HAL
archives-ouvertes.fr

Cyclopean geometry of binocular vision

Miles Hansard, Radu Horaud

To cite this version:

Miles Hansard, Radu Horaud. Cyclopean geometry of binocular vision. Journal of the Optical Society of America. A Optics, Image Science, and Vision, Optical Society of America, 2008, 25 (9), pp.2357-2369. 10.1364/JOSAA.25.002357 . inria-00435548

HAL Id: inria-00435548

https://hal.inria.fr/inria-00435548

Submitted on 24 Nov 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Cyclopean Geometry of Binocular Vision

Miles Hansard and Radu Horaud

miles.hansard@inrialpes.fr, radu.horaud@inrialpes.fr

Inria Rhône-Alpes, 655 Avenue de l'Europe,

Montbonnot, France 38330.

Josa A, 25(9), 2357–2369,
2008.

Abstract

The geometry of binocular projection is analyzed in relation to the primate visual system. An oculomotor parameterization, which includes the classical vergence and version angles, is defined. It is shown that the epipolar geometry of the system is constrained by binocular coordination of the eyes. A local model of the scene is adopted, in which depth is measured relative to a plane containing the fixation point. These constructions lead to an explicit parameterization of the binocular disparity field, involving the gaze angles as well as the scene structure. The representation of visual direction and depth is discussed, with reference to the relevant psychophysical and neurophysiological literature.

1 Introduction

Information about the 3-D structure of a scene is present in binocular images, owing to the spatial separation of the two viewpoints. If the projections of corresponding points can be identified in the left and right views, then the 3-D information can, in principle, be recovered. In addition to the image data, the process of recovery involves the parameters of the binocular projection; in particular, the relative orientation of the eyes is important. If some, or all, of the projection parameters remain unknown, then the 3-D information that can be recovered may be limited to affine or projective properties of the scene [1, 2, 3].

Psychophysical evidence suggests that non-visual information about the current orientation of the eyes is very limited [4]. Hence, in order to facilitate the 3-D interpretation of the binocular disparity field, it would be desirable to keep the eyes stationary with respect to the head. Human vision, however, involves frequent eye movements, of several different types [5]. For example, the eyes may be moved in order to direct the field of view, or to foveate an object of interest. Eye movements are also used to stabilize the retinal image with respect to head movements[6], and to track moving visual targets. It would be undesirable to suspend these functions, which are essentially monocular, during the binocular analysis of a scene.

The geometry of binocular stereopsis is complicated by movements of the eyes, as described above. However, the two eyes typically move in a coordinated fashion, such that a single point in the scene is fixated. This can be achieved, in particular, by vergence eye-movements, which are driven by binocular disparity [7]. These coordinated eye movements benefit stereopsis, as they align the two retinal images at the respective foveas. It follows that the amount of disparity around the fixation point tends to be reduced, assuming that the scene is locally smooth. This is important, given the relatively short range of biological disparity detectors [8, 9]. It should, however, be noted that stereopsis also exists in animals that do not move their eyes significantly, such as owls [10].

There may be other ethological reasons for the existence of binocular eye-movements, despite the resulting complication of stereopsis. It has been suggested that the evolution of binocular vision was motivated by the ability to detect camouflaged prey, by segmentation in depth, with respect to the background [11]. Another impetus may have been the improvement in image

quality that can be achieved by combining two views, especially in nocturnal conditions [12]. Both of these processes would benefit from binocular eye movements, which allow the scene to be scanned without moving the head, and which help to register the two views. These image-segmentation and enhancement processes do not require geometric reconstruction of the scene, and so the disadvantages of moving the eyes are limited.

It is clear that the binocular vision of humans (and other primates) has evolved beyond simple tasks such as camouflage-breaking. Psychophysical evidence shows that the geometric properties of a typical 3-D scene can be estimated by stereopsis, and that these estimates can be combined, as the eyes fixate successive visual targets [13]. Furthermore, it is clear that most types of eye movement are binocularly coordinated [5]. The combination of eye movements and stereopsis raises important questions about the oculomotor parameterization, the processing of the binocular disparity field, and the representation of the visible scene [14, 15]. These three questions are developed in more detail below, in sections 1.1, 1.2 and 1.3, respectively.

It will be emphasized, in this paper, that the structure of the disparity field depends on the *epipolar geometry* of the visual system. Furthermore, it will be shown that this can be obtained directly from the appropriate oculomotor parameterization. The combination of oculomotor and epipolar constraints leads, finally, to a simple model of the local scene-structure. The epipolar geometry of biological vision, based on the appropriate 'essential matrix', has not been developed elsewhere. The scope and novelty of the present approach is detailed in sections 1.4 and 1.5.

1.1 Oculomotor Parameterization

The first question to be addressed is: How should binocular eye movements be parameterized? This an important issue, because it determines the complexity of the control problem that the oculomotor system must solve. In particular, it is important to establish the minimal number of parameters that are compatible with the observed range of oculomotor behaviour. The combination of two sensors, each of which can rotate in space, results in a system which has six angular degrees of freedom. However, if Donders' law is obeyed[5], then the rotation of each eye *around* the corresponding line of sight is determined by the *direction* of that line of sight. This removes one degree of freedom from each eye. Furthermore, binocular fixation implies co-planarity of the visual axes, which removes one elevation angle from the parameterization. This leaves three degrees of freedom, which can be conveniently assigned to the elevation, azimuth and distance of the fixation point. These variables are most naturally specified in relation to the 'cyclopean point' [16], which, in the present work, is situated halfway along the inter-ocular axis. The trigonometry of this cyclopean parameterization is defined in section 3, and its relationship to the classical vergence/version model[17] is stated.

1.2 Disparity Processing

The second question to be addressed is: How does the orientation of the eyes affect the structure of the binocular disparity field? The difference in position between the left and right projections of a given scene point is, by convention, called the 'absolute disparity' of the point[18]. This is the quantity that can be measured most directly, by disparity-sensitive mechanisms[19]. It is important to note that the response of such a mechanism must depend on the orientation of the eyes. Indeed, for a typical scene, and a typical fixation point, it may be hypothesized that the relative orientation of the eyes will be the dominant source of absolute disparity.

It is important, for the reasons given above, to establish exactly how the disparity field is affected by eye movements. The question is approached in section 4, in which the horopter of the fixating system is defined; this is the set of scene points that project to the same location in each image [16]. The horopter is used in section 5 to construct the epipolar geometry[20] of the system, which is effectively parameterized by the vergence and version angles. If a projected point is identified in one image, then the epipolar constraint restricts the location of the corresponding point to a *line* in the other image. This important relationship can be expressed for any configuration of the eyes. In principle, the epipolar geometry could be used to 'rectify' the retinal images, thereby removing the effect of eye movements on the disparity field. However,

this would not be consistent with the observed dependence of early binocular processing on absolute retinal disparity[19]. Hence it is desirable to parameterize the disparity field, with respect to the orientation of the eyes. The epipolar geometry is the basis of such a parameterization.

1.3 Scene Representation

The two questions described above are part of a third, more general question: How can the geometric structure of the scene be represented by the visual system? This issue is complicated by the fact that the early mechanisms of primate binocular vision are sensitive to a quite limited range of disparities[8, 9]. The region of space that can be resolved in depth depends, consequently, on the relative orientation of the eyes. Specifically, only those points in Panum's area (which is centred on the fixation point) can be fused[21, 18]. It follows that any global representation of the scene must be assembled piecewise, over a series of fixations. It is natural to formulate this process as the measurement of scene-structure with respect to a reference surface, followed by an integration of the resulting local models[22, 23]. The plane that passes through the fixation point, and that is orthogonal to the cyclopean visual direction, is a convenient local model for binocular scene representation, as will be shown in section 6.

1.4 Scope and Assumptions

The word 'cyclopean' has, in the present context, several possible meanings. As described in section 1.1, the 'cyclopean point' is a notional centre of projection, located halfway along the inter-ocular axis (cf. Helmholtz[16]). It is convenient to use this point as the origin of the binocular coordinate system, although there is a useful alternative, as will be shown in section 3. The word 'cyclopean' is used elsewhere in a more general sense, with reference to visual information that is intrinsically binocular, such as the 'edges' that can be perceived in a random-dot stereogram (cf. Julesz[11]). The phrase 'cyclopean geometry', as used here, refers to the fact that the binocular configuration of a fixating visual system can be parameterized by the direction and distance of the fixation point, with respect to a *single* eye (cf. Hering[17]). Furthermore, it is convenient to make this parameterization with respect to the cyclopean point, as will be explained in section 3.

It will be assumed, in the present work, that the retinal projections can be described by the usual pin-hole camera equations, and that these projections are 'internally calibrated'. This means that the visual system is able to relate the monocular retinal separation of any two points to the angle between the corresponding optical rays [2]. A weaker assumption could be made, given that the visual system does not ultimately achieve a Euclidean representation of the scene[24]. Indeed, the main constructions developed here, including the horopter and the epipolar geometry, can be obtained directly in the projective setting, based only on the pin-hole model. [2, 3]. However, the effects of the oculomotor configuration on binocular vision are emphasized in the present analysis, and these effects are more readily studied by Euclidean methods.

A distinction should be made between *descriptions* and *models* of binocular vision. The present work aims to describe binocular geometry in the most convenient way. This leads to cyclopean parameterizations of visual direction and binocular disparity. Whether these parameterizations are actually used by the visual system is a further question[25]. In particular, it is not necessary to assume that the cyclopean representation has any biological reality. Discussion of the psychophysical and physiological evidence that can be used to make such claims is confined to section 7. The present work aims to provide a useful description of binocular geometry; not to construct a detailed model of biological stereopsis. For this reason, the *estimation* of scene and gaze variables is not considered in detail. Indeed, the present geometric account is compatible with a range of algorithmic models.

It will not be assumed that the orientation of the eyes is known. Rather, the binocular disparity field will be parameterized by a set of gaze variables, as well as by the scene structure. If the visual system is to recover the unknown gaze parameters from the observed disparity field, then this is the required representation. Although the orientation of the eyes is unknown, some

qualitative constraints on oculomotor behaviour will be observed. For example, it will assumed here that the left and right visual axes intersect at a point in space. This is approximately true, and moreover, in the absence of an intersection, it would be possible to define an appropriate chord between the left and right visual axes, and to choose a notional fixation point on this segment. In particular, it would be straightforward to extend the analysis of the disparity field (sec. 6) to allow for mis-alignment of the eyes. In addition to the fixation constraint, it will be assumed that each eye rotates in accordance with Donders' law, meaning that the cyclo-rotation of the eyes can be estimated from the gaze direction [5]. The 'small baseline' assumption (that the inter-ocular separation is small with respect to the viewing distance) will not be required here. Nor will it be assumed that the disparity function is continuous from point to point in the visual field.

1.5 Relation to Previous Work

The geometry of binocular vision has been analyzed elsewhere, but with different objectives, methods and assumptions. The present work will be contrasted with the principal existing approaches, which are recalled below. A more detailed summary of these models is given by Gårding, Porrill, Mayhew and Frisby [26].

It was shown by Koenderink and van Doorn [27] that the gradient of reciprocal-distance to a visible surface can be recovered from the first-order structure of the corresponding disparity field. This *differential* approach can be extended in several ways; for example, it is possible to recover measures of surface shape, from the second-order structure of the disparity field[28, 29, 30]. These models are essentially *local*, and require that the disparity field is (or can be made) continuous. The small-baseline assumption is also an important part of such models. The work that will be described here is not concerned with the differential structure of the disparity field, and so none of above assumptions are needed. The present analysis, unlike the differential approach, makes the epipolar geometry explicit, and does not involve derivatives of the disparity field. Although the results of section 6 can be extended to include surface orientation (as indicated in sec. 7.3), it would also possible to combine the differential and epipolar analyses. For example, the former could be used to estimate orientation and shape, and the latter to estimate the gaze parameters. The differential and epipolar methods are, in this sense, complementary.

