Kai Markus Schreiber

Sensorimotor Interaction in Stereopsis

SENSORIMOTOR INTERACTION IN STEREOPSIS

by Kai Markus Schreiber

A thesis submitted in conformity
with the requirements for the degree of Doctor of Philosophy
Graduate Department of Physiology
University of Toronto

ISBN 3-930171-28-7

Contents

List of Figures

Chapter 1

Introduction

Who'll know aught living and describe it well,
Seeks first the spirit to expel.
He then has the component parts in hand
But lacks, alas! the spirit's band.
Encheirisis naturae, Chemistry names it so,
Mocking herself but all unwitting though.
Mephistopheles[1]

1.1 The Unbearable Complexity of the Brain

The human brain is often said to be the single most complex object on earth, if not in the whole universe[2]. The brain weighs only 1400 grams on average, but its billions of neurons form a mind-boggling number of connections, giving rise to the rich and complex world of experience that most of us are familiar with from introspection. But on the flipside, this complexity and richness poses the very serious challenge to modern neuroscience of making sense of it all and find a convincing explanation of why there are minds at all and what it is they are doing.

Attempts at uncovering the secrets of the brain have been made for millennia now, with the origins of neuroscience vanishing in the mists of time.

[1] In Faust I by Johann Wolfgang Goethe. Translated by George Madison Priest (Goethe, 1952).

[2] This seems to be one of the sayings everyone knows, yet nobody knows who it is that said it first. I couldn't even locate a source for it using today's information fetish, the omnipotent internet

We can be quite sure that most of the great ancient civilizations had some distinct concept of the brain. And there is evidence that even earlier, trepanations were performed, indicating some realization of the special role of the brain.

However, in all this time, the brain has kept all but some of the most superficial of its secrets. We now know quite a bit about the execution of motor commands, and we have some understanding of the first steps in sensory perception, especially vision. But most of the higher brain function is still *terra incognita*. For most of the higher functions, not even a description so far has been agreed upon. There probably are about as many definitions of *consciousness* as there are scientists investigating it.

This situation certainly is not due to a lack of effort to make advances. In fact there appears to be a marked asymmetry between the efforts to understand the brain and the human mind throughout history, and the resulting body of hard fact. The reason for this asymmetry is arguably the abovementioned complexity of the organ under scrutiny.

1.2 A Devil's Problem

As with any system too complex to understand as a whole, a promising approach is to break it down into smaller subsystems. These subsystems are likely to be more manageable and understandable, with the hope being that the knowledge gained in the analysis of the subsystems shows a path to the understanding of the whole. Very often though, it doesn't.

If we try to understand what a car does, and how it does it, starting with the role of the motor might be a good idea, and we might, after carefully looking at the available evidence, conclude that cars have motors so they can make noise, produce smelly fumes and heat up the environment. It achieves this by rotating some metal parts at high speeds. Obviously, as soon as we connect this to the bigger concept of a car, it won't work any more. If, on the other hand, we start by looking at the whole of the car, we will never get much further than stating that cars are made to move, and they do this when someone presses down a pedal.

Enough of fuel-burning metaphors. The scientist's task is to identify systems that are complex enough to reveal interesting principles but simple enough to be manageable by the means he has at hand. An additional important concern is that they must not be interconnected with many other,

unknown, sitemaps, since this would make any results highly questionable.

Among the systems that have been extensively used to break up the scientific problem of the brain into smaller pieces, the visual and the oculomotor system both have a prominent role. The visual system has been a focus simply because as much as 30% of our brains are concerned with the processing of visual input, and also because the effects of vision are easily accessible by introspection and experimental stimuli to probe the system are relatively well defined and easy to construct.

The oculomotor system, on the other hand, is a useful model system for motor control. One reason is its kinematic simplicity. The eyes are close to spherical and their inertia is negligible, making the dynamics of movement much easier to describe than those of the arm or hands. Another reason is that while the kinematics of eye movements are simple, their behavioral patterns are surprisingly complex and rich.

Thirdly, both systems described are located on the surface of brain function, that is, close to the sensory input or the motor output systems. The structure of brain function further removed from these surface layers has remained largely unknown mainly because these parts of the brain have no well defined relation to sensory input or motor output, making it very hard to describe them as information processing units at all.

But even the oculomotor system, as this thesis will show, cannot be fully understood without taking into account the sensory side. Nor can the problem of vision—in the case of this thesis, of depth vision—be well defined without considering the effect of eye movements on the geometry of the visual scene.

So while the analysis of subsystems is a useful and valid approach, it has to be kept in mind that the results obtained may be true only for the subsystem in isolation, and that interaction with other parts of the brain can completely change the interpretation of the system's behaviour.

1.3 Optimization

Another possible and very fruitful approach in understanding complex neural systems is to ask what goal they are there to serve. In other words, to identify parameters the system under scrutiny optimizes. These parameters might for example be the speed of eye movements or highly abstract descriptions of complex interactions of behaviour and environment, like keeping the

organism as well fed as possible. But ultimately, all of these optimization principles and goals can be formulated in terms of neural signals available to the brain circuits involved. This will be done by specialized sensors, for instance the biochemical ones monitoring blood sugar levels involved in feeding behaviour.

As this thesis will show, the interaction between oculomotor kinematics and the retinal geometry of binocular correspondence can be understood by assuming that the system as a whole tries to minimize certain abstract items known in optimization theory as cost functions: the energy used for neuronal computations, the absolute number of neurons and the speed of motor execution.

Investigating a neural system from an optimization standpoint might not tell us how the system does the things it does, but it can tell us what these things are. From there functional models can be devised which in turn make predictions about the physiology.

1.4 The Dilemma of Binocular Vision

Most animals are bilaterally symmetric, having pairs of most external organs. Where there is just one organ, it usually is symmetric in itself, for example lower jaws and noses. This very likely allows for a great simplification in the genetic code. Thus its very likely that the fact that most animals have two eyes is just the reflection of this underlying symmetry. In this view two eyes originally evolved because the machinery to create a symmetrical organism already was there. Of course they also gave the organism the advantage of having a spare eye in case something bad should happen to one of them, raising its chances of survival.

Once there were two eyes, however, there was evolutionary pressure to make good use of them. Two eyes, for geometric reasons, can do significantly more than one eye. They can scan more of the external environment. And they can make implicit use of the laws of geometry to triangulate the location of objects in depth. A closer look at the anatomy involved shows, however, that these two uses are in competition. If one animal has a wider binocular field of view than another, the region of space where it can have binocular depth vision conversely must be smaller. This is shown in Figure 1.1.

Evolutionary pressure thus drove the eyes in the animal's head where the resulting structure of the field of view was best suited for the animal's

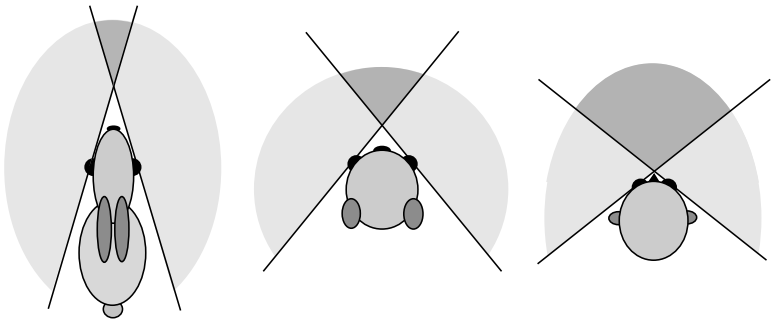


Figure 1.1: The field of view of binocular animals compared with the region in space where they see binocularly and thus can have depth vision. From left to right the overall field of view gets smaller while the binocular region, marked in dark grey, grows.

survival. While in general a lateral position of the eyes is best for prey animals, who need constantly to survey as much of their environment as possible, a frontal position of the eyes is best for predators. With accurate depth judgement, charges will be more on target, yielding higher kill rates, more food and ultimately higher survival rates.

However, having a large binocular overlap region means that the overall field of view of the animal will be rather small. To be able to effectively scan the visual environment the organism will need some mechanism of shifting gaze, either by moving the whole body, turning the head, rotating the eyes or a combination of the three: depth vision calls for the ability to shift gaze, the most effective method for small gaze shifts being eye movements.

As this dissertation will show, the optimal combination of mobile eyes and binocular depth vision requires both systems to interact and to adapt in order to suit the other system's needs.

Chapter 2

Historical Background

> If I have seen further, it is
> because I have been standing on the shoulder of giants.
> *Sir Isaac Newton*[1]

When looking at the history of knowledge and ideas, it is always a fascinating and at times an entertaining discovery to find that concepts that seem self-evident to us have eluded a long line of original and brilliant thinkers for centuries. Moreover, oftentime they will have adopted theories in their stead, that to us seem contrived and ridiculous. Rays emanating from the eye? Hollow nerves teeming with animal spirits? Evaporating phlogiston? What were they thinking?

But on closer evaluation the history of these concepts not only help us appreciate the enormous accomplishments of scientists past, but also reveals the deep roots our current understanding has in these ideas and in the minds of their authors. It is very easy to think of the scientific process as an objective and impersonal machine that takes a scientist's time (often lots of it) and converts it into objective knowledge. But more often than not this knowledge will depend on opinion as much as on fact, on history as much as on experiment. If anything scientists do allows them to see a bit further, it is because they have become giants by learning from those before them. And then they grew a little.

This is by no means an original thought. In the 12th century

> Bernard of Chartres used to compare us to [puny] dwarfs perched on the shoulders of giants. He pointed out that we see more and

[1] In a letter to Robert Hooke, 1676.

> farther than our predecessors, not because we have keener vision or greater height, but because we are lifted up and borne aloft on their gigantic stature[2]

Newton's metaphor is not original, but then that is its point.

2.1 Depth Vision

The history of depth vision may seem to be particularly astounding, because it took scholars and philosophers contemplating vision such a very long time to describe the basic problems the depth vision system solves. While today's average person might take it for granted that we can see objects in depth better with two eyes than with one, this major advantage of binocular vision was not widely recognized until the early 19th century. As usual, our acquaintance with these ideas masks the large effort it took to develop them[3].

2.1.1 Early Optics

Many of the early civilizations knew about the obvious role of the eyes in vision and about various visual disturbances. In Egyptian mythology there is a story in which the god Re tested Horus' vision after he suffered an eye injury. Re presented a brightly colored wall on which a line and a pig were painted in dark colour. Horus could see the pig but missed the line, from which Re concluded that Horus still had partial vision. Hammurabi's code contained several sections dealing with the eye, some of them stating what to charge a patient for operations on his eye, depending on his wealth and status. And in ancient India, Susrata, the author of the first known surgical treatise, described procedures used for couching cataracts[4]. But the first systematic studies of vision, its organs and the underlying principles were done in ancient Greece.

[2] As quoted in John of Salisbury's Metalogicon (Salisbury, 1955), p. 167.

[3] This section is based on the introduction to Ian P. Howard's and Brian J. Rogers' great book *Binocular Vision and Stereopsis* (Howard & Rogers, 1995), on the book on early optics by Lindberg (Lindberg, 1976), on *Origins of Neuroscience* by Stanley Finger (Finger, 1994) and *A Natural History of Vision* by Nicholas Wade (Wade, 1998).

[4] In couching, a sharp instrument is used to push the clouded lens from its position in the eye, letting it sink to the bottom of the eyeball. This is the first and for a long time was the only treatment for cataracts.

Figure 2.1: Euclid ($\star\sim$325 B.C., $\dagger$265 B.C.)

The Greeks

The earliest known book on human vision is Euclid's Optics (Euclid, 300 BC/1945), written in about 300 B.C. The book is mostly concerned with the geometry of vision, and bases its theorems on 7 postulates. Among these is the assumption that light emanates out from the eye in rays, and objects are visible only when they are hit by such a beam of light.

This theory, known as the emanation theory of sight, was by no means shared by all scholars at the time. Its merits, however, probably lay in explaining how it is possible for vision to sense objects that are far away from the observer. Like fingers reaching out and touching the surface of an object, the emanating light beams can interact with far away surfaces to enable perception. This idea masked the basic problem of depth vision, for there is nothing inherently surprising about solid objects in depth being perceived as such, if perception is achieved by sending out rays to touch them rather than by analysing flat images. Even so, Euclid was aware of the fact that each eye has a slightly different view of the world and that two eyes see more of an object than one eye alone.

The first known remark about binocular disparity was made before Euclid, by a philosopher opposed to the emanation theory, Aristotle. He believed that objects continuously emit images of themselves that travel in straight lines into the eyes. Aristotle described how an object appears double when pressure is applied to the side of one eyeball, causing the eye to change

Figure 2.2: Aristotle (⋆384 B.C., †322 B.C.)

convergence (Baere, 1931) (Beare, 1906).

The emanation theory persisted, however, and the great physician and anatomist Galen developed it further (Galen, 1968). His famous anatomical studies, done on cattle and many other animals (including even an elephant) but probably never on humans[5], led him to propose a theory of vision in which the lens is the main organ of sight, being located in the middle of the eye. Visual spirit travels down through the hollow optic nerves[6] to the lens, which is thought to be situated in the middle of the eyeball. From there the visual spirits emanate to the outside world. Later they return, conveying sensory information about distant objects back to the lens. From the lens

[5]While he had access to two dried human corpses, there is no evidence that he ever did an autopsy—a term he coined— on them.

[6]An misconception going back to Alcmaeon of Croton (⋆∼535 B.C., †unknown), who was the first known person to dissect and study humans. He discovered the optic nerves and described them as hollow tubes. Alcmaeon also believed in emanation theory and influenced if not inspired the later belief in it. Theophrastus, in his essay *On the senses* says that

> (Alcmaeon states that) eyes see through the water round about. And the eye obviously has fire within, for when one is struck (this fire) flashes out. (Theophrastus, 1917)

Thus, the visual sensation of coloured flashes resulting from pressure on the eyeball is at least one of the reasons for the widespread belief in emanation theory.

Figure 2.3: Galen ($\star\sim$130 A.D., $\dagger\sim$200 A.D.)

the visual spirits then travel back up to the brain's ventricles, where they mix with animal spirits, thus creating the visual perception.

Galen was the first to describe the anatomy of the optic chiasma, the crossing of the optic nerves from the two eyes. He thought that while the optic nerves from the two eyes did not cross at the chiasma, the visual spirits travelling in them communicated there, uniting the impressions from the two eyes, effectively creating a cyclopic eye at the chiasma's location, behind the two eyeballs and in the middle plane of the head. This idea, Galen says, was suggested to him by a god in a dream[7].

Galen also used Euclid's optics to describe binocular parallax and described how the two eyes see different parts of an object and unite the view in the cyclopean eye to allow the viewer to see more of the object. This was in his opinion the main advantage of binocular vision.

The Arabs

With the advent of the middle ages scientific enquiry in Europe came to an almost complete halt. Emperor Justinian disbanded the neo-Platonist school in Athens in 529 AD, and scholars fled to Khosru, the Shah of Persia and

[7]This joins Galen with James Watt, who not only invented the steam engine, but also persistently over several days dreamed of a new technique for producing lead shots, and with Friedrich August Kekulé, who dreamed the molecular structure of benzene.

Figure 2.4: Abu Ali al-Husain ibn Abdallah ibn Sina (Avicenna) ($\star$980 A.C., †1037 B.C.)

one of the leaders of the now emerging Arabic civilization. Both economic prosperity and intellectual freedom contributed to the thriving science of the Arabic world. The many books translated into Arabic or written during that time would eventually be translated back into Latin and find their way to Europe, laying the foundation for the Renaissance and ultimately modern science.

One of the first of the great scholars of Arabia was Abu Yusuf Yaqub ibn Ishaq al-Sabbah Al-Kindi, ($\star\sim$801 in Kufah, † in 873 in Baghdad). He wrote around 24 books, among them a highly influential book on optics that was translated into Latin as *De aspectibus*. In it he postulated that every object constantly emits rays of light in all directions. However, he combined this view with the emanation theory and did not follow the idea to its logical conclusion.

A further step was taken by Abu Ali al-Husain ibn Abdallah ibn Sina, or Avicenna ($\star$980 in Kharmaithen in today's Uzbekhistan, †1037 in Hamadan, now in Iran). Avicenna devoted several books to the refutation of the emanation theory, and combined Galen's anatomy with Aristotle's ideas about optics.

The biggest change in the theory of vision, however, can be attributed to Abu Ali Al-Hazan ibn Al-Hazan ibn Al-Haytham, known in the western world as Alhazen ($\star$965 in Basra, †$\sim$1040 in Cairo). He wrote 16 books on

optics, 7 of which form the great *Book of Optics (Kitab al-Manazir)*, which was translated into Latin in the 12th century under the title *Perspectiva* and since 1989 is available in English(Alhazen, 1989). In it Alhazen rejected the emanation theory of light, embracing instead the view that objects send light to the eyes. He also dismissed Aristotle's idea of objects sending out copies of themselves, adopting al-Kindi's idea of radiation instead. He thus proclaimed that every point on an object's surface sends light in all directions. The foundation of modern optics was laid.

Being a Muslim, he was forbidden to dissect animals. While he used the anatomical descriptions provided by Galen, he rejected Galen's ideas about the transmission of light and the nature of vision. Alhazen described how light entering through the cornea and the lens converges into their common centre of curvature and then diverges again to project upside down onto the rear surface of the lens. From there he thought the images travelled up the fluid-filled interior of the optic nerve, where the two eyes' views were combined into one.

Alhazen knew about pinhole cameras (*camera obscura*), where light enters through a tiny opening in the front of a box to project onto its rear wall[8], and was aware that images projected through larger apertures are blurry. To account for the eye's sharp focus, he considered mechanisms that would allow only rays perpendicular to the corneal surface to enter the lens and speculated about a possible role of refraction, thus coming very close to the solution of the problem.

With regard to the usefulness of having two eyes, Alhazen remarks

> The eyes are two and not one because of the mercy of the Artificer, be He exalted, and the foresight of nature – so that when one eye is harmed the other remains [intact] – and also because they beautify the appearance of the face[9]

He also describes the geometry of binocular vision, namely the fact that objects away from the fixation point will appear double, because their images

[8] As Hammond writes in his book on the history of the *camera obscura*, its invention is usually attributed to Roger Bacon (Hammond, 1981). The description of the pinhole camera in Alhazen's treatise *On the form of the ellipse* predates Bacon by more than 200 years. Hammond also notes, however, that both the ancient Chinese and Aristotle described the pinhole effect.

[9] Quote from (Alhazen, 1989), p. 102

do not fall onto corresponding points. With respect to an object located to the side of the binocular fixation target he remarks that

> (...) its form will not be distinct, because no point in it will be similarly situated relative to the two eyes, since no point in such an object will be equally distant either from the points in the surface of the eyes where its two forms will occur or from the axes. If, however, both eyes turn towards such an object so that the common axis may fall upon the object or close to it, its form will become distinct.[10]

And with respect to differences in the depth of objects he states:

> If, however, the two axes meet on a visible object, while the eyes perceive another object closer to or farther from them than that on which the two axes meet, while at the same time being located between those axes, then that object will be differently situated with reference to the eyes with respect to direction. Because if it lies between the two axes, then: it will be to the right of one of them and to the left of the other; ... The forms of such an object will occur in differently situated places in the eyes; the two forms produced by it in the eyes will proceed to two different places in the cavity of the common nerve that lie aside from the Centre, and thus will be two non-coincident forms[11].

While Alhazen was unaware that this non-coincidence allows the visual system to judge the depth of the perceived objects, he described in detail the geometry of binocular vision, and remarked on the tolerance of small differences in visual direction (this range of tolerance today is called Panum's fusional area).

The greatest achievement of Alhazen, besides his remarkable ideas and accurate descriptions, was perhaps his willingness to discard old ways of thinking that had blocked progress for centuries. Scholars counted themselves as Galenists, Euclideans or Aristotelians. Instead of citing the classics, Alhazen used physical and physiological argument to demolish what was wrong—the emanation theory of vision—and to unify what was right. In that sense he was one of the first scientists.

[10] op. cit., p. 236
[11] op. cit., p. 237

Figure 2.5: Abu Ali Al-Hazan ibn Al-Hazan ibn Al-Haytham (Alhazen) ($\star$965 in Basrah, $\dagger\sim$1040)

2.1.2 European Renaissance

In the thirteenth century interest in science started to reawaken in Europe. Works of the Arabian scholars were translated into Latin and the forgotten ancient philosophers also received new attention. Most of the books on vision, however, were mainly commentaries on Alhazen's writings.

Interest in perspective and projections was created by the flourishing Italian artists, and it was Leonardo da Vinci ($\star$1452, $\dagger$1519), in his manuscript *Trattoria della Pittura*, who first observed that differences in the two retinal images can be used for depth judgement:

> A painting, though conducted with the greatest art and finished to the last perfection, both with regard to its contours, its lights, its shadows and its colours, can never show a relievo equal to that of the natural objects, unless these be viewed at a distance and with a single eye[12].

The reason Leonardo gave for this is the fact that an object in the foreground will obscure different parts of the background when seen with the other eye. This had been described by Euclid before, but unlike Euclid Leonardo believed it to be a source of information about depth.

[12] Cited after Wheatstone (Wheatstone, 1838).

Figure 2.6: Leonardo da Vinci (⋆1452, †1519)

Optics, and with it the science of vision, took a big leap forward at the end of the 16th century, when Felix Platter, an anatomist in Basel, stated in 1583 (Platter, 1583) that the image forms on the retina rather than on the lens, and Kepler (⋆1571, †1630) described the principles of refraction governing its projection in his two works *Ad vitellionem paralipomena* (Kepler, 1604) and *Dioptrice* (Kepler, 1611). Christopher Scheiner (⋆1579, †1650) provided experimental proof for this idea by cutting away the back of a bull's eyeball and directly observing the inverted image (Scheiner, 1619).

This discovery brought to the fore the central problem of stereo vision: how can the brain reconstruct the depth of a visual scene from the two flat images projected onto the backs of the eyes? It was a long time before the answer began to take shape.

2.1.3 Aguilonius' *Opticorum Libri Sex*

François D'Aguillon (⋆1546 Brussels, †1617 Antwerp), Latinized— as was customary among scholars at the time—to Franciscus Aguilonius, was a Jesuit monk and teacher, and he also was charged by the king with organizing science in Belgium. In 1613 he published a collection of six books, called *Opticorum Libri Sex*, that was meant to be part one of a three-part work on optics. The aim of the work was to synthesize the work on optics that had been done before him, but Aguilonius died in 1617, before he could finish the two remaining parts. The first volume, of about 650 pages, is beautifully set

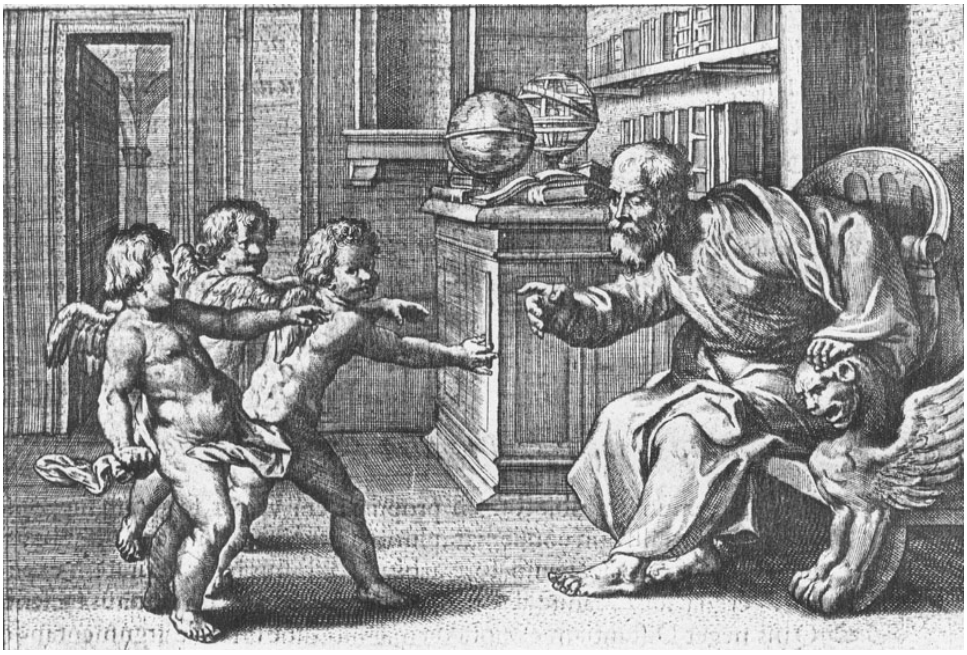


Figure 2.7: Opening illustration for Book III of Aguilonius' *Opticorum Libri Sex*, designed by Peter Paul Rubens. The one-eyed man underestimates the distance to the stick held out by the putti, demonstrating that binocular vision aids depth judgements.

and accompanied by a marvellous frontispiece and six engravings designed by Peter Paul Rubens.

Book II of the six, called *de radio optico et horoptere* contains Aguilonius' ideas about binocular vision. He knew that binocular vision improves depth perception. He adopted Galen's idea of a cyclopean eye located at the chiasma, and he introduced the term 'horopter'[13], although it didn't refer, as it does today, to the location in space where binocularly seen objects appear fused. Rather, Aguilonius thought the horopter, as he defined it, to be the frontoparallel plane[14] containing the fixation point (see Figure 2.8). He believed that the double images seen for objects at a depth different from fixation were perceived to lie in this plane:

[13]The word itself is derived from the Greek *horos*, meaning "boundary" and *opter*, meaning "observer".

[14]A frontoparallel plane is any plane parallel to the interocular axis and perpendicular to the ground.

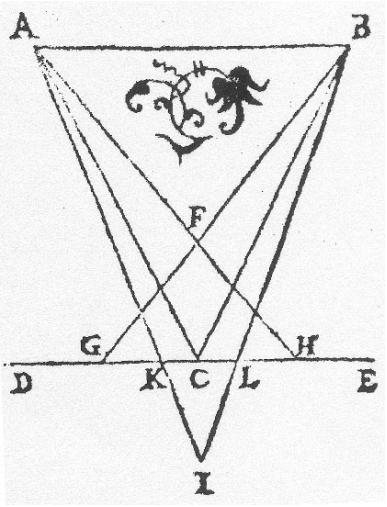


Figure 2.8: The horopter as defined by Aguilonius. The eyes at A and B are converged to look at the visual target at C. C is located in a frontoparallel plane indicated by DE, called the horopter. An object located at F is seen double, with both images being perceived in the horopter plane at locations G and H. Conversely, an object at I will be seen double, with the images being perceived at K and L respectively (Diagram from Aguilonius (Aguilon, 1613), p.111).

> For example, look at the visible object F. Even though the optical radii AF and BF come together in F, they carry the images[15] until they reach the horopter, like a common boundary and station, where the twin sites of H and G are. Likewise, for the longer radii extending to I, the horopter pulls the object I closer and into itself, locating it in the regions denoted by K and L. Thus, the horopter is the limit of all objects lying on either side of the optic axes, and in which all objects seen by single vision are located. [16]

Aguilonius also describes a circle, passing through the fixation target and the two eyes. He notes that for any object lying on this circle, the angle between the optical axis from each eye—the convergence angle—would be

[15] literally: *rei phantasiam*, the things of fantasy

[16] (Aguilon, 1613), book 2, page 111, my translation

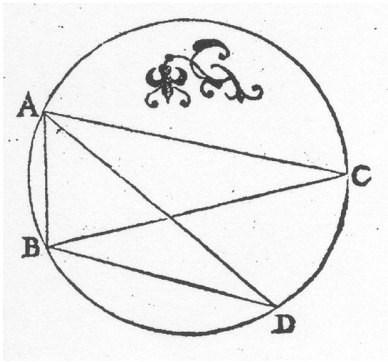


Figure 2.9: Aguilonius' isoconvergence circle. The two objects at C and D are seen under the same convergence angles from the eyes being in locations A and B (Diagram from Aguilonius (Aguilon, 1613), p.156).

the same (see Figure 2.9). He uses this as a proof for his opinion that this angle can not be used to judge depth, since objects on the circle will have the same convergence angle, while being located at different depths [17]

Aguilon had read Alhazen and quoted him several times in his book, but he didn't apply Alhazen's description of how differences in binocular visual angles will make objects appear double to his circle of isoconvergence. He would have found that the locus of isoconvergence is geometrically equivalent to the locus of objects seen in the same visual direction in both eyes. He missed this connection, and so the discovery of the modern concept of the horopter had to wait a little longer, until Vieth more than 200 years later published his *Über die Richtung der Augen* (Vieth, 1818)[18].

2.1.4 The Path to Modern Visual Science

Descartes, in his *Treatise on Man*, adopted Kepler's ideas about image formation, but also clung to the idea of visual spirits travelling to the ventricles.

[17] (Aguilon, 1613), book 3, p. 156.

[18] Howard and Rogers speculate as to why Aguilonius missed the implications for fusion of his isoconvergence circle ((Howard & Rogers, 1995), p. 15). Throughout his book Aguilonius represents the eyes as points, being unaware of the projection of light onto the retinas. Kepler's *Dioprice* had appeared 2 years before Aguilonius' book, but "it seems that Aguilonius was not aware of Kepler's work on the formation of the retinal image". Without a clear concept of retinal projection, the step from isoconvergence to equal visual angle might have been just too large to take.

