5)

torsional ocular postures as two additional degrees of freedom of the stereo camera pair, for any given vergence point.

7.1 The Listing's Law

In theory the eye could assume an infinite number of torsional positions for any gaze direction. Thus, there are infinitely many ways to fixate any given target. Donders discovered that, for steady fixation with the head upright, the actual positions of the eye are restricted in a way that there is only one eye position for every gaze direction [2]. In other words, Donders found that the eye is restricted to a two-dimensional subspace of the three-dimensional space of all possible orientations. He observed that there is only one torsional eye position for each combination of horizontal and vertical eye positions, and postulated that the torsional position of the eye is always the same, independent of how the eye reaches a particular gaze direction. Listing's law goes one step further, by specifying the amount of ocular torsion. It states that, when the head is fixed, there is an eye position called primary position, such that 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 primary position [3]. In other words, one can consider any given eye movement as caused by rotation about an axis. The collection of these axes for all rotations that start from the primary position constitutes Listing's plane, see figure 19.

This is valid if we consider a movement from and to the primary position. What happens if the eye starts its rotation from an eccentric eye position? The orientation of the eye is still determined by rotation about axes that lie in a plane (irrespective of the direction of movement), but this plane is no longer orthogonal to the line of sight; instead, it is tilted in the same direction as the line of sight but only half as much [3] [4], see figure 20. To summarize, Listing's law can be expressed in terms of any initial eye position, not only primary position. In this form the law 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 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).

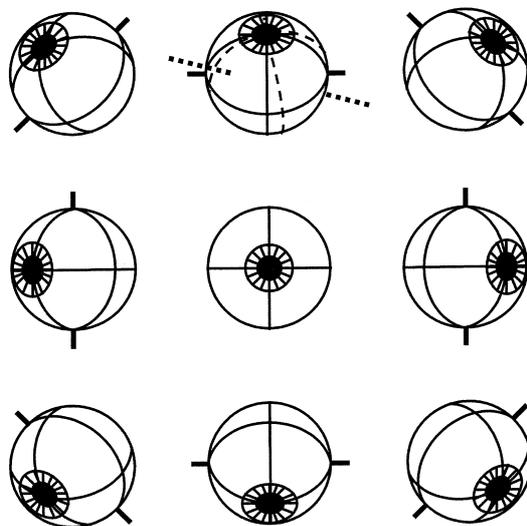


Figure 19: The nine orientations drawn in solid line according to Listing's Law: they are obtained by rotation to these positions from the primary position, about axes (thick solid lines) that lie on the Listing's plane (in this case represented by the paper plane). The position drawn in dashed lines at the top center does not obey Listing's Law, because the rotation to this position from primary position occurs about an axis (thick dotted line) that is tilted out the paper plane.

7.2 Binocular Listing's Law

Listing's law applies when the eye fixates a target at optical infinity. However, the torsional position of the eye changes when the eyes converge on a near object [5; 6; 7; 8; 9; 10; 11]. During convergence, the orientation of each eye is still determined by rotation about the axes that lie in a plane; however, this plane is rotated temporally and roughly symmetrically in each eye through an angle proportional to the vergence. These convergence-dependent changes of torsional position (i.e., orientation of Listing's plane) have been referred to as the binocular extension of Listing's law or L2. Note that L2 is a generalization of the original monocular Listing's Law, and reduces to it when the vergence angle is zero, as it occurs when the eyes fixate 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 change when vergence changes. The more the convergence there is, the more the temporal rotation of the plane there is, meaning that during convergence, there is a relative excyclotorsion on upgaze, and a rel-

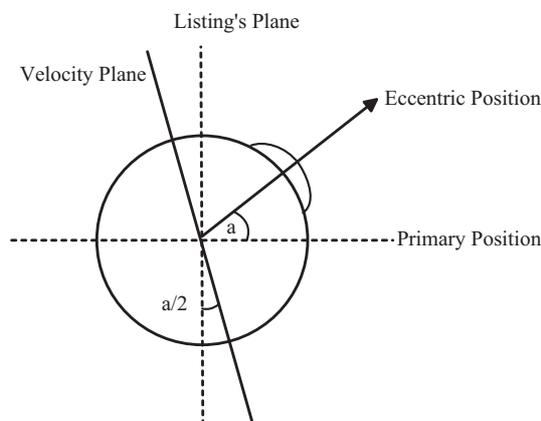


Figure 20: Listing’s half angle rule. The horizontal dashed line represents the line of sight when the eye is in primary position, and the vertical dashed line represents Listing’s plane, orthogonal to the line of sight. When the eye starts to move from tertiary position (a angle, solid arrow) the orientation of the eye is determined by rotation about the axes that lie on a plane rotated in the same direction, but only half as much as the line of sight, that is $a/2$.

ative incyclotorsion on downgaze, when one expresses torsion in Helmholtz coordinates.