An alternative, non-differential, approach to binocular vision was initiated by Mayhew and Longuet-Higgins[31, 32]. This approach is based on the fact that the horizontal and vertical components of the disparity field contain different information. It is, in particular, possible to estimate the viewing-distance and azimuth from the vertical component. The full scene-structure can then be approximated, by combining the estimated viewing-parameters with the horizontal component of the original disparity-field. Related decompositions have been described by Gårding et al. [26], and by Weinshall [33]. The present approach, quite unlike these models, represents each disparity as a scalar offset in a *variable* epipolar direction. Note that the epipolar direction is *not*, for finite fixation points, horizontal. The advantage of the epipolar decomposition is that the 'gaze' and 'structure' components of the disparity field can be identified directly, as will be shown in section 6. It may also be noted that the small-baseline assumption, which is used to simplify the horizontal/vertical decomposition, is not needed in the following work.

A large amount of psychophysical work has been based on the horizontal/vertical disparity decomposition [34, 35, 36, 37]. It should be emphasized that the present work is entirely compatible with this literature. Any geometrically possible disparity field can be represented in terms of horizontal and vertical components, or in terms of (variably oriented) epipolar lines and offsets. The main practical difference is that the epipolar model is much more compact, because it automatically incorporates the physical constraints which must otherwise be imposed on the vertical disparity field [32, 26].

Both the differential and horizontal/vertical decompositions are, like the present work, based on purely visual information. If additional (e.g. oculomotor) information about the orientation of the eyes is available, then the situation is *greatly* altered. This is because, given the viewing configuration, it is possible to directly triangulate points in 3-D space. Erkelens and van Ee develop this approach, which leads to the definition of *head-centric* disparity [38]. Unlike the

head-centric approach, the present work develops the disparity field in the *images*, without assuming that the orientations of the eyes are known. Nonetheless, it would be straightforward to incorporate oculomotor information in the present analysis; for example, initial estimates of the gaze-parameters could be based on efference-copy signals.

The present analysis is related to established approaches in computer vision [39, 2, 3, 40, 41]. The derivations however, are novel, and the details are specific to the biological context. The following results are of particular interest: I. The cyclopean parameterization of binocular orientation (equations 5,6,9); II. The identification of the midline horopter as an axis that passes through the pole of the visual plane (eqns. 17,18); III. The construction of the essential matrix from the epipoles and midline horopter (eqn. 24); IV. The symmetric parameterization of binocular correspondence (28) and V. the parameterization of binocular parallax as a function of deviation from the fixation plane (eqn. 47).

2 Projection Model

The notation and coordinate systems used in this work are described here. Points and vectors will be shown in bold type, for example $\boldsymbol{q}$, $\boldsymbol{v}$. The transpose of $\boldsymbol{v}$ is a row-vector $\boldsymbol{v}^\top$, and the Euclidean length is $|\boldsymbol{v}|$. The notation $(\boldsymbol{v})_3$ will be used to indicate the third component of the vector $\boldsymbol{v}$. Matrices are represented by upper-case letters, for example $\boldsymbol{M}$. Table 1 gives a summary of the notation used in this paper.

$\bar{\boldsymbol{c}}_\ell$, $\bar{\boldsymbol{c}}_r$.	Left, right optical centres.
$\boldsymbol{b}$, $\bar{\boldsymbol{c}}_b$.	Optical baseline, mid-point.
$\boldsymbol{R}_\ell$, $\boldsymbol{R}_r$.	Left, right eye rotation matrices.
$\bar{\boldsymbol{p}}_0$, $\boldsymbol{v}$.	Fixation point, fixation direction.
$\mathcal{P}$, $\bar{\boldsymbol{p}}$.	Fixation plane, point in plane.
α, β, ρ.	Elevation, azimuth, distance of $\bar{\boldsymbol{p}}_0$.
β_ℓ, β_r.	Left, right azimuths of $\bar{\boldsymbol{p}}_0$.
γ_ℓ, γ_r.	Left, right cyclo-rotation angles.
δ, ϵ.	Vergence, version angles.
ζ, η, $\bar{\boldsymbol{c}}$.	Centre, radius, rear-point of VM circle.
$\bar{\boldsymbol{q}}$; $\boldsymbol{q}_\ell$, $\boldsymbol{q}_r$.	Scene-point; left, right projections.
$\bar{\boldsymbol{c}}_a$, $\bar{\boldsymbol{q}}_a$.	Points on vertical horopter.
$\boldsymbol{a}$.	Image of vertical horopter.
$\boldsymbol{u}_\ell$, $\boldsymbol{u}_r$.	Left, right epipolar lines.
$\boldsymbol{e}_\ell$, $\boldsymbol{e}_r$; $\boldsymbol{E}$.	Left, right epipoles; essential matrix.
$\boldsymbol{d}_\ell$, $\boldsymbol{d}_r$.	Left, right disparity directions.
s.	Distance of $\bar{\boldsymbol{q}}$ from plane $\mathcal{P}$.
t_ℓ, t_r.	Left, right cyclopean parallax.

Table 1: Summary of the notation used in the text.

The 3-D Euclidean coordinates of a point will be distinguished by a bar, e.g. $\bar{\boldsymbol{q}}$. Note that the difference of two Euclidean points results in a vector, e.g. $\boldsymbol{v} = \bar{\boldsymbol{q}} - \bar{\boldsymbol{p}}$. The homogeneous image-coordinates of points and lines are written without a bar; for example, a point at retinal location $(x, y)^\top$ is represented by $\boldsymbol{q} = (\mu x, \mu y, \mu)^\top$, with $\mu \neq 0$. Note that the inhomogeneous coordinates can be recovered from $\boldsymbol{q}/\mu$. Scalar multiples of the homogeneous coordinates represent the same image point. For example, if $\boldsymbol{p} = (\lambda x, \lambda y, \lambda)^\top$ then $\boldsymbol{q}/\mu = \boldsymbol{p}/\lambda$; this relationship will be written as $\boldsymbol{p} \sim \boldsymbol{q}$.

A line in the image plane has homogeneous coordinates $\boldsymbol{n} = (a, b, c)^\top$, such that $\boldsymbol{q}$ is on $\boldsymbol{n}$ if $\boldsymbol{n}^\top \boldsymbol{q} = 0$. Scalar multiples represent the same line; if $\boldsymbol{m} = (\kappa a, \kappa b, \kappa c)^\top$, then $\boldsymbol{m} \sim \boldsymbol{n}$, with $\boldsymbol{m}^\top \boldsymbol{q} = 0$, as before. If $\boldsymbol{n}$ is defined as $\boldsymbol{n} = \boldsymbol{p} \times \boldsymbol{q}$, then $\boldsymbol{n}^\top \boldsymbol{p} = \boldsymbol{n}^\top \boldsymbol{q} = 0$; hence $\boldsymbol{n}$ is the line through the two points. Similarly, given any pair of lines $\boldsymbol{m}$ and $\boldsymbol{n}$, if $\boldsymbol{q} = \boldsymbol{m} \times \boldsymbol{n}$ then $\boldsymbol{m}^\top \boldsymbol{q} = \boldsymbol{n}^\top \boldsymbol{q} = 0$; hence $\boldsymbol{q}$ is the intersection point of the two lines [42].

The left and right optical centres are labelled $\bar{\boldsymbol{c}}_\ell$ and $\bar{\boldsymbol{c}}_r$, respectively. The difference between these locations defines the 'baseline' vector $\boldsymbol{b}$, while the cyclopean point $\bar{\boldsymbol{c}}_b$ is fixed halfway between the eyes[16, 27];

$$\boldsymbol{b} = \bar{\boldsymbol{c}}_r - \bar{\boldsymbol{c}}_\ell \tag{1}$$

$$\bar{\boldsymbol{c}}_b = \tfrac{1}{2}(\bar{\boldsymbol{c}}_\ell + \bar{\boldsymbol{c}}_r). \tag{2}$$

Only the *ratio* of the scene size to the baseline length can be recovered from the images, in the absence of other information. For this reason it is helpful to define the distance between the two optical centres as $|\boldsymbol{b}| = 1$, so that Euclidean coordinates are measured in units of inter-ocular separation. The location of the scene coordinate-system is immaterial, so it will be convenient to put the origin at the cyclopean point $\bar{\boldsymbol{c}}_b$. The coordinates of the optical centres, with reference to figure 1, will be

$$\bar{\boldsymbol{c}}_\ell = -\left(\tfrac{1}{2}, 0, 0\right)^\top \quad \text{and} \quad \bar{\boldsymbol{c}}_r = \left(\tfrac{1}{2}, 0, 0\right)^\top. \tag{3}$$

The baseline vector (1) is therefore parallel to the $\boldsymbol{x}$ axis, and a perpendicular axis $\boldsymbol{z} = (0, 0, 1)^\top$ will be taken as the head-centric outward direction. These two vectors define Cartesian coordinates in the horizontal plane. The downward normal of this plane is $\boldsymbol{y} = (0, 1, 0)^\top$, so that the axes $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$ form a right-handed system, as shown in figure 1.

The orientations of the left, right and cyclopean eyes are expressed by 3×3 rotation matrices $\boldsymbol{R}_\ell$, $\boldsymbol{R}_r$ and $\boldsymbol{R}$, respectively. A view of the scene is obtained by expressing each point $\bar{\boldsymbol{q}}$ relative to an optical centre, and applying the corresponding rotation. The homogeneous perspective projection into the left image $\mathcal{I}_\ell$ is, for example

$$\boldsymbol{p}_\ell \sim \boldsymbol{R}_\ell(\bar{\boldsymbol{q}} - \bar{\boldsymbol{c}}_\ell) \tag{4}$$

and similarly for the right image, $\mathcal{I}_r$. If the scale factor in this equation is known, then $\boldsymbol{p}_\ell = (x_\ell, y_\ell, z_\ell)^\top$, where the 'depth' z_ℓ is the distance to $\bar{\boldsymbol{q}}$ along the optical axis of the eye. The triple $(x_\ell, y_\ell, z_\ell)^\top$ will be called the (left) 'eye coordinates' of $\bar{\boldsymbol{q}}$.

The use of the above notation will now be illustrated, in a short example. Suppose that both eyes are looking straight ahead, with zero cyclo-rotation, $\boldsymbol{R}_\ell = \boldsymbol{R}_r = \boldsymbol{I}$. It follows that the projections of $\bar{\boldsymbol{q}} = (x, y, z)^\top$ can be computed easily; they are $\boldsymbol{q}_\ell \sim \left(x + \frac{1}{2}, y, z\right)^\top$ and $\boldsymbol{q}_r \sim \left(x - \frac{1}{2}, y, z\right)^\top$. Division by z gives the left and right coordinates $\left((x + \frac{1}{2})/z,\, y/z,\, 1\right)^\top$ and $\left((x - \frac{1}{2})/z,\, y/z,\, 1\right)^\top$, respectively. The difference between these points, taken in the 2-D image-plane, is the binocular disparity, $(1/z, 0)^\top$. Note that, because the visual axes are parallel, the disparity vector is confined to the horizontal direction. For general orientations of the eyes, disparity equations are more complicated, as will be seen in section 6.

3 Binocular Orientation

A cyclopean parameterization of binocular eye movements is introduced in this section, based on the azimuth, elevation and distance of the fixation point. The role of cyclo-rotation in the present work will also be discussed. The classical binocular vergence and version parameters are reviewed, and related to the present account. The parameterization will be used to construct the geometric horopter in section 4, and the epipolar geometry in section 5.