His idea was that the optic nerves project the retinal image, point by point, along fibres of the visual nerve onto the wall of the cerebral ventricle. From there, corresponding points of the two images are projected onto the same location on the pineal gland, where, according to Descartes, cognition took place. While his idea on the point by point transmission – essentially retinotopy – and eventual fusion of the two images was basically correct, his choice of target organ was not.

Galen's original idea of a union of the two images in the position of the optic chiasma had been replaced by that time by the belief that the two nerves travelled separately to the brain, where the two images were integrated.

It was Isaac Newton, in his 1704 *Opticks*, who first proposed that visual inputs split at the chiasma in the way we now know they do: the right half of both images travelling to the left side of the brain and vice versa. However, he also believed that corresponding points fused at the chiasma, and that only one image eventually reached the brain. Although there was anatomical evidence that nerve fibres from the two eyes remain separate until they reach the brain, the issue was not settled until Bernhard von Gudden published his findings in 1874 (von Gudden, 1874). Basing his conclusions on microscopic studies of sliced preparations and on studies of nerve atrophy, he stated that

> in all animals with separate fields of view, the optic nerves cross completely[19]

whereas

> In all animals whose fields of view coincide (and also in humans), the optic nerves cross only partially.[20]

[19](von Gudden, 1874), my translation. Von Gudden is also a fascinating historical character. He reformed the psychiatric systems, liberalizing the strict practices of physical force and educational strictness, stating that proper treatment required communal social life and personal freedom for the patients. In 1875 he was appointed physician to the mad king Ludwig II of Bavaria. On June 9th of 1886 he and three others delivered their opinion on the state of Ludwig II, diagnosing paranoia. On the basis of this diagnosis the king was relieved of all his duties and sent to Schloss Berg, which was to become his residence. Three days later, he took a walk with von Gudden to Lake Starnberg where both of them were found drowned soon afterwards. To this day, the dispute continues as to whether Ludwig II murdered his physician and committed suicide afterwards, or whether both were assassinated.

[20](von Gudden, 1874)

The understanding of the geometry of binocular vision meanwhile was advanced by several writers, developing the thought that binocular vision improves depth vision. Joseph Harris (⋆1704, †1764), in his *Treatise of Optics*, which was not published until 1775, might have been the first to refer to the difference in appearance in the object as seen by each eye as a source of depth. He writes

> For an object is not a mere coloured extension, contained within a certain out-line; but a figure, wherein are comprehended, besides their colour and out-lines, the particular figures, *relievos* or swellings, hollowings, & c. of every visible part. (...) In a face, the nose is seen projecting before the rest. (...)
>
> And by the parallax on account of the distance betwixt our eyes, we can distinguish besides the front, part of the two sides of a near object (...) and this gives a visible *relievo* to such objects, which helps greatly to raise or detach them from the plane, on which they lie: Thus, the nose on a face, is the more remarkably raised, by seeing each side of it at once.[21]

He also referred to the horopter as the locus of objects perceived single and noted that objects "a little out of the plane [of the horopter], may yet appear single"[22]. This essentially repeats Alhazen's much earlier observation of what today is known as Panum's fusional areas—the fusional system's tolerance of small binocular differences in visual angle.

The geometrical properties of these differences in visual angle were stated clearly by Vieth in 1818 (Vieth, 1818). He started by reviewing the notion that a frontoparallel surface containing the point of fixation was the locus of objects seen single, and concluded that neither experience nor experiment had given any indication of this being true. He goes on to say that

> the claim that only objects located in the horopter appear single, contradicts the reason given for a single or double appearance. It is said, and this indeed is in accord with experience, that single vision has its origin in the images of the object being cast onto *corresponding points* on the retina, and the reason for double vision is the opposite[23]

[21] (Harris, 1775)
[22] op. cit., p. 113
[23] (Vieth, 1818), my translation, emphasis by author.

He then goes on to point out that the definition of corresponding points is somewhat vague, but that any object is seen under the same visual angle in both eyes,

> when [the object] lies in the *circumference of a circle* containing [the fixation target] and [the two eyes][24]

In 1826 Johannes Müller published a similar and independent analysis (Müller, 1826), which is why the circle today is known as the Vieth-Müller Circle. I discuss its exact geometry in Section 3.2. In 1867 Hermann von Helmholtz (⋆1821, †1894) will generalize the geometry of the horopter over the visual field (Helmholtz, 1867).

2.1.5 Wheatstone's Stereoscope

Sir Charles Wheatstone, born in 1802 in Gloucester, England, became a professor of experimental physics at King's College in London in 1834, but he made significant contributions to a great many scientific disciplines. His many inventions include the concertina, the use of ticker tape (a predecessor of punch cards) to store data, and the Playfair cipher code, used by the English during the Boer War and World War I.

But most importantly in this context, in 1838 Wheatstone published a paper on the principles underlying binocular depth perception (Wheatstone, 1838). He begins with a long passage about how everybody missed the obvious fact that differences in perspective give the impression of depth.

> After looking over the works of many authors who might be expected to have made some remarks relating to this subject, I have been able to find but one, which is (...) Leonardo da Vinci[25].

He quotes Leonardo's statements about occlusion of different parts of the background scenery by an sphere seen binocularly, then continues:

> Had Leonardo da Vinci taken, instead of a sphere, a less simple figure for the purpose of his illustration, a cube for instance, he

[24] (Vieth, 1818), my translation, emphasis by author. Brackets indicate replacement of names of dots X, O, U and P by description of the items represented by the dots. Note that this is the isoconvergence circle Aguilonius wrote about

[25] (Wheatstone, 1838), p. 372. For Leonardo's quote see above

Figure 2.10: Charles Wheatstone (⋆1802, †1875)

> would not only have observed that the object obscured from each eye a different part of the more distant field of view, but the fact would also perhaps have forced itself upon his attention, that the object itself presented a different appearance to each eye. He failed to do this and no subsequent writer within my knowledge has supplied the omission; the projection of two obviously dissimilar pictures on the two retinas when a single object is viewed, while the optic axes converge, must therefore be regarded as a new fact in the theory of vision.[26]

This is not quite true, as we have seen that Joseph Harris had some understanding of depth perception from differences in retinal images more than 60 years before Wheatstone. But Wheatstone did much more than just state his theory. If depth perception, he argued, is enhanced by differences in the retinal images, it should be possible to create a percept of depth artificially by controlling the images seen by each eye separately. He argued that this could be achieved by providing two images and crossing the eyes, so that the left eye looked at the right image and vice versa, a technique now known as cross-fusion. He also describes the shortcomings of this technique:

> The adaptation of the eye, which enables us to see distinctly

[26] op. cit. p. 372 f.

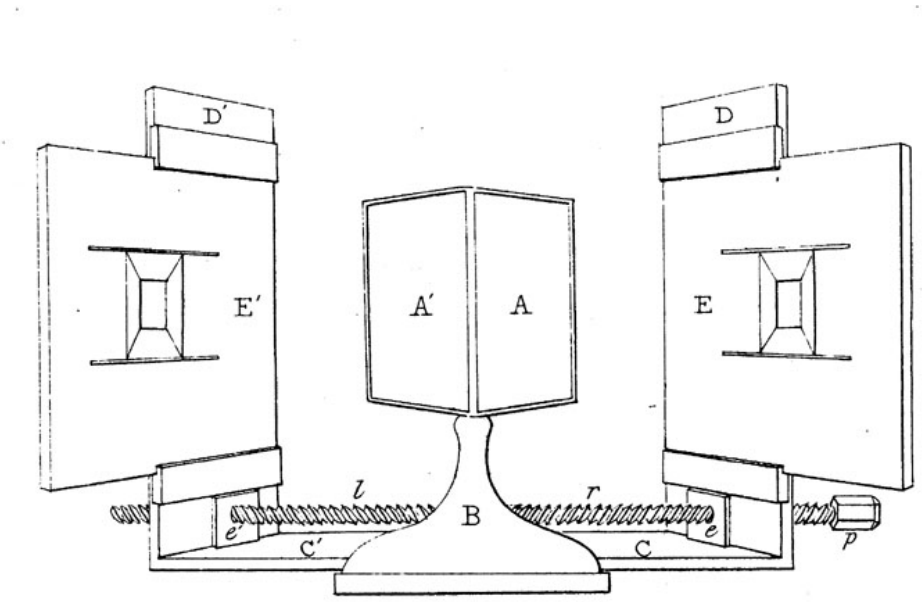


Figure 2.11: The Wheatstone mirror stereoscope. The two mirrors at A and A' are positioned so that the images at E and E' are seen in the mirrors when the person using the stereoscope is looking straight ahead. The two projections of the cube will then fuse and create the percept of a solid cube floating behind the mirrors.

> at different distances, and which habitually accompanies every different degree of convergence of the optic axes, does not immediately adjust itself to the new and unusual condition; and to persons not accustomed to experiments of this kind, the pictures will either not readily unite, or will appear dim and confused. Besides this, no object can be viewed according to either mode when the drawings exceed in breadth the distance of the two points of the optic axes, in which their centres are placed[27].

He then goes on to describe an instrument that was built for him by opticians in London. "To indicate its property of representing solid figures" he proposes this instrument to be called a stereoscope. He also provides twelve line drawings in his paper, which can be viewed in his stereoscope or

[27]op. cit. p. 374

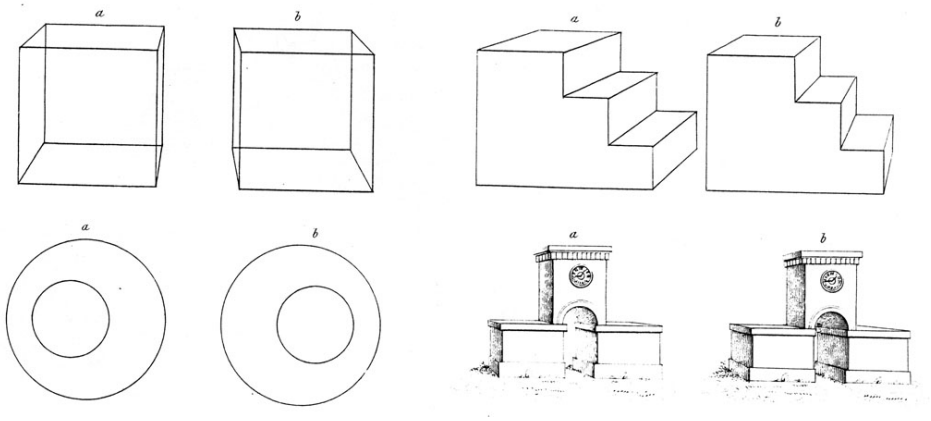


Figure 2.12: Four of the twelve stereopairs Wheatstone gave in his article. These drawings, when cross-fused or viewed in a stereoscope, will create the impression of depth. From (Wheatstone, 1838).

by cross-fusion to create the percept of depth.

Wheatstone also stated clearly "that objects whose pictures do not fall on corresponding points of the two retinas may still appear single", a statement elaborated upon by Panum 20 years later (Panum, 1858).

2.1.6 Stereographic Imaging

With Wheatstone's article, the central concepts of binocular vision had been established and its main problems stated: differences in the two retinal images give rise to the perception of depth. Objects which project onto or close to corresponding retinal points appear single. And the system enabling this can be tested by supplying different images to the two eyes.

One huge problem remained, however, making it difficult to probe the stereoptic system: every stimulus given to a subject would always contain cues to depth other than pure differences in location. It was the advent of random-dot stereograms that solved this problem. In a random-dot stereogram, described in detail in Section 3.5, the only source of information about the shape of objects is binocular disparity. Thus, they are the ideal tool for probing the stereoptic system, with the caveat that they do not represent stimuli actually occurring in the natural environment[28]. This issue and the

[28] Though the often used example of a jungle scene with thousands of overlapping leaves

Figure 2.13: The world's first photographic image (bar the Turin shroud, possibly), taken in 1827, shows a building on the left, a tree just to the right of it, and a barn in the foreground. The exposure lasted eight hours.

history of random-dot stereograms is further discussed in Section 3.5.

I will end this review of the history of binocular vision with Wheatstone, and review the developments of the late 19th and all of the 20th century as well as the neurophysiology of stereopsis in Chapter 4 while discussing current opinions and concepts.

I would like to conclude by giving a brief overview on the impact the invention of the stereoscope and the stereogram have had outside of vision science, on popular culture and the general public.

The first photographic picture was produced in 1827 by Joseph Nicephore Niepce[29], who soon teamed up with Louis Daguerre. Only four years after the first daguerrotype was produced, Wheatstone made the first stereophotograph, a portrait of Charles Babbage.

In 1851 Sir David Brewster presented his prism stereoscope, a hand-held device employing prisms for viewing stereophotographs. Serial production of these stereoscopes began and in a few years several hundreds of thousands of them were sold, along with stereophotographs of all parts of the world (see Figure 2.14 for an example from Toronto).

In 1895 the Lumière brothers held the first public screening of a moving

of similar color and shape may come very close to a random-dot stereogram

[29]There is an entertaining dispute about whether or not the shroud of Turin is actually the first photograph. Nicholas Allen of the Port Elizabeth Technikon, a theological university, claims so in his book *The Turin Shroud and the Crystal Lens* (Allen, 1998).

Figure 2.14: Historic stereophotograph of the Queen's Park Legislative Building, Toronto.

picture in Paris, reportedly driving their audience into panicking shrieks as a train rolled into a station on screen. This new intensity of images soon replaced the excitement of stereophotography, and sales declined.

Twenty years later the first 3D movie was screened in New York. It used the anaglyph technique (see Section A.3, where two films are shown through red and green filters and viewed with the familiar red and green glasses; owing to poor image quality it was not a success. Apart from a short rush of 3D movies produced in the early 50s— 27 were made at its height in 1953—3D movies have never become popular.

In 1939 stereophotography was revived with the commercial introduction of the View Master. The View Master carried several stereoscopic photographs on a reel, and by the use of the then new Kodachrome film also allowed for vivid colour. The View Master originally was aimed at an adult audience. The company producing it changed ownership several times throughout its history. With each change in ownership, the media produced focused on a younger audience. Today the View Master is owned by Mattel and run in Mattel's pre-school division, marketed under Mattel's Fisher-Price imprint.

At the beginning of the 1990s, another brief fad occurred, when autostereograms appeared on the market. Autostereograms are single images, com-

posed of vertical stripes of complex patterns. The arrangement of the items in the pattern can change from stripe to stripe. When the eyes converge or diverge so that the left and the right eyes fixate neighbouring stripes, these positional changes in the patterns create a depth impression. This is also known as the wallpaper-effect, since the stripes of wallpaper give a vivid depth impression when neighboring pattern stripes are fused[30]

While the wallpaper-effect has been described by Brewster (Brewster, 1844a; Brewster, 1844b), and autostereograms have been created by a number of people after him[31], it was Tyler and Clarkes demonstration of the fact that almost any stereoscopic shape could be encoded in an autostereogram that started the public success story of these images (Tyler & Clarke, 1990).

Today anaglyphs and stereographic imaging, using shutter glasses, are an important tool in computer-aided design and are commonly used to display molecular structure in chemistry and especially biochemistry, where 3D-structure of complex molecules is especially important.

2.2 The History of Random-Dot Stereograms

The earliest reference to something resembling a random-dot stereogram was made by Santiago Ramon Y Cajal (⋆1852, †1934), the famous physiologist who shared the 1905 Nobel prize with Camillo Golgi for work on the neural structure of the nervous system. Four years before that, Ramon y Cajal had published a short paper in a photographic journal. He wrote about the recreational value of stereophotography, and noted

> During my stereoscopic honeymoon, that is to say, long ago between the years [18]70 and [18]72, I was absorbed in imagining new fancies and recreations of this genre. My aim was to achieve mysterious writing, which could only be deciphered with the stereoscope and usable for those people who don't want to divulge their own matters. My little invention is, in fact, a puerile

[30] A particularly good place to try this is the subway of Prague. Many of the subway stations there are tiled with square tiles with either an elevated or depressed half- sphere in the middle. Crossfusing these tiles can be almost as good a reason to go to Prague than the Hradschin or Wenceslas square. When these tiles are crossfused the impression of an iridescent pattern floating in depth above the tracks can be so captivating, however, that it is unwise to try this while on a schedule. You might miss your train.

[31] See (Howard & Rogers, 1995), p. 28

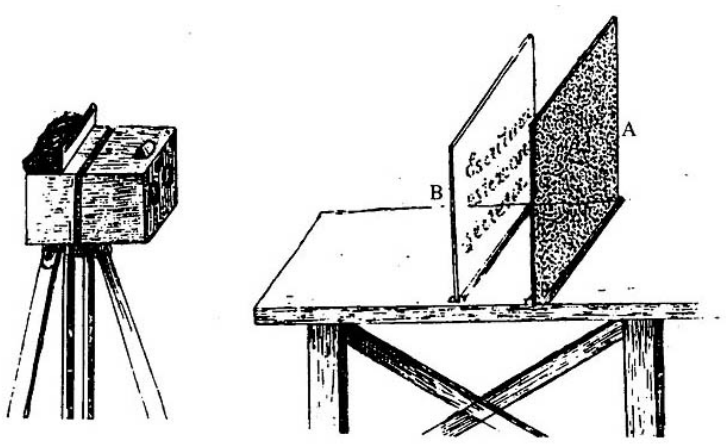


Figure 2.15: The apparatus suggested by Cajal for his "puerile play": one glass surface carried a secret message, written in dots. The second surface contained just random dots of a matching size. The stereo-camera would the take two pictures from slightly different vantage points. These pictures, when fused, would show the message standing out in depth above the random dot background.

> game unworthy of publishing, but it really amused me at that time, and it could be possible that others enjoy this pastime as well. That is the reason why we are publishing it[32].

Ramon y Cajal did not follow this path further, and apparently was not aware that the "puerile game" he designed was a stimulus visible only by means of stereo vision, and thus would be of great worth to the scientist investigating it.

In 1954, random-dot stereograms were invented for the second time, this time by Claus Aschenbrenner, who was concerned with means of reading information from aerial photographs. He writes

> The last problem to be discussed here concerns the seemingly random distribution of density specks which has been mentioned before. If this random distribution is caused by brightness differences on the ground and not by the graininess of the emulsion, it is at least the same in two subsequent pictures forming a stereo

[32] (Ramon y Cajal, 1901), quoted from (Bergua & Skrandies, 2000), their translation

> pair. (...) To the photo-interpreter, such an area in one picture may not reveal any meaningful pattern at all, but he, too, had better take a stereoscope, because hidden behind this apparently random arrangement of density specks, there might be valuable information which will become visible only in the third dimension[33].

While both Ramon y Cajal and Aschenbrenner clearly described the principle of random-dot stereograms, it was Bela Julesz (⋆1928-) in 1960, then working at Bell laboratories, who, after inventing them for the third time, routinely used random-dot stereograms created on some of the first computers to probe the stereovisual system (Julesz, 1960) (Julesz, 1964). This marked a turning point in binocular vision research. For the first time it had become possible to probe the binocular system's without providing any other cues to shape or depth.

2.3 The Laws of 3D Eye Movements

Eye movements can be divided into two main categories. There are voluntary eye movements, under control of the conscious mind and used to direct gaze to a visual target. And there are several automated mechanisms less subject to conscious control, such as smooth pursuit and the vestibular reflexes.

While voluntary eye movements are both easy to generate and to observe in others, and thus were described quite early in the history of science, the rest of the oculomotor system had to wait for a long time before it received much attention.

This account on the history of eye movements focuses on the central aspects to this thesis, namely ocular torsion and Listing's law. An overview of eye movements relevant to this thesis will be given in Chapter 5.

2.3.1 Early Ideas

Aristotle remarks that "when one eye moves, the common source of sight is also set in motion; and when this moves, the other eye moves also", referring to the binocular direction of gaze.

[33](Aschenbrenner, 1954), p. 401

The first detailed account of both the anatomy of eye movements and the rotation axes used was provided by Galen:

> (...) since there are four movements of the eyes, one directing them in toward the nose, another out toward the small corner, one raising them up toward the brows, and another drawing them down toward the cheeks, it was reasonable that the same number of muscles should be formed to control the movements (...) Since it was better that the eye should also rotate, Nature made two other muscles that are placed obliquely, one at each eyelid, extending from above and below toward the small corner of the eye, so that by means of them we turn and roll our eyes just as readily in every direction[34]

Galen was not entirely right on the role the six extraocular muscles play in eye movements, and he failed to provide a reason for the rotation of the eyes around their line of sight, but he was already close to the truth. A possible role for this ocular torsion was not to be supplied for many centuries.

Nothing much was added to the knowledge of eye movements for a long time. Alhazen stated that the two eyes always move together and by an equal amount, so as to make the visual axes converge on the fixation point.

Interest in eye movements awoke in the late 17th century, when Descartes postulated that the extraocular muscles work in pairs and that while one of a pair pulls on the eye the other relaxes.

John Hunter[35] (⋆1728, †1793) in 1786 (Hunter, 1786) categorized eye movements as follows:

> we find these three modes of action produced; first, the eye moving from one fixed object to another; then the eye moving along with an object in motion; and last, the eye keeping its axis to an

[34] Galen, (Galen, 1968), p. 483

[35] John Hunter was an eminent anatomist and surgeon and is considered the father of modern scientific surgery. He is also infamous for an experiment he allegedly did on himself, in order to prove his theory that syphilis and gonorrhoea were aspects of the same disease. He is said to have punctured his glans with a lancet to infect himself with gonorrhoea. Unbeknownst to him, the lancet also carried syphilis. When Hunter came down with both diseases, he wrongly concluded his theory to be right. He died of a syphilitic heart. (Dobson, 1969)

> object, although the whole eye, and the head, of which it makes a part, are in motion[36].

He thus described two of the major eye movement systems, namely the saccadic and the smooth pursuit system, and gave a description of what today is known as the vestibulo-ocular reflex, or VOR. He goes on to describe the action of the VOR when the head rotates torsionally, or about a forward-backward axis:

> Thus when we look at an object, and at the same time move our heads to either shoulder, it is moving in the arch of a circle whose centre is the neck, and of course the eyes would have the same quantity of motion on this axis, if the oblique muscles did not fix them upon the object. When the head is moved towards the right-shoulder, the superior oblique muscle of the right-side acts and keeps the right-eye fixed on the object, and a similar effect is produced upon the left-eye by the action of the inferior oblique muscle; when the head moves in the contrary direction, the other oblique muscles produce the same effect[37].

This effect of head tilt was described in more detail by Alexander Friedrich von Hueck. He claimed that counterroll keeps the eye's vertical diameter "vertical [in space] if the head is inclined sideways by up to 28°, and therefore during this movement the image maintains its location unchanged."[38]

Hueck's findings have been impossible to reproduce, and counterroll is now known to correct only partially for head tilt. Its gain—the ratio between head tilt and compensatory eye torsion—averages about 10% in humans (Zee & Leigh, 1983).

2.3.2 Donders and Listing

The attention on this torsional dimension of eye rotations led a Dutch military doctor named Franciscus Donders (⋆1818, †1889) to investigate the change of this orientation with gaze direction. By careful examination of the after-images obtained in different gaze directions, he found that eye torsion

[36](Hunter, 1786), p. 210
[37](Hunter, 1786), p. 212
[38]Cited from (Hueck, 1838), p. 31 after (Wade, 1998)

Figure 2.16: Johann Benedict Listing (⋆1808,†1882)

varied with gaze direction, and that the amount of torsion depended on the horizontal and vertical gaze angles. It was Hermann von Helmholtz who called this discovery *Donders' law*.

The actual amount of torsion for each gaze direction depends on the coordinate system used (see Section 3.1 for details). Johann Benedict Listing (⋆1808,†1882) in 1845 formulated what was to be known as Listing's law (Listing, 1845). It received its name when Ruete (Ruete, 1846; 2nd edition, Vol. 1, 1853) in 1853 included it in his textbook on Ophthalmology, saying

> The normal eye position mentioned above (starting position, primary position) of the eye is changed into another, secondary position, by a cooperative action of the six muscles. This change of position is done so that it can be viewed as a rotation about an axis through the optical centre of the eye which at any given time will be perpendicular to both the primary and secondary positions of the optic axis. Every secondary eye position therefore has a relation to the primary position making the projection of the rotation onto the optic axis equal zero[39].

In simpler words: the torsional orientation of the eye in any given gaze

[39] (Ruete, 1846; 2nd edition, Vol. 1, 1853), p.36-37, quoted from (Simonsz & den Tonkelaar, 1990), my translation

position will be the same as it would have been had the eye rotated to this gaze position from the primary position along the shortest possible route.

2.3.3 Beyond Listing's Law

Another way of describing Listing's law is by stating that the rotation vectors for reaching any eye position lie in a plane. This plane is known as Listing's plane, and from Ruete's account above it is obvious that Listing's plane is the plane perpendicular to the primary gaze direction.

This primary position varies over time (Schubert, 1927) and changes with head position (Fisher, 1922). While all this was known at the beginning of the 20th century, it was not until its end that the systematic deviance from Listing's law for convergence was described in detail.

Chapter 3

Mathematical Background

> Philosophy is written in this grand book of the universe, which stands continually open to our gaze. But the book cannot be understood unless one first learns to comprehend the language and to read the alphabet in which it is composed. It is written in the language of mathematics.
>
> *Galileo Galilei*[1]

One of the pillars of modern science is the notion that the laws of nature can be accurately expressed in the system of mathematical notation. This is by no means a trivial assumption. As the physicist Eugene Wigner put it in 1963 "The miracle of the appropriateness of the language of mathematics for the formulation of the laws of physics is a wonderful gift which we neither understand nor deserve." (Wigner, 1960)

But whatever the reason for the success of mathematics in describing the world, it entails a reverse that cannot be stressed enough: none of the scientific laws of nature can be understood without the proper mathematical vocabulary. This chapter provides the necessary words and grammar to handle eye movements, retinal correspondence and psychophysical measurements.

All computations of binocular disparities both for stimulus creation and for simulation made two simplifying assumptions. In this simplified description, the eye is a sphere rotating about its geometric centre, with the nodal point for optical projections coinciding with this rotation centre.

[1] In *Opere Il Saggiatore*, 1623.

This simplified description will be used for the simulations in Chapter 7. As I will point out in Section 5.2, the actual kinematics of the eye and position of the nodal point are different from this idealization.

I have used the simplifications for the sake of clarity and mathematical simplicity. A more precise modelling of the oculomotor system and the optical geometry would not change the qualitative results of the simulations and would have very little impact on the quantitative results. The experimental results are not affected at all by the model used.

3.1 Coordinate Systems

3.1.1 A Word of Caution

If someone tells you Queen's Park is right of University College, the statement is clearly neither right nor wrong but incomplete, because it omits the coordinate system. The statement becomes meaningful—and wrong—if we include a coordinate system by saying Queen's Park is right of University College when you stand in front of it facing the main entrance.

Though this is intuitively obvious in the case of directions, its validity extends to all cases where numerical data are given, and the more complex the described data are, the more important becomes a clearly defined frame of reference. When it comes to eye movements and retinal disparities, a failure to clarify the coordinate system will render all communication meaningless.

Unfortunately at least four major coordinate systems are in wide use, each having its merits in certain contexts, and severe shortcomings in others. If we look at a figure showing data on binocular gaze direction, and we find that the vertical angles of the two eyes are equal, it is easy to think we have learned something about the eyes' positions, but really the statement that the vertical angles are equal is as incomplete as the statement that Queen's Park is right of University College. If the two eyes' vertical angles are equal in Helmholtz coordinates, then we can correctly deduce that the visual axes of the two eyes lie within a single plane and therefore that if the eyes are crossed, their lines of sight will intersect at some binocular fixation target. But if the two eyes' vertical angles are equal in Fick coordinates, the same conclusion is unjustified and likely false. In short, **Any description of movements or geometrical relations is meaningless, unless its coordinate system**

is specified.[2]

Eye movements are described as rotations with respect to a reference gaze position[3]. Thus the mathematical description and theory of rotations is of particular interest for eye movement research.

3.1.2 Coordinate Systems for Rotations

In 1765 Euler published his theorem stating that any rotation of a solid object in space could be described by exactly three numbers, which are in his honour called Euler angles. There are many possible ways to arrange the axes of rotation for these angles, but as long as there are three independent axes[4], any rotation can be broken down into three consecutive rotations about these three axes.

I use a right-handed Euclidean coordinate system[5], as shown in Figure 3.1, naming the three axes x, y and z and naming the angles of rotation about the x-, y- and z-axis ψ, ϕ and θ respectively. This Euclidean coordinate system's orientation is fixed to the orientation of the head. I will therefore call these coordinates the *head-fixed coordinates*.

But these three angles are not yet sufficient to describe a rotation. For if we specify $\psi = 0°$, $\phi = -90°$ and $\theta = 90°$, the actual rotation described by these angles depends upon the order in which these rotations are carried out, as the next section will show.