7.3 The meaning of Listing’s Law

The oculomotor system follows precise laws in order to move and to rotate the eye. But we can ask what the real advantage is of following them. Concerning Listing’s Law it is possible to state that it enhances motor efficiency by minimizing the rotational eccentricity of the eye. The theory is based on the fact that some eye rotations are more efficient than others when it comes to move the gaze line [10]. Now let us suppose there is some “special”, central eye position and that we want to direct our gaze in all directions using the smallest possible rotation displacement from the centre; i.e with the smallest possible 3D eye eccentricity. To this goal, the eye rotates back and forth to the centre about the axes orthogonal to the vector gaze direction (\mathbf{g}_c). That means, the eye takes only positions that can be reached from the centre by rotating about an axis that lies in the plane orthogonal to \mathbf{g}_c . If we rename the centre position as *primary position* and call the plane *Listing’s plane* we see that Listing’s Law yields, consequently, the minimization of eccentricity. Let us attempt a mathematical proof of this concept [10]. First of all we write the quaternion q associated with a particular eye position defined by

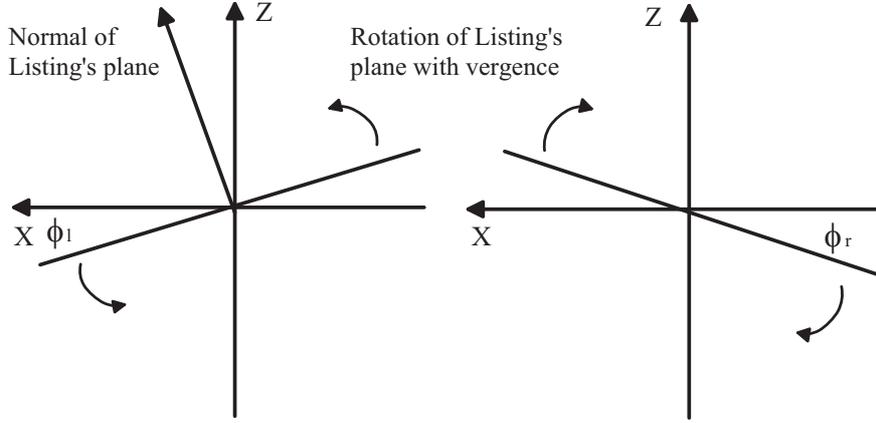


Figure 21: Binocular extension of Listing's Law. During convergence, the Listing's plane is rotated temporally and symmetrically in each eye about an angle ϕ proportional to the vergence angle.

the three Helmholtz angles V, H and T:

$$\begin{aligned}
 q &= q_V \circ q_H \circ q_T \\
 &= (c_{V/2}c_{H/2}c_{T/2} - s_{V/2}s_{H/2}s_{T/2}) + \\
 &+ \mathbf{i}(c_{V/2}s_{H/2}s_{T/2} + s_{V/2}c_{H/2}c_{T/2}) + \\
 &+ \mathbf{j}(c_{V/2}s_{H/2}c_{T/2} - s_{V/2}c_{H/2}s_{T/2}) + \\
 &+ \mathbf{k}(c_{V/2}c_{H/2}s_{T/2} + s_{V/2}s_{H/2}c_{T/2})
 \end{aligned} \tag{7.1}$$

where $c_{V/2}$ is the cosine of half the elevation angle V and $s_{H/2}$ is the sine of half the azimuth angle H, etc; and \mathbf{i} , \mathbf{j} , \mathbf{k} are (head-fixed) unit vectors pointing along the X (left), Y (up) and Z (forward) axis. Since the primary position coincides with the Z axis, and since Listing's Law states that the rotation axis must lie on a plane orthogonal to the primary position, this means that the component of q along the versor \mathbf{k} must always be equal to zero, $q_z = 0$. This leads to the following equation:

$$c_{V/2}c_{H/2}s_{T/2} + s_{V/2}s_{H/2}c_{T/2} = 0 \tag{7.2}$$

Divided by $c_{V/2}c_{H/2}c_{T/2}$ this simplifies to

$$\tan(T/2) = -\tan(V/2)\tan(H/2) \tag{7.3}$$

Now let us consider the scalar component of q :

$$q_0 = (c_{V/2}c_{H/2}c_{T/2} - s_{V/2}s_{H/2}s_{T/2}) \tag{7.4}$$

This is the eccentricity of the rotation, hence the angle by which to rotate around the axis represented by the vector part of q . If we derive equation 7.4 by T and equal the result to zero - in order to obtain the T angle that minimizes eccentricity - we obtain again equation 7.2. Obeying Listing's law brings many functional advantages, improving motor efficiency. Though, our eyes violate Listing's law on near fixation, as we observed above. So, there must be some even greater functional advantage, incompatible with Listing's law, to make the brain violate it so markedly on near gaze.