As was noted in section 1.1, the degrees of freedom of the binocular system can be reduced from six, to three. The reduction is achieved by imposing the fixation constraint, together with Donders' law. An appropriate parameterization will be developed, based on the cyclopean azimuth, elevation and distance of the fixation point. It will be shown that this representation complements the classical vergence/version coordinates[17].

Suppose that the point $\bar{\boldsymbol{p}}_0$ is to be fixated. The scene coordinates of this point can be specified by a head-fixed direction $\boldsymbol{v}$ from the cyclopean origin, in conjunction with a distance, ρ, along the corresponding ray;

$$\bar{\boldsymbol{p}}_0 = \rho \boldsymbol{v}. \tag{5}$$

The direction $\boldsymbol{v}$ is a unit-vector, and the positive scalar ρ will be called the *range* of the fixation point $\bar{\boldsymbol{p}}_0$. The cyclopean direction may be written in terms of the elevation and azimuth angles α and β respectively;

$$\boldsymbol{v} = (\sin\beta,\ -\sin\alpha\,\cos\beta,\ \cos\alpha\,\cos\beta)^\top \tag{6}$$

where $\cos\beta$ is the projected length of $\boldsymbol{v}$ in the mid-sagittal plane $x = 0$, which divides one side of the head from the other. Note that the elevation α is positive for points above the horizontal plane ($y < 0$), and that the azimuth β is positive for points to the right of the mid-sagittal plane ($x > 0$). These visual angles will each be in the range $[-\pi/2, \pi/2]$, so that any point with $z \geq 0$ can be identified, as shown in figure 1. If the fixation point $\bar{\boldsymbol{p}}_0 = (x, y, z)^\top$ is given in Cartesian coordinates, then the corresponding range and direction are

$$\rho = |\bar{\boldsymbol{p}}_0| \tag{7}$$

$$\boldsymbol{v} = \bar{\boldsymbol{p}}_0/\rho \tag{8}$$

respectively. The elevation and cyclopean azimuth angles can be obtained from the equations $\tan\alpha = -y/z$ and $\sin\beta = x/\rho$ respectively. The vector $(\alpha, \beta, \rho)^\top$ contains the *Helmholtz coordinates* of the point $\bar{\boldsymbol{p}}_0$.

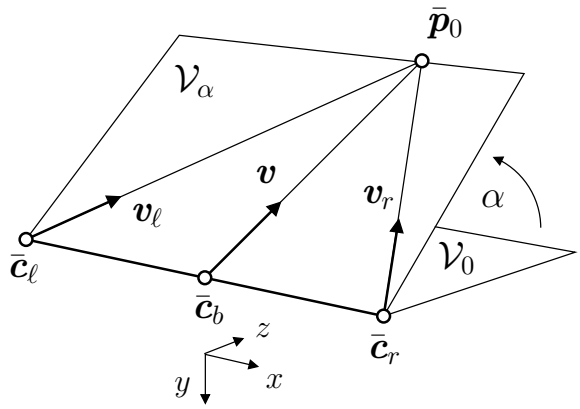

Figure 1: Visual directions. A visual plane $\mathcal{V}_\alpha$ is defined by the optical centres $\bar{\boldsymbol{c}}_\ell$ and $\bar{\boldsymbol{c}}_r$, together with the fixation point $\bar{\boldsymbol{p}}_0$. The visual directions $\boldsymbol{v}$, $\boldsymbol{v}_\ell$ and $\boldsymbol{v}_r$ lie in this plane, which has an elevation angle α. The scene coordinates are located at the cyclopean point $\bar{\boldsymbol{c}}_b = (0, 0, 0)^\top$, such that $\mathcal{V}_0$ coincides with the x, z plane.

In addition to the cyclopean visual axis $\boldsymbol{v}$, defined above (8), there exist left and right axes $\boldsymbol{v}_\ell$ and $\boldsymbol{v}_r$, respectively. If the eyes are fixating the point $\bar{\boldsymbol{p}}_0$, as described above, then $\boldsymbol{v}_\ell$ and $\boldsymbol{v}_r$ can be derived from $\boldsymbol{v}$ and ρ, as will be shown below. The optical centres $\bar{\boldsymbol{c}}_\ell$ and $\bar{\boldsymbol{c}}_r$, together with the fixation point $\bar{\boldsymbol{p}}_0$ define a *visual plane*, $\mathcal{V}_\alpha$ as shown in figure 1. The three visual axes intersect at $\bar{\boldsymbol{p}}_0$, and so $\boldsymbol{v}_\ell$, $\boldsymbol{v}_r$ and $\boldsymbol{v}$ lie in $\mathcal{V}_\alpha$. All of the possible visual planes contain the baseline $\boldsymbol{b}$, and may be parameterized by the dihedral angle α between $\mathcal{V}_\alpha$ and the horizontal plane $\mathcal{V}_0$. The azimuth angles β, β_ℓ and β_r will now be defined *in* the visual plane $\mathcal{V}_\alpha$.

Firstly it will be shown that, if the eyes are fixating, then the left and right visual directions can be simultaneously parameterized by the cyclopean direction and distance of the fixation point. It is convenient to begin by assuming that the fixation point is in the horizontal plane, such that $\alpha = 0$. The role of this assumption will be discussed subsequently. It can be seen, with reference to figure 2, that if the baseline separation is $|\boldsymbol{b}| = 1$, then $\tan\beta_\ell$ and $\tan\beta_r$ are equal to $(\rho\sin\beta \pm \frac{1}{2})/(\rho\cos\beta)$. Some re-arrangement leads to the definitions

$$\tan\beta_\ell = \tan\beta + \frac{\sec\beta}{2\rho} \quad \text{and} \quad \tan\beta_r = \tan\beta - \frac{\sec\beta}{2\rho}. \tag{9}$$

It is clear from these equations that, for a given cyclopean azimuth β, the visual directions become more equal as the fixation distance, ρ, increases. It may also be noted that if $\beta = 0$,

then the fixation is symmetric, with left and right azimuths $\pm\tan^{-1}(\frac{1}{2}/\rho)$, as is commonly assumed in the literature. If the fixation point $\bar{\boldsymbol{p}}_0$ is in $\mathcal{V}_\alpha$, then the matrices representing the orientation of the eyes are easily constructed. For example, the matrix $\boldsymbol{R}_\ell$ in (4) is

$$\boldsymbol{R}_\ell = \begin{pmatrix} \cos\beta_\ell & 0 & -\sin\beta_\ell \\ 0 & 1 & 0 \\ \sin\beta_\ell & 0 & \cos\beta_\ell \end{pmatrix}. \tag{10}$$

The analogous definitions are made for the matrices $\boldsymbol{R}$ and $\boldsymbol{R}_r$, with angles β and β_r, respectively.

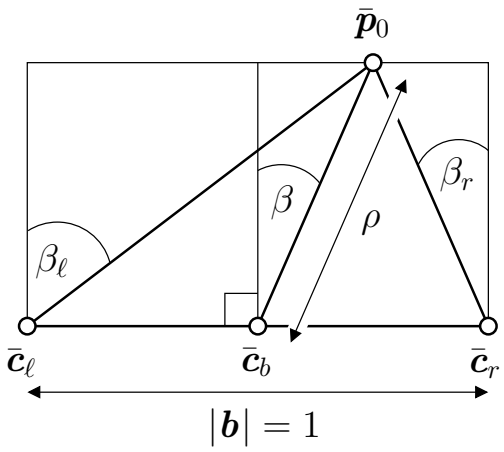

Figure 2: Binocular coordinates. The fixation point $\bar{\boldsymbol{p}}_0$, in the visual plane $\mathcal{V}_\alpha$, is shown. The optical centres are indicated by $\bar{\boldsymbol{c}}_\ell$ and $\bar{\boldsymbol{c}}_r$. The azimuth angles β and β_ℓ are positive in this example, whereas β_r is negative. The cyclopean range of the fixation point is ρ.

Although the Helmholtz coordinates are convenient for specifying visual directions, the eyes do not, in general, rotate around the corresponding axes. An important characteristic of actual eye movements is that, for general fixation points, each eye will be cyclo-rotated around the corresponding visual direction. Although the observed cyclo-rotation angles γ_ℓ and γ_r are non-zero, Donders' law states that they are completely determined by the corresponding visual directions; hence there exist functions $\gamma_\ell(\alpha, \beta_\ell)$ and $\gamma_r(\alpha, \beta_r)$. The definition of these functions can be obtained from Listing's law and its extensions [43, 44, 45].

Cyclo-rotation, like the azimuth and distance of the fixation point, has a significant effect on the binocular disparity field [27, 15]. The angles γ_ℓ and γ_r are, however, determined by the cyclopean parameters α, β and ρ. This follows from Donders' law, via (6) and (9), as indicated above. Hence, in order to develop a *minimal* oculomotor parameterization, it is convenient to make the simplifying assumption

$$\gamma_\ell(\alpha, \beta_\ell) = \gamma_r(\alpha, \beta_r) = 0 \tag{11}$$

which is (trivially) consistent with Donders' law. The practical advantage of this restriction is that any dependence on the elevation angle α is removed from the analysis. This makes it possible to study the binocular geometry with respect to fixation points in a *single* visual plane. Furthermore, Listing's law (including the 'L-2' extension) agrees with (11) when the elevation α is zero [44, 45]. This makes it is useful, as well as convenient, to choose the horizontal plane $\mathcal{V}_0$ for further investigation [14].

The above approximation (11) is good for $\alpha \approx 0$ and, in general, it is straightforward to incorporate any cyclo-rotation model (e.g. L-2) into the *geometric* framework described below. For example, in section 6, the scalar binocular disparity is defined, at each retinal point, in the direction of the epipole. Both the point and the epipole can be cyclo-rotated as a function of the fixation point. Furthermore, note that these rotations do not change the magnitude of the

disparity vectors. Although this procedure can be used to *describe* the effect of cyclo-rotation, it does not say how the visual system should cope with it. Some suggestions will, however, be made in section 7.

The vergence angle, δ, will be defined as the angle between the lines of sight at the fixation point; the version angle, ϵ, will be defined as the average gaze azimuth. In relation to the Helmholtz coordinates, this means that

$$\delta = \beta_\ell - \beta_r \tag{12}$$

$$\epsilon = \tfrac{1}{2}(\beta_\ell + \beta_r). \tag{13}$$

The vergence angle δ is non-negative, owing to the inequality $\beta_r \leq \beta_\ell$, which follows from the signs and limits of β_ℓ and β_r as defined above. The equality $\beta_\ell = \beta_r$ occurs for infinitely distant fixation points, for which $\delta = 0$. These definitions are illustrated in figure 3.

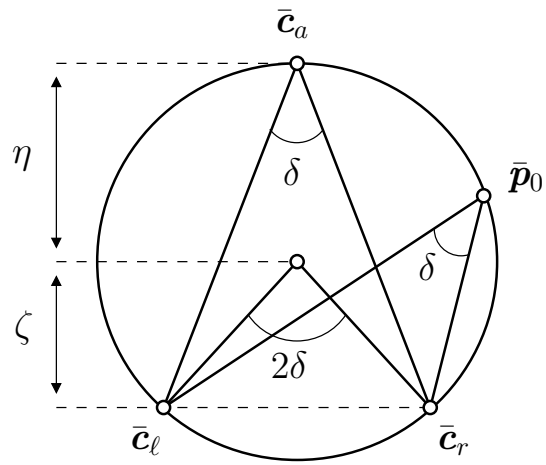

Figure 3: Vergence geometry. The Vieth-Müller circle is defined by the positions of the optical centres $\bar{\boldsymbol{c}}_\ell$ and $\bar{\boldsymbol{c}}_r$, together with the fixation point, $\bar{\boldsymbol{p}}_0$. The forward ($z > 0$) arc of the circle intersects the mid-sagittal plane at the point $\bar{\boldsymbol{c}}_a$. The vergence angle, δ, is inscribed at $\bar{\boldsymbol{p}}_0$ by $\bar{\boldsymbol{c}}_\ell$ and $\bar{\boldsymbol{c}}_r$. The same angle is inscribed at all other points on the circle, including $\bar{\boldsymbol{c}}_a$.