Noncommutativity of Three-Dimensional Rotations

If the z axis points from your toes toward your head and you are facing in the direction of the positive x-axis, the above angles correspond to a rotation to the left and up. Now, if you hold your hand out in front of you, pointing your thumb up in positive z-direction and your index finger straight ahead, imagine you are a gunman and you want to hit something whose direction from you is described by the above angles. So you turn your arm to the left

[2] I apologize for the bold characters, but it is a rather important point.

[3] This assumes that the eye moves around a single center of rotation. For a discussion of the appropriateness of this assumption see 5.2.1

[4] Meaning that rotation about none of the axis can be accomplished by consecutive rotations about the other two axis

[5] Right-handed refers to the arrangement of the axis, where the x-, y- and z-axis are represented by the thumb, index finger and middle finger of the right hand, respectively. In a left- handed system the z-axis would point in the opposite direction.

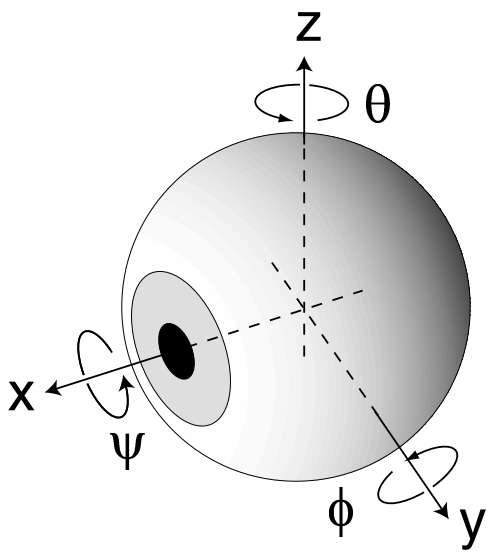


Figure 3.1: Rotation coordinate system and rotation angles. See text for details.

90 degrees, followed by a counterclockwise rotation about the y-axis through another 90 degrees. If you have done this, your index finger should point to your left and your thumb should point behind. If you want you can make some shooting movements now, but consider this: if you had started with the rotation about the y-axis, and then rotated about the z-axis, you would have ended up pointing overhead, with the thumb pointing to your right.

If you have done all this while you were reading it, not only were you displaying great acrobatic skills, your were also experiencing the noncommutativity of rotations and the difference between Fick- and Helmholtz- angles using your own body. Wasn't it amazing?

Helmholtz and Fick Angles

The difference between the two is the order in which horizontal and vertical rotations are carried out. For Helmholtz angles the order of rotation is ψ, θ, ϕ (Helmholtz, 1863), while for Fick angles it is ψ, ϕ, θ (Fick, 1854). As you have seen by pointing your fingers at an imaginary target, this little swap makes all the difference in the world.

There is an additional complication, because we can also specify whether

we want to rotate only the object or the coordinate system we are using along with it. If we rotate only the object, rotation axes are space fixed, but if we rotate the coordinate system the rotations carried out first influence the rotation axes of subsequent rotations and the rotation axes become fixed to the object that is being rotated.

If rotation axes are space fixed, and you turn from an upright position first 90 degrees left about the z-axis and then 90 degrees 'clockwise' about the x-axis, you will end up lying on your back. But if the axes are body-fixed, and you turn 90 degrees left about your z-axis and then 90 degrees clockwise about your x- axis, you will end up lying on your right ear.

Fortunately there is a simple relationship between a sequence of rotations with space-fixed axes and the corresponding sequence with object-fixed rotation axes: one is the reverse of the other; e.g. turning 90 degrees left then 90 degrees clockwise with space-fixed axes is the same as turning 90 degrees clockwise then 90 degrees left with object-fixed axes. The order given above for the Helmholtz and Fick system was for space-fixed rotation axes.

A Polar Coordinate System

Polar coordinates use angles that reflect the interest in the magnitude of changes in position independent of their direction. Rotations in polar coordinates are specified using a torsion angle ψ, a direction angle ϕ and an angle of gaze eccentricity θ (see Figure 3.2).

In a polar coordinate system Listing's law takes the very simple form of $\psi = 0$, since the rotation axis specified by ϕ has exactly the properties mentioned in Listing's law, namely it is perpendicular to both the primary gaze position and the new gaze position (in Figure 3.2 it would therefore be perpendicular to the triangle spanned by the x-axis and s).

Rotation Vectors

Another way to describe rotations is by giving the orientation of an axis in space and the angle of rotation about this axis. This information can be provided by a single vector $\vec{r}$, whose orientation will indicate the axis of rotation and whose length is the tangent of half the rotation angle. The convention for the sign of this rotation angle is known as the right hand rule: when the thumb of the right hand points in the direction of the vector, the other fingers point in the direction of the rotation.

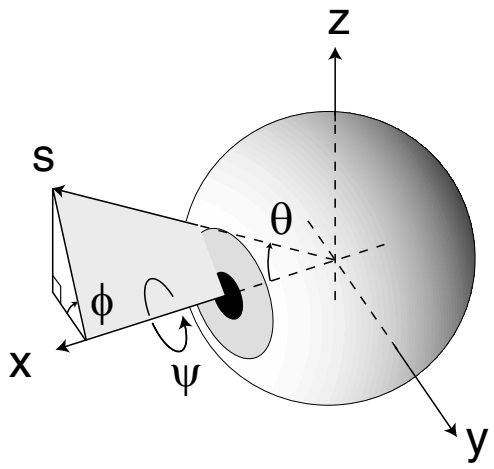


Figure 3.2: Polar coordinate system. Rotations are described by specifying the direction ϕ and the rotation angle θ. ψ specifies eye torsion. S indicates the new direction of gaze after the rotation.

This means that a rotation vector indicating an upward rotation in our example above will point along the negative y-axis, a rotation vector indicating a leftward rotation will point along the positive z-axis. With $\tan(\frac{90}{2}°) = 1$ and $\tan(\frac{-90}{2}°) = -1$, these two rotations in rotation vector form are

$$\vec{v}_1 = \begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix} \tag{3.1}$$

$$\vec{v}_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \tag{3.2}$$

A common way of presenting rotation vectors is by projecting their components along the underlying coordinate systems axes, usually in the horizontal, vertical and torsional (or yaw, pitch and roll, respectively) directions. In the case of the two vectors given here, this would yield a vertical component of $-1 \equiv -90°$ for the first and a horizontal component of $1 \equiv 90°$ for the second vector.

It is very easy to confuse these components with Helmholtz or Fick rotation angles, which they are not. To illustrate this point, assume the two

rotations above are executed in the order given, i.e. first a rotation upward, then one to the left. This could be described by $\phi = -90°, \theta = 90°$ in Helmholtz coordinates. If you start with your arm pointing forward, thumb pointing up, and execute these two rotations, you will end up pointing leftward, your thumb pointing back.

The corresponding rotation vector, however, if constructed like this

$$\vec{v} = \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} \tag{3.3}$$

lies obliquely in the y-z plane, pointing to your right and up. Also, the length of this vector is $\sqrt{2}$, so the rotation angle is $2\arctan(\sqrt{2}) \approx 109.5°$.

So this rotation vector, if you again start with your arm pointing forward, thumb up, will have you point to your left, upwards and backwards after the rotation. Clearly this position is very different from the endposition for the Helmholtz rotation described above, again illustrating the point that components of different coordinate systems, even though they might have the same name in data panels (e.g. vertical eye position), are not the same and must not be confused.

Quaternions

Another system frequently used in eye movement research (Westheimer, 1957) (Tweed & Vilis, 1987) makes uses of mathematical entities called quaternions. They were invented by William Rowan Hamilton ($\star$1805, †1865), while he was walking with his wife in Dublin[6].

[6]The account he gives in a letter to his son provides some insight into the mind of a mathematician:

> But on the (...) Council day of the Royal Irish Academy I was walking in to attend and preside, and your mother was walking with me, along the Royal Canal, to which she had perhaps driven; and although she talked with me now and then, yet an under-current of thought was going on in my mind, which gave at last a result, whereof it is not too much to say that I felt at once the importance. An electric circuit seemed to close; and a spark flashed forth, the herald (as I foresaw, immediately) of many long years to come of definitely directed thought and work, by myself if spared, and at all events on the part of others, if I should even be allowed to live long enough distinctly to communicate the discovery. Nor could I resist the impulse—unphilosophical as it may have been—to cut with a knife on a stone of Brougham Bridge,

Mathematicians have a tendency to take concepts already in wide use and extend their range, creating new entities with the original ones being contained in the new definitions as special cases. One such example was the extension of real numbers by the concept of complex numbers, another one is the extension of rotation vectors into Quaternions[7]. A rotation around a unit vector $\vec{r}$ by an angle ω translates into the quaternion

$$\begin{aligned} q &= [s, \vec{v}] = [s, v_1, v_2, v_3] && (3.4)\\ &= s + v_1 i + v_2 j + v_3 k && (3.5)\\ &= \sin(\frac{\omega}{2}) + \cos(\frac{\omega}{2})(r_1 i + r_2 j + r_3 k) && (3.6) \end{aligned}$$

with

$$\begin{aligned} i^2 &= j^2 = k^2 = ijk = -1 && (3.7)\\ ij &= -ji = k && (3.8)\\ jk &= -kj = i && (3.9)\\ ki &= -ik = j && (3.10) \end{aligned}$$

Note that the last three components of the quaternion divided by the first yield the corresponding rotation vector:

$$\frac{\cos\left(\frac{\omega}{2}\right)}{\sin\left(\frac{\omega}{2}\right)} \begin{pmatrix} r_1 \\ r_2 \\ r_3 \end{pmatrix} = \tan\left(\frac{\omega}{2}\right) \vec{r} = \vec{t} \qquad (3.11)$$

Since $\vec{r}$ is a vector of unit length, the length of the rotation vector $\vec{t}$ is $\tan(\frac{\omega}{2})$. $\vec{t}$ thus describes a rotation by angle ω around the rotation axis defined by $\vec{r}$.

To rotate a vector $\vec{p}$ by a quaternion q we write the vector as a quaternion p=[0, $\vec{p}$] and compute

---

as we passed it, the fundamental formula with the symbols, i, j, k; namely, $i^2 = j^2 = k^2 = ijk = -1$, which contains the Solution of the Problem, but of course, as an inscription, has long since mouldered away.

(Cited from (Graves, 1882))

[7]Interestingly, quaternions can also be viewed as an extension of complex numbers.

$$p' = q \cdot p \cdot q^{-1} \tag{3.12}$$

where q^{-1} is the multiplicative inverse of q.

3.1.3 Rotation Matrices

To once more stress the difference between rotations given in different coordinate systems, I will here provide the rotation matrices for rotations specified in three of the coordinate systems described above. The rotation matrix for quaternions can be easily obtained by constructing the corresponding rotation vector.

Fick- and Helmholtz-Coordinates

When rotations are described using Euler angles and the coordinate axes described above, the three elementary rotations can be written using rotation matrices:

$$\mathcal{M}_\theta = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{3.13}$$

$$\mathcal{M}_\phi = \begin{pmatrix} \cos\phi & 0 & \sin\phi \\ 0 & 1 & 0 \\ -\sin\phi & 0 & \cos\phi \end{pmatrix} \tag{3.14}$$

$$\mathcal{M}_\psi = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\psi & -\sin\psi \\ 0 & \sin\psi & \cos\psi \end{pmatrix} \tag{3.15}$$

The full rotation in Fick angles then is $\mathcal{M}_{Fick} = \mathcal{M}_\theta \cdot \mathcal{M}_\phi \cdot \mathcal{M}_\psi$; in Helmholtz angles it is $\mathcal{M}_{HH} = \mathcal{M}_\phi \cdot \mathcal{M}_\theta \cdot \mathcal{M}_\psi$. On multiplication this yields

$$\mathcal{M}_{Fick} = \begin{pmatrix} c\theta c\phi & c\theta s\phi s\psi - s\theta c\psi & c\theta s\phi c\psi + s\theta s\psi \\ s\theta c\phi & s\theta s\phi s\psi + c\theta c\psi & s\theta s\phi c\psi - c\theta s\psi \\ -s\phi & c\phi s\psi & c\phi c\psi \end{pmatrix} \tag{3.16}$$

$$\mathcal{M}_{HH} = \begin{pmatrix} c\theta c\phi & s\phi s\psi - s\theta c\phi c\psi & s\theta c\phi s\psi + s\phi c\psi \\ s\theta & c\theta c\psi & c\theta s\psi \\ -c\theta s\phi & s\theta s\phi c\psi + c\phi s\psi & c\phi c\psi - s\theta s\phi s\psi \end{pmatrix} \tag{3.17}$$

with $s = \sin$ and $c = \cos$ as abbreviations.

Rotation Vectors

For a given rotation vector

$$\vec{v} = \begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix} \tag{3.18}$$

the corresponding rotation angle and matrix will look like this:

$$\begin{aligned}
\alpha &= \sqrt{v_x^2 + v_y^2 + v_z^2} && (3.19)\\
u &= 1 - cos\alpha && (3.20)\\
n_x &= \frac{v_x}{\alpha} && (3.21)\\
n_y &= \frac{v_y}{\alpha} && (3.22)\\
n_z &= \frac{v_z}{\alpha} && (3.23)\\
\mathcal{M}_{rotv} &= \begin{pmatrix} n_x^2 u + c\alpha & n_x n_y u - n_z s\alpha & n_x n_z u + n_y s\alpha \\ n_x n_y u + n_z s\alpha & n_y^2 u + c\alpha & n_y n_z u - n_x s\alpha \\ n_x n_z u - n_y s\alpha & n_y n_z u + n_x s\alpha & n_z^2 u + c\alpha \end{pmatrix} && (3.24)
\end{aligned}$$

3.1.4 Pseudotorsion

Since the actual values of the gaze angles for all of these coordinate systems will differ for any given eye position, a change of the coordinate system will generally change the value of the torsion angle for this eye position. As mentioned above, Listing's law has the form $\psi = 0$ in polar coordinates, but in Helmholtz coordinates it is $\psi = -\frac{\theta\phi}{2}$. Donders called any torsion that was due only to a change in coordinate systems pseudotorsion. He did not believe that another kind of torsion existed (Simonsz & den Tonkelaar, 1990). While this illustrates once more the fact that a change in coordinate system can dramatically change the superficial structure of a system's behaviour, it is just Donders' law in disguise: when ocular torsion is determined by gaze direction, as Donders' law states, any apparent change in torsion cannot reflect physical movement but must be due to the description.

3.1.5 Helmholtz Coordinates in Binocular Vision

From the rotation matrices as provided above the transformations between the coordinate systems can be easily computed. I will not provide the formulae here, since all the rotations given from here on will be in Helmholtz coordinates as described above.

This is because Helmholtz coordinates are uniquely suited for binocular vision. To see why, observe that whenever both eyes converge onto a fixation target, the optical centres of the eyes and the target will lie in a plane, known as the plane of regard. Suppose that when both eyes are in primary position, the gaze vectors lie in the horizontal plane, which is the plane of regard for this position. In Helmholtz coordinates, the first two rotations are the torsional and horizontal rotations, both of which do not move the gaze vector out of this plane. The third, vertical rotation will rotate the vectors into a new plane. If the angle for this rotation, ϕ, is the same in both eyes, the gaze vectors will once again lie in a common plane, as shown in Figure 3.3. Thus, in Helmholtz coordinates whenever the two eyes are fixating the same target, their vertical Helmholtz position ϕ will be the same. Conversely, when the two eyes' vertical Helmholtz angles are different, they cannot be fixating a common target.

Additionally, when Helmholtz torsion in the two eyes is zero, the plane of regard will project onto the horizontal meridians of both retinas, so Helmholtz torsion is simply related to binocular correspondence, in a way that Fick, rotation-vector and quaternion torsion are not.

A further advantage of the Helmholtz coordinate system with respect to binocular correspondence and disparity will be discussed in the next section. In short, the Helmholtz coordinate system is uniquely suited to epipolar geometry.

3.1.6 Retinal Coordinate Systems

Since the retina is a two-dimensional surface, two coordinates are needed to describe the retinal location of an image. There is a considerable variety of coordinate systems that can describe retinal coordinates. Since retinal location geometrically corresponds to a visual direction, any coordinate system for describing eye position without torsion can be used for the purpose. For this thesis I use a Helmholtz-like coordinate system (see Figure 3.4).

The direction $\lambda = \mu = 0$ is the centre of the foveola, the retinal spot

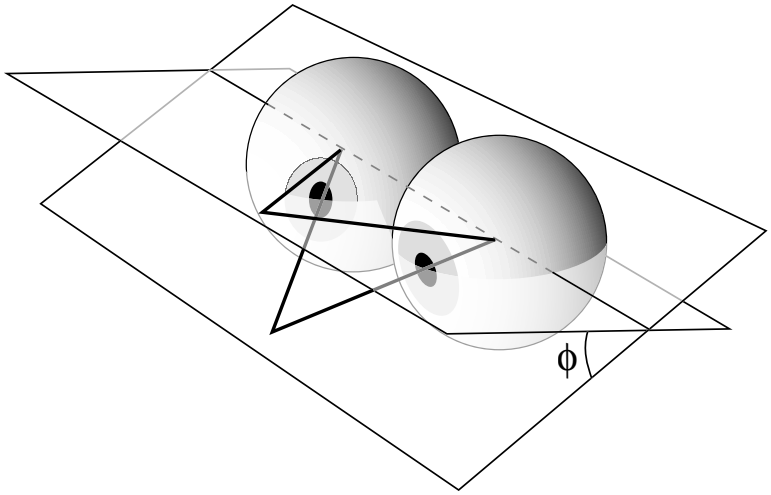


Figure 3.3: The advantage of the Helmholtz system is that the plane of regard in which binocular fixations lie is described by $\phi_{lefteye} = \phi_{righteye} = const.$ Shown are two fixations at different elevations and the corresponding fixation planes.

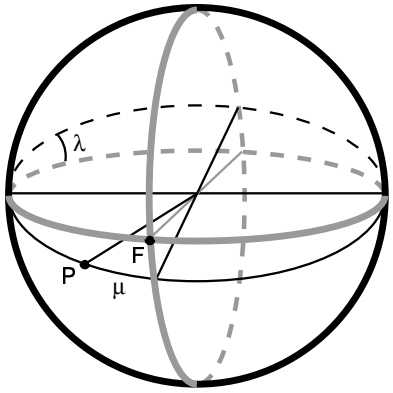


Figure 3.4: The retinal coordinate system. The coordinates of a retinal location P are the elevation angle λ and the horizontal azimuth μ.

projecting in the direction of gaze, $\lambda > 0$ corresponds to points *above* the horizontal meridian of the eye, $\mu > 0$ corresponds to leftward gaze and thus to positions to the *right* of the vertical meridian.

These retinal coordinates are ideally suited to describe retinal locations. However, as we will see in Section 3.4.2, the description of the epipolar lines for eccentric gaze is difficult in these coordinates. We therefore define an additional eye-fixed Euclidean coordinate system, the x-axis pointing in the direction of gaze, the y axis lying in the plane of the horizontal meridian and the z-axis lying in the plane of the vertical meridian.

A point with retinal coordinates λ and μ has the following components in the eye-fixed Euclidean coordinate system

$$\vec{x}_{ret} = r_{eye} \begin{pmatrix} -\cos(\lambda)\cos(\mu) \\ \sin(\mu) \\ \sin(\lambda)\cos(\mu) \end{pmatrix} \tag{3.25}$$

Above I have called the head-fixed Euclidean coordinate system *head-fixed coordinates*. In analogy I will call these eye-fixed Euclidean coordinates *eye-fixed coordinates*.

When the eye is looking straight ahead, this eye-fixed coordinate system is aligned with the head-fixed coordinate system. When the eye is rotated, the head- fixed coordinates for a retinal location are obtained by rotating the eye-fixed coordinates, yielding

$$\vec{x}_{head} = \mathcal{M}_\phi \cdot \mathcal{M}_\theta \cdot \mathcal{M}_\psi \cdot r_{eye} \cdot \begin{pmatrix} -\cos(\lambda)\cos(\mu) \\ \sin(\mu) \\ \sin(\lambda)\cos(\mu) \end{pmatrix} \tag{3.26}$$

when the eye is in Helmholtz position (θ, ϕ, ψ).

3.1.7 Vergence and Binocular Helmholtz Coordinates

Both gaze direction and the direction of a visual target can be represented by a set of horizontal and vertical Helmholtz coordinates for each eye separately. Mathematically equivalent is the description in what we call binocular Helmholtz coordinates, where conjugate horizontal angle and the binocular subtense angle are given (see Figure 3.1.7). The binocular subtense of the fixation point is called the vergence angle (γ), its conjugate horizontal angle is called the version angle (δ).

The relation between version and vergence and horizontal Helmholtz coordinates is

$$\gamma = \theta_L - \theta_R \tag{3.27}$$

$$\delta = \frac{\theta_L + \theta_R}{2} \tag{3.28}$$

That is, the conjugate horizontal angle or version, is the mean of the two monocular horizontal angles and the binocular subtense or vergence is their difference.

3.2 Binocular Correspondence and Retinal Disparity

The theoretical definition of binocular correspondence was first given in clear form by Vieth in 1818. Two spots on the retinas will coincide if their retinal coordinates (relative to the foveae of the two eyes) are the same.

The difference in retinal positions is called binocular disparity. This difference can be represented by a two-dimensional vector on the retina, starting at the location of the visual feature in one eye's retina and ending at the retinal location of the same feature in the other eye. If two objects appear in the same visual direction in both eyes, their binocular disparity will be zero.

3.3 The Theoretical Horopter

The locus in space of objects that will project onto corresponding retinal positions and thus have zero binocular disparity is called the horopter. In the plane of regard, this is a circle going through the optical centres of the eyes and the current fixation target. This is also known as the horizontal theoretical horopter.

Outside of the plane of regard the only possible locations in space for objects projecting onto corresponding retinal positions are on the median plane (the plane located in the middle of the interocular axis and perpendicular to it). That this must be true can be visualized using Helmholtz coordinates. Any point H' located outside of the plane of regard will project onto a retinal spot with nonzero vertical Helmholtz angle. If the distance of H' to one eye

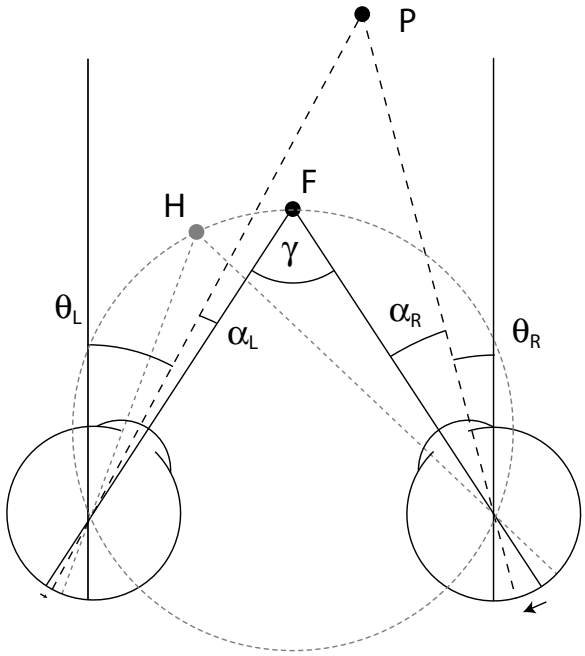


Figure 3.5: Binocular angles in the plane of regard. Shown are a fixation target F and the lines of sight connecting both eyes' foveae with it. An additional target at P projects into the two eyes along the dotted lines, falling on noncorresponding retinal positions (arrows behind the two retinas indicate direction and magnitude of the angles between visual axis and projection of P, α_L and α_R. P projects slightly counterclockwise of F in the left eye and a few degrees clockwise of F in the right eye). Point H lies on the Vieth- Müller-Circle and therefore projects onto corresponding retinal positions. Also shown is the vergence angle for this eye position, γ. See text for details.

is larger than to the other, the vertical angle of its projection will be smaller in this eye. Thus only points H' that have the same distance to both eyes can project onto corresponding retinal points. The median plane is the locus of such points. Since these points must also appear under the same binocular subtense, the only points outside of the plane of regard that are part of the theoretical horopter are on a vertical line in the median plane that crosses the plane of regard in the Vieth-Müller-circle (though not necessarily in the fixation point, which will not be in the median plane for eccentric fixation).

So, the theoretical horopter for fixation of a target consists of the Vieth-Müller-circle and a vertical line in the median plane. When the two eyes are fixating at infinity, making binocular gaze parallel, any two corresponding points' projections will meet in the frontoparallel plane at infinity. This is the only case when the theoretical horopter defined by corresponding points can ever be a surface in space.

This description of the theoretical horopter assumes that single vision occurs for corresponding points on congruent visual angles and that the perceived targets are points. This is a useful assumption, since the horopter for parallel binocular gaze then includes every object that is far away compared to the interocular distance. In other words, the horopter is a surface at infinity, which corresponds to our subjective impression: we don't have double vision when we look at a far away landscape.

This is, however, a simplification, since it does not take into account the shape of the retina, and it also ignores the correspondence of more complex visual objects.

For a more complete discussion of the geometry of horopters therefore see Chapter 2 of (Howard & Rogers, 1995) and (Tyler, 1991).

3.4 Epipolar Geometry

The binocular disparity defined above is the difference in visual angle for a pair of retinal projections. This can be visualized as a vector on the retinal surface, a two-dimensional entity. This can be quantified in any retinal coordinate system by two numbers. However, for geometric reasons, the binocular disparity can be broken down in a natural way into two one-dimensional components, commonly referred to as horizontal and vertical disparity. The coordinate system for this breakdown is specified by the geometry of the epipolar lines—epipolar geometry.

3.4.1 Epipolar Lines

For any visual target T in space, the plane containing the object's location and the two eyes' nodal points is called an epipolar plane[8]. The intersections of this plane with the two eyes' retinal spheres are called (a pair of) epipolar lines. In Helmholtz coordinates the epipolar planes can be described by

$$\phi = const. \tag{3.29}$$

For gaze straight ahead, this translates directly into the retinal coordinates of the epipolar lines:

$$\lambda = \phi = const. \tag{3.30}$$

Any object projecting onto an epipolar line in one eye's retina must be located in the corresponding epipolar plane and thus will project onto the paired epipolar line in the other eye's retina (see Figure 3.6).

It is important to note that in this description the nodal points are head-fixed because they are located in the rotation centre of the eyes. The epipolar planes then are also head-fixed. They are not, however, retina-fixed, since any rotation will move the eye with respect to the epipolar planes. The location of the epipolar line on the retina of an eye depends on eye position.

3.4.2 Retinal Motion of Epipolar Lines

As mentioned above, the epipolar plane specified by $\phi = const$ in Helmholtz coordinates will be intersecting the retina in the epipolar line specified by $\lambda = \phi$ in retinal coordinates when gaze is straight ahead. The coordinates of any point on this epipolar line in eye-fixed Euclidean coordinates are then

$$\vec{x}_{epi} = r_{eye} \begin{pmatrix} -\cos(\lambda_{el})\cos(\mu) \\ \sin(\mu) \\ \sin(\lambda_{el})\cos(\mu) \end{pmatrix} \tag{3.31}$$

These are also the coordinates of a point on this epipolar line in head-fixed coordinates, since the eye-fixed and the head-fixed coordinate systems are aligned for gaze straight ahead.

When the eye rotates to a new position, specified by Helmholtz angles θ, ϕ and ψ, the epipolar line remains fixed in space, so the head-fixed coordinates

[8] The epipolar plane for the fixation target thus coincides with the plane of regard.

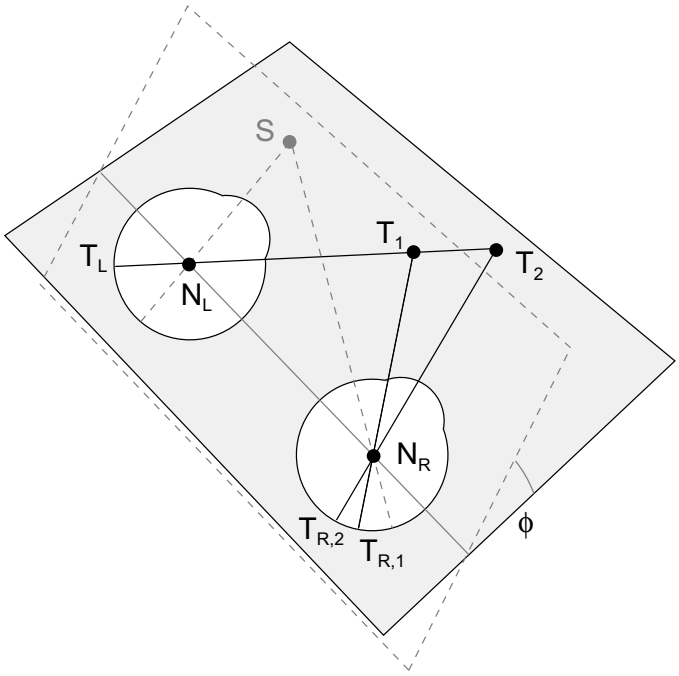


Figure 3.6: Epipolar geometry of visual targets. An object lying somewhere in the grey plane (the epipolar plane) spanned by the object's location and the two eyes' nodal points will project onto the intersections of this plane with the two eyes' retinas. These intersections are called epipolar lines. Conversely, an object projecting onto an epipolar line in one eye must project onto the paired epipolar line in the other eye. An object projecting onto T_L, for instance, must lie in space on the line connecting T_L and N_L. The projection into the right eye of any of these objects, say T_1 in the figure, will lie on the intersection of the epipolar plane with the right eye's retina and thus on T_L's paired epipolar line. Also shown by the grey dashed lines is a second epipolar plane, rotated by an angle ϕ with respect to the first and an object S in that plane, to which the same arguments apply.

do not change. However, both the retinal coordinate system and the eye-fixed coordinate system are moving with the eye. So the description of the same epipolar line will change with eye position, reflecting the fact that epipolar lines are not retina-fixed.