7.4 The meaning of binocular Listing's Law or L2

Listing's Law states that each eye's cyclorotation is proportional to the product of its horizontal and vertical angles in radians, $T = -HV/2$. This means that when the eyes are fixating a distant object (at infinity) the torsional angle for both eyes are equal; but this is not true when they are converging on a near object. Indeed, when vergence is not zero, the azimuth angles for the right and the left eye respectively, are different. This implies different values of T_r and T_l . This difference gives rise to a cyclovergence that has to be reduced since it afflicts directly the stereopsis. This is probably the reason why the eyes violate Listing's Law and they follow the L2 law on near vision: to reduce the cyclovergence and to restrict the motion of the epipolar line, thus permitting stereo matching to work with smaller search zones [10] [12]. Also in this case we try to give a mathematical explanation of the fact that by following the L2 law it is possible to nullify the cyclovergence, $T_r - T_l$. Empirically, it is found that each rotation axis for the left eye lies on a plane rotated temporally by an angle ϕ_l , and each rotation axis for the right eye lie on a plane rotated temporally by an angle ϕ_r , where the angles ϕ are linear functions of vergence ν : $\phi_l = \mu\nu$; $\phi_r = -\mu\nu$. This means that each rotation axis for the left eye is perpendicular to the normal of the plane $[\sin(\phi_l), 0, \cos(\phi_l)]$ and each rotation axis for the right eye is perpendicular to the normal of the plane $[\sin(\phi_r), 0, \cos(\phi_r)]$. If we define q_l and q_r respectively, the quaternion representing the position for the left and right eye, the L2 law requires [13]:

$$V(q_l) \cdot [\sin(\phi_l), 0, \cos(\phi_l)] = 0 \quad (7.5)$$

$$V(q_r) \cdot [\sin(\phi_r), 0, \cos(\phi_r)] = 0 \quad (7.6)$$

where $V(q)$ represents the vector part of the quaternion q , and \cdot the dot product. The solution of this equation yields the following equations that

give the torsion angle as function of fixation:

$$\begin{aligned}\tan(T_l/2) &= -\tan(V_l/2) \left[\frac{\tan(H_l/2) + \tan(\mu\nu)}{\tan(H_l/2) \tan(\mu\nu) + 1} \right] \\ \tan(T_r/2) &= -\tan(V_r/2) \left[\frac{\tan(H_r/2) + \tan(-\mu\nu)}{\tan(H_r/2) \tan(-\mu\nu) + 1} \right]\end{aligned}\tag{7.7}$$

Then from equation 7.8 we can derive the value for μ that equal the torsion for both the eye:

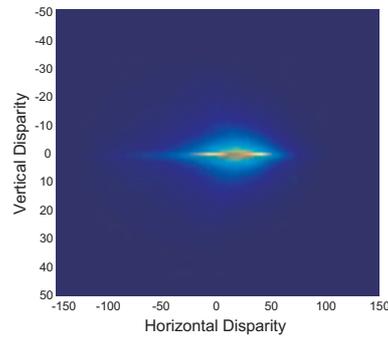
$$\mu = -\frac{\arcsin(\sec(\frac{H_l+H_r}{2}) \sin(\frac{H_l-H_r}{2}))}{2(H_r - H_l)}\tag{7.8}$$

It is worth noting that, since Helmholtz torsion represents a rotation relative to the visual plane, minimizing cyclovergence brings the two eyes into torsional alignment relative to the visual plane (and thus relative to the visual world). In other words, the torsion prescribed by L2 keeps the classical theoretical horopter, a circle in the visual plane and a vertical line in the midsagittal plane, making it's basic shape invariant across gaze movements. Actually, it has been reported [7] [10] [11] that human eye movements do not actually follow L2 precisely. It has been argued that the actual angle of rotation of Listing's planes with vergence strikes a compromise between the motor advantages associated with Listing's Law, and the improvement of binocular alignment that L2 brings about [14].

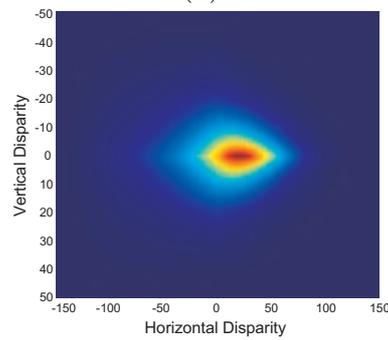
7.5 Implications on the disparity patterns

The eyes follow particular and different strategies of movements in order to fixate objects in the world around us. Specifically, for far and near fixations, the Listing's Law and its binocular extension pose constraints which influence the torsional components of eyes movements. For implementing these two types of ocular movements it is necessary to include, besides the classical azimuth H and elevation V angles, also a torsional T angle along the line of sight to describe the eye positions. This further degree of freedom affects directly on how an object projects on our retina (or CCD) and obviously this strikes again the disparity pattern, defined as the difference between the projections of an object on the left and the right retina. Hence the idea of a disparity statistics arises in order to understand the influence of eye movements on depth perception of the 3D peripersonal world [15] [16] [17] [18].