The properties of the vergence and version parameters can be understood with reference to the Vieth-Müller circle[46], which is defined by the two optical centres $\bar{\boldsymbol{c}}_\ell$ and $\bar{\boldsymbol{c}}_r$, together with the fixation point $\bar{\boldsymbol{p}}_0$. The vergence, δ, is the inscribed angle at $\bar{\boldsymbol{p}}_0$, being opposite the inter-ocular axis $\boldsymbol{b}$. The law of sines gives the diameter of the circumcircle as $1/\sin\delta$, with $|\boldsymbol{b}| = 1$, as usual. The angle subtended by $\boldsymbol{b}$ from the centre of the circle is 2δ, being twice the inscribed angle. The isosceles triangle formed by $\bar{\boldsymbol{c}}_\ell$, $\bar{\boldsymbol{c}}_r$ and the centre of the circle can be split into two right-angled triangles, such that $\tan\delta = \frac{1}{2}/\zeta$, where ζ is the z coordinate of the centre. It follows that the Vieth-Müller circle is centred at the point $(0, 0, \zeta)^\top$, and has radius η, where

$$\zeta = \tfrac{1}{2}\cot\delta \tag{14}$$

$$\eta = \tfrac{1}{2}\csc\delta. \tag{15}$$

The optical centres, $\bar{\boldsymbol{c}}_\ell$ and $\bar{\boldsymbol{c}}_r$, divide the Vieth-Müller circle into two arcs, according to the sign of z. The forward ($z \geq 0$) arc contains the fixation point, $\bar{\boldsymbol{p}}_0$, with inscribed angle δ. Furthermore, the inscribed angles at all other points $\bar{\boldsymbol{q}}_{\mathrm{VM}}$ on this arc must be equal; hence the Vieth-Müller circle contains the locus of iso-vergence.

The version angle ϵ gives the azimuth of $\bar{\boldsymbol{p}}_0$ from a cyclopean point $\bar{\boldsymbol{c}} = (0, 0, \zeta - \eta)^\top$, which lies at the back of the Vieth-Müller circle, as shown in figure 4. Evidently the location of the point $\bar{\boldsymbol{c}}$ varies according to the vergence angle, as the radius of the circle is determined by the latter. This is one reason for *deriving* the (δ, ϵ) parameterization from the (β, ρ) parameterization, as above. The present analysis has a fixed reference point $\bar{\boldsymbol{c}}_b = (0, 0, 0)^\top$, meaning that visual

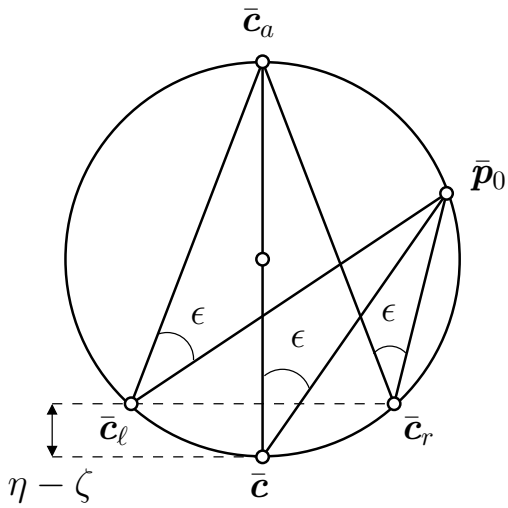

Figure 4: Version geometry. The points $\bar{\boldsymbol{c}}_a$ and $\bar{\boldsymbol{p}}_0$ inscribe the version angle, ϵ, at an optical centre, $\bar{\boldsymbol{c}}$, which is located on the backward ($z < 0$) arc of the Vieth-Müller circle. The same angle is inscribed at $\bar{\boldsymbol{c}}_\ell$ and $\bar{\boldsymbol{c}}_r$. It follows that as $\bar{\boldsymbol{p}}_0$ is fixated, $\bar{\boldsymbol{c}}_a$ lies in the same visual *direction* from each eye. Furthermore, the triangle defined by $\bar{\boldsymbol{c}}_\ell$, $\bar{\boldsymbol{c}}_r$, and $\bar{\boldsymbol{c}}_a$ is isosceles, and so the point $\bar{\boldsymbol{c}}_a$ is at the same *distance* from each eye.

information can easily be combined as the eyes re-fixate. Furthermore, the range parameter ρ can be interpreted directly, whereas the vergence parameter δ is measured in relation to the oculomotor system. Nonetheless, the vergence and version parameters are essential to the geometry of visual fixation, as will be shown in the following sections.

4 Fixed Points

In this section it will be shown that, for a given vergence angle, certain points in the scene project to the same location in each image. These points constitute the geometric *horopter* of the fixating system. The constructions that are given here will be used to construct the epipolar geometry of the two images, in section 5. For this purpose, it will be convenient to study the horopter in the absence of cyclo-rotation.

It was shown in section 3 that the Vieth-Müller circle is defined by the optical centres $\bar{\boldsymbol{c}}_\ell$ and $\bar{\boldsymbol{c}}_r$, together with the fixation point $\bar{\boldsymbol{p}}_0$. Consider another scene-point point, $\bar{\boldsymbol{q}}_{\text{VM}}$, that lies on the forward section of the Vieth-Müller circle. This point is in the visual plane $\mathcal{V}_0$, and therefore satisfies the equation $y = 0$, as well as the conditions

$$\begin{aligned} x^2 + (z - \zeta)^2 &= \eta^2 \\ z &\geq 0. \end{aligned} \tag{16}$$

The two points $\bar{\boldsymbol{p}}_0$ and $\bar{\boldsymbol{q}}_{\text{VM}}$, both of which are on the forward section of the Vieth-Müller circle, must inscribe equal angles at the optical centres. The point $\bar{\boldsymbol{p}}_0$ is being fixated, and therefore appears in the fovea of each image. Hence the projected point $\boldsymbol{q}_{\text{VM}}$ is 'fixed' with respect to the mapping between the left and right images [40, 42]. It appears on the horizontal meridian of each retina, at the same angular offset from the corresponding fovea. This can be re-stated in eye-coordinates as $x_\ell/z_\ell = x_r/z_r$ and $y_\ell/z_\ell = y_r/z_r = 0$, for any point on the Vieth-Müller circle.

The Vieth-Müller circle does not, however, constitute the complete horopter. The remaining points can be found, in this case, by solving the equations $x_\ell = x_r$, $y_\ell = y_r$ and $z_\ell = z_r$. Any scene-point that satisfies these equations is fixed with respect to the rigid-body transformation between the left and right eyes, as well as with respect to the mapping between the images.

Recall that the Euclidean coordinates of $\bar{\boldsymbol{q}} = (x, y, z)^\top$ in the left and right eye frames are $\bar{\boldsymbol{q}}_\ell = \boldsymbol{R}_\ell(\bar{\boldsymbol{q}} - \bar{\boldsymbol{c}}_\ell)$ and $\bar{\boldsymbol{q}}_r = \boldsymbol{R}_r(\bar{\boldsymbol{q}} - \bar{\boldsymbol{c}}_r)$, respectively. The point $\bar{\boldsymbol{q}}$ is fixed with respect to the left/right transformation if $\bar{\boldsymbol{q}}_\ell = \bar{\boldsymbol{q}}_r$, which in turn implies that $|\bar{\boldsymbol{q}}_\ell|^2 = |\bar{\boldsymbol{q}}_r|^2$. The squared-lengths are preserved by the rotation matrices $\boldsymbol{R}_\ell$ and $\boldsymbol{R}_r$, and so $|\bar{\boldsymbol{q}} - \bar{\boldsymbol{c}}_\ell|^2 = |\bar{\boldsymbol{q}} - \bar{\boldsymbol{c}}_r|^2$. From the definition of $\bar{\boldsymbol{c}}_\ell$ and $\bar{\boldsymbol{c}}_r$ in (3), this is equivalent, in scene-coordinates, to the condition $(x + \frac{1}{2})^2 = (x - \frac{1}{2})^2$. Hence it can be seen that any such point must lie in the mid-sagittal plane $x = 0$, that divides one side of the head from the other. Substituting $x = 0$ into (16) leads immediately to $z = \zeta + \eta$, leaving y free to vary. In general, $y_\ell = y_r$, because the axis of the vergence rotation is perpendicular to the visual plane $\mathcal{V}_\alpha$. This argument has established that there is an axis of points $\bar{\boldsymbol{q}}_a$ that are fixed with respect to the rigid-body transformation between the left and right eyes. This axis intersects the Vieth-Müller circle at a point $\bar{\boldsymbol{c}}_a$, and is perpendicular to the visual plane. If, as previously supposed, $\alpha = 0$, then the coordinates of these points are

$$\bar{\boldsymbol{c}}_a = (0, 0, \zeta + \eta)^\top \tag{17}$$

$$\bar{\boldsymbol{q}}_a = \bar{\boldsymbol{c}}_a + (0, y, 0)^\top. \tag{18}$$

This axis of points, $\bar{\boldsymbol{q}}_a$, which has been identified elsewhere[46, 47], is the geometric *midline horopter*. The point $\bar{\boldsymbol{c}}_a$ is the *pole* of the planar transformation induced by the translation $\boldsymbol{b}$ and vergence rotation $\boldsymbol{R}_r\boldsymbol{R}_\ell^\top$. The points $\bar{\boldsymbol{q}}_a$ lie on the associated screw-axis[39, 40].

It will be useful to compute the *image* coordinates of the axis, which are common to both eyes, as shown in figure 4. The points $\bar{\boldsymbol{p}}_0$ and $\bar{\boldsymbol{c}}_a$ inscribe equal angles at $\bar{\boldsymbol{c}}_\ell$, $\bar{\boldsymbol{c}}_r$ and $\bar{\boldsymbol{c}}$; moreover, the angle at $\bar{\boldsymbol{c}}$ is, by definition, the binocular version, ϵ. Having established that the angular direction of $\bar{\boldsymbol{q}}_a$ from either optical centre is ϵ, the common distance of this point will also be computed. The points $\bar{\boldsymbol{c}}_\ell$, $\bar{\boldsymbol{c}}_r$ and $\bar{\boldsymbol{c}}_a$ form an isosceles triangle, from which it can be seen that $|\bar{\boldsymbol{c}}_a| \sin(\delta/2) = \frac{1}{2}$. It follows that, in the coordinates of either eye, the axis is specified by

$$\boldsymbol{c}_a = \tfrac{1}{2}\csc(\delta/2)\,(-\sin\epsilon, 0, \cos\epsilon)^\top \tag{19}$$

$$\boldsymbol{q}_a = \boldsymbol{c}_a + (0, y, 0)^\top. \tag{20}$$

These image points lie on a vertical line $\boldsymbol{a}$, which has the same coordinates in each eye. The equation of the line is $\boldsymbol{q}_a^\top \boldsymbol{a} = 0$, and so it follows from (19) that the coordinates of the line are determined by the version angle ϵ;

$$\boldsymbol{a} \sim (\cos\epsilon, 0, -\sin\epsilon)^\top. \tag{21}$$

The results of this section can be summarized as follows. If a scene point $\bar{\boldsymbol{q}}$ is on the geometric horopter, then the image coordinates of the corresponding points are equal, $\boldsymbol{q}_\ell \sim \boldsymbol{q}_r$. The geometric horopter, in the absence of cyclo-rotation, consists of the forward part of the Vieth-Müller circle, together with the midline component. Furthermore, the image coordinates (21) of the midline part are determined by the binocular version angle. It will be seen, in the following section, that the epipolar geometry can be constructed via the vertical horopter. The epipolar geometry extends the cyclopean parameterization out of the visual plane, and leads to geometric constraints that are defined across the entire left and right images.