To find the eye-fixed coordinates of a point with head-fixed coordinates $\vec{x}_{epi}$, we have to rotate the eye and the point back to straight ahead position, aligning eye-fixed and head-fixed coordinates. The eye-fixed coordinates of the epipolar line given above, for any gaze direction, thus are

$$\vec{x}_{eye} = \mathcal{M}_{-\psi} \cdot \mathcal{M}_{-\theta} \cdot \mathcal{M}_{-\phi} \cdot \vec{x}_{epi} \tag{3.32}$$

To obtain retinal coordinates for this epipolar line, we invert Equation 3.31, yielding:

$$\mu = \arcsin(y_{eye}) \tag{3.33}$$

$$\lambda = \arccos(-x_{eye} \cos(\arcsin(y_{eye}))) \tag{3.34}$$

The formula resulting from substituting the components of $\vec{x}_{eye}$ in these formulæ with the results of the matrix multiplication in Equation 3.32 would be very complicated, without yielding much insight. However, I have written a simulation visualizing the resulting geometry of the epipolar lines in retinal coordinates depending on eye position. This simulation is described in Appendix E

3.4.3 Horizontal and Vertical Disparity

The disparity of any two retinal projections can now be divided into a component parallel to the direction of the epipolar lines and a component orthogonal to them like this: Let N_L and N_R be the left and right eye's retinal locations for which we want to compute the disparity.

We then call the smallest distance from the projection N_R to the epipolar line in the right eye corresponding to N_L the *vertical disparity* of the pair (see Figure 3.7). The distance on this epipolar line from the point corresponding to N_L to the base of the projection of N_R onto the epipolar line is called *horizontal disparity*. Since the epipolar lines move on the retina with changes in eye position, horizontal and vertical disparities are eye position dependent.

When gaze is straight ahead, the coordinates of the epipolar lines become very simple in retinal coordinates. If (μ_L, λ_L) and (μ_R, λ_R) then are the

retinal coordinates of N_L and N_R, their horizontal and vertical disparities are

$$\delta_{hor} = \mu_L - \mu_R \tag{3.35}$$
$$\delta_{ver} = \lambda_L - \lambda_R \tag{3.36}$$

The general formula for horizontal and vertical disparities depending on gaze angles can in principle be derived from the description of the epipolar lines in retinal coordinates given above.

Vertical Disparity Patterns

When this convention is used, no real world object can have vertical disparity, because no real world object will ever project outside of its epipolar plane (unless some optical apparatus alters its projections). Vertical disparity can, however, be created in random-dot stereograms or in any display that manipulates the left and right eye's views independently.

The horizontal and vertical disparities thus defined are relative to the retinal direction of the epipolar lines. Since the latter depends on eye position, so do the former:

Gaze-dependent vertical disparity will be zero for all real objects only with respect to actual eye position. In other words, when an epipolar analysis of binocular disparity is done, gaze dependent vertical disparity will disappear for all correct matches when the correct eye position is used. This fact could help the system both simplify matching and determine eye position, as we will see in Section 9.1.

3.4.4 Disparities in the Frontoparallel Plane

The epipolar planes are by definition a family of planes rotated in the vertical Helmholtz direction about the interocular axis. This means that when the head is upright, the intersection of any epipolar plane with any frontoparallel plane will be a line parallel to the horizon. A common way of visualizing binocular disparity is to project the two retinal images onto a frontoparallel plane and then shift one of the projections so that the projections of the two foveolas coincide. Any other visual features whose projections coincide in this overlap have no binocular disparity[9].

[9]The objects causing them thus are located on the theoretical horopter.

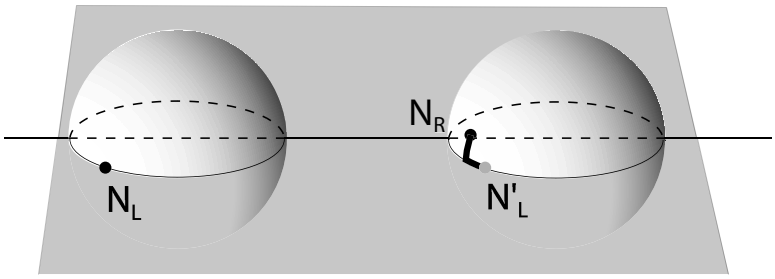


Figure 3.7: The breakdown of binocular disparity into horizontal and vertical components. The grey plane is the epipolar plane for the left eye's retinal position N_L. The binocular disparity is the distance from N_R to the position corresponding to N_L in the right eye, N'_L. The vertical disparity of the pair is the vertical distance from N_R to the epipolar line, the horizontal disparity is the distance from N'_L along the epipolar line to the base of the projection.

The binocular disparity of non-coinciding visual features, as mentioned before, breaks down into the horizontal disparity component along the epipolar lines and the vertical disparity component orthogonal to the epipolar lines. Since the epipolar lines project into the frontoparallel plane as horizontal lines, horizontal and vertical disparities are projected into the frontoparallel plane as horizontal and vertical position differences.

In the literature, the term *vertical disparity* is used somewhat inconsistently, most commonly though it is referring to this projection of disparities into the frontoparallel plane with gaze straight ahead. This disguises the fact that vertical disparity is gaze dependent. I have more to say on this in the review of the literature on vertical disparity in Section 4.3.

3.5 Random-Dot Stereograms

Random-dot stereograms consist of two images composed of picture elements, usually dots, in a random spatial distribution. The locations of these elements can be represented using two-dimensional coordinate vectors $\vec{x}_{L,i}$ and $\vec{x}_{R,i}$. L and R denote the elements of the left and right image, respectively, while i is the index into the elements.

In some studies, the picture elements used are not dots, but gaussian

patches or Gabor patches[10].

A very simple random dot stereogram is defined by $\vec{x}_{R,i} = \vec{x}_{L,i}$ for all i. None of these elements has binocular disparity, the resulting surface created by these elements would upon fusion appear to be flat.

To encode stereoptic shape, the coordinates of a subset of these elements are manipulated to have binocular disparity. For instance, to create a circle floating in front of the background, all the elements in the circle are shifted horizontally[11]:

$$\vec{x}_{R,i} = \begin{cases} \vec{x}_{L,i} + \begin{pmatrix} \delta \\ 0 \end{pmatrix} & , \quad \|\vec{x}_{L,i}\| < r_c \\ \vec{x}_{L,i} & , \quad \|\vec{x}_{L,i}\| \geq r_c \end{cases} \tag{3.39}$$

For a visual example of the creation of a stereogram containing a square see Figure 3.8.

3.6 The Ψ-Method

A common variable to measure in psychophysical experiments is the sensitivity threshold, that is the stimulus intensity at which a stimulus just becomes detectable. Threshold is usually detected in forced choice experiments, where subjects have to decide which of a predefined set of stimulus configurations is present in the current trial. When $\Psi(x)$ is the probability for the subject to correctly perceive the stimulus at intensity x, the probability of a correct answer at that intensity when there are n different stimuli is

[10]Named after Johann Carl Friedrich Gauss (⋆1777, †1855) and Denis Gabor (⋆1918, †1970), respectively. The formulæ for the intensities for these functions are

$$I_{Gauss} = e^{-\frac{(x-x_0)^2}{2\sigma^2}} \tag{3.37}$$

$$I_{Gabor} = e^{-\frac{(x-x_0)^2}{2\sigma^2}} \cdot e^{-\frac{(x-x_0)^2}{2\sigma^2}} \cdot \cos\left(\frac{x-x_0}{T}\right) \tag{3.38}$$

[11]When depth is encoded by horizontal shifts in the random dot stereogram and the stereogram is presented in a frontoparallel plane, the binocular disparities of its elements will be the same as they would be if there were a set of dotlike physical objects floating in space.

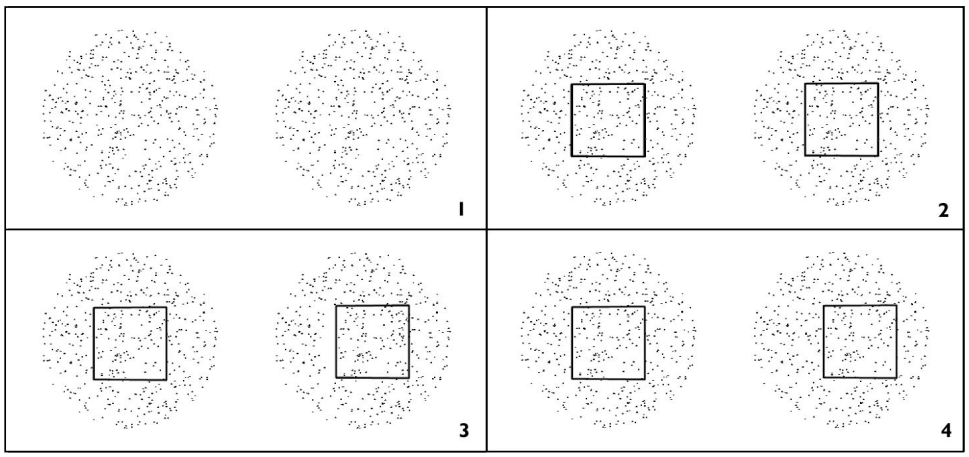


Figure 3.8: The construction of a simple stereogram, encoding a square floating in depth before the circle of dots.
1: Two identical sets of random dots.
2: The same square region is selected on both stereograms.
3: The square region is shifted to the right in the right stereogram, creating a gap on its left and a region of double density on the right.
4: The gap is filled with random dots of appropriate density. The double dots to the right of the square are removed. Note that the square outlines are only present in this figure for illustratory purposes and are not part of the stereogram.

$$p(x) = \Psi(x) + \frac{1 - \Psi(x)}{n} \tag{3.40}$$

The second term reflects the fact that even when subjects do not perceive the stimulus, it is still possible for them to guess correctly.

Usually in psychophysics, Ψ is modelled as a sigmoid function:

$$\Psi(x)_{a,m} = \frac{1}{1 + e^{-a(x-m)}} \tag{3.41}$$

Ψ is called the psychometric function. Common intensity parameters varied for threshold detection include stimulus contrast or stimulus noise.

The typical problem in psychophysical experiments is to determine the stimulus intensity c where the perception probability reaches some predetermined value, the threshold. This can be done by simply delivering a great number of stimuli at a range of preset stimulus intensities, and then fitting a sigmoid function to the performance data obtained. More efficient are so-called staircase methods, where the stimulus intensity is varied depending on the correctness of the last answer given by the subject. A common staircase protocol could for instance increase the stimulus intensity if a wrong answer was given and decrease intensity for a right answer. If the step sizes are chosen carefully, stimulus intensity will stay close to the threshold intensity.

The first version of the study presented in Chapter 6 used such a staircase method. The efficiency of psychophysical testing can, however, be increased significantly, if the choice of intensity of each new stimulus is based on all the answers given before in the experiment.

Therefore in most of the studies in this dissertation the Ψ-method, developed by Kontsevich and Tyler (Kontsevich & Tyler, 1999), was used to increase the effectiveness of the experiments. In brief, the Ψ-method keeps track of the posterior probability for an array of possible threshold values and determines, before each trial, which stimulus intensity would be expected to decrease the entropy of this probability distribution the most.

3.6.1 Entropy Minimization

The Ψ-method uses Bayes' theorem to compute a threshold estimate from the already completed trials. This is done by updating the posterior probability distribution for the psychometric functions. These functions can be described

by their two parameters a and m (see equation above), with $p_n(a, m)$ giving the probability for the psychometric function $\Psi_{a,m}$ after n trials.

The method evaluates this probability distribution by means of its entropy

$$\mathcal{H}_n = \sum_{a,m} p_n(a, m) \log(p_n(a, m)) \tag{3.42}$$

The smaller the entropy of the distribution, the higher the confidence as to which set of parameters (a,m) best describes the actual psychometric function. The smallest possible value of $\mathcal{H}_n$ is zero, corresponding to the case where one particular $p(a, m)$ equals one and all others are zero.

The basic idea of the Ψ-method is to compute the expected Entropy for all the possible intensities for the next trial, and then place the trial at the intensity where the expected entropy becomes minimal.

3.6.2 Calculations

Before each new trial is placed, the expected probability for getting correct and wrong answers is computed by summing up the products of the probability for that answer for each psychometric function with the overall probability for that psychometric function:

$$p_n(correct, x) = \sum_{a,m} p_{a,m}(x) p_n(a, m) \tag{3.43}$$

$$p_n(wrong, x) = \sum_{a,m} (1 - p_{a,m})(x) p_n(a, m) \tag{3.44}$$

Now, using Bayes' theorem, the posterior probability of each psychometric function is computed, assuming that the next trial will produce a correct or wrong response R:

$$p_n(a, m|x, R) = \frac{p_n(a, m) p(R|a, m, x)}{\sum_{a,m} p_n(a, m) p(R|a, m, x)} \tag{3.45}$$

Using this distribution we can compute the entropy of the probability density function p_n, assuming that the next trial at intensity x will produce response R:

$$\mathcal{H}_n(R, x) = \sum_{a,m} p_n(a, m|R, x) \log(p_n(a, m|R, x)) \tag{3.46}$$

Using the probability for each response and the resulting entropy for each response we can compute the expected entropy for each test intensity:

$$[\mathcal{H}_n](x) = \mathcal{H}_n(correct, x)p_n(correct, x) + \mathcal{H}_n(wrong, x)p_n(wrong, x) \quad (3.47)$$

Now we simply place the next trial at the intensity x_n for which the expected entropy becomes minimal. This ensures that the next trial yields the maximum available information about the psychometric function.

After the trial, the probability distribution is updated to reflect the result R_n of the trial:

$$p_{n+1}(a, m) = p_n(a, m|x_n, R_n) \quad (3.48)$$

Then a new trial begins unless some condition for finishing the experiment is met. The final estimates for the parameters of the psychometric function are

$$a = \sum_{a,m} a p_n(a, m) \quad (3.49)$$

$$m = \sum_{a,m} m p_n(a, m) \quad (3.50)$$

For further details see (Kontsevich & Tyler, 1999).

3.7 Cylinder Projection

In Chapter 7 we use computer simulations to see whether different patterns of binocular coordination might allow the stereoptic system to operate with smaller search regions. To compute the areas of the simulated search regions, we projected them from the curved retina onto a flat surface using an area-preserving cylindrical projection. In this projection, each point on a sphere is projected outwards onto the surface of a cylinder with the same radius as the sphere. The direction of the projection is perpendicular to the axis of the cylinder.

This means that any point with retinal coordinates λ and μ

$$\vec{x}_{ret} = r_{eye} \begin{pmatrix} -\cos(\lambda)\cos(\mu) \\ \sin(\mu) \\ \sin(\lambda)\cos(\mu) \end{pmatrix} \quad (3.51)$$

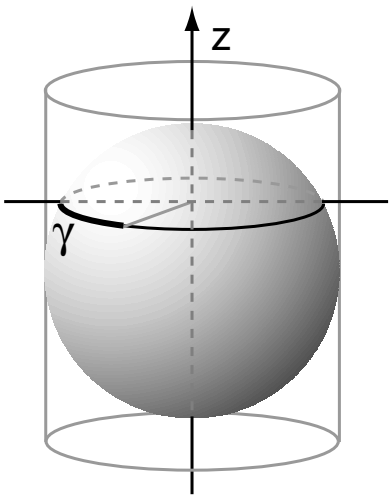


Figure 3.9: In cylinder projections, a point on the surface of a sphere is projected onto the surface of a cylinder wrapped around the sphere. The two resulting coordinates are z and the polar angle γ in radians.

will be projected[12] onto

$$\vec{x}_{cyl} = \begin{pmatrix} \frac{-\cos(\lambda)\cos(\mu)}{\cos^2(\lambda)\cos^2(\mu)+\sin^2(\mu)} \\ \frac{\sin(\mu)}{\cos^2(\lambda)\cos^2(\mu)+\sin^2(\mu)} \\ \sin(\lambda)\cos(\mu) \end{pmatrix} r_{eye} \tag{3.52}$$

The cylindrical coordinates of that point then are the polar angle γ and the vertical position z (see Figure 3.9):

$$\vec{x}_{cyl} = \begin{pmatrix} \arctan(\frac{\sin(\mu)}{-\cos(\lambda)\cos(\mu)}) + \pi \\ \sin(\lambda)\cos(\mu) \end{pmatrix} \tag{3.53}$$

3.8 The Moore-Penrose Pseudoinverse

A physical object projects onto the retinas at the retinal positions P_L and P_R, where both retinal positions are given in a head-fixed coordinate system

[12] The z-coordinate is unchanged, the x and y coordinate get scaled by the ratio of the radius of the retinal sphere and the radius of the circle parallel to the x-y-plane through the retinal point

with the origin in the nodal point of the respective eye. The object O must be at the intersection of the projections of these locations into space. Or in Galileo's language of nature:

$$\vec{O} = k_L \vec{P_L} = \vec{I} + k_R \vec{P_R} \tag{3.54}$$

With the interocular vector $\vec{I}$ connecting the two eyes' nodal points. This equation can be written as

$$\vec{I} = \left(\vec{P_L}\vec{P_R}\right) \cdot \begin{pmatrix} k_L \\ -k_R \end{pmatrix} \tag{3.55}$$

$$= \mathcal{M} \cdot \begin{pmatrix} k_L \\ -k_R \end{pmatrix} \tag{3.56}$$

Which yields

$$\begin{pmatrix} k_L \\ -k_R \end{pmatrix} = \mathcal{M}^{-1} \cdot \vec{I} \tag{3.57}$$

where $\mathcal{M}^{-1}$ is the inverse of the matrix $\mathcal{M}$ composed of the two retinal location vectors. In general, however, this matrix will not be invertible, signifying the fact that in general there will be no possible location in space projecting on any two given retinal locations.

In a case where the two projection lines do not meet, it might still be desirable to find the point in space where they come closest to meeting. With use of the Moore- Penrose pseudoinverse instead of the inverse the above equations yield the points on each of the projection lines closest to the other projection line. I will apply this technique in Chapter 7.

Chapter 4

Binocular Depth Vision

He had but one eye,
and the popular prejudice runs in favor of two.
Charles Dickens[1]

The task of binocular depth vision is to reconstruct the three-dimensional structure of a visual scene from the two two-dimensional retinal images. While there remain many questions as to how this is done in detail, a lot of the neuronal substrate as well as the functional properties of the systems involved has been revealed by recent research. This chapter will give an overview on current knowledge and opinion.

Depth vision refers to the ability of the visual system to use visual cues to reconstruct the depth of a visual object. There are several known cues the brain can and does use to achieve this goal. Binocular disparity is one of them, and thanks to extensive research and the popularity of random-dot stereograms may be the one that comes most readily to mind. But depth can also be reconstructed using colour, motion of the visual scene, occlusion of objects and visual parallax from a change of viewpoint[2].

4.1 The Structure of Stereopsis

Stereopsis solves the problem of reconstructing depth of visual objects from two images. This complex problem can be broken down into smaller steps,

[1] In *Nicholas Nickleby*.

[2] These cues become more important in animals or people with little or no stereopsis. See (Land, 1999) for an interesting account on the parallax techniques of jumping spiders.

each of which might or might not be realized in separate neuronal systems.

The first step is to find matching features in the two retinal images. Once these matches are identified, their absolute retinal disparity can be computed. Comparison between these disparities yields relative retinal disparities. To compute depth from disparities, information about the position of the eyes is necessary. Either proprioception, efference copy of eye muscle commands or visual input can be used for this step, called relief transformation. After relief transformation, the loci of visual objects in space are in principle available.

It is very likely that there is not one, but several physiologically separate stereoptic systems, processing different types of information. One distinction is between transient and sustained stereoptic pathways, another is between coarse and fine stereopsis.

4.2 The Hierarchy of Stereopsis

4.2.1 Stereo Matching

The task of stereo matching is to identify visual features in the two eye's retinal images that have been cast by the same physical object. This becomes a problem especially when many similar or identical objects are present: If a visual scene contains n identical objects, there are n^2 possible pairings between the left and the right image, only n of which are correct. In a random-dot stereogram which presents 5000 dots to each eye, the 5000 correct matches are opposed by a staggering 5000!-5000 possible false matches. Stereo matching solves random-dot stereograms with comparable and higher complexity within a few hundred milliseconds.

Seeing Ghosts

Stereo matching can be a real problem, however. This is demonstrated neatly by the double nail illusion, published by Krol and van de Grind in 1980 (Krol & van de Grind, 1980). In this illusion, shown in Figure 4.1, the human visual system can be made to mismatch visual image features even in a very simple case where there is only a total of two possible matches.

For a physical demonstration of this phenomenon, I have built an apparatus that is described in Appendix C.

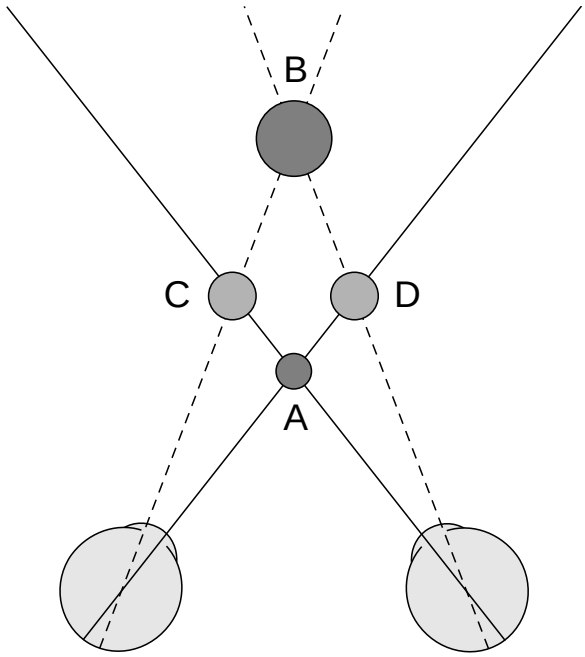


Figure 4.1: The double nail illusion. The two eyes fixate a point between the two metal bars A and B. The correct matches are between the images projecting along dashed lines and the images projecting along the solid lines, leading to the accurate perception of the bars. However, if the matching algorithm matches the left eye's dashed image with the right eye's solid image and vice versa, the bar's positions are reconstructed wrongly and the bars are seen in positions C and D instead. If other cues to the depth of the bars are carefully removed, observers indeed see the "ghosts" in C and D.

The Primitives of Matching

A second important question connected to stereo matching concerns the so-called matching primitives. What image features are used for identifying matches? Marr and Poggio in 1979 proposed that the algorithm is searching for matches between regions where the second spatial derivative of normalized luminance has a zero- crossing (Marr & Poggio, 1979). These zero crossings occur at positions where the luminance gradient has a maximum, i.e. at the edges where dark and light patches meet in the image. In Marr and Poggio's model these zero-crossings were matched in the output of four different spatial filters, to minimize ambiguities in the matching process.

Mayhew and Frisby in 1981 produced evidence that the matching primitives cannot be zero crossings exclusively, by creating stereograms that contained no disparity between zero crossings but only between luminance peaks (Mayhew & Frisy, 1981).

More recently, the energy model of cortical binocular cells developed by Ohzawa and others has provided more insight into which image features are matched. I will review this in Section 4.5.2.

The Epipolar Constraint

Stereo matching is a very complex task. Therefore it has commonly been assumed that matching eliminates false matches by making use of the geometric fact that correct matches must lie on epipolar lines. This has been called the epipolar constraint. The epipolar constraint reduces the matching problem from a two dimensional task to a one dimensional one. If the epipolar lines have a very small width, there will rarely be false matches.

The assumption that the visual system uses the epipolar constraint was called into question by Rogers and Bradshaw in 1996 (Rogers & Bradshaw, 1996), when they found that the apparent motion between two stereograms displayed alternatingly was independent of the vergence state. They carefully concluded that

> Assuming that minimal dichoptic motion is a valid indicator of optimal binocular correspondence, we found little evidence to suggest that the visual system is able to compensate for the change in the location of epipolar lines with the changes in vergence state.

Panum's Fusional Areas

For each retinal stimulus position in one eye, there is a range of retinal positions in the other eye where matches can be found. This range is commonly referred to as Panum's fusional area.

4.2.2 Computing Disparities

Once matching retinal features have been found, the retinal disparities can be computed. The first step is the computation of absolute retinal disparity. However, as Westheimer showed in 1979, stereopsis depends on relative rather than absolute disparity (Westheimer, 1979). Erkelens and Collewijn in 1985 and Regan *et al.* in 1986 confirmed this by demonstrating that large changes in absolute disparity of a wide field display produce no sensation of motion-in-depth. (Erkelens & Collewijn, 1985) (Regan *et al.*, 1986).

The relative disparity of two pairs of matched features is the difference in their absolute disparities. The length of this difference vector is independent of retinal position and thus does not change with eye position. This robustness with respect to a change in eye position or errors in the eye position signals is believed to be the reason for the stereoscopic system relying on relative disparity rather than on absolute (Cumming & Parker, 1999). It should be noted, however, that the *direction* of the relative disparity vector on the retina does change when eye position changes: For two objects on the Vieth-Müller-circle the absolute disparities are oriented along the horizontal meridian of the eyes when both eyes are looking straight ahead. Since the absolute disparities are along the horizontal meridian, the relative disparity is, too. Suppose now both eyes rotate 90° counterclockwise around their lines of sight. The disparities will now lie along the vertical meridians of the eyes.

4.2.3 Relief Transformation

Absolute or relative disparities do not by themselves allow the computation of object depth. An eye-position signal is needed to interpret the disparities (Howard & Rogers, 1995). This signal can come from either proprioception, efference copy or the visual images themselves. Mechanism for this have been proposed by Mayhew and Longuet-Higgins in 1982 (Mayhew & Longuet-Higgins, 1982) and by Gårding *et al.* in 1995 (Gårding *et al.*, 1995). These

models will be discussed in detail in Section 4.3.2

4.3 Vertical Disparity

4.3.1 Theoretical Considerations

When I defined the terms horizontal and vertical disparity in Section 3.4.3, I mentioned the fact that my definition differs from the usage of the two terms in the literature. The reason is that there seems to be no generally agreed upon definition of vertical disparity in the literature. Depending on the phenomenon studied and the stimuli used, vertical disparities are defined using different coordinate systems from study to study. Each of those coordinate systems will serve its purpose for the study it is used in, but the unfortunate fact that all these description are filed under the term vertical disparity leads to potential confusion.

Usually the fact that the two eyes are separated horizontally is quoted as the reason that horizontal disparities are the most important signal for encoding depth. This may sound appealing and convincing, but as I have shown in the discussion on coordinate systems in Section 3.4, the actual geometry is more complicated.

When horizontal disparity is defined as the disparity encoding the depth of real world objects, its retinal direction is eye-position dependent, since the epipolar lines giving the horizontal disparities' retinal direction are head fixed. This means that a retinal disparity that is purely horizontal for one eye position will have some vertical component for another eye position.

We can avoid this complication by giving a reference eye position, with gaze straight ahead[3] being an obvious, if somewhat arbitrary, candidate. Any retinal disparity then can be described in terms of horizontal and vertical disparity *relative to gaze straight ahead*.

I have shown in Section 3.4.3 that disparities of real objects must follow the epipolar geometry, that is, any disparity vector can only connect a point on one eye's epipolar line with any point of the other eye's paired epipolar plane. I have also shown that the disparity of any object when projected

[3]In the oculomotor literature, straight ahead gaze can be a complicated entity. For this dissertation, straight ahead is equivalent to the direction along the x-axis of the head-fixed coordinate system. This is just a convenient convention to describe a spatial direction and does not have implications for ocular kinematics.

into the frontoparallel plane must indeed be a horizontal vector, since the epipolar planes intersect with any frontoparallel plane in a horizontal line. In other words, there can be no vertical disparities *relative to the epipolar lines for the current eye position.*

There can, however, be vertical disparities relative to the reference eye position chosen, and the pattern of these disparities will allow for the computation of eye position *relative to this reference.*

To clarify the dependence of horizontal and vertical disparities on eye position, consider the following: for any eye position, any real world object's retinal disparity must be purely horizontal. Now rotate both eyes 90° around their lines of sight. This rotation has the disparity vector rotate on the retina[4] by 90 degrees, so now the same object, with the eyes having the same direction of gaze, has purely vertical disparity, *but only relative to the original eye position without the torsion.* Relative to the new epipolar lines for the new eye position, which have also been rotated on the retinal by 90°, this disparity is still horizontal.