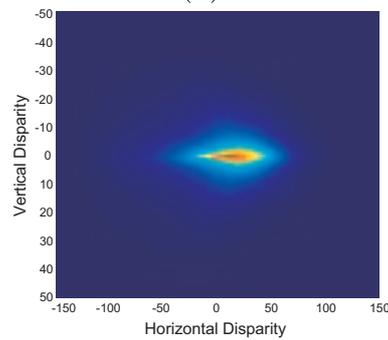
What we carried out is a comparative analysis of the behaviour of three types of stereo-head:



(a)



(b)



(c)

Figure 22: Comparative analysis between different stereo-head systems, (a) a classical Tilt/Pan system; (b) a system following Listing's Law; (c) a system following L2. Probability distribution of horizontal and vertical disparities. From these figures it is possible to note how a system implementing the Listing's Law is characterized by a widespread distribution of the disparity, while a system implementing L2 have a behavior very close to that of a classical Tilt/Pan stereo-head. The difference in the torsional angles of both eyes (characterizing the Listing's Law) determines an increase of the vertical disparities while in the other two cases the distribution has a horizontal elongation.

- a classical Tilt/Pan/Vergence system
- a system implementing the Listing's Law
- a system implementing the binocular extension of Listing's Law

In our simulation the 3D world was modelled as a set of random points distributed uniformly through space. Each scene consisted of 25000 points. The distance of the points from the observer were restricted between 30 cm and 100 cm. The distributions presented here are based on 50 such worlds. For each of these worlds we considered a grid of fixation points. These are characterized by a gaze eccentricity that varied between $\pm 30^\circ$ by step of 3° and by a distance from the observer ranging between 30 cm and 100 cm by step of 5 cm. For every fixation point we collected the disparity of the points in the world whose projections fall on both the retinas, and then we interpolated the disparity pattern on a 21×21 grid of retinal sample. As a first analysis we computed the distribution of the disparities. This is represented in figure 22, where the 2D histogram of the disparity occurrence is drawn.

7.6 Open issues

It is interesting to consider the kinematics of the classical robotic head systems tilt/pan/vergence. These systems are characterized by a common tilt axis, directed along the interaural axis, and two separated pan axes, one for each camera. There are no possibilities for the cameras to rotate around the visual axis. This means that besides nullifying the cyclovergence, these systems are characterized by zero cyclovergence. One can object that Listing's Law is only a way to obtain a binocular alignment which in the classical tilt/pan systems is structurally guaranteed (for free), by mechanical constraints that they are subject to. Of course this is true and, from a perceptual point of view tilt/pan heads would be optimal. However, in such systems there is no possibility of adaptation. Instead, it has been recently demonstrated that the control of ocular torsion can be changed by a cyclodisparity stimulus [19] [20]. This suggests a scenario where ocular torsions are dynamically controlled to optimize binocular image alignment to simplify the perception of slanted surfaces. From this point of view, the presence of further degrees of freedom would not increase the redundancy of the system, but it could increase its efficiency. The way by which the eyes (or the cameras) fixate a point on a surface could change accordingly and adaptively with the characteristics of the surface itself, eventually improving the depth perception.

References

- [1] C. Samson, M. Le Borgne, and B. Espiau. *Robot Control: The task function approach*. Oxford science publications, Oxford, England, clarendon press edition, 1991.
- [2] F.C. Donders. Beitrag zur lehre von den bewegungen des menschlichen auges. *Holland Beit Anatom Physiolog Wiss*, 1:105–145, 1848.
- [3] D. Tweed and T. Vilis. Geometric relations of eye position and velocity vectors during saccades. *Vision Res.*, 30(1):111–127, 1990.
- [4] D. Tweed, W. Cadera, and T. Vilis. Computing three-dimensional eye position quaternions and eye velocity from search coil signals. *Vision Res.*, 30(1):97–110, 1990.
- [5] P. Bruno and A.V. Van den Berg. Relative orientation of primary position of the two eyes. *Vision Res.*, 37(7):935–947, 1997.
- [6] A.W.H. Mincken and J.A.M. Van Gisbergen. A three dimensional analysis of vergence movements at various level of elevation. *Exp. Brain Res.*, 101(2):331–345, 1994.
- [7] D. Mok, A. Ro, W. Cadera, J.D. Crawford, and T. Vilis. Rotation of listing’s plane during vergence. *Vision Res.*, 32:2055–2064, 1990.
- [8] J. Porrill, J.P. Ivins, and J.P. Frisby. The variation of torsion with vergence and elevation. *Vision Res.*, 39:3934–3950, 1999.
- [9] R.A.B. Somani, J.F.X. Desouza, D. Tweed, and T. Vilis. Visual test of listing’s law during vergence. *Vision Res.*, 38(6):911–923, 1998.
- [10] D. Tweed. Visual-motor optimization in binocular control. *Vision Res.*, 37(14):1939–1951, 1997.
- [11] L.J. Van Rijn and A.V. Van den Berg. Binocular eye orientation during fixations: Listing’s law extended to include eye vergence. *Vision Res.*, 33(5/6):691–708, 1993.
- [12] K. Schreiber, J.D. Crawford, M. Fetter, and D. Tweed. The motor side of depth vision. *Nature*, 410:819–822, 2001.
- [13] J.C.A. Read and B.G. Cumming. Does depth perception require vertical-disparity detectors? *J. Vis.*, 6(12):1323–1355, 2006.