It should be noted that, in the presence of cyclo-rotation, the geometric horopter takes the form of a twisted cubic curve[16]. This curve coincides with the Vieth-Müller circle as it passes through the optical centres, and has asymptotes at $\bar{\boldsymbol{c}}_a \pm (0, y, 0)^\top$.

5 Epipolar Geometry

It was established, in the preceding section, that certain scene points have the same coordinates in both images. The related *epipolar constraint* is weaker, but much more useful, as it applies to *all* scene points. The epipolar geometry of the fixating system will now be described; in particular, the image of the midline horopter (21), will be used to construct the appropriate *essential matrix* [20].

The epipolar constraint is as follows: Given an image point $\boldsymbol{q}_\ell$ in $\mathcal{I}_\ell$, the corresponding point $\boldsymbol{q}_r$ in $\mathcal{I}_r$ must lie on a known epipolar line $\boldsymbol{u}_r$, such that $\boldsymbol{u}_r^\top \boldsymbol{q}_r = 0$. The geometric interpretation of this is that the scene point $\bar{\boldsymbol{q}}$ must be located on the ray defined by the optical centre $\bar{\boldsymbol{c}}_\ell$ and the image point $\boldsymbol{q}_\ell$; the ray projects to a line $\boldsymbol{u}_r$ in the other view, and so $\boldsymbol{q}_r$, being another image of $\bar{\boldsymbol{q}}$, must lie on the line. Furthermore, note that the optical centre $\bar{\boldsymbol{c}}_\ell$ is common to all such rays, and so the resulting lines $\boldsymbol{u}_r$ must intersect at a single point in $\mathcal{I}_r$. This point is the right *epipole*, $\boldsymbol{e}_r$. Similar arguments can be used to introduce the left epipole $\boldsymbol{e}_\ell$, as well as the associated lines $\boldsymbol{u}_\ell$ in $\mathcal{I}_\ell$.

Suppose that the point $\boldsymbol{q}_\ell$ is given; then, with reference to figure 5, $\boldsymbol{u}_\ell = \boldsymbol{e}_\ell \times \boldsymbol{q}_\ell$. Furthermore, this line intersects the projection $\boldsymbol{a}$ of the midline horopter (21) at the image point $\boldsymbol{q}_a = \boldsymbol{a} \times \boldsymbol{u}_\ell$. Any point on $\boldsymbol{u}_\ell$ must be the projection of a scene-point in the plane defined by $\bar{\boldsymbol{c}}_\ell$, $\bar{\boldsymbol{c}}_r$ and $\bar{\boldsymbol{q}}_a$. This scene point must, therefore, also project onto the other epipolar line, $\boldsymbol{u}_r$. Hence $\boldsymbol{u}_r$ can be constructed from $\boldsymbol{e}_r$ and the point in $\mathcal{I}_r$ that corresponds to $\boldsymbol{q}_a$. Furthermore, $\boldsymbol{q}_a$ is a fixed point (being on $\boldsymbol{a}$), and so its coordinates are unchanged in $\mathcal{I}_r$. It follows that $\boldsymbol{u}_r = \boldsymbol{e}_r \times \boldsymbol{q}_a$. The preceding construction may be summarized as

$$\boldsymbol{u}_r \sim \boldsymbol{e}_r \times \big(\boldsymbol{a} \times (\boldsymbol{e}_\ell \times \boldsymbol{q}_\ell)\big). \tag{22}$$

This equation will now be put into a more useful form. Suppose that $\boldsymbol{w} = (x, y, z)^\top$; then the cross product $\boldsymbol{w} \times \boldsymbol{p}$ can be expressed as a matrix-vector multiplication, $(\boldsymbol{w}\times)\boldsymbol{p}$, where

$$\big(\boldsymbol{w}\times\big) = \begin{pmatrix} 0 & -z & y \\ z & 0 & -x \\ -y & x & 0 \end{pmatrix} \tag{23}$$

is a 3×3 antisymmetric matrix, constructed from the components of $\boldsymbol{w}$. Consider the part of equation (22) that does not depend on the particular choice of point $\boldsymbol{q}_\ell$; the equivalence (23) can be used to express this as a transformation

$$\boldsymbol{E} \sim \big(\boldsymbol{e}_r\times\big)\big(\boldsymbol{a}\times\big)\big(\boldsymbol{e}_\ell\times\big) \tag{24}$$

which is the 3×3 essential matrix[20]. Given a point $\boldsymbol{q}_\ell$, the corresponding point $\boldsymbol{q}_r$ must be on a certain epipolar line $\boldsymbol{u}_r$, as described above. This constraint is expressed via the essential matrix as

$$\boldsymbol{q}_r^\top \boldsymbol{E} \boldsymbol{q}_\ell = 0 \tag{25}$$

where $\boldsymbol{u}_r \sim \boldsymbol{E}\boldsymbol{q}_\ell$. The analogous constraint, $\boldsymbol{q}_\ell^\top \boldsymbol{E}^\top \boldsymbol{q}_r = 0$ applies in the opposite direction, the epipolar line being $\boldsymbol{u}_\ell \sim \boldsymbol{E}^\top \boldsymbol{q}_r$ in this case. The epipoles, as described above, are each the image of the 'other' optical centre. This means that $\boldsymbol{e}_\ell \sim \boldsymbol{R}_\ell(\boldsymbol{b})$, and $\boldsymbol{e}_r \sim \boldsymbol{R}_r(-\boldsymbol{b})$, where $\boldsymbol{b}$ is the vector between the optical centres. Equations (1), (3) and (4) can be used to show that the epipoles are simply

$$\boldsymbol{e}_\ell \sim (\cos\beta_\ell,\, 0,\, \sin\beta_\ell)^\top \quad \text{and} \quad \boldsymbol{e}_r \sim (-\cos\beta_r,\, 0,\, -\sin\beta_r)^\top. \tag{26}$$

These equations can be combined with the definition of the geometric midline horopter (21), to give a parametric structure to the essential matrix. The non-zero terms E_{ij} in the matrix product (24) are found to be $E_{12} = -E_r \sin\beta_r$, $E_{21} = E_\ell \sin\beta_\ell$, $E_{23} = -E_\ell \cos\beta_\ell$ and $E_{22} = E_r \cos\beta_r$, where $E_\ell = \cos\beta_r \cos\epsilon + \sin\beta_r \sin\epsilon$ and $E_r = \cos\beta_\ell \cos\epsilon + \sin\beta_\ell \sin\epsilon$. The factors E_ℓ and E_r are seen to be the angle-difference expansions of $\cos(\beta_\ell - \epsilon)$ and $\cos(\beta_r - \epsilon)$, respectively. Furthermore, by reference to (12,13) the arguments $\beta_\ell - \epsilon$ and $\beta_r - \epsilon$ are equal to $\pm\delta/2$, and so it follows from the even-symmetry of the cosine function that $E_\ell = E_r = \cos(\delta/2)$. The essential matrix is defined here as as a homogeneous transformation (cf. 25), and so this common scale-factor can be disregarded, which leaves

$$\boldsymbol{E} \sim \begin{pmatrix} 0 & -\sin\beta_r & 0 \\ \sin\beta_\ell & 0 & -\cos\beta_\ell \\ 0 & \cos\beta_r & 0 \end{pmatrix}. \tag{27}$$

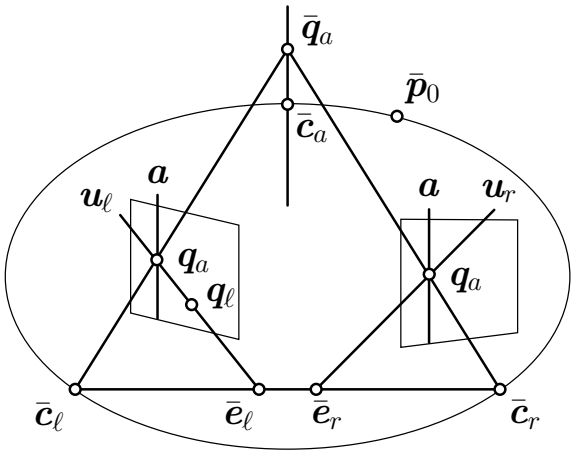

Figure 5: Construction of the epipolar geometry. Point $\boldsymbol{q}_\ell$ is given, so the epipolar line in $\mathcal{I}_\ell$ is $\boldsymbol{u}_\ell = \boldsymbol{q}_\ell \times \boldsymbol{e}_\ell$. This line intersects the image $\boldsymbol{a}$ of the midline horopter in $\mathcal{I}_\ell$ at $\boldsymbol{q}_a = \boldsymbol{a} \times \boldsymbol{u}_\ell$. The point $\boldsymbol{q}_a$ is on $\boldsymbol{a}$, and is therefore fixed, having the same coordinates $\boldsymbol{q}_a$ in $\mathcal{I}_r$. It follows that the epipolar line in $\mathcal{I}_r$ is $\boldsymbol{u}_r = \boldsymbol{e}_r \times \boldsymbol{q}_a$. The location of $\boldsymbol{q}_r$, which corresponds to $\boldsymbol{q}_\ell$, is unknown, but it must lie on $\boldsymbol{u}_r$. The Vieth-Müller circle, determined by the fixation point $\bar{\boldsymbol{p}}_0$, is shown in the figure, as is the midline-horopter, which passes through points $\bar{\boldsymbol{c}}_a$ and $\bar{\boldsymbol{q}}_a$.

It is straightforward to verify that $\boldsymbol{E}$ is indeed an essential matrix, having one singular-value equal to zero, and two identical non-zero singular-values (here equal to unity).

The same result can be obtained from a variant of the more traditional definition of the Essential matrix [20, 48], $\boldsymbol{E} \sim \boldsymbol{R}_r(\boldsymbol{b}\times)\boldsymbol{R}_\ell^\top$. This definition is, however, not specific to the case of visual fixation, and offers correspondingly less insight into the present configuration. The essential matrix has, in general, five degrees of freedom; three for relative orientation, and three for translation, minus one for overall scaling [3]. The essential matrix obtained above (27) has just two parameters, β_ℓ and β_r. This simplification of the epipolar geometry is due to the fixation constraint, in conjunction with Donders' law.