4.3.2 Empirical Evidence

The Induced Effect

In the 1930s Ogle described that when one eye's retinal image was magnified in the vertical direction, using a cylindrical lens, a frontoparallel surface appeared to the observer to be tilted around a vertical axis (Ogle, 1938) (Ogle, 1939) . He called this the induced size effect and in his book *Researches in Binocular Vision* gave the following hypothesis for the reason for it:

> Now we can make the hypothesis that the induced effect, in which the image of one eye is magnified in the vertical meridian, corresponds to the geometric situation in which the eyes are converged asymmetrically, and the result in both cases is a change in the egocentric stereoscopic spatial localization[5]

Ogle's idea thus was that the vertical disparities of the image suggested to the visual system a certain eye position and a viewing geometry associated with it, which in turn influenced the interpretation of the horizontal disparities coding depth.

[4] And also change position, which is irrelevant for this argument.

[5] Cited from (Ogle, 1950), page 223f.

Westheimer in 1978 reported that he was unable to reproduce the induced effect with a different stimulus configuration (Westheimer, 1978), using vertically expanded line stereograms rather than real surfaces and lenses, as Ogle had done. Westheimer speculated that for oblique lines, a vertical magnification of one image would actually be equivalent to horizontal magnification of the other image. Arditi *et al.* in 1981 expanded on this, proposing that the induced size effect arises because vertical magnification of images containing oblique components gave rise to horizontal disparities (Arditi *et al.*, 1981). Arditi in 1982 tested this theory, and found that oblique components indeed were necessary to elicit the induced effect (Arditi, 1982). He also found evidence that the matching leading to the detection of the horizontal disparities was not done in a horizontal direction, but that instead nearest neigbors on the oblique lines present were matched.

Models for the Induced Effect

Also in 1982, Mayhew and Longuet-Higgins put forward a model of depth perception that explicitly computed the viewing geometry from the vertical disparities of a view image points, and correctly predicted the induced effect (Mayhew & Longuet-Higgins, 1982). They also predicted that the induced effect should be a whole field phenomenon, since different estimates for the viewing geometry for different parts of the retinal image seemed unlikely to occur.

Gillam and Lawergren in 1983 proposed a similar model, and also noted that the discomfort commonly associated with the induced effect can be attributed to the fact that the vertical disparities signal a certain viewing geometry, which the extraretinal signals contradict (Gillam & Lawergren, 1983).

Stenton *et al.* in 1984 applied vertical disparities not to all of the retinal image, but to varying fractions of points present in it, and found that the size of the induced effect varied linearly with the percentage of points included (Stenton *et al.*, 1984). They concluded that the induced effect was based on a mechanism for pooling vertical disparity. They argue that this is in keeping with theories proposed to explain the effect, since the effect of viewing geometry is not local, but global.

Rogers and Koenderink in 1986 proposed a different explanation, based on experiments where they varied the vertical extent of a surface seen monocularly, while the subjects head was moved side to side. This configuration

also led to the perception of a slanted surface, but, as Rogers and Koenderink argued, not for the reasons put forward in Longuet-Higgins and Gillam and Lawergren's models (Rogers & Koenderink, 1986). They also found that it was possible to have a different induced effect in the left and the right half of the retinal image, contradicting the pooling hypothesis. They proposed that the induced effect as well as the motion parallax effect described by them was based on the analysis of image gradient patterns, which have been described in a review paper on optic flow by Koenderink (Koenderink, 1986). This, they argued, could be done without estimating the viewing geometry.

Vertical Disparity Signals

The hypothesis that vertical disparities are used to judge the viewing geometry not only makes predictions about the slant of surfaces, of course, but also about the overall shape of objects. In 1991, Cumming *et al.* asked subjects to judge the shape of objects encoded in random-dot stereograms, when either the vertical disparity of the dots was manipulated or actual eye position was changed (Cumming *et al.*, 1991). They found no influence of vertical disparity on the shape judgements, but a change in perception with changing eye position, and concluded that vertical disparity is not used to scale disparities for distance, but extraretinal signals likely are.

Rogers and Bradshaw in 1993 replied, by demonstrating the influence of vertical disparity gradients on the perception of surfaces in experiments they did, and by suggesting that the failure of Cumming *et al.* to find an influence of vertical disparity was due to their stimuli being too small (Rogers & Bradshaw, 1993).

Polar Disparity Model

In 1994 Liu *et al.* presented a different approach to binocular disparity, describing it not by means of horizontal and vertical disparity directions, but as a two-dimensional vector, giving its length and angular direction on the retina (Liu *et al.*, 1994). They showed that the spatial distribution of the angle of disparity could describe the slant of a planar surface relative to gaze, and that geometric and induced effect can be explained by the distortion of the disparity angle map by the magnifier.

Regional disparity correction

In 1995 Gårding *et al.* put forward a complex model that uses the overall retinal disparity pattern to achieve two steps of image analysis they call *disparity correction* and *disparity normalization* (Gårding *et al.*, 1995). Disparity correction is the process of determining the relief structure of the scene, i.e. determining which parts are closer to the observer. Disparity normalization then computes the actual metric structure of the scene. They review the existing models for vertical disparity processing, grouping them into two categories according to the scale at which these disparities were taken into account. Local models use the vertical disparity information to compute the depth of visual objects in the immediate neighborhood, while regional models assume a range of pooling. They cite evidence for regional processing, but also point out that the existing regional models, including that of Mayhew and Longuet-Higgins, were incompatible with the results on perception of three-dimensional shape. The solution they propose is a model that uses vertical disparity information pooled over retinal regions for disparity correction, i.e. to retrieve the relative depth structure of the scene, but not for disparity normalization. Thus they can explain the induced effect as well as the failure of vertical disparities to influence shape perception.

In the first part of their paper they give an extensive overview on the geometry and mathematics of binocular vision[6], describing epipolar geometry and providing visualizations of the disparity fields for different viewing geometries. They also remark that the use of epipolar geometry in stereomatching would require the system to either know eye position or estimate

[6]Here they claim that

> In order to obtain a numerical representation of disparity we must choose a way of parameterizing retinal position, or equivalently, visual direction relative to the each eye. This can be done in a number of different but essentially equivalent ways, e.g. by spherical of projective coordinates. Clearly this choice is only a matter of convenience which has nothing to do with the physical shape of the retina[7].

While this is true, it is also misleading, for their choice of coordinate system, while not interfering with the physical shape of any object, leads them to name the horizontal and vertical components of their projection h and v. As I have argued before, this suggests the implicit assumption that disparities are estimated relative to a reference gaze position. The choice of coordinate system, as indeed every choice made in describing a complex system, is not a matter of convenience, but a way of expressing hidden assumptions.

it from the visual information—two of the main themes of this dissertation.

Adams *et al.* followed up on this model in 1996 with a study where they looked at the effect of vertical disparity manipulation on estimates of the amplitude of a dome shape encoded in a random dot stereogram (Adams *et al.*, 1996). While they found a significant effect of vertical disparity, it was less than predicted by either the theory of Mayhew and Longuet-Higgins or Gårding *et al.*. Additionally they found that when they manipulated the vertical disparity of half of the dots, leaving the other half unchanged, the resulting response could be predicted from pooling the vertical disparity information, unless the manipulation of half of the dots was very strong, in which case these points tended to be ignored for the pooling.

Ideal Observer Theory

The study of the effect of vertical disparities on perception was mated to Bayesian theory in a study by Porrill *et al.* from the same lab, published in 1999 (Porrill *et al.*, 1999a). Again half of the dots of a random-dot stereogram were given vertical disparity, suggesting two different estimates of viewing geometry within the stereogram. They specify ideal observers for this task[8], differentiating between non-robust and robust ideal observers. The robust ideal observer is "an observer that has optimal performance in the presence of a suitably specified population of outliers"[9]. Monte-Carlo simulations using the robust ideal observer led to the prediction of a bifurcation in subject's responses: when the difference in vertical disparity, and conversely the difference in encoded viewing geometries, becomes too big, the observer switches from a pooling strategy to a strategy where he discards half of the dots and bases his depth estimate on only one half of the dots. The human observers of Porrill *et al.* indeed showed this bifurcation behaviour, indicating that the human stereovisual system can be described by optimal observer theory.

In a conference presentation in 2002, Frisby *et al.* reported that when they combined vertical disparity information with changes in vergence, the performance of their subjects could be described by a maximum likelihood analysis of viewing geometry, with the weight for the vergence cue being smaller than the weight for the disparity cue, indicating that the vergence

[8] The ideal observer is a theoretical entity which uses all the available information and Bayesian statistical decision theory to compute the probabilities of physical events given the observations made. These probabilities form the posterior probability distribution.

[9] Cited from (Porrill *et al.*, 1999a), page 63

signal is noisier.

Matching Vertical Disparities

The studies described so far were all concerned with the effect of vertical disparities on the perception of depth. But a more fundamental question, that has also been raised by Gårding *et al.*, concerns the epipolar constraint: does the stereo-matching system take into account epipolar geometry and restrict the search for matching features to retinal bands aligned with the epipolar lines?

In 1996 Rogers and Bradshaw reported that they were unable to produce evidence for the use of the epipolar constraint, when they asked observers to identify dichoptic image pairs that created less apparent motion in different vergence positions (Rogers & Bradshaw, 1996). This finding, however, was complicated by the fact that it is not immediately clear that reduction of dichoptic motion is a valid indicator for good binocular correspondence. Also, the results were complicated by the fact that their subjects were not converging completely.

If the epipolar constraint is not used (I will argue in Chapter 6 that indeed it is not), then matching must become two- dimensional. In other words, Panum's fusional areas must have vertical extent on the retina. In 1997 Stevenson and Schor set out to measure the shape and extent of the fusional areas on the retina, finding that while the horizontal extent of Panum's area was at about 1° somewhat larger than their vertical extent, vertical disparities of up to 45 minutes of arc did still allow near/far judgements. This means that the stereoptic system clearly is able to match outside of the epipolar lines.

This was confirmed in 1998 in a study by Farell, who found that stereo-matching occurs in all directions of two-dimensional retinal space (Farell, 1998).

Headcentric Disparity

In 1998, Erkelens and van Ee published a model for disparity processing, that uses headcentric disparity (Erkelens & van Ee, 1998). In their model, an extraretinal signal for eye position is used to recompute retinal disparity into headcentric disparity. Any remaining vertical disparities then must be due to an error in the eye position signals. The six different possible eye

position errors (three muscle pairs in two eyes) lead to different patterns of horizontal and vertical disparities. These patterns are then used by the model to correct the faulty extraretinal signals. The distance of visual objects can then be computed by horizontal disparity alone. This model uses the global disparity field to recalibrate its eye position signals and local disparity information to compute depth. This model is especially attractive, since it completely removes the implicit arbitrary reference eye position that comes with the usual conception of vertical disparity. In that respect it is similar to the earlier polar disparity model by Liu *et al.* mentioned above.

Slant Perception

In 1999 Backus *et al.* published an extensive study on slant estimation based on a number of cues, including extraretinal information about vergence state, vertical disparity (Backus *et al.*, 1999). They found that both cues are used to interpret horizontal disparity, but that vertical disparity information was given more weight when the two signals were in conflict. However, when the vertical disparity signal was difficult to obtain, as in stimuli composed of vertical lines, the system relies on extraretinal information. They could explain the data they found using a model based on the weighted average of the predictions based on vertical disparity and extraretinal information. The weights, however, changed with different viewing positions, indicating that a complex estimation of the reliability of the signals is involved.

Anisotropy of Noise Tolerance

In 2001, Palmisano *et al.* examined the detection of stereoscopic corrugations in the presence of horizontal and vertical additive disparity noise, and found that while at low amplitudes of corrugations and noise the tolerance for vertical disparity noise was much greater than the tolerance for horizontal disparity noise, the relation reversed for large amplitudes of corrugations and noise (Palmisano *et al.*, 2001). This, they concluded, added to the existing evidence for stereopsis involving complex two-dimensional matching rather than simple one-dimensional processing. They also concluded that the detection of corrugation surfaces probably involved extensive post-matching processing.

Judgements of Distance and Direction

The notion that the visual system uses the patterns of vertical disparities to compute absolute depth has the system, at least implicitly, compute an estimate of eye position and fixation distance. In principle it would be possible to use these estimates for judgements of binocular eye position, which in turn could be used to determine absolute distance to objects and eccentricity of gaze.

Brenner *et al.* found in 2001 that a horizontal gradient of vertical disparities indeed is used to determine the distance to an object subjects are looking at (Brenner *et al.*, 2001). But they also concluded that the mechanism that explained their results was distinct from the mechanism causing the induced effect, indicating that the analysis of vertical disparities is possibly done separately for different tasks.

Banks *et al.* in 2002 reported that they didn't find any evidence for the use of vertical disparity in the judgement of azimuth (Banks *et al.*, 2002).

4.4 The Empirical Horopter

In Section 3.3 I described the so-called theoretical horopter. Its derivation was based on the assumption that equal visual angles in the two eyes are a prerequisite for binocular single vision. While this is intuitively a reasonable assumption, leading to a well-defined entity, the actual empirical horopter differs from it in important aspects.

The theoretical horopter was constructed based on the assumption that corresponding retinal points in the two eyes are those having the same retinal coordinates[10]. But experiments have shown that the actual region of space where objects appear single—the so-called empirical horopter—differs from the Vieth-Müller-circle, which implies that the corresponding retinal points do not have the same retinal coordinates.

Ogle showed that the empirical horizontal horopter is a conic section lying between a frontoparallel line and the Vieth-Müller-circle (Ogle, 1950). The curvature of this deviation, named the Hering-Hillebrand deviation, after the

[10]There is another definition of the horopter where corresponding points are defined by congruency of the retina, taking into account the actual shape and size of the two retinas. To distinguish the two theoretical horopters, Howard and Rogers call the latter the theoretical point horopter and the one defined above theoretical optic-array horopter. For further details see (Howard & Rogers, 1995), p. 48ff.

two researchers who discovered it, can be explained by the assumption that the nasal eccentricity is larger than the temporal eccentricity for every pair of corresponding points, with this difference increasing with eccentricity.

Another empirical deviation of the horopter from the predicted theoretical form concerns the vertical horopter. The assumption of equal visual angles comprising corresponding retinal points predicts that the vertical horopter should be a vertical line. However, as Volkmann (Volkmann, 1859) and Helmholtz (Helmholtz, 1867) have found, corresponding meridians in the vertical direction are both rotated outwards by about 1° from the true vertical. Since no such tilt exists for horizontal meridians (as would be expected if the tilt represented ocular torsion), this deviation is best described as a shear of retinal correspondence (Tyler, 1991). The result of this shear is that the vertical horopter is not a vertical line in space but is tilted with its top away from the observer.

In 1999, Siderov *et al.* published an experimental study investigating the reason for the backwards tilt of the vertical horopter (Siderov *et al.*, 1999). They found that no cyclovergence eye movements occurred during their measurements. This in combination with the fact that the backwards tilt of the horopter increases with viewing distance led them to conclude that a shear in the binocular correspondence indeed is the reason for this tilt.

4.5 Neurophysiology

The widespread belief in the 19th and early 20th century was that stereopsis was a feature of vision made possible by higher cortical circuits. Helmholtz wrote

> We therefore learn that two distinct sensations are transmitted from the eyes, and reach consciousness at the same time and without coalescing; that accordingly the combination of these two sensations into a single perceptual picture of the external world is not produced by any anatomical mechanism of sensation, but by a mental act[11].

Ramon y Cajal proposed that inputs from corresponding regions of the two retinas converge on what he called "isodynamic cells", and that this

[11] (Helmholtz, 1863), cited after (Howard & Rogers, 1995)

mechanism forms the basis of binocular fusion (Ramon y Cajal, 1995). This idea was verified when Hubel and Wiesel reported in 1959 and 1962 that cells in the cat's visual cortex had receptive fields in corresponding positions in the two retinas (Hubel & Wiesel, 1959) (Hubel & Wiesel, 1962).

In 1967 Barlow *et al.* described the behaviour of disparity-sensitive cells in the cat's visual cortex (Barlow *et al.*, 1967), finding that certain binocular cells responded selectively to lines and bars having binocular disparity. This was confirmed by Nikara *et al.* in 1968 (Nikara *et al.*, 1968).

4.5.1 Transient and Sustained Pathways

There is a clear anatomical and functional distinction between the visual parvocellular (P) and magnocellular (M) pathways from the retina through the primate lateral geniculate nucleus to the cortex (For a review see (Schiller *et al.*, 1991)). The existence of these two pathways suggests that there may be perceptual tasks specific for each of these pathways, and in 1987 and 1988 Livingstone and Hubel proposed that primate stereopsis was mediated solely by the M pathway (Livingstone & Hubel, 1987) (Livingstone & Hubel, 1988).

M cells typically have large receptive fields and are selectively sensitive to transient stimuli, while P cells have substantially higher spatial resolution and typically have lowpass temporal sensitivity (Derrington & Lennie, 1984). In 2000, Kontsevich and Tyler tested the hypothesis by presenting stimuli with dynamic random noise stimuli (Kontsevich & Tyler, 2000). They found that the stereoscopic system is more sensitive to sustained random-dot stimuli than to transient ones, indicating that stereopsis in humans is mainly influenced by the parvocellular input from the P stream.

4.5.2 Primary Visual Cortex

In 1997 Cumming and Parker demonstrated that neurons in V1 of the awake monkey are selective to absolute disparity (Cumming & Parker, 1997). They also showed, using anticorrelated random-dot stereograms[12], that these neurons can detect disparities that do not give rise to a percept of depth. In 2000 the same authors showed that V1 neurons of the macaque responded

[12] Anticorrelated stereograms are stereograms where the contrast between the two retinal images is inverted. Thus, if one eye sees a display of white random dots on a black background, the other eye sees black dots on a white background. Anticorrelated stereograms cannot be fused by human observers.

to local disparity patterns rather than to the depth perceived by the animal (Cumming & Parker, 2000). Prince *et al.* showed that while the precision of disparity detection can be supported with signals from a small number of V1 neurons, these neurons can also consistently outperform the observer (Prince *et al.*, 2000).

Recently Büchert *et al.* demonstrated that binocular interactions are present in V1 in humans, using fMRI (Buchert *et al.*, 2002).

In V1 Hubel and Wiesel found two main cell types, termed simple and complex cells (Hubel & Wiesel, 1962), both of which can detect binocular disparities.

Simple Cells

A defining characteristic of simple cells is that they show linear spatial summation (Hubel & Wiesel, 1962). Their responses to monocular stimuli can be summarized by a receptive field map that describes the cells response to small bright and dark spots presented at different locations in space. These maps are well described by a Gabor function. The response of a simple cell can be predicted by convoluting the receptive field profile with the visual pattern (Cumming & DeAngelis, 2001). The firing pattern of the binocular cell is then predicted by the sum of the convolutions in each eye. The key to understanding disparity selectivity lies in differences in the two receptive fields.

Ohzawa and Freeman in 1986 pointed out that the underlying mechanism leading to the detection of disparities could be either a difference in the receptive field position of a binocular cell in the two eyes, or by the receptive fields in the two eyes having different shapes (Ohzawa & Freeman, 1986b). This is termed a phase difference, since the mechanism by which stimulus disparity is detected is a difference in phase of the Gabor function between the two receptive fields, resulting in a shift of the maximum of the receptive field sensitivity within the receptive field.

In 1999 Anzai *et al.* showed that indeed disparities are detected mainly by phase differences (Anzai *et al.*, 1999a; Anzai *et al.*, 1999b; Anzai *et al.*, 1999c).

Complex Cells

Complex cells respond to oriented contours over a range of positions. With monocular sinusoidal gratings, complex cells are selective for the spatial frequency, but are insensitive to the spatial phase. Interestingly, when such gratings are presented binocularly, the cells are sensitive to the phase difference between the eyes, even though they are insensitive to absolute phase in monocular presentation (Ohzawa & Freeman, 1986a).

The disparity energy model, put forward by Ohzawa *et al.* in 1990 (Ohzawa *et al.*, 1990), explains this by proposing that a complex cell is constructed from several simple cells. All of the constituent simple cells have the same disparity tuning, but their receptive field shapes are distributed so that one of them will fire wherever a stimulus falls anywhere in the complex cell's receptive field. More specifically, their receptive field profiles will be arranged so that the whole array is indifferent to monocular phase. Since at least one of them will fire when a stimulus at the preferred disparity is present, the complex cell will be selective for correct stimulus disparity and hence for phase differences (Ohzawa, 1998). This model explains the behaviour of V1 complex cells very well and can also predict the response to anticorrelated random-dot stereograms (see (Cumming & DeAngelis, 2001)).

However, recently Cumming published data showing that neurons in V1 of awake monkeys can show tuning for horizontal disparities, i.e. they show sensitivity for a greater range of horizontal than vertical disparities (Cumming, 2002). This cannot be explained by the disparity energy model in its current form.

Recent evidence reviewed by Cumming and De Angelis indicates that phase mechanisms as well as positional differences between receptive fields are used to compute disparity. It is unclear why both coding mechanisms are used (Cumming & DeAngelis, 2001), although Erwin and Miller in 1995 offered a model they termed the subregion correspondence model, based on an earlier developmental model of simple cell formation, that would predict a linear relation between simple cell phase and position differences (Erwin & Miller, 1999).

Disparity Tuning Classes

Neurons in V1 can be roughly divided into three distinct classes (see (Poggio, 1995)):

- *Tuned excitatory* neurons respond maximally to zero or small disparities and have a symmetrical response profile. These cells can be subdivided further into tuned zero cells, responding to zero disparities, tuned near cells, responding to crossed disparities, and tuned far cells, responding to uncrossed disparities.

- *Tuned inhibitory* neurons are similar, but have an inverted profile, showing maximal suppression for small disparities.

- *Near* and *far* neurons have asymmetrical response profiles, responding only to crossed and uncrossed disparities respectively. These cells are usually described as having an extended rather than tuned profile, but Cumming and De Angelis argue that this might be due to the stimuli used to measure the profiles, and that they find the profiles of these cells to be well described by Gabor functions (Cumming & DeAngelis, 2001).

4.5.3 Higher Cortical Areas

Much less is known about the processing of stereoscopic information outside of V1. Disparity information has been found in many visual areas, including V2, V3, V3A, V4, V5/MT and MST, as reviewed by Gonzalez and Perez in 1998 (Gonzalez & Perez, 1998).

An fMRI study done in 2001 by Backus *et al.* found that area V3A was more sensitive to stereoscopic stimuli than other visual areas, indicating a special role for this area in the processing of stereoscopic stimuli (Backus *et al.*, 2001). Since activity in V3 in their study also correlated with subjects' conscious responses to depth, this area might play a role in the perception of stereoscopic depth. Iwami *et al.*, in another fMRI study in 2002, showed activation in the parieto-occipital cortices by dynamic and static random-dot stimuli (Iwami *et al.*, 2002).

Thomas *et al.* showed in 2002 that there is sensitivity for relative disparity in V2 of the macaque (Thomas *et al.*, 2002).

4.6 Development of Stereopsis

Adultlike binocular vision is not present in human infants for the first three or four month of life. This was found by Braddick *et al.* in 1980, using correlated

and anticorrelated stereograms and recording visually-evoked potentials from infants (Braddick *et al.*, 1980). Several more studies in the early eighties confirmed their results, with similar methods (Held, 1991).

In 1986, Granrud showed that this onset of disparity detection corresponds to a preference for the nearer of two objects in grasping, suggesting that the detection of disparity is accompanied by true depth discrimination (Granrud, 1986).

Birch *et al.* in 1982 found evidence for the fact that crossed disparity can be detected several weeks before uncrossed disparities (Birch *et al.*, 1982), and Gwiazda *et al.* demonstrated that this preference for crossed disparities continues throughout childhood (Gwiazda *et al.*, 1989).

Birch *et al.* in 1983 showed that the reason for improved stereopsis was not improved vergence, but most likely some change in the central nervous system.

Chapter 5

The Eye in Motion

> Eyes are bold as lions, –roving, running, leaping here and there, far and near. They speak all languages. They wait for no introduction; they are no Englishmen.
>
> *Ralph Waldo Emerson*[1]

Human eye movements can be classified into seven main groups[2]:

- *Visual fixation*, holds the image of a stationary object on the fovea
- *Vestibular*, holds the image steady during head rotations
- *Optokinetic*, hold images steady during prolonged head rotations
- *Smooth pursuit*, holds the image of a moving target on the fovea
- *Nystagmus, quick phases*, resets the eye during prolonged rotation
- *Saccades*, bring images of objects of interest onto the fovea
- *Vergence*, moves the two eyes in opposite directions to place objects of interest on both foveae simultaneously.

The neural substrates of these different kinds of movements are quite different, as is the role the specific sort of eye movement plays in establishing undisturbed visual perception of the environment.

[1] In *Conduct of Life: Behavior*

[2] Following (Zee & Leigh, 1983)

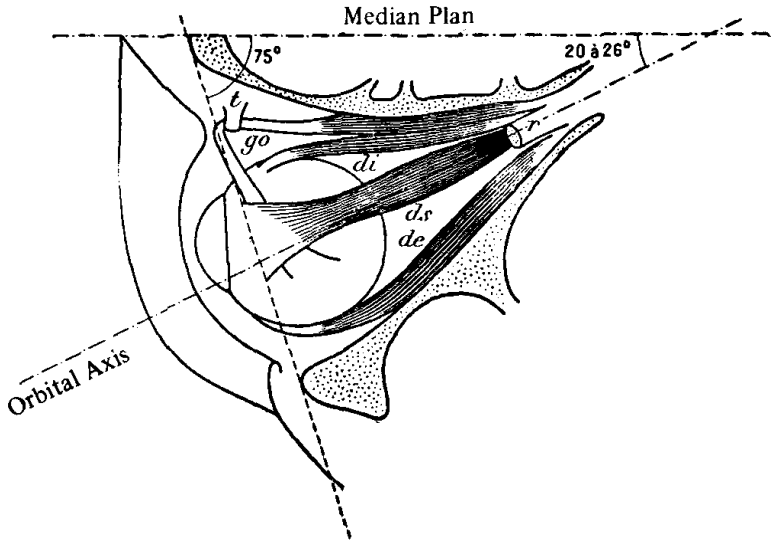


Figure 5.1: Anatomy of the eye and orbit. Visible are four of the six extraocular muscles of the left eye in a top view. They are: *superior oblique* (go), *superior rectus* (ds), *medial rectus* (di), *lateral rectus* (de). Not visible are the *inferior rectus* and the *inferior oblique*, both being obscured by the eye. Partly visible is the *levator* (r), responsible for lifting the eyelid. From (LeGrand & ElHage, 1980)

For this study, however, fixation and vergence are the most important systems, driving the eyes to binocularly fixate a target in depth.

Any binocular gaze position has six degrees of freedom, three angles of rotation for each of the eyes. When gaze is fixed onto a binocular target, four of these angles are no longer free. The two remaining degrees of freedom describe the two eyes' rotation around their lines of sight. Their connection to gaze direction during fixation is the subject of this chapter.

5.1 Anatomy of the Eye and the Extraocular Muscles

The human eye is located in a cavity in the skull called the orbit. It is moved by three pairs of extraocular muscles, the anatomy of which is depicted in Figure 5.1.

Muscle	Primary Action	Secondary Action	Tertiary Action
Medial rectus	Adduction	-	-
Lateral rectus	Abduction	-	-
Superior rectus	Elevation	Intorsion	Adduction
Inferior rectus	Depression	Extorsion	Adduction
Superior oblique	Intorsion	Depression	Abduction
Inferior oblique	Extorsion	Elevation	Abduction

Table 5.1: The primary actions of the six extraocular eye muscles in primary position. From (Leight & Zee, 1991).

The direction of rotation for each of these three muscle pairs depends on current gaze position, making the complete description of their action quite a complicated matter. The principal directions of each of the six muscles pulling on the eyeball in primary position are summarized in Table 5.1.

Abduction in this table is an eye movement in the horizontal plane away from the nose, adduction a horizontal movement toward the nose. Elevation and depression describe vertical eye movements. Intorsion is a rotation of the upper pole of the eye toward the nose.

5.2 Approximating Life

It has been assumed in the calculations and simulations in this thesis that the eye moves about a single centre of rotation and that the optical nodal point of the retinal projection is at the position of this centre of rotation. This assumption is common in binocular research (Howard & Rogers, 1995), since it greatly simplifies the geometry and removes eye position dependencies in the retinal projections. The real geometry is given in some more detail here.

5.2.1 The Rotation Center of the Eye

Adolf Friedrich Volkmann (⋆1801, †1877) in 1836 wrote

> All eye movements are rotations around one point, which is simultaneously the intersection of the optical axis and the visual axis[3].

[3] (Volkmann, 1836), p. 35, quoted from (Wade, 1998)

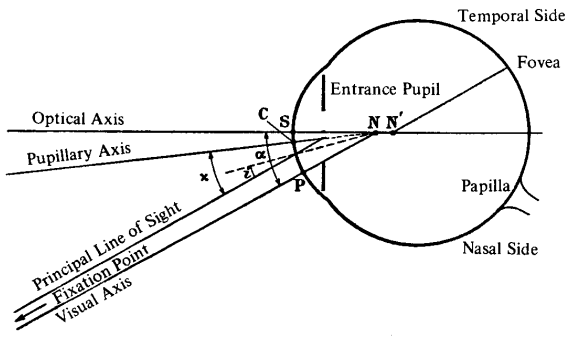


Figure 5.2: The optical geometry of the eye.