- [14] K.M. Schreiber, D.B. Tweed, and C.M. Schor. The extended horopter: Quantifying retinal correspondence across changes of 3d eye position. *J. Vis.*, 6:64–74, 2006.
- [15] Z. Yang and D. Purves. A statistical explanation of visual space. *Nat. Neurosci.*, 6(6):632–640, 2003.
- [16] Y. Liu, A.C. Bovik, and L.K. Cormack. Disparity statistics in natural scenes. *J. Vis.*, 8(11):1–14, 2008.
- [17] M. Hansard and R. Horaud. Pattern of binocular disparity for a fixating observer. In *BVAI 2007*, volume 4729 of *LNCS*, pages 308–317, 2007.
- [18] M. Hansard and R. Horaud. Cyclopean geometry of binocular vision. *J. Opt. Soc. Am. A*, 25(9):2357–2369, 2008.
- [19] J.S. Maxwell, E.W. Graf, and C.M. Schor. Adaptation of torsional eye alignment in relation to smooth pursuit and saccades. *Vision Res.*, 41:3735–3749, 2001.
- [20] C.M. Schor, J.S. Maxwell, and E.W. Graf. Plasticity of convergence-dependent variations of cyclovergence with vertical gaze. *Vision Res.*, 41:3353–3369, 2001.
- [21] J.M. Miller and D.A. Robinson. A model of the mechanics of binocular alignment. *Computers and Biomedical Res.*, 17:436–470, 1984.
- [22] T. Haslwanter. Mathematics of three-dimensional eye rotations. *Vision Res.*, 35(12):1727–1739, 1995.
- [23] C.M. Schor, J.S. Maxwell, and S.B. Stevenson. Isovergence surfaces: the conjugacy of vertical eye movements in tertiary positions of gaze. *Ophthalmic Physiol. Opt.*, 14:279–286, 1994.
- [24] J.M. Miller, J.L. Demer, V. Poukens, D.S. Pavlovski, H.N. Nguyen, and E.A. Rossi. Extraocular connective tissue architecture. *J. Vis.*, 3(3):240–251, 2003.
- [25] J.M. Miller. Understanding and misunderstanding extraocular muscle pulleys. *J. Vis.*, 7(11):1–15, 2007.
- [26] J.L. Demer, J.M. Miller, V. Poukens, H.V. Vinters, and B.J. Glasgow. Evidence for fibromuscular pulleys of recti extraocular muscles. *Invest. Ophthalmol. Vis. Sci.*, 36:1125–1136, 1995.

A Extraocular muscles and their actions

The direction of gaze of each globe is determined by a delicate and extremely precise balance of rotational tractions exerted by a group of muscles that, by virtue of their respective attachments to the surface of the eye globes, allow their rotation. This is the group of the extraocular muscles and they comprise four rectus muscles for each eye (medial **MR**, lateral **LR**, superior **SR** and inferior **IR**) and the two oblique ones (superior **SO** and inferior **IO**), see figure 23.

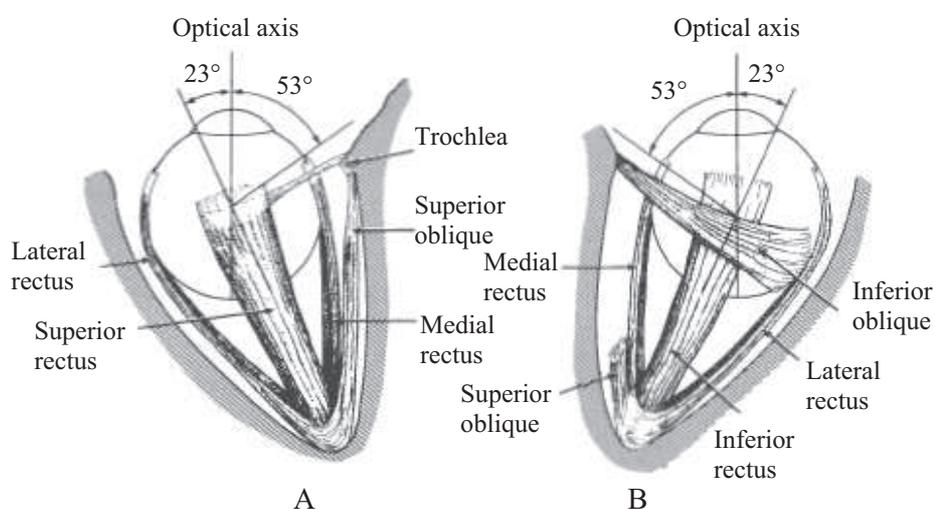


Figure 23: Left eye position and its muscles. (A) top view. (B) bottom view.