6 Binocular Disparity

It was established in the preceding section that projected points in the left and right images must obey the epipolar constraint. In particular it was shown that, given a point $\boldsymbol{q}_\ell$ in $\mathcal{I}_\ell$, the corresponding point $\boldsymbol{q}_r$ must lie on a line $\boldsymbol{u}_r \sim \boldsymbol{E}\boldsymbol{q}_\ell$ in the other image, $\mathcal{I}_r$. The structure of the scene displaces the left and right image points along the corresponding lines, which pass through the left and right epipoles, respectively. These image displacements are quantified in this section. In particular, it is shown that a scene-point $\bar{\boldsymbol{q}}$ which is in cyclopean direction $\boldsymbol{p}_c$ will be projected to corresponding image-points $\boldsymbol{q}_\ell$ and $\boldsymbol{q}_r$, where

$$\boldsymbol{q}_\ell = \boldsymbol{p}_\ell + t_\ell(s)\,\boldsymbol{d}_\ell \quad \text{and} \quad \boldsymbol{q}_r = \boldsymbol{p}_r + t_r(s)\,\boldsymbol{d}_r. \tag{28}$$

The unit-vectors $\boldsymbol{d}_\ell$ and $\boldsymbol{d}_r$ point towards the respective epipoles $\boldsymbol{e}_\ell$ and $\boldsymbol{e}_r$. The relation to the epipolar geometry developed in section 5 is that $\boldsymbol{q}_\ell$ and $\boldsymbol{p}_\ell$ are on the same epipolar line; hence, in addition to $\boldsymbol{q}_r^\top \boldsymbol{E}\boldsymbol{q}_\ell = 0$, as in (25), it is true that $\boldsymbol{q}_r^\top \boldsymbol{E}\boldsymbol{p}_\ell = 0$. The formulation (28) makes an approximate correspondence between points $\boldsymbol{p}_\ell$ and $\boldsymbol{p}_r$, which is corrected by parallax functions $t_\ell(s)$ and $t_r(s)$. The common parameter s is the signed orthogonal distance of $\bar{\boldsymbol{q}}$ from the fronto-parallel plane $\mathcal{P}$ which passes through the fixation point $\bar{\boldsymbol{p}}_0$. This representation makes it possible, in principle, to estimate a 'local' cyclopean depth map at each fixation point. The local depth is a function $S(\boldsymbol{p}_c; \beta, \rho)$, where $\boldsymbol{p}_c$ parameterizes the cyclopean field of view, given the azimuth and range (β, ρ) of the fixation point.

The decomposition (28) has three important properties. Firstly, the unknown parallax variables are scalars; there is no need to consider horizontal and vertical disparities separately. Secondly, each image correspondence is parameterized by a single variable s, which has a direct interpretation as a Euclidean distance in the scene. Thirdly, for points close to the fixation plane, the predictions $\boldsymbol{p}_\ell$ and $\boldsymbol{p}_r$ will be close to $\boldsymbol{q}_\ell$ and $\boldsymbol{q}_r$ respectively. In particular, the predicted correspondence will be exact if the point $\bar{\boldsymbol{q}}$ lies in the fixation plane; $t_\ell(0) = t_r(0) = 0$. The decomposition (28) will now be described in detail, with respect to figure 6.

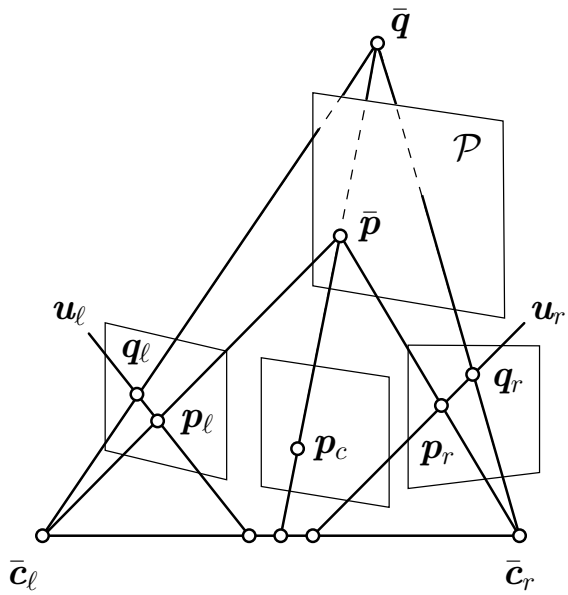

Figure 6: Geometry of cyclopean parallax. The fixation plane $\mathcal{P}$ is defined by the fixation point, and is parallel to the cyclopean image plane. Any point $\boldsymbol{p}_c$ defines a cyclopean ray, which intersects the fixation plane at $\bar{\boldsymbol{p}}$, and the scene at $\bar{\boldsymbol{q}}$. The scene point $\bar{\boldsymbol{q}}$ has depth s with respect to $\mathcal{P}$. The predicted image-projections of $\bar{\boldsymbol{q}}$ are at $\boldsymbol{p}_\ell$ and $\boldsymbol{p}_r$. The true projections, $\boldsymbol{q}_\ell$ and $\boldsymbol{q}_r$, are displaced along the corresponding epipolar lines, $\boldsymbol{u}_\ell$ and $\boldsymbol{u}_r$, respectively. The displacement can be parameterized by s, as described in the text.

The fixation plane $\mathcal{P}$, by definition, passes through the fixation point $\bar{\boldsymbol{p}}_0$, and is perpendicular to the cyclopean gaze direction. Hence the plane has an outward normal vector $\boldsymbol{v}$, as defined in (8). The plane consists of scene points in the set

$$\mathcal{P} = \Big\{\bar{\boldsymbol{p}} : \boldsymbol{v}^\top(\bar{\boldsymbol{p}} - \bar{\boldsymbol{p}}_0) = 0\Big\}. \tag{29}$$

The orthogonal distance from $\mathcal{P}$ to the cyclopean origin is equal to the range, ρ, of the fixation point, as defined in (7). The orthogonal distance from $\mathcal{P}$ to the scene point $\bar{\boldsymbol{q}}$ will be s; hence

$$\rho = \boldsymbol{v}^\top \bar{\boldsymbol{p}}_0 \tag{30}$$

$$s = \boldsymbol{v}^\top(\bar{\boldsymbol{q}} - \bar{\boldsymbol{p}}_0). \tag{31}$$

The range, ρ, is strictly positive, whereas s is negative, positive or zero, according to whether $\bar{\boldsymbol{q}}$ is closer than, further than, or in the plane $\mathcal{P}$, respectively. Note that s represents the *structure* of the scene with respect to $\mathcal{P}$. Equations (30) and (31) can now be used to decompose the cyclopean depth z_c of the point $\bar{\boldsymbol{q}}$ as follows;

$$z_c = \boldsymbol{v}^\top \bar{\boldsymbol{q}} = \rho + s. \tag{32}$$

The cyclopean ray through $\bar{\boldsymbol{q}}$ intersects the fixation plane at $\bar{\boldsymbol{p}}$. Hence the cyclopean coordinates of $\bar{\boldsymbol{q}}$ can be expressed as $z_c\boldsymbol{p}_c = \boldsymbol{R}\bar{\boldsymbol{q}}$, where $\boldsymbol{p}_c$ has been normalized such that $(\boldsymbol{p}_c)_3 = 1$, and

$\boldsymbol{R}$ encodes the orientation of the cyclopean eye (defined by the angle β, cf. equation 10). The scene-coordinates of points on the corresponding visual ray, parameterized by z_c, can be obtained by inverting this equation. In particular, the intersection $\bar{\boldsymbol{p}}$ of the ray with the fixation plane $\mathcal{P}$ can be obtained, as can the original scene-point $\bar{\boldsymbol{q}}$;

$$\bar{\boldsymbol{p}} = \rho \boldsymbol{R}^\top \boldsymbol{p}_c \tag{33}$$

$$\bar{\boldsymbol{q}} = z_c \boldsymbol{R}^\top \boldsymbol{p}_c. \tag{34}$$

These two points, which lie on the same cyclopean ray, will now be projected into the left image, $\mathcal{I}_\ell$, and the difference $t_\ell(s)\boldsymbol{d}_\ell$ between the two projections will be evaluated. The analogous derivation applies to the other image, $\mathcal{I}_r$, with subscripts 'ℓ' and 'r' exchanged. The left coordinates are

$$\rho_\ell \boldsymbol{p}_\ell = \rho \boldsymbol{R}_\ell \boldsymbol{R}^\top \boldsymbol{p}_c + \tfrac{1}{2}\boldsymbol{e}_\ell \tag{35}$$

$$z_\ell \boldsymbol{q}_\ell = z_c \boldsymbol{R}_\ell \boldsymbol{R}^\top \boldsymbol{p}_c + \tfrac{1}{2}\boldsymbol{e}_\ell \tag{36}$$

where $\frac{1}{2}\boldsymbol{e}_\ell = \boldsymbol{R}_\ell(-\boldsymbol{c}_\ell)$, as the left, right and cyclopean optical centres are collinear. Note that the image-points are normalized such that $(\boldsymbol{p}_\ell)_3 = (\boldsymbol{q}_\ell)_3 = 1$, as in the case of $\boldsymbol{p}_c$. It is necessary to consider the third components of the vectors $\boldsymbol{R}_\ell \boldsymbol{R}^\top \boldsymbol{p}_c$ and $\frac{1}{2}\boldsymbol{e}_\ell$ in (35) and (36) respectively; these are

$$\lambda_\ell = (\boldsymbol{R}_\ell \boldsymbol{R}^\top \boldsymbol{p}_c)_3 = x_c \sin(\beta_\ell - \beta) + \cos(\beta_\ell - \beta) \tag{37}$$

$$\mu_\ell = (\tfrac{1}{2}\boldsymbol{e}_\ell)_3 = \tfrac{1}{2}\sin\beta_\ell. \tag{38}$$

The actual image point $\boldsymbol{q}_\ell$ will now be decomposed into a sum of the predicted point $\boldsymbol{p}_\ell$, plus a scalar *parallax* in the direction of a unit vector $\boldsymbol{d}_\ell$. The vector $\boldsymbol{d}_\ell$ should be in the direction of $\boldsymbol{q}_\ell$ with respect to the epipole $\boldsymbol{e}_\ell$. Furthermore, the vector $\boldsymbol{d}_\ell$ must lie in the image plane, $(\boldsymbol{d}_\ell)_3 = 0$. However, it is desirable to avoid defining $\boldsymbol{d}_\ell$ from $\boldsymbol{p}_\ell - \frac{1}{2}\boldsymbol{e}_\ell/\mu_\ell$, because $\mu_\ell = 0$ whenever $\beta_\ell = 0$, as is clear from equation (38). Hence it is better to use

$$\boldsymbol{d}_\ell = \big(\mu_\ell \boldsymbol{p}_\ell - \tfrac{1}{2}\boldsymbol{e}_\ell\big)/\kappa_\ell \tag{39}$$

where $(\mu_\ell \boldsymbol{p}_\ell)_3 = (\frac{1}{2}\boldsymbol{e}_\ell)_3 = \mu_\ell$, and hence $(\boldsymbol{d}_\ell)_3 = 0$. The scalar $\kappa_\ell = \big|\mu_\ell \boldsymbol{p}_\ell - \frac{1}{2}\boldsymbol{e}_\ell\big|$ has been introduced, so that $\boldsymbol{d}_\ell$ is a unit vector. This is not strictly necessary, but has the advantage of imposing the original unit of measurement, $|\boldsymbol{b}|$, on the parallax function $t_\ell(s)$ that is associated with each scene point. The function $t_\ell(s)$ will now be derived. It follows from (35) and (36), along with the requirement $(\boldsymbol{p}_\ell)_3 = (\boldsymbol{q}_\ell)_3 = 1$, that the depth variables ρ_ℓ and z_ℓ can be expressed as affine functions of the corresponding cyclopean parameters ρ and z_c;

$$\rho_\ell = \lambda_\ell \rho + \mu_\ell \tag{40}$$

$$z_\ell = \lambda_\ell z_c + \mu_\ell \tag{41}$$

where λ_ℓ and μ_ℓ are the scalars identified by (37) and (38). A solution for λ_ℓ can be obtained from either of these equations, and substituted into the other. The resulting expression can then be solved for μ_ℓ;

$$\mu_\ell = \frac{\rho_\ell z_c - \rho z_\ell}{z_c - \rho}. \tag{42}$$

The equation (42) is independent of the point $\bar{\boldsymbol{q}}$ that is associated with depths z_c and z_ℓ. The result $\boldsymbol{q}_\ell = \boldsymbol{p}_\ell + t_\ell(s)\,\boldsymbol{d}_\ell$, as in (28), can now be derived in full. Equations (35) and (36) are used to express the actual projection $\boldsymbol{q}_\ell$ as a function of the predicted projection $\boldsymbol{p}_\ell$;

$$\rho z_\ell \boldsymbol{q}_\ell = \rho_\ell z_c \boldsymbol{p}_\ell - (z_c - \rho)\tfrac{1}{2}\boldsymbol{e}_\ell. \tag{43}$$