While this certainly is a quite good approximation of the true pattern of eye movements (and considered good enough for all of the work presented here), it is probably not the entire truth. Park (Park & Park, 1933) showed in 1933 that the assumed centre of rotation changes its position by as much as a millimetre during eye movements, and Fry and Hill demonstrated in 1962 (Fry & Hill, 1962), that a fixed center of rotation was incompatible with the pattern of eye movements they found.

5.2.2 The Optical Geometry of the Eye

With the fovea not being in the centre of the eyeball, but a few degrees temporal, it is already clear that the above simplified description cannot be true. Figure 5.2 shows the actual geometry more accurately. The optical axes of the three optical surfaces of the eye, the outer and inner surface of the cornea and the lens, are roughly co-linear, yielding a single optical axis, running approximately through the centre of the eyeball. The line going through the centre of the pupil and crosses the cornea at a right angle is called the pupillary axis and deviates from the optical axis.

The visual axis is the line projecting the nodal point onto the fovea. Finally, the principal line of sight connects the center of the pupil and the fixation point. The angles between all these axes in Figure 5.2 have been exaggerated to clarify the image.

5.3 Vergence Eye Movements

In the classic scheme for eye movements, going back to Karl Ewald Konstantin Hering (⋆1834, †1918) (Hering, 1868), eye movements are accomplished by two separate systems. One for versional movements, where vergence is unaltered, and a second system changing vergence. Hering writes

> ...we can think of both eyes as a single imaginary eye which lies midway between the two real eyes. If such an eye had to be innervated to turn to the left, right, above or below, the two real eyes would always be equally innervated, and if such an eye had to be innervated to accommodate for greater nearness or distance, both eyes would be innervated not only for an internal accommodation but also for an external bifixation of both lines of sight for nearness and distance[4].

This not only describes Hering's principle of equal innervation but also the concept of the cyclopean eye discussed before. His idea that the two eyes are guided not by separate input changing their gaze direction but by input changing version and vergence separately is known as Hering's law. Its validity is today in much doubt, but it is certain that there is a distinct system for changing vergence. One distinguishing characteristic of this system is that whereas the saccadic system has a rapid onset and high velocities, vergence eye movements are much smoother and take a longer time to complete.

Vergence eye movements can be grouped in three main classes: accommodative, tonic and disparity-driven. Of these, disparity-driven vergence is the only one of interest for this study, so I will just briefly touch upon the first two.

Tonic vergence refers to the vergence angle the eyes assume when the subject is kept in darkness. It averages about 3° of convergence and changes with gaze elevation, increasing in downward gaze and decreasing in upward gaze (Judge, 1991).

Accommodative vergence refers to the coupling of accommodation of the eyes and vergence angle. It was first demonstrated by Müller in 1826 (Müller, 1826) by occluding one eye while the other eye fixated a target. When a negative lens was suddenly placed before the seeing eye, the covered eye converged by an amount linearly related to the change in accommodation.

[4] Cited after (Judge, 1991)

There also is a converse influence, driving accommodation depending on the current state of binocular vergence. This mechanism is the main reason people have trouble seeing random-dot stereograms, where the vergence angle calls for an amount of accommodation that makes it impossible to focus the image.

5.3.1 Disparity-Driven Vergence

Binocular disparity can elicit vergence movements, horizontally and vertically as well as torsionally. Since ocular torsion is dealt with in a separate section, this section will only be concerned with horizontal and vertical disparity-driven vergence.

Disparity-driven vergence was first described by Rashbass and Westheimer in 1961 (Rashbass & Westheimer, 1961). They used a Wheatstone stereoscope displaying targets with disparities. Targets whose disparity indicated that the object depicted was closer than fixation elicited a convergence response, targets which appeared to be farther away elicited divergence. This indicated that there was negative feedback loop in place, trying to minimise disparity and thus maintain eye alignment.

Erkelens and Collewijn in 1991 (Erkelens & Collewijn, 1991) elicited vergence movements using a small random-dot stimulus (of about 1° subtense), and showed that the horizontal disparity of the selected target is used to guide vergence movements. They found a latency for the induced vergence movements of about 200 ms. Busettini *et al.* in 2001 used larger random-dot stimuli, subtending 85°, and found latencies of the vergence reaction of about 80 ms (Busettini *et al.*, 2001). They also found that for disparities of about 7° and more, the elicited vergence response became generic and did not depend on the direction of the disparity any more. From this they concluded that the vergence response was driven by a local matching system and that for large disparities this system detected false matches from a residual correlation (Busettini *et al.*, 2001).

Vergence movements induced by vertical disparity have similar characteristics to horizontal vergence movements (Howard *et al.*, 1997); they however are slightly slower and their gain is smaller (Busettini *et al.*, 2001).

Both vertical and horizontal vergence gains depend on the size of the stimulus area as well as the spatial frequency content of the stimulus (Howard *et al.*, 2000).

In 1997 Masson *et al.* reported that disparity steps applied to anticorrelated random-dot stereograms also elicited vergence eye movements. Since Cumming and Parker have shown in 1997 (Cumming & Parker, 1997) that neurons in V1 are sensitive to anticorrelated stereograms as well, the conclusion of Masson *et al.* was that disparity-driven vergence uses early stages of disparity detection, before depth percepts are elaborated.

5.4 Ocular Torsion

Ocular torsion in general is the rotation of the eyeball around the line of sight. This rotation does not change gaze direction. As I have pointed out in Section 3.1.4, the choice of coordinate system can introduce (or remove) eye torsion from the eye's position. This has been called pseudotorsion by Donders (Simonsz & den Tonkelaar, 1990).

Donders' law states that for any gaze direction the amount of ocular torsion is always the same. Listing's law specifies the amount of torsion and says that any secondary eye position can be reached by a rotation from the primary position through a rotation axis perpendicular to the gaze lines in both the primary and the target positions. This means that the rotation axes all lie in a common plane, namely the plane perpendicular to the primary gaze direction (see Figure 5.3).

The formula for Listing's law in Helmholtz coordinates is

$$\psi = -\frac{\theta\phi}{2} \tag{5.1}$$

For a long time it was widely believed that the eyes always followed Listing's law, making ocular torsion a parameter of very limited interest. But in fact, neither does the direction for primary gaze stay the same under all conditions, nor do the eyes always follow Listing's law.

5.4.1 Deviations from Listing's Law

During saccadic eye movements the eyes systematically deviate from positions ascribed by Listing's law. Tweed and Vilis in 1987 showed, in a three dimension analysis of the mathematics involved, that the noncommutative nature of rotations required the system to frequently use rotation axes outside of Listing's plane in order to follow Listing's law during fixation (Tweed &

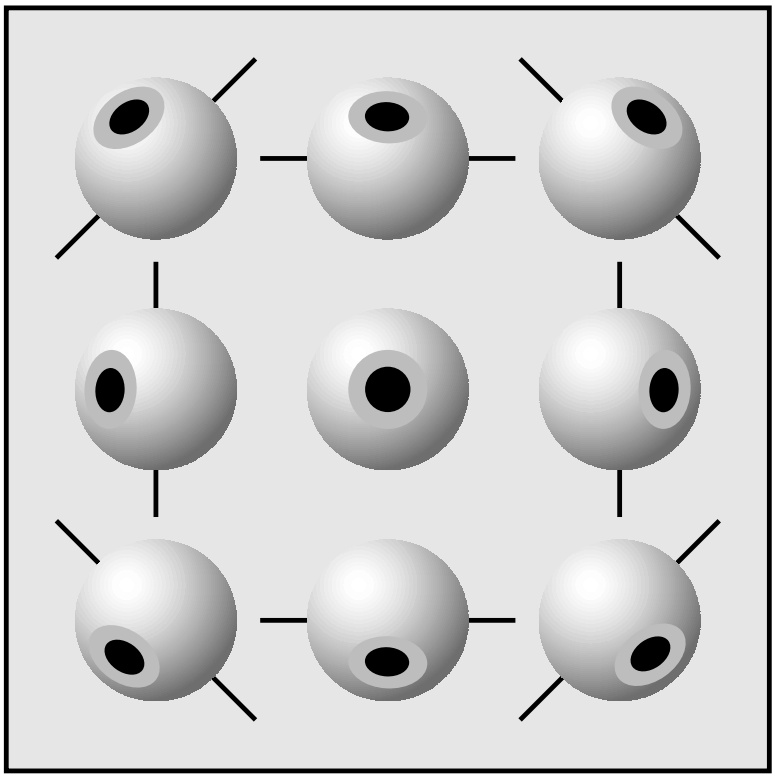


Figure 5.3: Listing's plane (shaded square) is perpendicular to primary gaze direction (represented by the central eye staring out of the paper at you. Listing's plane contains all the rotation axis for secondary gaze positions.

Vilis, 1987). Tweed *et al.* in 1998 showed that Listing's law in humans in fact was violated when the head is moved, to optimize retinal image stabilization (Tweed *et al.*, 1998).

5.4.2 Visually Evoked Cyclovergence

Another deviation from Listing's law is visually induced cyclovergence, where a torsional disparity in the two retinal images elicits a torsional vergence movement. This was first postulated by Nagel in 1868 (Nagel, 1868) and later picked up by Helmholtz and Hering[5].

Kertesz and Jones were the first to carry out objective measurements of this cyclofusional response (Kertesz & Jones, 1970). They found, using stimuli subtending 10°, that subjects were able to fuse orientation disparities of several degrees without any apparent cyclovergence movements. Kertesz also found that the amount of cyclofusion depended upon stimulus complexity, and that Panum's fusional areas therefore should not be considered constant, but that they depended on stimulus parameters (Kertesz, 1972). But the Kertesz group also stated that practically all of the observed tolerance to cyclodisparity was of a sensory nature. Their subjects showed almost no cyclovergence. Kertesz and Jones concluded that motor cyclofusion did not exist.

In 1971 R. A. Crone reported in a letter to the editors of Vision Research that he had indication of a cyclovergence response to cyclodisparity, and in 1975, more than 100 years after the existence of cyclovergence had first been postulated by Nagel, Crone and Everhard-Halm were able to show its existence (Crone & Everhard-Halm, 1975). They used stimulus displays subtending 25° and found that stimulus size as well as stimulus complexity affected the strength of the motor response, with stimulus size having the greatest effect. Their findings were confirmed by Hooten *et al.* in 1979, who additionally found that cyclovergence conformed to Hering's law of equal innervation (Hooten *et al.*, 1979). This finding, that the torsion angles of the two eyes are equal and of opposite direction, even when the stimulus itself is asymmetrical, i.e. the image in one eye is rotated more than the image in the other, had also been shown by Sullivan and Kertesz (Sullivan & Kertesz, 1978).

Howard and Zacher in 1992 investigated the gain of visually evoked cy-

[5]See review by Ogle (Ogle, 1950), p. 101f and (Ogle & Ellerbrock, 1946).

clovergence using stimuli spanning a range of temporal frequencies and amplitudes, and found it to be greatest for small temporal frequencies and amplitudes. The gain depended on the amplitude of the cyclodisparity, indicating the system to be nonlinear. Another peculiarity of visually evoked cyclovergence was reported by Howard *et al.* in 1994, when they showed that cyclovergence was evoked by stimuli in the far periphery just as well as by large central stimuli (Howard *et al.*, 1994).

5.4.3 The Binocular Extension of Listing's Law

When humans converge, the primary gaze directions of the two eyes move laterally, rotating Listing's planes temporally for each eye. While several studies have documented this tilt (Allen, 1954), the angle of the plane's rotation they found differed. Minken and van Gisbergen in 1994 found $\frac{1}{4}°$ of rotation of Listing's plane for each degree of vergence (Minken & van Gisbergen, 1994), Mok *et al.* in 1992 found 0.16° (Mok *et al.*, 1992) and van Rijn and van den Berg in 1993 reported 0.5° (van Rijn & van den Berg, 1993). The motor behaviour of rotating Listing's planes with convergence has been called the *binocular extension of Listing's law* or shorter, L2. The ratio between the vergence angle and the rotation of Listing's plane is known as μ; it is sometimes called the gain of L2.

The formula for L2 for each eye in Helmholtz coordinates is (Tweed, 1997b):

$$\phi_L = -\mu V\phi - \frac{\theta_L\phi}{2} \tag{5.2}$$

$$\phi_R = \mu V\phi - \frac{\theta_R\phi}{2} \tag{5.3}$$

Note that this reduces to Listing's law when the vergence angle is $V = 0$, as it should.

We can also express this law in horizontal and vertical Helmholtz angles for the two eyes instead of the vergence angle. Since vergence is the difference between the right and left horizontal angle

$$V = \theta_R - \theta_L \tag{5.4}$$

We find

$$\psi_L = -\mu(\theta_R - \theta_L)\phi - \frac{\theta_L\phi}{2} \quad (5.5)$$

$$= \left(-\mu\theta_R - \left(\frac{1}{2} - \mu\right)\theta_L\right)\phi \quad (5.6)$$

$$\psi_R = \mu(\theta_R - \theta_L)\phi - \frac{\theta_R\phi}{2} \quad (5.7)$$

$$= \left(-\mu\theta_L - \left(\frac{1}{2} - \mu\right)\theta_R\right)\phi \quad (5.8)$$

From L2 it follows that each eye's torsion should depend on the gaze direction of the other eye. In 1999 Porrill *et al.* measured binocular torsion in a situation where the target's position was changed along the line of sight of the right eye (Porrill *et al.*, 1999b). They reported that while cycloversion varied greatly between subjects and between convergence and divergence, cyclovergence followed L2.

Mok *et al.* proposed a tentative theory as to what the advantages of L2 were, arguing that it served to align the planes of angular velocity vectors. However, in 1997 Tweed published an analysis of L2, showing that while the planes of the velocity vectors were aligned, the eye's velocity vectors themselves were not, making it hard to find an advantage of the plane's alignment (Tweed, 1997a). He argued that the most likely reason for L2 was the optimization of motor control and binocular alignment. van Rijn and van den Berg had shown in their 1993 study that L2 with $\mu = \frac{1}{4}$ had the advantage of setting Helmholtz torsion equal in the two eyes:

$$\psi_L = \left(-\frac{1}{4}\theta_R - \left(\frac{1}{2} - \frac{1}{4}\right)\theta_L\right)\phi \quad (5.9)$$

$$= -\frac{(\theta_R + \theta_L)\phi}{4} \quad (5.10)$$

$$= -\frac{\theta\phi}{2} \quad (5.11)$$

$$\psi_R = \left(-\frac{1}{4}\theta_L - \left(\frac{1}{2} - \frac{1}{4}\right)\theta_R\right)\phi \quad (5.12)$$

$$= -\frac{(\theta_L + \theta_R)\phi}{4} \quad (5.13)$$

$$= -\frac{\theta\phi}{2} \quad (5.14)$$

With θ being the conjugate horizontal angle: $\theta = \frac{\theta_L + \theta_R}{2}$.

In 1999 Kapoula *et al.* presented a study where they systematically changed the stimulus cue driving vergence, using disparity-driven vergence, accommodative vergence and a combination of disparity-driven vergence with the presentation of a stereogram containing a depth stimulus. They found that the gain was greatest for the combined stimulus, doubling the gain relative to simple disparity-driven vergence (Kapoula *et al.*, 1999). Vergence driven by accommodation showed idiosyncratic behaviour and didn't allow the computation of a gain across their subjects. From this they concluded that L2's main purpose was to serve binocular depth vision. The same conclusion was reached by Schor *et al.* in 2001. They showed that the coupling of gaze direction to cyclovergence is plastic, suggesting that the actual pattern of L2 might be calibrated to serve binocular vision (Schor *et al.*, 2001).

In a study published in 2002, Steffen *et al.* adapted subjects to vertical disparities created by prisms and found that the vertical vergence this induced led to a change in the orientation of Listing's planes, tilting them vertically (Steffen *et al.*, 2002).

The relation between visually evoked cyclovergence and L2 has been investigated by Hooge and van den Berg in 2000 (Hooge & van den Berg, 2000). They found that cyclotorsional angles from visual stimuli and L2 add approximately linearly and are independent of each other. Since, however, Schor *et al.* also used cyclorotated visual stimuli to elicit their adaptation of L2, this is probably only true when the cyclorotated stimuli are not presented for an extended period of time.

5.5 Pulleys and Noncommutativity

In 2002 Joseph Demer published a paper titled "A revolution in concepts of orbital anatomy"[6] where he presents arguments and evidence for the existence of the orbital pulley system and proposes the *active pulley hypothesis*. Pulleys and models using them have been described mainly by Demer himself, C. Quaia, T. Raphan and D. Tweed (Quaia & Optican, 1998) (Raphan, 1998) (Demer, 2002) (Demer *et al.*, 2000) (Tweed *et al.*, 1999). They are connective tissue structures in the orbit that constrain the extra ocular muscle

[6] In the article he calls this revolution a paradigm shift, which allows him to cite Kuhn (Kuhn, 1996). It is always a good idea to cite philosophers, even though the implicit claims made by these references are hard to falsify (Popper, 1935).

paths, first described by Demer *et al.* in 1995 (Demer *et al.*, 1995).

In 1987 Tweed and Vilis had presented their quaternion model for three dimensional saccade generation, which they based on the fact that three dimensional rotations are noncommutative. They pointed out that the steering of a noncommutative plant required noncommutative neural control mechanisms. Their model thus contained noncommutative rotational operators (Tweed & Vilis, 1987). In 1994 Schnabolk and Raphan proposed that since innervation signals determine muscle torque and not eye position, a noncommutative operator in the brain was unnecessary. Tweed *et al.* replied in 1994, pointing out that the Schnabolk-Raphan model did not fit experimental evidence, while the Tweed-Vilis quaternion model did, and concluded that this was because Schnabolk and Raphan had tried to model noncommutative kinematics with a commutative model. Here Tweed et al. presented a 'linear-plant' model of the eye muscles which is equivalent to later 'pulley' models.

In 1997 Raphan published a model that incorporated extraocular muscle pulleys (Raphan, 1997), and Quaia and Optican in 1998 followed this with a theoretical analysis (Quaia & Optican, 1998). They argued that the position dependency introduced by the pulleys allowed the neural system generating the saccadic neural pulse to be commutative, even when the generated movements are not.

Tweed *et al.* in 1999 showed experimentally that the human vestibulo-ocular reflex drives the eyes in a way "unattainable by any commutative system" (Tweed *et al.*, 1999) and that the system driving the eyes therefore must contain noncommutative operators, independent of whether or not it contains extraocular muscle pulleys.

The *active pulley hypothesis* claims that muscle pulleys are responsible for taking the two dimensional output of the brain circuits creating eye movements and creating the complex three dimensional behaviour of Listing's law, the deviations from it and, during vergence, L2. The pulleys do this by actively changing their positions for different kinds of eye movements. It is claimed that:

> While the ocular motor system cannot ever be entirely commutative, a fortunate arrangement of the [extraocular muscles] and orbital connective tissue makes it appear, for all practical purposes, commutative to the brain[7].

[7] From (Demer, 2002).

But this is incorrect, because the combined brain system that contracts the muscles and moves the pulleys must still be non-commutative. As Tweed et al. observed in 1999, the non-commutativity of the human vestibulo-ocular reflex requires non-commutative neural processing even if the pulleys are actively controlled

Steffen *et al.* in 2000 provided more evidence for central innervation being responsible for Listing's law, when they demonstrated that adaptive changes in L2 can be independent of eye position (Steffen *et al.*, 2000).[8]

5.6 Matching with Moving Eyes

As I have shown in Section 3.4, a change in eye position changes the retinal location of the epipolar lines. If stereo matching uses the epipolar constraint, it can only do so by monitoring eye position and adjusting the location of the retinal search lines, i.e. the narrow bands of retina used for matching.

Alternatively, the matching system would have to search larger retinal areas for matches. These areas would have to be large enough to cover all the retinal regions the epipolar lines can slide through. In the next chapter I will describe an experiment designed to find out which solution the visual system uses.

[8]The ongoing pulley-debate is very important to understanding how the brain achieves Listing's law and other, more complex movement patterns of the eyes. Since this dissertation, however, is only concerned with the effects of the movement patterns themselves on perception, the question of how the patterns are engineered in the oculomotor plant and the motor nuclei has no direct relevance to the studies presented here.

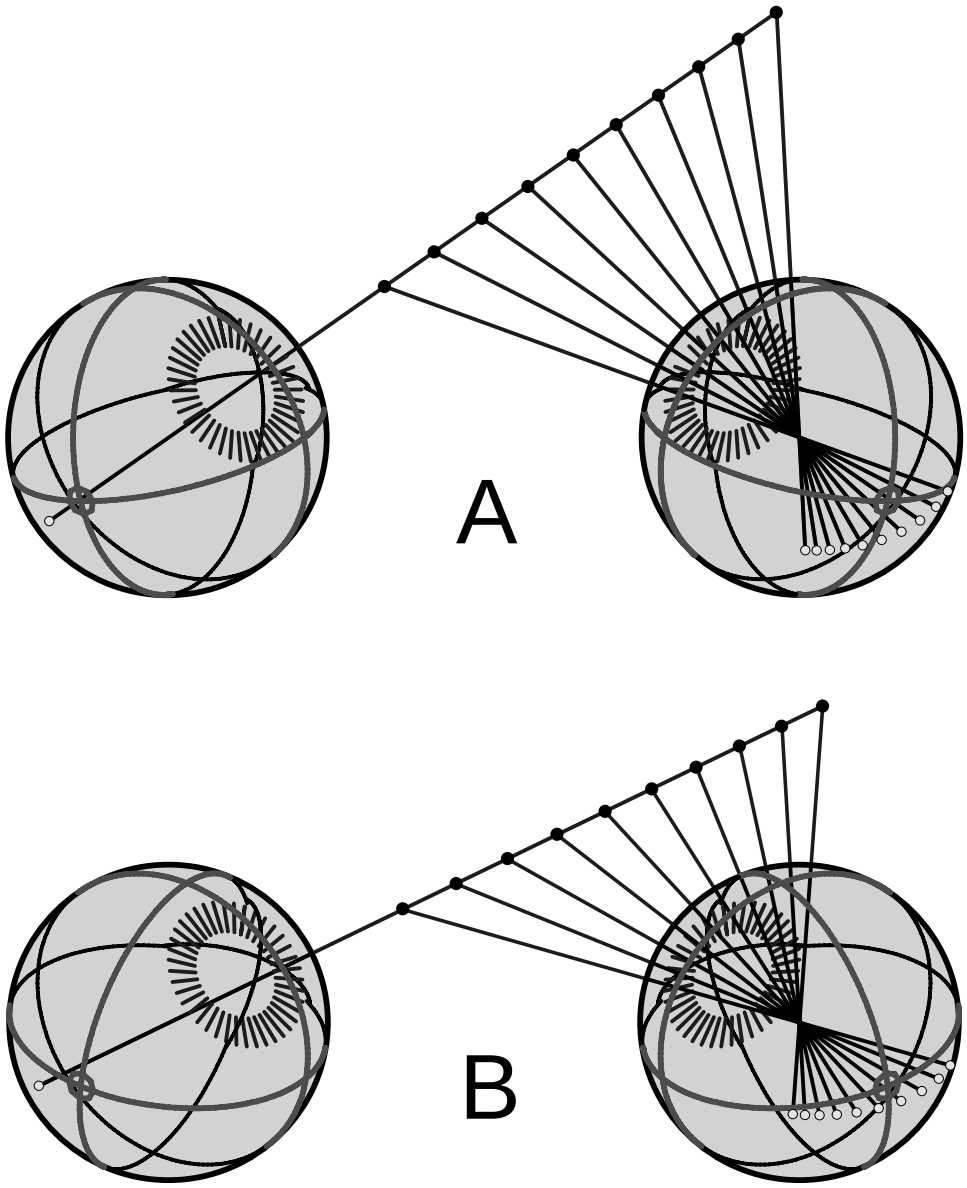


Figure 5.4: A way of visualizing the moving epipolar lines. A retinal image in the left eye has to lie along the depicted line in space, projecting somewhere along the epipolar line in the other eye. The eyes have been rotated around their lines of sight from A to B, both shifting and rotating the epipolar line on the right eye's retina.

Chapter 6

Positional Stereoblindness

In common things that round us lie
Some random truths he can impart,–
The harvest of a quiet eye
That broods and sleeps on his own heart.
William Wordsworth[1]

In my first experimental study I investigated the solution the stereoptic system has found for the problem of sliding epipolar lines. If the matching system uses the epipolar constraint, any change in the location of the epipolar lines on the retina must be accounted for by the matching system. Stereo-matching then does not depend on eye position.

If, on the other hand, the migration of the epipolar lines across the retina is not corrected for, the perceptibility of stereograms should change with eye position. It should then be possible to create random-dot stereograms that are better visible in certain eye positions. In an extreme case there could be stereograms that are visible only in certain eye positions and become invisible in others. By constructing such stereograms and demonstrating this positional stereoblindness, I show that the stereo matching system does not correct for the shift of the epipolar lines.

The question addressed in this chapter is:

Does stereo-matching use eye-position information to adjust the retinal search regions to the location of the epipolar lines?

[1] In *A Poet's Epitaph*. Unfortunately, experts have very little to say on the question whether Wordsworth was referring to positional stereoblindness when he said "a quiet eye" in this poem. It does seem a little unlikely.

6.1 Cyclorotated Stimuli

In keeping with the binocular extension of Listing's law, the eyes rotate around their lines of sight when gaze elevation changes in vergence. We can visually mimic this by rotating the two halves of a random-dot stereogram in opposite directions, creating cyclodisparity in the stereogram.

Subjects can tolerate several degrees of this cyclodisparity, but eventually stereo-matching fails and subjects lose fusion. By encoding a stereoscopic shape in the stereogram and asking the subjects to report which shape they perceived, I mapped out the angles for ex- and incyclorotation of the stimuli where the probability of the subject perceiving the stereoscopic shape dropped to 50%.

I then shifted the whole display arrangement to elevated or depressed angles (see Figure 6.2). Since with L2 the eyes excyclorotate on upgaze, subjects should be able to see excyclorotated stereograms better when they look up. Conversely, the eyes incyclorotate on downgaze, making incyclorotated stereograms easier to perceive.

6.2 Experimental Setup

The subject's heads were stabilized using a bitebar. The stereograms were displayed on a laptop computer screen. The laptop was supported by a cardboard apparatus that allowed it to be placed in three different positions, requiring the subjects to look straight ahead, 30° down and 30° up, respectively, in order to fixate the fixation cross and the stereograms displayed on the screen. The screen was 35 cm away from the subject's eyes in all three positions. The distance between the two halves of the stereogram was set so that fixating them required 30° of vergence.

Stereograms comprised 2000 white dots in each eye's image on a black screen and spanned 12° of visual angle. Each dot was a square subtending about 5 minutes of arc[2]. Each stereogram contained one of four different disparity-defined shapes—a pointer, like a clock hand, pointing in each of four oblique directions. The shapes were presented in random order.

The stereograms were flashed for 200 ms on the screen, so that subjects would not have enough time to cycloverge their eyes.

[2]These values were chosen to make the stereograms easy to fuse for our subjects even when they were flashed.

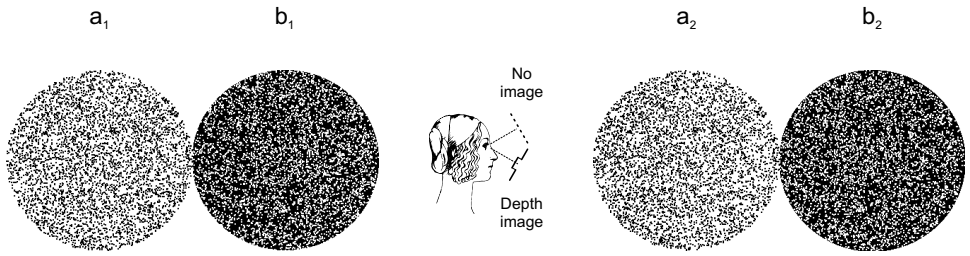


Figure 6.1: The experimental stimulus. Hold the page at a distance of about 30 centimeters below your face, then cross your eyes and fuse the two discs labeled a_1 and a_2. When you move the page upward, you should lose perception of the encoded shape. If you see the shape in all viewing directions, your retinal search zones are too large for this stimulus. Try fusing b_1 and b_1 instead.

For each trial, subjects then chose the direction in which they thought the clock hand was pointing. An interactive program varied the cyclorotation angles of the stereograms depending on the correctness of previous answers, mapping out the threshold ranges. Because subjects viewed the stereograms cross-eyed, their lines of sight hit the flat computer screen at a 15° slant. To eliminate the optical effects of this slant, I warped the random-dot patterns on the screen in a precise way so that each retinal projection was of a disc seen face-on, without slant-induced disparities, cyclorotated about the line of sight. But as slants are present in real life when one views near objects, I also performed the experiments without the de-slanting procedure. In these with-slant experiments, the stereograms were presented for 8 s rather than 200 ms, to show that the positional stereoblindness also occurs with prolonged viewing, and I used 20 disparity-defined shapes, such as discs, squares, triangles, instead of a clock hand, to show that the effect is not specific to judgements of orientation. The order of shapes presented again was randomized, with subjects indicating which shape they had perceived in the last stimulus presented. If they were uncertain, subjects were required to guess.