Each extraocular muscle rotates the globe in specific directions, also dependent on the current position of the eye, see figure 24. Nasal movement of the eye is adduction, temporal movement is abduction. Up and down movements are elevation and depression, respectively. Torsional eye movements rotate the eye around its visual axis, whereby intorsion is nasal rotation of the vertical meridian and extorsion is temporal rotation of the vertical meridian [21].

MR and **LR** have horizontal actions, even if they acquire a vertical action when the globe is elevated or depressed. **SR** has a vertical action when the eye is abducted by 23°. When the globe is adducted starting from 23° of abduction the action of **SR** changes in a combination of adduction and intorsion. On the contrary, when eye is abducted more than 23° its action becomes a combination of abduction and extorsion. **IR** behaves in the same way but with opposite actions. When the globe is adducted **SR** and **IR**

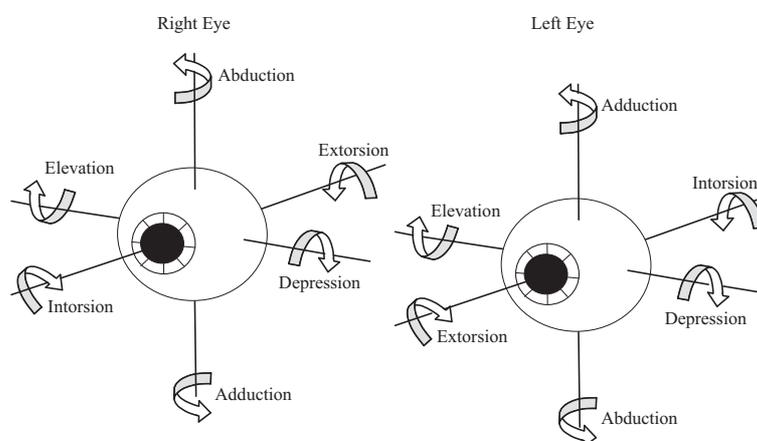


Figure 24: The eye's rotations

are less or not effective in elevating or depressing the eye. These become the functions of the oblique muscles which in adduction are very efficient, **SO** for depression and the **IO** for elevation. In full abduction **SO** causes only abduction (assisting the **LR**) and intorsion. The actions of the **IO** correspond with and balance those of the **SO** [21].

B Mathematics of eye movements

The movement of the globe approximately corresponds to a rotation of an object in the three dimensional space around a certain axis. The globe center can be regarded as the rotation center. Eye positions usually are classified in three groups: primary, secondary and tertiary position. In primary position the eye looks straight ahead and in this position the muscles exhibit the minimum force. From primary position any rotation about the vertical or the horizontal axis bring the eye in secondary position. In this case it 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 is in tertiary position. The current eye position is defined by characterizing the 3D rotation from a somewhat arbitrarily chosen reference position to the current eye position. This reference position is usually defined as the position the eye assumes when the subject is looking straight ahead, while the head is kept upright [22]. To describe the 3D orientation of the eye, Euler's theorem can be applied: it states that for every two orientations of an object, the object can always move from one to the other by a single rotation about a fixed axis. The rotation from the reference position to the current eye position

can be described, other than this single rotation, also by three consecutive rotations about well defined, hierarchically nested axes. A combination of a horizontal and a vertical rotation of the eye is a well defined sequence, uniquely characterizing the direction of the line of sight. However, this does not completely determine the 3D eye position, since the rotation around the line of sight is still unspecified. A third rotation is needed to completely determine the orientation of the eye. The sequence of rotation plays an important role, since the execution of rotations specifying the same angles but in different order, leads to a different final orientation of the rotated object. Helmholtz coordinate systems are widely adopted. The Helmholtz system uses a head-referenced horizontal axis for describing the vertical component of eye rotation and an eye referenced vertical axis for describing the horizontal component of eye position. In this system the eye position is characterized first by a vertical rotation \mathbf{V} around the X (head fixed) axis, then a horizontal rotation \mathbf{H} around the Y (eye fixed) axis and, finally, a torsional rotation \mathbf{T} around the Z (eye fixed) axis ¹⁴, see figure 25.