The quantity $\rho z_\ell \boldsymbol{p}_\ell$ is now subtracted from both sides of (43), and the resulting equation is re-arranged as follows;

$$\begin{aligned}
\rho z_\ell(\boldsymbol{q}_\ell - \boldsymbol{p}_\ell) &= (\rho_\ell z_c - \rho z_\ell)\boldsymbol{p}_\ell - (z_c - \rho)\tfrac{1}{2}\boldsymbol{e}_\ell \\
&= (z_c - \rho)\left(\frac{\rho_\ell z_c - \rho z_\ell}{z_c - \rho}\,\boldsymbol{p}_\ell - \tfrac{1}{2}\boldsymbol{e}_\ell\right) \\
&= (z_c - \rho)\big(\mu_\ell \boldsymbol{p}_\ell - \tfrac{1}{2}\boldsymbol{e}_\ell\big),
\end{aligned} \tag{44}$$

where the substitution of μ_ℓ has been made with reference to equation (42). Both sides of (44) are now divided by ρz_ℓ, and comparison with (39) leads to

$$\begin{aligned} \boldsymbol{q}_\ell - \boldsymbol{p}_\ell &= \frac{z_c - \rho}{\rho z_\ell} \left(\mu_\ell \boldsymbol{p}_\ell - \tfrac{1}{2}\boldsymbol{e}_\ell\right) \\ &= \frac{\kappa_\ell (z_c - \rho)}{\rho z_\ell}\, \boldsymbol{d}_\ell \\ &= \frac{\kappa_\ell s}{\rho z_\ell}\, \boldsymbol{d}_\ell \end{aligned} \tag{45}$$

where (32) has been used above, to make the substitution $s = z_c - \rho$. The practical problem with (45) is that in addition to the free parameter s, the variable z_ℓ is apparently unknown. This is resolved by making the substitution $z_\ell = \lambda_\ell(\rho + s) + \mu_\ell$, which follows from (32) and (41). Therefore, if $\boldsymbol{p}_\ell$ is added to both sides of (45), then the result is

$$\boldsymbol{q}_\ell = \boldsymbol{p}_\ell + t_\ell(s)\, \boldsymbol{d}_\ell \tag{46}$$

with

$$t_\ell(s) = \frac{\kappa_\ell (s/\rho)}{\lambda_\ell(\rho + s) + \mu_\ell}. \tag{47}$$

The analogous definitions are made for $\mathcal{I}_r$, with subscripts 'ℓ' and 'r' exchanged. Equations (45), (46) and (47) can be interpreted as follows. It is assumed that the cyclopean coordinates $(\boldsymbol{v}, \rho)$ of the fixation point $\bar{\boldsymbol{p}}_0$ are known. Then, given the cyclopean direction $\boldsymbol{p}_c$ of another point, it is possible to compute the predicted point $\boldsymbol{p}_\ell$ (35), as well as the vector $\kappa_\ell \boldsymbol{d}_\ell$ (39). The scalars λ_ℓ and μ_ℓ are obtained from (37) and (38), respectively. The unknown parallax, $t_\ell(s)$ is proportional to s/z_ℓ; this ratio is the depth of the scene point $\bar{\boldsymbol{q}}_\ell$ with respect to the fixation plane $\mathcal{P}$, divided by the depth of the point with respect to the left viewing direction.

For points that are on the fixation plane, $s = 0$, and therefore it is clear from (47) that $t_\ell(s) = 0$. It follows from (28) that $\boldsymbol{q}_\ell = \boldsymbol{p}_\ell$ and $\boldsymbol{q}_r = \boldsymbol{p}_r$. This makes it interesting to consider the relationship between $\boldsymbol{p}_\ell$ and $\boldsymbol{p}_r$. It can be shown that the points $\boldsymbol{p}_\ell$ can be mapped onto the corresponding points $\boldsymbol{p}_r$ by a projective transformation

$$\boldsymbol{p}_r = \boldsymbol{H}\boldsymbol{p}_\ell \tag{48}$$

where $\boldsymbol{H}$ is the *homography* induced by the fixation plane $\mathcal{P}$. If w_ℓ is the perpendicular distance from $\bar{\boldsymbol{c}}_\ell$ to $\mathcal{P}$, then the transformation is represented by the 3×3 matrix

$$\boldsymbol{H} = \boldsymbol{R}_r \left(\boldsymbol{I} - \frac{\boldsymbol{b}\boldsymbol{v}^\top}{w_\ell} \right) \boldsymbol{R}_\ell^\top. \tag{49}$$

In the general case, $s \neq 0$, equations (28), (48) and (49) can be combined, leading to the well-known 'plane plus parallax' decomposition[41, 3]

$$\boldsymbol{q}_r = \boldsymbol{H}\boldsymbol{p}_\ell + t_r(s)\boldsymbol{d}_r. \tag{50}$$

The symmetric representation (28) is, however, preferable in the present context. This is because it encodes the depth map directly in cyclopean coordinates, $S(\boldsymbol{p}_c; \beta, \rho)$. It is interesting to note that any essentially 2-D transformation, such as a relative cyclo-rotation or a change of focal-length, can be readily absorbed into the homographic part $\boldsymbol{H}$ of the mapping (50). This is convenient, because it means that the analysis of the binocular disparities $t_\ell(s)$ and $t_r(s)$ is unchanged by such image-transformations.

7 Discussion

Three problems of binocular vision were identified in section 1, concerning oculomotor parameterization, disparity processing, and scene representation. A unified geometric treatment of these problems has been given in sections 2–6. The main psychophysical and physiological findings that relate to each of the proposed geometric solutions will now be reviewed.

7.1 Oculomotor Parameterization

As described in section 3, the binocular orientation of the eyes can be specified, in the visual plane, by the vergence and version angles, δ and ϵ. Furthermore, these variables are convenient for the specification of a coordinated eye movements from one fixation point to another. It was suggested by Hering that the oculomotor system actually encodes binocular eye movements in terms of vergence and version [17, 5]. Specifically, Hering's law of *equal innervation* states that the eyes are oriented by signals (δ, ϵ) where, according to (12,13), the corresponding azimuths are $\beta_\ell = \epsilon + \delta/2$ and $\beta_r = \epsilon - \delta/2$. Each eye moves according to the sum of the appropriate vergence and version components, which may cancel each other. The existence of physiological mechanisms that encode pure vergence has been demonstrated in the midbrain [49], and it has been suggested that the vergence/version decomposition is used to represent the difference between the current and target fixation points. However, the actual trajectories of large binocular eye movements are not consistent with the simple vergence/version decomposition [50]. Furthermore, it has been found that the visual axes can be significantly mis-aligned during REM sleep [51], which seems more consistent with the independent parameterization of each eye.

It may be that the application of Hering's law is limited to those situations in which the existence of a 3-D fixation point can be ensured by the foveal correspondence of the images. This condition would, for example, distinguish the vergence response from the disconjugate component of a binocular saccade. This is because vergence is driven by visual feedback[7], which is not generally available during the course of a binocular saccade. Likewise, when the eyes are closed, there is no visual information to ensure the existence of a 3-D fixation point. In summary, the evidence for Hering's law of equal innervation is mixed, but it seems clear that there are situations in which it is not a satisfactory model.

7.2 Disparity Processing

It is possible to separate the initial estimation of image correspondences from the *interpretation* of the resulting disparity field; this distinction is made in several computational models of stereopsis[52, 13, 31]. It is supposed, in these models, that the correspondence problem is first solved, independently of the gaze parameters. The latter are then recovered from the estimated disparity field[31], and the two types of information are combined, leading to a 3-D (though not necessarily Euclidean) interpretation of the scene[27, 26]. This scheme is compatible with the physiological basis of stereopsis; for example, it has been demonstrated that the initial binocular mechanisms in area V1 are tuned to absolute disparity[19], as described in section 1.2. This finding indicates that the low-level mechanisms of stereopsis do not 'compensate' for any disparity that is imposed by the relative orientation of the eyes.

The biological feasibility of a general solution to the binocular correspondence problem will now be considered. Individual disparity-tuned cells, in area V1, typically respond to a small range of absolute disparities, centred on some preferred value. The distribution of preferred disparities, over the population of cells, can also be inferred from the experimental data[53]. This arrangement suggests the following difficulty: It seems likely that, for any particular scene, and any particular fixation point, a large proportion of the V1 disparity response may be spurious. This is because the occurrence of the preferred absolute disparity, for a given detector, effectively depends on the orientation of the eyes, as well as on the scene structure. Hence the possibility of detecting a 'false match' is exacerbated by the fact that, owing to the variable orientation of the eyes, the true match may not even be in the range of a given detector.

One way to address this problem involves the use of prior knowledge about the typical structure of the scene. For, example, it might be assumed that the scene is approximately planar in the neighbourhood of the fixation point. Such a model could then be used to define *sets* of disparity detectors that are effectively tuned to the same surface in depth. This is done by computing the image-to-image mapping induced by the surface model, and comparing it to the disparity response. Note that this process is entirely consistent with detectors that respond to absolute disparity, as the scene model does not influence the output of the individual mechanisms. Rather, the local scene model is used to *identify* the relevant part of the V1 response.

If this approach is to be effective, then the induced image-to-image mapping must be appropriately parameterized[54]. It is important to note that an appropriate parameterization would allow prior knowledge to be used in estimating the gaze parameters, as well the scene structure. For example, image-to-image mappings associated with extreme configurations of the eyes might be penalized, and some mappings might be excluded altogether (e.g. those associated with non-intersecting visual axes). It is emphasized that this approach does not require non-visual information about the current gaze parameters; rather, it assumes that a model of gaze-variation can be learned from the image data. The geometric constraints described in sections 3 and 5 would seem to make this a biologically feasible task. For example, if the scene is assumed to be approximately perpendicular to the cyclopean gaze direction, then the appropriate scene/gaze model has just three parameters; α, β and ρ. Hence the disparity field induced by the fixation plane, including the full epipolar geometry, can be predicted from any hypothesized fixation point. An appropriate cyclo-rotation model, such as L-2, can easily be incorporated [44, 45]. The form of the induced image-to-image mapping is as described in section 6.

7.3 Scene Representation

It was shown in section 6 that binocular disparity can conveniently be measured with respect to the fixation plane, $\mathcal{P}$. This plane passes through the fixation point, and is orthogonal to the cyclopean viewing direction. This construction is commonly used in physiological studies of stereopsis. In particular, the shape of the disparity tuning profile, with respect to the depth of the fixation plane, has been used to classify binocular neurons[55]. Subsequent experiments have suggested that there exists a continuum of different disparity tuning curves [9], rather than a number of distinct types. Nonetheless, the continuum of disparity tuning is clearly organized around the fixation plane; for example, the majority of cells are tuned to disparities close to that of the plane[56].