The cyclorotation angle for each trial was selected dynamically during the experiment using the $|psi$-method described in Section 3.6. The Ψ-method also dynamically decided upon the number of trials presented and ended the experiment when a predefined accuracy of the estimate of the perception thresholds was reached (standard deviation of current estimate below 0.3 degrees).

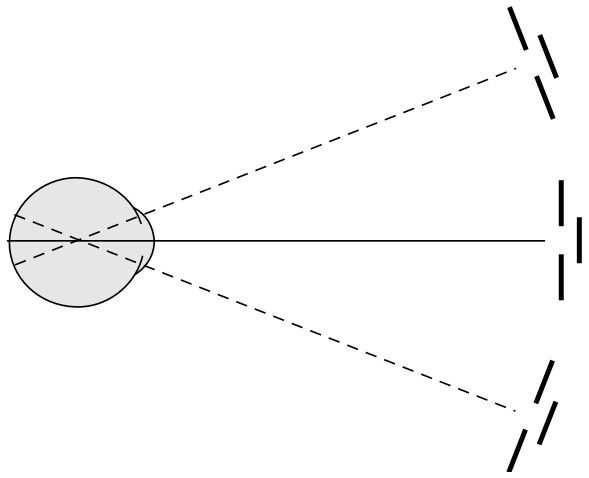


Figure 6.2: Display of the stereograms at three different gaze elevations.

6.3 Cyclovergence Measurements

I measured the cyclovergence of my subjects using the scleral search coil technique (Robinson, 1963), in which subjects sit within magnetic fields wearing a ring-shaped contact lens. Wire coils embedded in the lens register the orientation of the eye because they develop voltages proportional to their angles with respect to the fields. But the results with this method were imprecise owing to coil slip, signal drift and cross-talk. Much better accuracy was obtained by the nonius method.

6.3.1 The Nonius Method

Subjects were presented with a stereogram consisting of a binocular fixation circle and two monocular narrow lines (called nonius-lines[3]) in oblique positions above and below the circle. When the subjects fixated the binocular circles, they saw the monocular lines on either side of the circle. One of these lines was fixed at an oblique angle. The subjects task was to rotate the second line until both lines appeared parallel. The angle between the two

[3]The word nonius comes from the latinized form of the name of Pedro Nunes (1502-1577), a Portuguese mathematician. It was first used for the graduations of an instrument described by Nunes is his work *De Crepusculis liber unus* (Nunes, 1542), and from there has aquired the meaning of *very fine line* it has attained in the nonius-method described here.

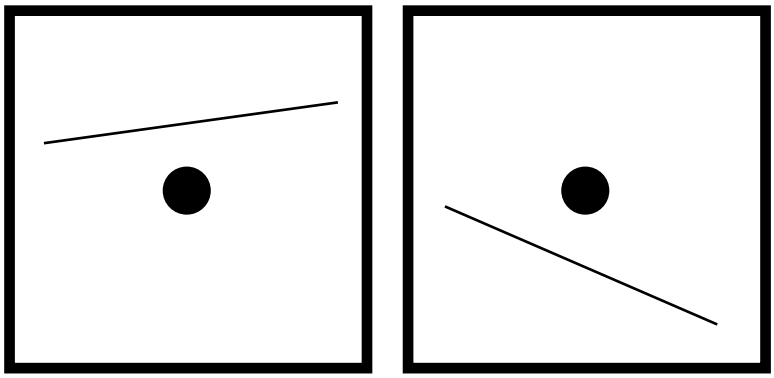


Figure 6.3: The nonius stereograms. Subjects had to adjust the rotation angle of one of the two lines so that the two lines appeared parallel. While doing this they kept the central circle fused.

nonius lines on the screen then reflected the relative orientation of the two eyes. This measurement was repeated twice before and twice after the block of trials and once in the middle of the block.

This method was first used systematically by Hofmann and Bielschowsky in 1900 (Hofmann & Bielschowsky, 1900). Crone and Everhard-Halm in 1975 arrived at the conclusion that the nonius method for static stimuli agreed well with objective measurements (Crone & Everhard-Halm, 1975). Howard *et al.* in 1993 found that the nonius technique overestimated cyclovergence by about a degree, but they placed the nonius display within a cyclodisparate surround, and attributed the inaccuracy to interference from the surround. In my study the nonius lines were presented without any other stimulus, making a systematic measurement error unlikely. Also, my search coil data, in cases where slip and other problems were minimal, agreed with the nonius measurements, further increasing the confidence in these data.

6.4 Results

Figure 6.4 shows the results from a typical subject. This subject clearly shows the predicted shift in the perceptible range of stereogram cyclorotation angles and therefore the predicted positional stereoblindness: for this subject, a stereogram where the two halves had been excyclorotated by 10° was almost always visible when it was viewed in upgaze, but was not visible at all in

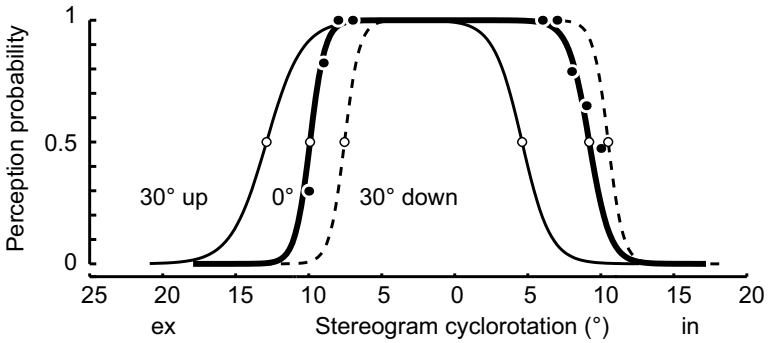


Figure 6.4: Stereopsis depends on gaze elevation. For this typical subject, black dots show the fraction of images that were correctly reported when the subject converged 30° and looked level. Fitted curves plot performance at three gaze elevations. Open circles mark perception thresholds, i.e. angles at which stereograms were reported correctly 62.5% of the time; this is the performance expected when the subject perceives the image 50% of the time and on the remaining 50% performs at chance level, guessing correctly a quarter of the time.

downgaze.

Figure 6.5 shows the results for all my subjects. In the figure cyclorotation thresholds are plotted against the cyclovergence position of the two eyes. All subjects show the predicted shift in thresholds with eye position.

If the retinal search zones are retina-fixed, the shift in thresholds should correspond exactly to the change in eye position. Namely, when the eyes excyclorotate by 1°, the perception range should shift by one degree toward excyclorotation.

The dashed line in Figure 6.5 has a slope of one. As is visible in the figure, the subject's data cluster closely around this line. A regression line fitted to the threshold average against cyclovergence has a slope of 1.06.

From this I conclude that retinal search regions indeed are eye fixed. No correction for a change in eye position is made. This means that the epipolar constraint cannot be used. Stereo matching is essentially two-dimensional.

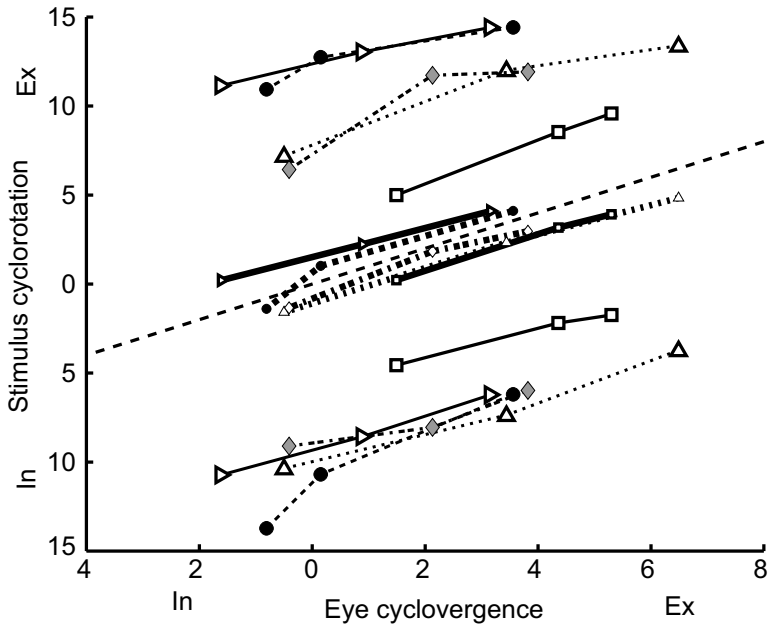


Figure 6.5: Result for all measured subjects. Plotted are the cyclorotation thresholds for ex- and incyclorotation (big symbols, thin lines) and the average of these two thresholds (small symbols, thick lines), against the cyclovergence position of the two eyes. Same symbols signify same subject.

6.5 Visually Evoked Cyclovergence

I also created dynamic, non-flashed versions of my stimuli, where stimulus cyclodisparity changed smoothly, growing in size to a maximum and then decreasing again. The position of random dots in the display changed between frames. The stereograms also contained a shape encoded in disparity, which changed at random intervals of 1-3 seconds. Subjects were asked to indicate when they saw the shape change.

I measured cyclovergence during these stimulus displays and found that subjects' cyclovergence followed the stimulus cyclovergence closely up to a maximum cyclovergence angle. While the stimulus cyclodisparity increased further, the eyes remained at this cyclovergence. Once the stimulus cyclodisparity had come down into the fusible range again, the eyes once more followed.

This seems to conflict with the findings of Kertesz and Jones (Kertesz & Jones, 1970), who found no cyclovergence response for stimuli subtending 10°. The fact that my subjects, with stimuli just 2° larger followed the stimulus cyclodisparity with a cyclovergence of several degrees might be due to the fact that my stimuli were more complex than Kertesz and Jones' line stimuli. Another possibility is that since my subjects not only had to fuse the images but also were asked to detect changes in the encoded shape, their attention boosted their visually evoked cyclovergence. A third possibility is that visually induced cyclovergence shows some variance between individuals, with my subjects being more responsive to the stimuli.

Chapter 7

Motor Control and Stereo-Matching

With the search zones for stereo matching being retina fixed, they cannot be covering just the epipolar line for the current gaze position but must be large, two-dimensional patches. They must be large enough to cover all the positions the epipolar lines will slide through during normal eye movement, in order to enable stereo matching for all eye positions.

This being the case, motor control acquires a crucial role: the sliding of the epipolar lines depends on the pattern of eye movement. To find the oculomotor pattern that has the least motion, I simulated many different patterns and computed the minimum size the search zones could have, given each pattern, and still allow stereopsis.

The questions addressed in this chapter are:

What are the shapes of the retinal search zones? Which motor-program would minimize their size? What are the optimization principles governing the movement of the eyes?

7.1 The Binocular Extension of Donders' Law

Donders' law specifies that for each gaze direction there is a specific value of ocular torsion. Put mathematically: for each eye there exists a function $\psi = \psi(\theta, \phi)$ describing eye torsion ψ based on horizontal and vertical Helmholtz angles θ and ϕ.

We have seen, however, that human oculomotor control is in fact more

flexible than this. The torsion of one eye depends on the gaze directions of *both* eyes. In other words: for each eye there exists a function $\psi = \psi(\theta_L, \phi_L, \theta_R, \phi_R)$.

This can be simplified a bit by restricting the two eyes to equal vertical Helmholtz angles. In addition, as we have seen before, we can describe binocular horizontal eye positions using either the two horizontal Helmholtz angles, or their average and difference, horizontal version and vergence. So the function for the binocular extension of Donders' law is $\psi = \psi(\alpha, \beta, \phi)$, with α being horizontal version and β being horizontal vergence.

For the purpose of these simulations I selected a subset of all possible such functions, namely the functions depending on polynomials of up to second order of these three angles:

$$\begin{aligned} \psi_{\vec{l}}(\alpha, \beta, \phi) &= l_1 + l_2\alpha + l_3\beta + l_4\alpha\beta + l_5\alpha^2 + \\ &\quad l_6\beta^2 + l_7\phi + l_8\beta\phi + l_9\alpha\phi + l_{10}\phi^2 \qquad (7.1) \\ \psi_{\vec{r}}(\alpha, \beta, \phi) &= r_1 + r_2\alpha + r_3\beta + r_4\alpha\beta + r_5\alpha^2 + \\ &\quad r_6\beta^2 + r_7\phi + r_8\beta\phi + r_9\alpha\phi + r_{10}\phi^2 \qquad (7.2) \end{aligned}$$

There is no linear term for ϕ, since a change in vertical position does not alter the position of the epipolar lines or the geometry of the problem.

There are two such functions, since in principle it is possible that the two eyes have different kinematic behaviour. An analysis of the geometry of the matching problem, however, greatly restricts the freedom for the parameter vectors $\vec{l}$ and $\vec{r}$:

If the search zone size is minimal for any set of parameters $(\vec{l}, \vec{r})$, it must be minimal also when the whole system is reflected on the median plane. That is, the left and right eye swap positions and the horizontal version angle α and the torsion angle ψ change sign.

A change of sign in Ψ is equivalent to a change of sign of the parameter vectors, a change of sign in α changes the sign in the second, fourth and ninth component of the vectors (the ones linear in α), and the swap of left and right eye finally gives this transformation as:

$$\begin{pmatrix} l_1 \\ l_2 \\ l_3 \\ l_4 \\ l_5 \\ l_6 \\ l_7 \\ l_8 \\ l_9 \\ l_{10} \end{pmatrix} = \begin{pmatrix} -r_1 \\ r_2 \\ -r_3 \\ r_4 \\ -r_5 \\ -r_6 \\ -r_7 \\ -r_8 \\ r_9 \\ -r_{10} \end{pmatrix} \tag{7.3}$$

Since this transformation leaves the property of being an optimal motor program unchanged, we can use it to eliminate ten parameters, leaving us with only one set of ten.

7.2 Simulating the Zones

To obtain the retinal search zones needed for each motor program, I collected segments of the epipolar lines for a large number of eye positions and then computed the area of the retina covered by these segments.

This was done for 17 retinal locations, depicted in Figure 7.3 as small circles. For each retinal location, described by the retinal coordinates λ and μ, I computed the head centered coordinates for both eyes, as given in Section 3.1.6. Note that the horizontal Helmholtz angle for each eye is computed from version and vergence.

$$\vec{x}_L = \mathcal{M}_\phi \cdot \mathcal{M}_{\alpha-\beta/2} \cdot \mathcal{M}_\psi \cdot r_{eye} \cdot \begin{pmatrix} -\cos(\lambda)\cos(\mu) \\ \sin(\mu) \\ \sin(\lambda)\cos(\mu) \end{pmatrix} \tag{7.4}$$

$$\vec{x}_R = \mathcal{M}_\phi \cdot \mathcal{M}_{\alpha+\beta/2} \cdot \mathcal{M}_\psi \cdot r_{eye} \cdot \begin{pmatrix} -\cos(\lambda)\cos(\mu) \\ \sin(\mu) \\ \sin(\lambda)\cos(\mu) \end{pmatrix} \tag{7.5}$$

$$\tag{7.6}$$

Since the nodal points coincide with the center of the eyeball, which is also the origin of the head-fixed coordinate systems centered on each eye, these vectors also are the projection vectors for the retinal locations.

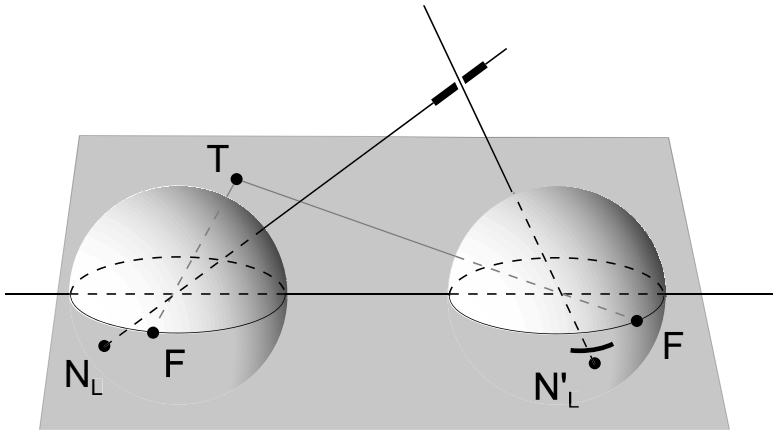


Figure 7.1: When both eyes fixate a visual target T, the projections from retinal locations with the same retinal coordinates N_L and N'_L outside of the plane of regard will not in general intersect in space. Using the Moore-Penrose-pseudoinverse we can find the point on the projection line from N_L where it most closely approaches the line from N'_L. A small segment (thick line) around this point is then projected onto the right retina. This is a portion of the epipolar line corresponding to N_L in this eye position.

An object projecting onto the left eye's retina in (λ, μ) therefore must be on the line in space defined by

$$\vec{x}_{ob} = k_L \cdot \vec{x}_L \tag{7.7}$$

This line projects onto the right eye's retina into the right eye's epipolar line, which in general will not run through the location (λ, μ) on the right eye's retina. To find the location on the right eye's epipolar line closest to (λ, μ), we use the Moore-Penrose pseudoinverse to solve

$$k_L \cdot \vec{x}_L = \vec{I} + k_R \cdot \vec{x}_R \tag{7.8}$$

where $\vec{I}$ is the interocular vector. $k_L \cdot \vec{x}_L$ then is the location in space O that projects to (λ, μ) on the left eye's retina and closest to (λ, μ) in the right eye's retina.

A small segment of the left eye's projection line centered on O is then projected into the right eye's retina. This is the segment of the right eye's retinal search line closest to the corresponding retinal location (see also Figure 7.1).

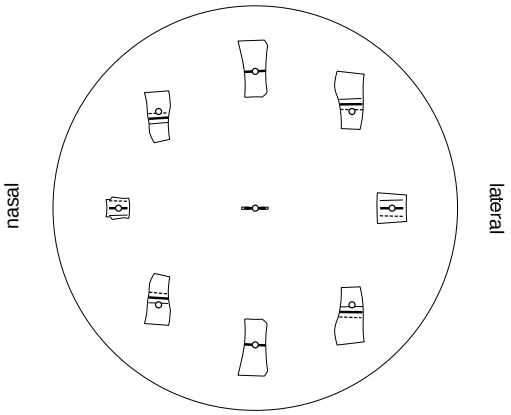


Figure 7.2: Three epipolar line segments, corresponding to three different binocular eye positions. The dots are retinal locations in the left eye. The thick lines are the epipolar line segments corresponding to a vergence angle of 30° with gaze straight ahead. Thin lines correspond to 30° vergence angle, gaze down, dashed lines to 30° vergence angle and gaze up. The nasal/lateral labels refer to the right eye, seen from behind.

This projection was done for a range of gaze directions ($\pm$ 40°horizontally and vertically in steps of 1°), collecting a large number of epipolar line segments for each retinal location (three such segments for each point are shown in Figure 7.2.).

For each point I then computed the area of the retinal patch (outlined areas in Figure 7.2) covered by all its epipolar segments. The area computation was simplified by projecting each patch onto the surface of a cylinder using a cartographic area-preserving projection (see Section 3.7). Areas for all 17 points were summed to yield a measure of the minimum retinal area that would have to be searched if no matches were to be missed. The same was repeated for many different patterns of eye motion, to see how this minimal search area depended the pattern of eye coordination, and I used gradient descent to find the patterns of eye coordination—from among the family of patterns encompassed by Equations 7.4 and 7.5 above—that minimized the search area.

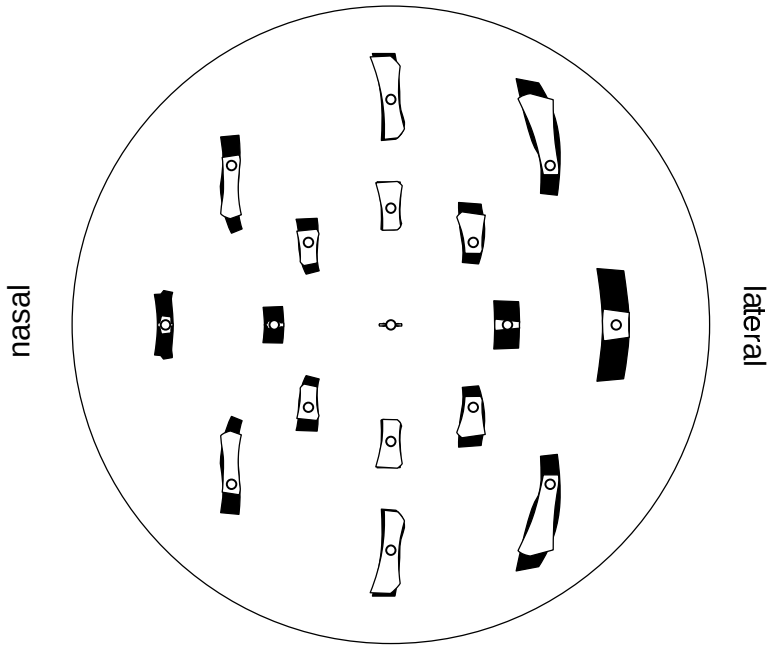


Figure 7.3: The retinal search regions for Listing's law (black) and for the motor program humans actually use (white). Dots are retinal positions in the left eye, the corresponding search regions are on the right eye's retina. The lateral/nasal labels refer to the right eye, seen from behind.

7.3 Results

Because of Equations 7.4 and 7.5 the space of possible motor patterns is ten-dimensional, making it impossible to visualize its full structure. However, the most interesting subspace is the one containing Listing's law, L2 as well as the optimal program. This is the plane spanned by the parameters for $\alpha\phi$ and $\beta\phi$. Figure 7.4 shows the total size of the retinal search regions for the 17 targets defined above for the motor programs within this plane. It should be noted that according to equation 5.3 (page 102) the value for x/l_8 in Figure 7.4 corresponds to L2's μ-value. Figure 7.5 presents a cut through this plane parallel to the l_8-axis.

The optimal motor program for binocular coordination is defined by $\psi = 0$, keeping Helmholtz torsion zero in both eyes for all gaze directions. This is shown by sphere 3 in Figure 7.4. As can be seen from the figures, the motor program used by humans does not minimize the size of the retinal

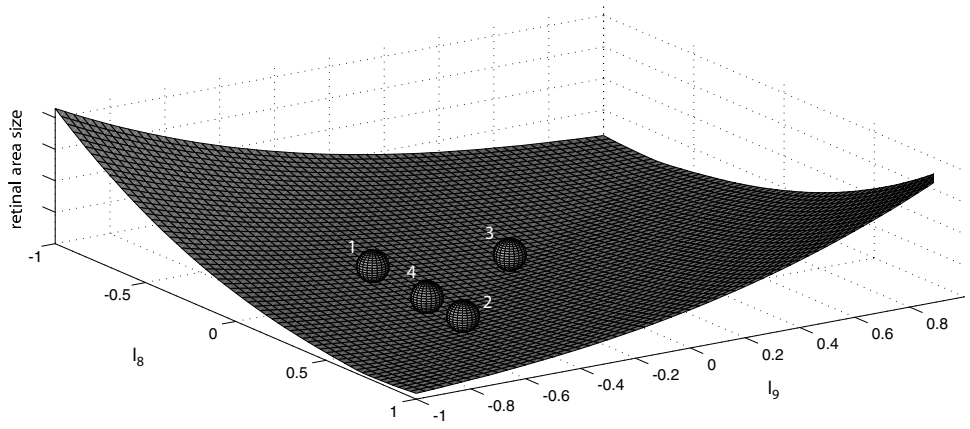


Figure 7.4: The overall size of retinal search regions for motor programs $\Psi = x\alpha\phi + y\beta\phi$ (see text). The four spheres represent 1 - Listing's law ($l_8 = 0, l_9 = -0.5$), 2 - pure L2 ($l_8 = 0.5, l_9 = -0.5$), 3 - optimal solution ($l_8 = l_9 = 0$), 4 - a realistic mix between L2 and Listing's law used by humans ($l_8 = 0.3, l_9 = 0.5$).

search zones. But it does strike a compromise between the optimal solution and Listing's law, which has been described as optimizing motor efficiency (Tweed, 1997b).

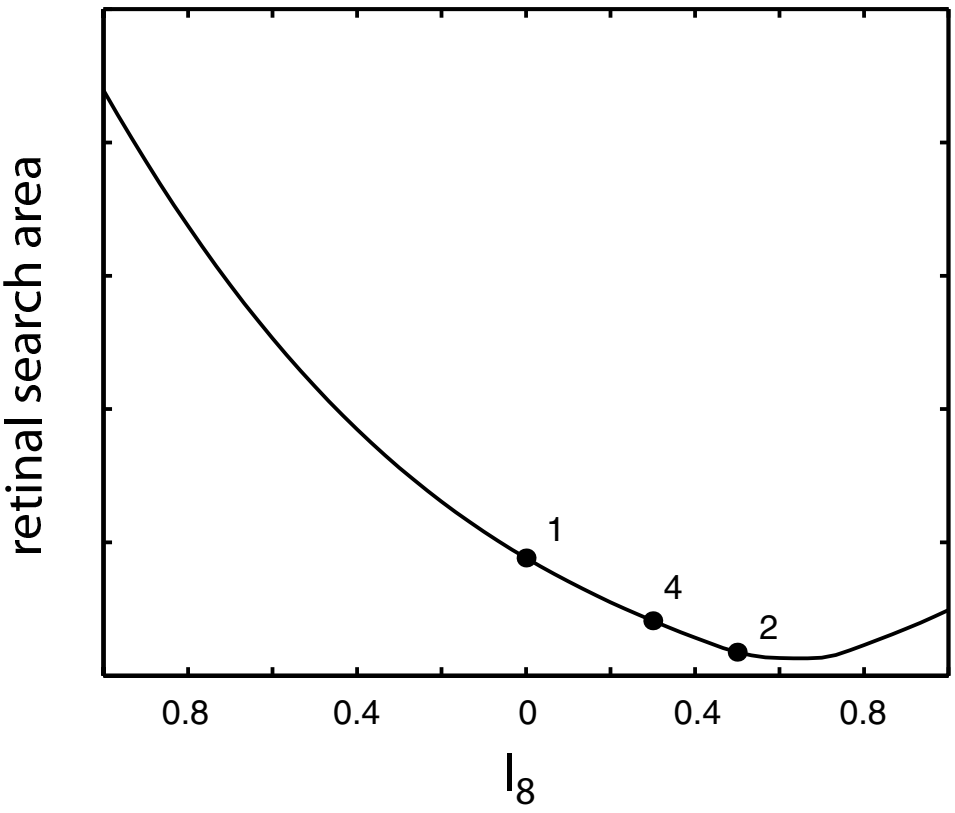


Figure 7.5: Overall size of the retinal search regions for different values of l_8. Same numbering used as in Figure 7.4.

Chapter 8

The Extended Theoretical Horopter

The theoretical horopter described in Section 3.3 is a useful tool for visualizing the geometry of binocular disparities. While it is relatively easy to correct for the Hering-Hillebrand deviation and the shearing of binocular correspondence in the vertical meridians, the second shortcoming of the theoretical horopter is more severe.

The basic assumption in constructing the theoretical horopter was that there exist pairs of retinal locations in the two eyes, so that an object projecting to these two locations will be perceived single. However, due to the geometry of rotations, the projection lines corresponding to most of these retinal location pairs will not meet for most eye positions. This, as mentioned before, is the geometrical reason for the theoretical horopter consisting just of a circle in the plane of regard and a vertical line for most eye positions.

However, it is well known that retinal correspondence is not, in fact, a correspondence between points but rather between retinal regions called Panum's fusional areas. When these areas are large enough, we can expect single vision even when lines projecting from paired retinal locations do not meet.

I therefore propose an extension of the concept of the theoretical point horopter: as with the prior definition, corresponding retinal points are points having the same retinal coordinates in both eyes. The point on the horopter corresponding to such an image pair is the point in space nearest to both optical projection lines.

This definition yields the old theoretical horopter in the plane of regard,

since the projections there will meet on the Vieth-Müller circle.[1] Outside of the plane of regard, this definition gives the location of points in space closest to a given pair of retinal locations in both eyes.

The question for this chapter thus is

What is the full shape of the theoretical horopter in three dimensions?