C Time derivative of unit geometric vectors

Let $\bar{\mathbf{u}} = \mathbf{u}/|\mathbf{u}|$ be a unit vector then:

$$\frac{d}{dt}\bar{\mathbf{u}} = \frac{d}{dt} \frac{\mathbf{u}}{|\mathbf{u}|} = \frac{1}{|\mathbf{u}|}\dot{\mathbf{u}} - \frac{\mathbf{u}}{|\mathbf{u}|^2} \frac{d}{dt} |\mathbf{u}| \quad (\text{C.1})$$

¹⁴The Helmholtz system is just one of the possible coordinate system that can be used. Another possibility is the so called Fick system: it uses a head-referenced vertical axis of rotation to describe the horizontal component of eye position and an eye referenced horizontal axis to describe the vertical component of eye position. In this system the eye position is characterized first by a horizontal rotation \mathbf{H} around the Y (head fixed) axis, then a vertical rotation \mathbf{V} around the X (eye fixed) axis and, finally, a torsional rotation \mathbf{T} around the Z (eye fixed) axis. Another example would be to use a coordinate system in which all the components of eye position are described by eye-referenced axes (Harms system), or on the contrary a system that uses only head fixed axes (Hess system) [23]. These four coordinate systems described above yield very different descriptions of eye movements responses to targets in tertiary positions of gaze; the problem is that they do not indicate the actual coordinate system used by the oculomotor system. The empirical question is which one best describes the movements of the two eyes, but the answer is not well-defined, because the oculo-motor system does not adhere to any one of the four systems, neither in terms of internal representation of target position, nor in terms of a mechanical gimbaling of the eye in the orbit. Recent evidence suggests that the extraocular muscles are mechanically constrained by tissue that connects muscles sheaths to the wall of the orbit [24] [25] [26]. Given this arrangement, the axes of rotation will tend to be either eye fixed or head fixed depending on whether the muscle sheaths act as pulleys or pivot points for the extraocular muscles.

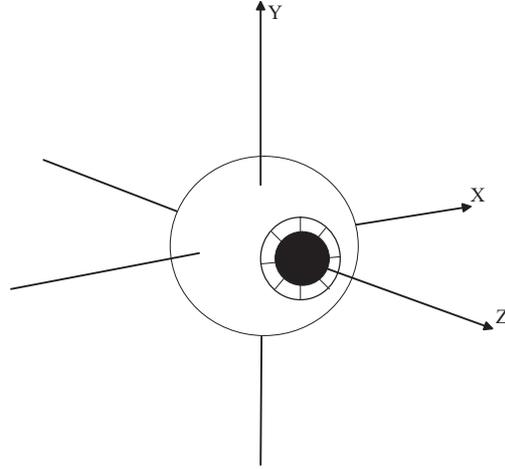


Figure 25: Eye fixed reference frame

Recalling that $|\mathbf{u}| = \sqrt{\mathbf{u} \cdot \mathbf{u}}$, equation (C.1) can be rewritten as:

$$\begin{aligned} \frac{d}{dt} \bar{\mathbf{u}} &= \frac{1}{|\mathbf{u}|} \left[\dot{\mathbf{u}} - \frac{\mathbf{u}}{|\mathbf{u}|} \frac{1}{2\sqrt{\mathbf{u} \cdot \mathbf{u}}} \frac{d}{dt} (\mathbf{u} \cdot \mathbf{u}) \right] = \\ &= \frac{1}{|\mathbf{u}|} \left[\dot{\mathbf{u}} - \frac{\mathbf{u}}{|\mathbf{u}|} \frac{(\mathbf{u} \cdot \dot{\mathbf{u}})}{|\mathbf{u}|} \right] = \frac{1}{|\mathbf{u}|} [\dot{\mathbf{u}} - (\bar{\mathbf{u}} \cdot \dot{\mathbf{u}}) \bar{\mathbf{u}}] \end{aligned} \quad (\text{C.2})$$

Definition 1 Given a unit geometric vector $\bar{\mathbf{u}}$ the orthogonal projection vector operator is defined as:

$$\mathcal{P}_{\bar{\mathbf{u}}}^{\perp} \mathbf{a} = \mathbf{a} - (\bar{\mathbf{u}} \cdot \mathbf{a}) \bar{\mathbf{u}} \quad \forall \mathbf{a} \quad (\text{C.3})$$

Remark 10 If $\bar{\mathbf{u}}$ and \mathbf{a} are projected onto the same reference frame $\langle 0 \rangle$, then the orthogonal projector $\mathcal{P}_{\bar{\mathbf{u}}}^{\perp}$ can be expressed in matrix form as:

$${}^0\mathcal{P}_{\bar{\mathbf{u}}}^{\perp} = [\mathbf{I} - {}^0\bar{\mathbf{u}} {}^0\bar{\mathbf{u}}^T] \quad (\text{C.4})$$

where \mathbf{I} is the 3×3 identity matrix.