The importance of the fixation plane is also reflected in psychophysical studies of stereopsis. As noted in section 1.3, only those points in space that are in Panum's area can be binocularly fused [21]. It has further been demonstrated, using simple stimuli, that stereo acuity is highest for targets located close to the fixation plane [21, 18]. However, the representation of more complex binocular stimuli raises further questions. In particular, it has been shown that judgment of the *relative* depth between nearby targets is much more accurate than judgment of the deviation of a singe target from the fixation plane [57]. Furthermore, the surface that is perceived in depth is, in some cases, an *interpolation* of the point-wise disparity stimulus [58]. These observations suggest that the representation of binocular disparity may depend on the local scene-structure, as well as on the gaze parameters [23]. The present model is compatible with this approach, and indeed, the fixation plane $\mathcal{P}(\rho, \boldsymbol{v})$ can be interpreted as the zeroth-order approximation of the local scene structure. It is straightforward to substitute the first-order model $\mathcal{P}(\rho, \boldsymbol{v}; \theta, \phi)$, in which the angles θ and ϕ represent the local surface orientation, and to repeat the derivation of binocular disparity in section 6.

The integration of visual information across the larger scene will now be considered. The representation described in section 6 allows the estimated viewing distance to be combined directly with the cyclopean depth map, because both are measured along the cyclopean gaze direction, as in equations (30) and (31). The global structure of the scene could therefore be encoded in a collection of local depth-maps, of the type described above. Although each depth-map would be associated with a different fixation point, it would be geometrically straightforward to combine the encodings, based on the corresponding gaze parameters.

Finally, it can be argued that the cyclopean representation is consistent with the perception of visual space[17]. Human binocular vision results, at least subjectively, in a single view of the scene. If this synthetic view has a meaningful centre of projection, then it may be hypothesized that it is located at the cyclopean point[59].

7.4 Conclusion

A cyclopean parameterization of binocular vision has been developed in detail. The parameterization has been used to construct the horopter and the epipolar geometry of a fixating visual system. Furthermore, the effect of the oculomotor parameters on the binocular disparity field has been described. It is clear that the interpretation of the disparity field is complicated by the variable orientation of the eyes. However, it has been argued here that this complication is minimized by binocular coordination of the eyes. The geometric and computational appeal of the cyclopean representation has been emphasized, and the biological relevance of the model has been indicated.

Acknowledgments

The authors thank Peter Sturm and Andrew Glennerster for their comments on the manuscript. This work is part of the *Perception on Purpose* project, supported by EU grant 027268.

References

[1] J. J. Koenderink & A. J. van Doorn. "Affine structure from motion". J. Opt. Soc. Am. 8(2), 377–385 (1991).

[2] O. Faugeras. "Stratification of 3-D vision: Projective, affine, and metric representations." J. Opt. Soc. Am. 12(3), 465–484 (1995).

[3] Q-T. Luong & T. Viéville. "Canonical representations for the geometries of multiple projective views." Computer Vision and Image Understanding 64(2), 193–229 (1996).

[4] E. Brenner & W. J. M. van Damme. "Judging distance from ocular convergence." Vis. Res. 38(4), 493–498 (1998).

[5] R. H. S. Carpenter. *Movements of the Eyes.* (Pion, 1988).

[6] G. L. Walls. "The evolutionary history of eye movements." Vis. Res. 2, 69–80 (1962).

[7] C. Rashbass & G. Westheimer. "Disjunctive eye movements." J. Physiol. 159, 339–360 (1961).

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

[9] S. J. D. Prince, B. G. Cumming & A. J. Parker. "Range and mechanism of encoding of horizontal disparity in macaque V1." J. Neurophysiol. 87, 209–221 (2002).

[10] J. D. Pettigrew, "Binocular visual processing in the owl's telencephalon." Proc. R. Soc. Lond. B-204, 435–454 (1979).

[11] B. Julesz. *Foundations of Cyclopean Perception.* (University of Chicago Press, 1971).

[12] J. D. Pettigrew, "Evolution of binocular vision." In *Visual Neuroscience*, eds. J. D. Pettigrew, K. J. Sanderson & W. R. Levick. 208–22 (Cambridge University Press, 1986).

[13] J. M. Foley. "Binocular distance perception". Psych. Review 87, 411-434 (1980).

[14] D. Tweed. "Visual-motor optimization in binocular control." Vis. Res. 37, 1939–1951 (1997).

[15] J. C. A. Read & B. G. Cumming. "Understanding the cortical specialization for horizontal disparity." Neural Computation 16, 1983–2020 (2004).

[16] H. L. F. von Helmholtz. *Treatise on Physiological Optics, vol. III* (Opt. Soc. Am. 3rd ed. 1910, trans. J. P. C. Southall, 1925).

[17] E. Hering. *The Theory of Binocular Vision* (Englemann 1868; ed. & trans. B. Bridgeman & L. Stark, Plenum Press, 1977).

[18] C. Blakemore. "The range and scope of binocular depth discrimination in man." J. Physiol. 211, 599–622 (1970).

[19] B. G. Cumming & A. J. Parker. "Binocular neurons in V1 of awake monkeys are selective for absolute, not relative disparity." J. Neurosci. 19, 5602–5618 (1999).

[20] H. C. Longuet-Higgins. "A computer algorithm for reconstructing a scene from two projections." Nature 293, 133–135 (1981).

[21] K. N. Ogle. *Researches in Binocular Vision* (W. B. Saunders, 1950).

[22] J. R. Bergen, P. Anandan, K. J. Hanna & R. Hingorani. "Hierarchical model-based motion estimation". Proc. European Conf. Computer Vision, 237–252 (1992).

[23] A. Glennerster, S. P. McKee & M. D. Birch. "Evidence for surface-based processing of binocular disparity". Current Biol. 12, 825–828 (2002).

[24] J. J. Koenderink, A. J. van Doorn, A. M. L. Kappers & J. T. Todd. "Ambiguity and the 'mental eye' in pictorial relief." Perception 30, 431–448 (2001).

[25] B. Julesz. "Cyclopean perception and neurophysiology". Investigative Ophthalmology and Visual Science 11, 540–548 (1972).

[26] J. Gårding, J. Porrill, J. E. W. Mayhew, & J. P. Frisby. "Stereopsis, vertical disparity and relief transformations". Vis. Res. 35(5), 703–722 (1995).

[27] J. J. Koenderink & A. J. van Doorn. "Geometry of binocular vision and a model for stereopsis" Biol. Cybernetics 21(1), 29–35 (1976).

[28] J. J. Koenderink & A. J. van Doorn. "Second-order optic flow". J. Opt. Soc. Am. 9(4), 530–538, (1992).

[29] B. Rogers & R. Cagenello. "Disparity curvature and the perception of three-dimensional surfaces." Nature 339, 135–137 (1989).

[30] A. J. Noest, R. van Ee & A. V. van den Berg. "Direct extraction of curvature-based metric shape from stereo by view-modulated receptive fields." Biol. Cybernetics 95, 455–486 (2006).

[31] J. E. W. Mayhew. "The interpretation of stereo-disparity information: The computation of surface orientation and depth." Perception 11, 387–403 (1982).

[32] J. E. W. Mayhew & H. C. Longuet-Higgins. "A computational model of binocular depth perception." Nature 297, 376–379 (1982).

[33] D. Weinshall. "Qualitative depth from stereo, with applications." Computer Vision, Graphics, and Image Processing 49(2), 222–241 (1990).

[34] B. Gillam, D. Chambers & B. Lawergren. "The role of vertical disparity in the scaling of stereoscopic depth perception: An empirical and theoretical study". Perception and Psychophysics 44, 473–484 (1988).

[35] B. J. Rogers & M. F. Bradshaw. "Vertical disparities, differential perspective and binocular stereopsis." Nature 361, 253–255 (1993).

[36] B. T. Backus, M. S. Banks, R. van Ee & J. A. Crowell. "Horizontal and vertical disparity, eye position, and stereoscopic slant perception." Vis. Res. 39, 1143–1170 (1999).

[37] E. M. Berends & C. J. Erkelens. "Strength of depth effects induced by three types of vertical disparity". Vis. Res. 41, 37–45 (2001).

[38] C. J. Erkelens & R. van Ee. "A computational model of depth perception based on head-centric disparity." Vis. Res. 38, 2999–3018 (1998).

[39] S. Maybank. "Theory of Reconstruction from Image Motion." (Springer Verlag, 1993).

[40] M. Armstrong, A. Zisserman & R. I. Hartley. "Self-calibration from image triplets." Proc. European Conf. Computer Vision, vol. 1, 3–16 (1996).

[41] A. Shashua & N. Navab. "Relative affine structure: Canonical model for 3-D from 2-D geometry and applications." IEEE Trans. Pattern Analysis and Machine Intelligence 18(9), 873–883 (1996).

[42] R. I. Hartley & A. Zisserman. "Multiple View Geometry in Computer Vision." Cambridge University Press (2000).

[43] K. Hepp "Oculomotor control: Listing's law and all that." Current Opinion in Neurobiology 4, 862–868 (1994).

[44] D. Mok, A. Ro, W. Cadera, J. D. Crawford & T. Vilis. "Rotation of Listing's plane during vergence." Vis. Res. 32, 2055–2064 (1992).

[45] L. J. Van Rijn & A. V. Van den Berg. Binocular eye orientation during fixations: Listing's law extended to include eye vergence." Vis. Res. 33(5), 691–708 (1993).

[46] C. W. Tyler. "The horopter and binocular fusion" In *Vision and Visual Disorders, vol. 9; Binocular Vision*. ed. D. Regan, 19–37 (Macmillan, 1991).

[47] M. L. Cooper & J. D. Pettigrew. "A neurophysiological determination of the vertical horopter in the cat and owl." J. Comparative Neurology 184, 1–25 (1979).

[48] M. J. Brooks, L. de Agapito, D. Q. Huynh & L. Baumela. "Towards robust metric reconstruction via a dynamic uncalibrated stereo head." Image and Vision Computing, 16, 989–1002 (1998).

[49] L. E. Mays. "Neural control of vergence eye movements: Convergence and divergence neurons in midbrain." J. Neurophysiol. 51 (5), 1091–1108 (1984).

[50] H. Collewijn, C. J. Erkelens & R. M. Steinman. "Trajectories of the human binocular fixation point during conjugate and non-conjugate gaze-shifts." Vis. Res. 37(8), 1049–1069 (1997).

[51] W. Zhou & W. M. King. "Binocular eye movements not coordinated during REM sleep." Experimental Brain Res. 117, 153–160 (1997).

[52] D. Marr & T. Poggio. "Cooperative computation of stereo disparity." Science 194 (4262), 283–287 (1976).

[53] A. Anzai, I. Ohzawa & R. D. Freeman. "Neural mechanisms for encoding binocular disparity: Receptive field position versus phase." J. Neurophysiol. 82, 874–890 (1999).

[54] S. J. Maybank & P. F. Sturm. "MDL, collineations and the fundamental matrix." Proc. British Machine Vision Conf. 53–62 (1999).

[55] G. F. Poggio & B. Fischer. "Binocular interaction and depth sensitivity in striate and prestriate cortex of behaving rhesus monkeys." J. Neurophysiol. 40, 1392–1407 (1977).

[56] B. G. Cumming & G. C. DeAngelis. "The physiology of stereopsis." Ann. Rev. Neurosci. 24, 203–238 (2001).

[57] G. Westheimer. "Cooperative neural processes involved in stereoscopic acuity." Exp. Brain Res. 36(3), 585–597 (1979).

[58] G. J. Mitchison & S. P. McKee. “Interpolation in stereoscopic matching.” Nature 315, 402–404 (1985).

[59] H. Ono & A. P. Mapp. “A re-statement and modification of Wells-Hering’s laws of visual direction.” Perception 24(2), 237–252 (1995).