Following the definition above, (C.2) can be rewritten as:

$$\dot{\mathbf{u}} = \frac{1}{|\mathbf{u}|} \mathcal{P}_{\bar{\mathbf{u}}}^{\perp} \dot{\mathbf{u}} \quad (\text{C.5})$$

D Simplified analysis of convergence properties of the decoupled control

In this section we provide an informal proof of the convergence of the simplified control method discussed in section 5.3. Assume here that the cameras have the same *elevation*, therefore the control strategy must perform a pure *vergence* task (the situation sketched in figure 4 corresponds to this planar control problem). It is easy to see in this case that any *small* rotation $\boldsymbol{\omega}_L dt$ of the left eye which tends to reduce the angle θ_L also produces a reduction of θ_R . Now, the angular velocity of the left eye proposed in (5.20) (or (5.28)) minimizes (monotonically) θ_L ; furthermore, by using the argument above, it also reduces θ_R . This means that assuming that only the left eye is controlled, then both θ_L and θ_R go to zero. Then, in the worst case, that the term $\mathbf{e}_R \cdot \mathbf{f}_R(\boldsymbol{\omega}_L)$ must go to zero faster than \mathbf{e}_L . A symmetric argument can be applied for the right eye, and since both control strategies (5.20) and (5.21) (or (5.28) and (5.29)) jointly produce a reduction of the task errors (in *cooperative* way) this means that the proposed control strategies (the smooth or the non-smooth) ensure the convergence to zero of the task errors as required.

E Solving Equation (6.6)

By projecting equation (6.6) along all the vectors forming the frame (6.9) we obtain:

$$\mathbf{k}_L \cdot [z_L \mathbf{k}_L - z_R \mathbf{k}_R] = \mathbf{k}_L \cdot \mathbf{b} \quad (\text{E.1})$$

$$\mathbf{k}_R \cdot [z_L \mathbf{k}_L - z_R \mathbf{k}_R] = \mathbf{k}_R \cdot \mathbf{b} \quad (\text{E.2})$$

$$\mathbf{n} \cdot [z_L \mathbf{k}_L - z_R \mathbf{k}_R] = \mathbf{n} \cdot \mathbf{b} \quad (\text{E.3})$$

where the last equation is identically 0 by definition of \mathbf{n} , while the others lead to:

$$z_L - (\mathbf{k}_R \cdot \mathbf{k}_L)z_R = \mathbf{k}_L \cdot \mathbf{b} \quad (\text{E.4})$$

$$(\mathbf{k}_R \cdot \mathbf{k}_L)z_L - z_R = \mathbf{k}_R \cdot \mathbf{b} \quad (\text{E.5})$$

Recall that $(\mathbf{k}_R \cdot \mathbf{k}_L) = \cos \alpha$ then the equations above can be written in matrix form as:

$$\begin{bmatrix} 1 & -\cos \alpha \\ \cos \alpha & -1 \end{bmatrix} \begin{bmatrix} z_L \\ z_R \end{bmatrix} = \begin{bmatrix} \mathbf{k}_L \cdot \mathbf{b} \\ \mathbf{k}_R \cdot \mathbf{b} \end{bmatrix} \quad (\text{E.6})$$

The system (E.6) can be solved in closed form as follows:

$$\begin{bmatrix} z_L \\ z_R \end{bmatrix} = \frac{1}{\sin^2 \alpha} \begin{bmatrix} 1 & -\cos \alpha \\ \cos \alpha & -1 \end{bmatrix} \begin{bmatrix} \mathbf{k}_L \cdot \mathbf{b} \\ \mathbf{k}_R \cdot \mathbf{b} \end{bmatrix} \quad (\text{E.7})$$

In particular:

$$z_L = \frac{1}{\sin^2 \alpha} [(\mathbf{k}_L \cdot \mathbf{b}) - (\mathbf{k}_R \cdot \mathbf{b}) \cos \alpha] \quad (\text{E.8})$$

$$z_R = \frac{1}{\sin^2 \alpha} [(\mathbf{k}_L \cdot \mathbf{b}) \cos \alpha - (\mathbf{k}_R \cdot \mathbf{b})] \quad (\text{E.9})$$

then since $|\sin \alpha| = |\mathbf{k}_R \times \mathbf{k}_L|$ and $\cos \alpha = \mathbf{k}_R \cdot \mathbf{k}_L$ it is possible to compute the following quantities which have been used through the text:

$$\frac{1}{z_L} = \frac{|\mathbf{k}_R \times \mathbf{k}_L|^2}{(\mathbf{k}_L \cdot \mathbf{b}) - (\mathbf{k}_R \cdot \mathbf{k}_L)(\mathbf{k}_R \cdot \mathbf{b})} \quad (\text{E.10})$$

$$\frac{1}{z_R} = \frac{|\mathbf{k}_R \times \mathbf{k}_L|^2}{(\mathbf{k}_R \cdot \mathbf{k}_L)(\mathbf{k}_L \cdot \mathbf{b}) - (\mathbf{k}_R \cdot \mathbf{b})} \quad (\text{E.11})$$