8.1 Mathematical Description

To find these locations in space, I use the Moore-Penrose pseudoinverse described in Section 3.8. For each set of retinal coordinates λ and μ, in head-fixed coordinates the projection vector is

$$\vec{v} = \begin{pmatrix} \cos(\lambda)\cos(\mu) \\ \sin(\lambda) \\ -\cos(\lambda)\sin(\mu) \end{pmatrix} \tag{8.1}$$

When the eye is in position (θ, ϕ, ψ), this vector becomes

$$\vec{v}_{\theta,\phi,\psi} = \mathcal{M}_\phi \mathcal{M}_\theta \mathcal{M}_\psi \vec{v} \tag{8.2}$$

The equation for intersections of the two projections from the two eyes then is

$$a_1 \vec{v}_l - a_2 \vec{v}_r = \vec{\imath} \tag{8.3}$$

or

$$\underbrace{\begin{pmatrix} \vec{v}_l & \vec{v}_r \end{pmatrix}}_{\mathcal{N}} \vec{a} = \vec{\imath} \tag{8.4}$$

The solution to this equation can be computed using the Moore-Penrose pseudoinverse of the matrix $\mathcal{N}$:

$$\vec{a} = \mathcal{N}^\dagger \vec{\imath} \tag{8.5}$$

The point of the horopter corresponding to λ, μ is then

[1] It is a simple matter to revise my new definition to yield the Hering-Hillebrand-deviation, but for clarity I will present the new ideas in the simplest form, without the deviation.

$$\vec{P}_h = \frac{a_1\vec{v}_l + \vec{i} + a_2\vec{v}_r}{2} \tag{8.6}$$

This denotes the point halfway between the points on each of the two projections closest to the other projection, i.e. the point in the middle between the two projection lines, closest to both.

8.2 Images

Using Matlab, I computed the shape of the extended theoretical horopter for a number of different eye positions (see Figures 8.1 to 8.7).

The theoretical horopter becomes smaller and more curved for increasing vergence angles.

In general, it is largest in the plane of regard, where it coincides with the Vieth-Müller circle, and becomes smaller above and below the plane.

While Figures 8.1 to 8.4 show the horopter for retinal locations of $\lambda, \mu = \pm 50°$, Figures 8.5 to 8.7 show the full shape of the theoretical horopter, excluding points with a z-coordinate of more than 10.

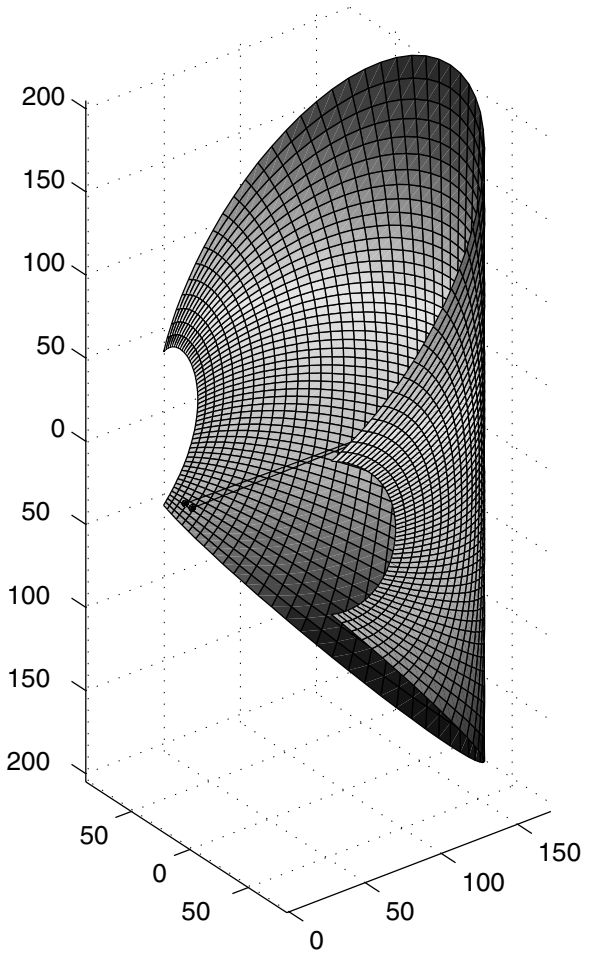


Figure 8.1: The extended theoretical horopter. Shown are the two eyes, the lines of sight and the surface of the horopter for retinal locations of $\pm$ 50°. Eyes are converged symmetrically by 1°. Units are centimetres.

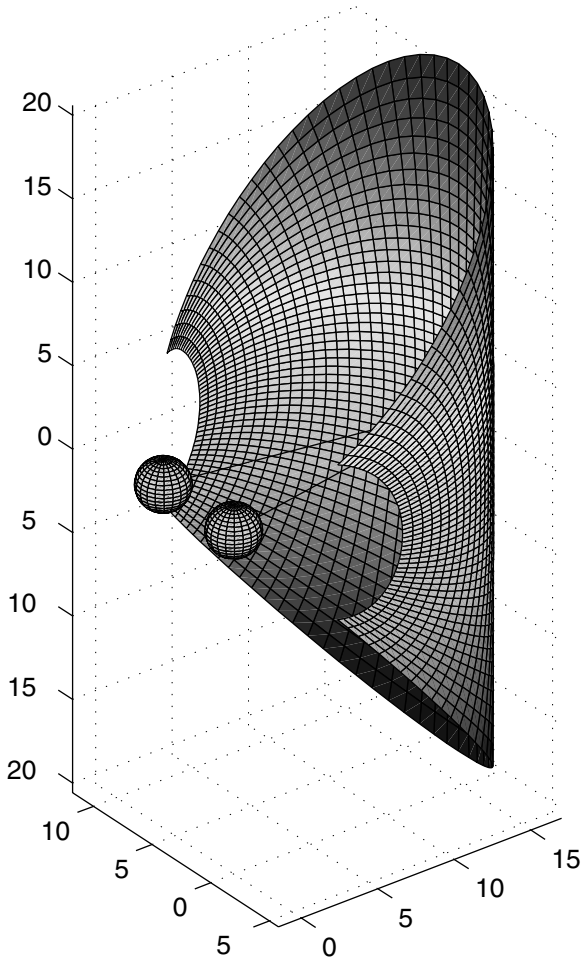


Figure 8.2: The extended theoretical horopter. Shown are the two eyes, the lines of sight and the surface of the horopter for retinal locations of $\pm$ 50°. Eyes are converged symmetrically by 10°. Units are centimetres.

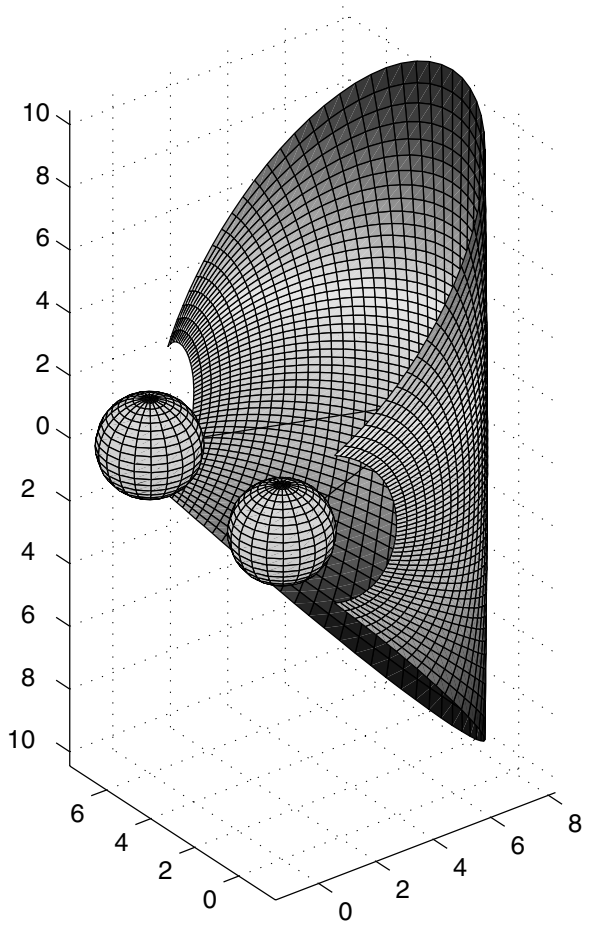


Figure 8.3: The extended theoretical horopter. Shown are the two eyes, the lines of sight and the surface of the horopter for retinal locations of ± 50°. Eyes are converged symmetrically by 20°. Units are centimetres.

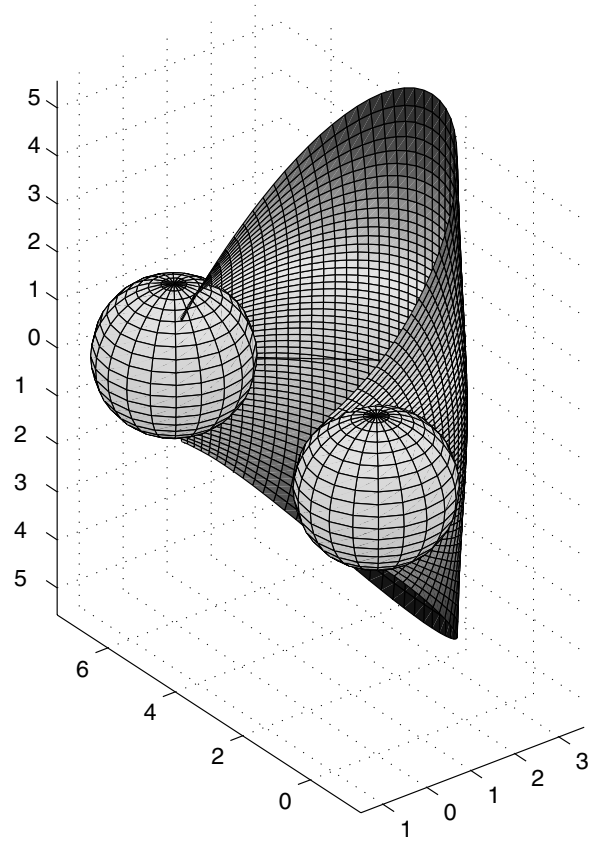


Figure 8.4: The extended theoretical horopter. Shown are the two eyes, the lines of sight and the surface of the horopter for retinal locations of $\pm$ 50°. Eyes are converged symmetrically by 40°. Units are centimetres.

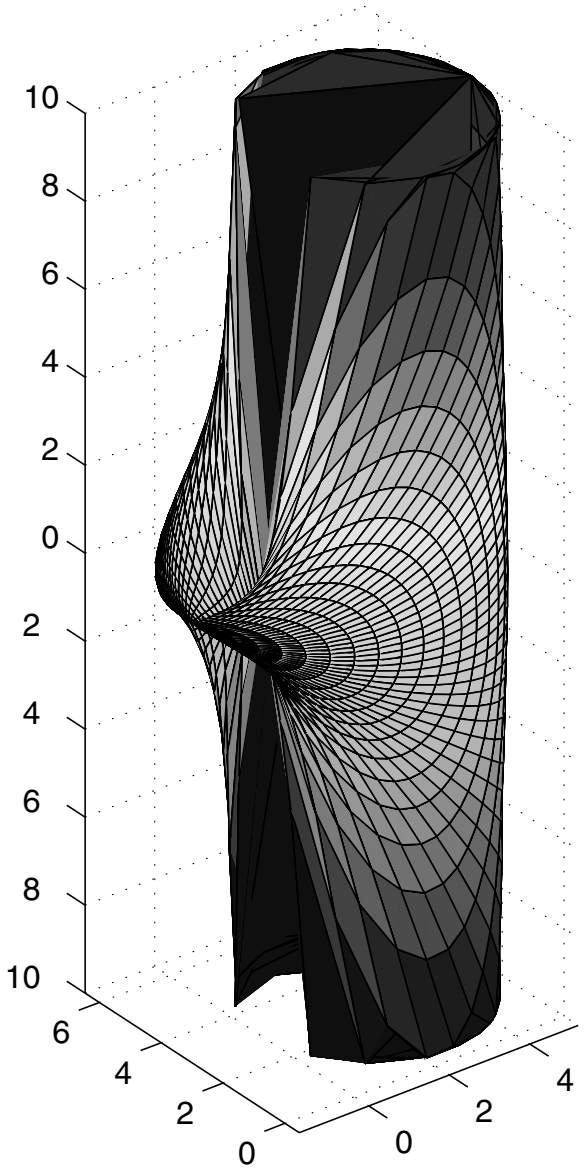


Figure 8.5: Surface of the extended theoretical horopter for retinal locations of ± 180°. Eyes are converged symmetrically by 30°. Units are centimetres. Oblique view.

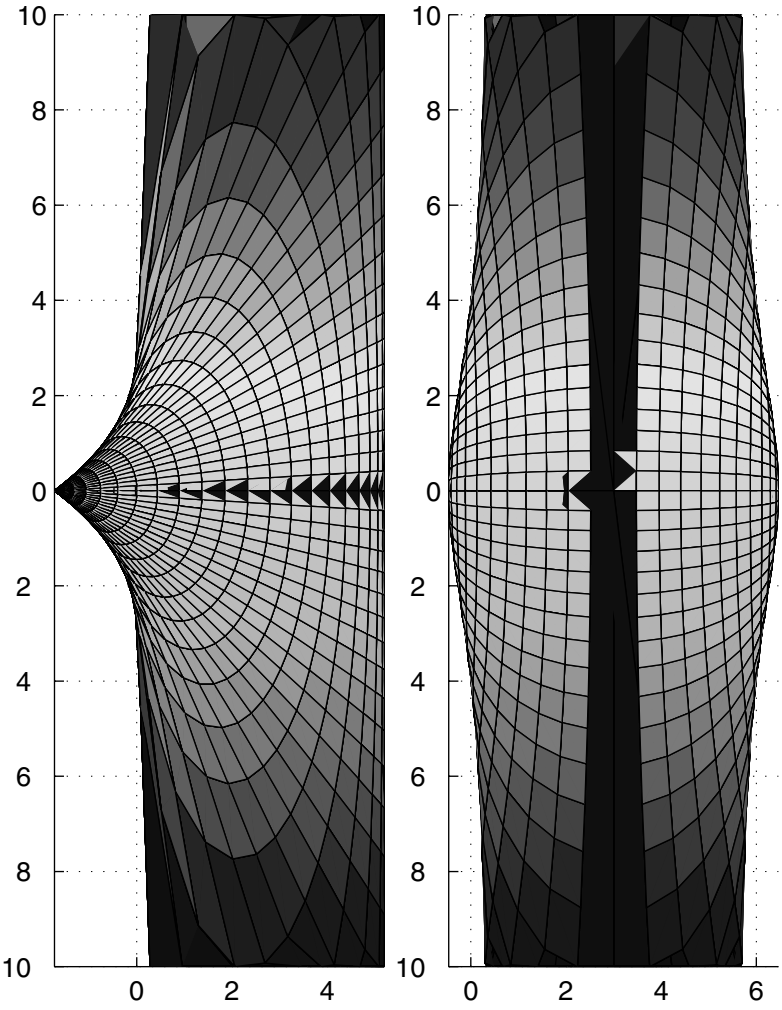


Figure 8.6: Surface of the extended theoretical horopter for retinal locations of $\pm$ 180°. Eyes are converged symmetrically by 30°. Units are centimetres. Left: Side view. Right: Behind view.

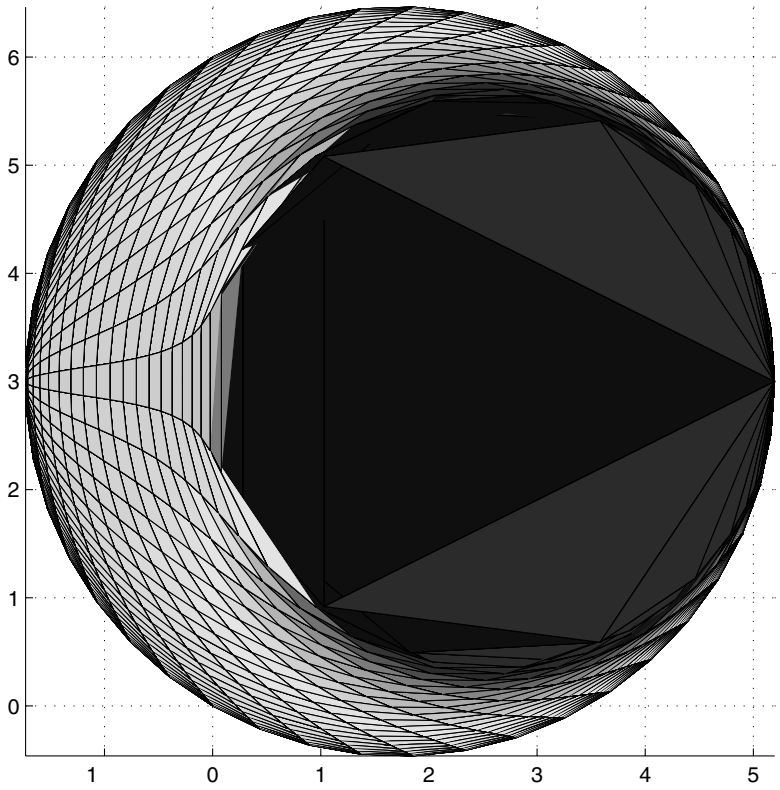


Figure 8.7: Surface of the extended theoretical horopter for retinal locations of $\pm$ 180°. Eyes are converged symmetrically by 30°. Units are centimetres. Above view.

Chapter 9

Visual Bootstrapping

> Another time I wanted to jump over a swamp, which at first didn't appear to be as wide as I found it when I was in the middle of the jump. Hovering in the air I therefore turned around to go back and take a longer run-up. In spite of this I jumped short on the second attempt, as well, and fell into the mud up to my neck. I would surely have died there, had not the strength of my own arm pulled me and my horse, which I grabbed firmly between my knees, out again by my own pigtail.
>
> *The Baron Munchausen*[1]

Since eye position signals are not used to determine the epipolar geometry of disparities, stereo matching is once again stuck with the problem of eliminating the false matches present within Panum's fusional areas.

[1] In Gottfried August Bürgers *Wunderbare Reisen zu Wasser und Lande, Feldzüge und Lustige Abenteuer des Freiherrn von Münchhausen* (Bürger, 1786), my translation. This story is often cited as the origin of the phrase 'pull yourself up by your bootstraps' which inspired the name 'bootstrapping' for a certain statistical procedure, and the term 'booting' for starting a computer. The citations are usually vague, most ascribing the story to Rudolph Erich Raspe, who published the first collection of Münchhausen adventures in 1785 in English. His collection, however, does not contain the story above. It was written by Bürger and has Münchhausen pulling himself up by his hair, not his bootstraps. The origin of the bootstrap phrase is unknown, which would mean that the word actually pulled itself out of the void by its, well, bootstraps.

9.1 The Idea of Bootstrapping

However, there is another way that the visual system might exploit the epipolar constraint: the visual images themselves contain the necessary information about eye position to compute the location of the retinal search lines.

This problem of computing relative camera position from two images is known in mathematics as the relative orientation problem and is fairly well understood (Horn, 1990). From five correct matches the relative orientation of the two eyes can be computed, and using this information the visual system could compute the location of the epipolar lines and then restrict its search for possible matches to these lines, simplifying the remainder of the search.

Since this proposed mechanism involves the use of matches to find matches, I named it visual bootstrapping.

Does stereo-matching use visual bootstrapping, i.e. does it use information about eye position to employ the epipolar constraint, even when search zones are retina-fixed?

If the visual system indeed uses visual bootstrapping, it should be possible to confuse it by presenting stereograms that contain inconsistent information about the viewing geometry. These stereograms would make the computation of an eye position estimate impossible. Such stereograms should therefore be harder to see.

There are two possible scenarios for this mechanism. Either a single eye position estimate is computed from the overall geometry of the scene, and this estimate then is used to constrain matching. Or the estimation of eye position is done locally, based on the distribution and direction of the disparities present on the retina. A local feedback mechanism could enforce the detection of disparities whose direction is geometrically consistent with each other and inhibit the detection of disparities from different geometries. In this local mechanism, separate estimates of eye position could coexist for distant retinal areas.

9.2 Rotated Stereograms

I first tested this hypothesis using cyclorotated stereograms. Cyclorotation of binocular stimuli should be a very common stimulus for the visual system, since L2 can create several degrees of cyclovergence under normal viewing conditions. If the visual system tries to compute eye position, image excy-

clorotation should tell it that the eyes must be converged and looking up. Conversely, image incyclorotation should suggest a state of convergence and gaze depression.

9.2.1 Local and Global Geometry

To test for both the local and global hypothesis of eye position estimation, I created two types of stereograms containing inconsistent information about eye position.

The globally inconsistent stimuli were created by having the left halves of each disc of the stereogram excyclorotated and the right halves incyclorotated (See figure 9.1. Thus, while no global eye position estimate was possible from these stereograms, local mechanisms would still be able to use the geometric information present almost everywhere in the image.

In contrast, for the locally inconsistent stimuli, I put an excyclorotated and an incyclorotated stereogram on top of each other. One of these stereograms contained the left stimulus circle, if present, the other one contained the right stimulus circle. Since the dots from one stereogram would be in the same position as the stimulus circle in the other stereogram, the stimulus circles would be seen as lying behind a circular patch in the main plane of the stereogram where the dot density was halved. For the locally consistent stimuli I encoded the stimulus circles using half the dots at the location of the stimulus circle, thus making the inconsistent and the consistent stimuli equivalent.

In the locally inconsistent stimuli, neighboring dots would contain different geometric information and their binocular disparities would indicate different viewing geometries. These stimuli could not be used to compute eye position estimates in any way.

9.3 Methods

In the experiments I presented the stimuli comprising 4800 dots in a circle with an angular subtense of 12°, on a computer screen at a distance of 50 centimetres. Each of the five subjects was placed on a chinrest and had to crossfuse the stereograms to obtain fusion. The vergence angle required for fusion was 15°. Stereograms were flashed for 200 ms, and contained either no stereoscopic shape, a small circle located behind the main plane of the

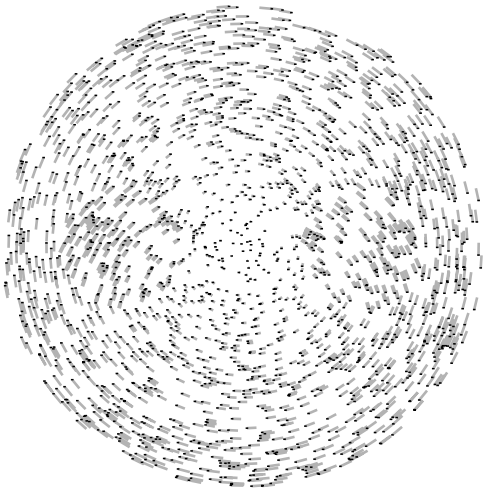


Figure 9.1: The disparity field for the consistent stimuli. The dots mark the positions of the random dots in one eye while the line traces give the disparity for each dot.

stereogram and to the left, a small circle located behind the main plane of the stereogram and to the right, or both small circles. Subjects indicated for each stereogram which of these four configurations they perceived.

In the first part of the experiment, the thresholds for perception of the stereograms were mapped using the Ψ-method (see Section 3.6) (Kontsevich & Tyler, 1999). Once the thresholds were identified, subjects were presented with 100 stereograms for each of four conditions in random order. The conditions were: (consistent, excyclorotated), (consistent, incyclorotated), (inconsistent, excyclorotated) and (inconsistent, incyclorotated). All the stimuli were presented at the threshold angles measured. Since this meant that the inconsistent stimuli contained two different cyclorotation angles, the following method was used to compare performance for global and local stimuli.

I calculated the percentage of correct answers for the left and right halves of the consistent stimuli, both for ex- and incyclorotation. Since the inconsistent stimuli would consist either of an excyclorotated left half and an incyclorotated right half (p_1) or vice versa (p_2), I computed expected performance values for the inconsistent stimuli using the percentages of correct answers for the consistent halves:

$$p_1 = p_{in,left} \cdot p_{ex,right} \quad (9.1)$$

$$p_2 = p_{ex,left} \cdot p_{in,right} \quad (9.2)$$

I then compared this expectation value with the actual percentage of right answers for the inconsistent stimuli.

9.4 Confounding Factors

The results of these experiments were surprising. In the global condition, I found that the *inconsistent* stimuli were significantly easier to match than the *consistent* ones (T-test, $p<0.01$, Figure 9.2). Since the two types of stimuli contained exactly the same disparities in their left and right halves, and only differed in their pairing, this was quite surprising. It seemed to indicate that disparity geometry was evaluated globally, but the evaluation did not seem to be what I thought it to be.

I felt the likeliest explanation was that while the inconsistent stimuli were indeed strictly inconsistent with any binocular configurations–that is, their pattern of disparity could never arise in any visual surroundings or eye position in normal life—nevertheless they were approximately consistent with a situation where the two eyes are vertically diverged. Perhaps the stereoptic system was undisturbed by the inconsistent stimuli because they were judged 'close enough' to consistent ones. And perhaps this approximation to vertical divergence was actually easier to deal with than the torsional misalignment implied by the consistent stimuli. The solution to this problem is described below in Section 9.5.

In the local condition, I found the predicted effect, the inconsistent stimuli being much harder to perceive ($p<0.01$, Figure 9.3). But further tests were necessary to check whether this effect was truly due to contradictory eye-position information, or to another factor which I called cluster-breaking: when one overlays two inconsistent stereograms, as I did in the locally inconsistent stimuli, one breaks up the clusters of dots in the stereograms: points that are close to each other in the left eye's image, but not part of the same substereogram will appear moved in opposite directions in the right eye's image, due to the opposite direction of image rotation of the substereograms. In short, clusters of points that are present in the left eye's image are destroyed in the right eye's image and vice versa.

So if stereopsis involves the brain recognizing similar clusters in the two eyes' images, then cluster-breaking rather than inconsistent eye-position information may account for my subjects' poor performance on the inconsistent stimuli.

To test this idea, we modified the consistent stimuli, so that they had the same degree of cluster-breaking as the inconsistent stimuli. To do this I horizontally shifted half of the dots in the consistent stereograms. The amount of the shift was the same as the vertical difference in disparities between the two dot subsets in the inconsistent stereograms, so that the resulting disparity difference between neighboring dots differed only in direction and not in length between the consistent and inconsistent conditions.

Unfortunately, this horizontal shift also, inevitably, complicated the consistent stimuli in another way: when fused, these stimuli now appeared as a transparent flat disc of dots lying in front of another disc with, possibly, smaller discs on the left or right removed even farther in depth. The foremost, transparent disc was not present in the inconsistent stimuli, so in that sense they were now simpler. Likely for that reason, subjects tested on these new stimuli found the consistent stereograms harder to perceive than the inconsistent. This result suggests that greater complexity made the consistent stereograms harder to perceive, but it still does not settle the question of whether the contradictory eye-position information in the inconsistent stimuli hindered stereopsis. In general, the problem is that it is impossible to construct locally consistent and inconsistent stimuli that differ only in the consistency of their eye-position information, without confounding differences in the other factors: cluster-breaking and image complexity. If these other factors had turned out to have little influence on stereopsis, this series of experiments would have sufficed to prove or disprove the bootstrapping hypothesis, but as it was I needed a new approach.

9.5 Going Horizontal

I found that it is impossible to devise locally inconsistent stimuli (where the contradictory information about eye position is present in every part of the image) without affecting other factors that turned out to be important in stereopsis, namely clusters and image complexity. Therefore I concentrated on devising new globally inconsistent stereograms, where the contradictory information is segregated in different regions of the image. The problem

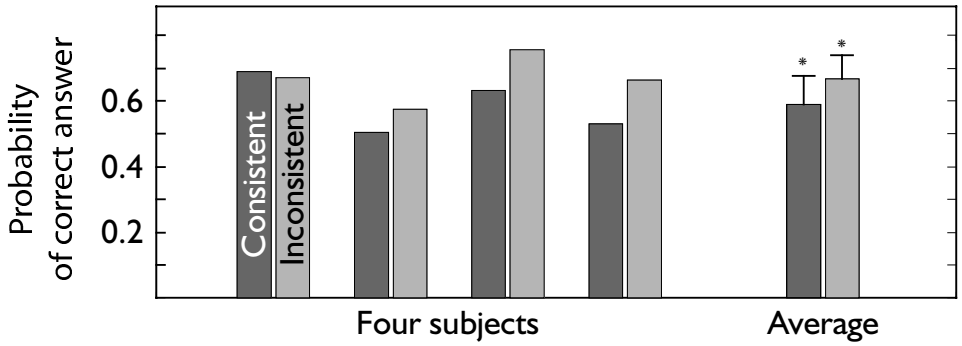


Figure 9.2: Experimental results for the global stimuli. Dark bars are the computed performance expectation values, light bars are actual performance on the inconsistent stimuli.

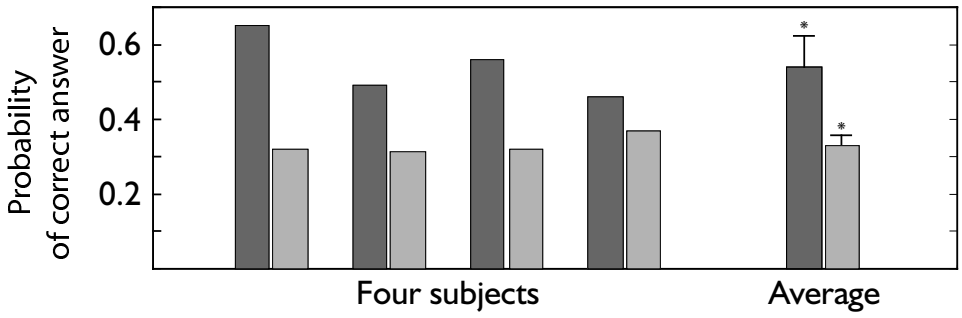


Figure 9.3: Experimental results for the local stimuli.

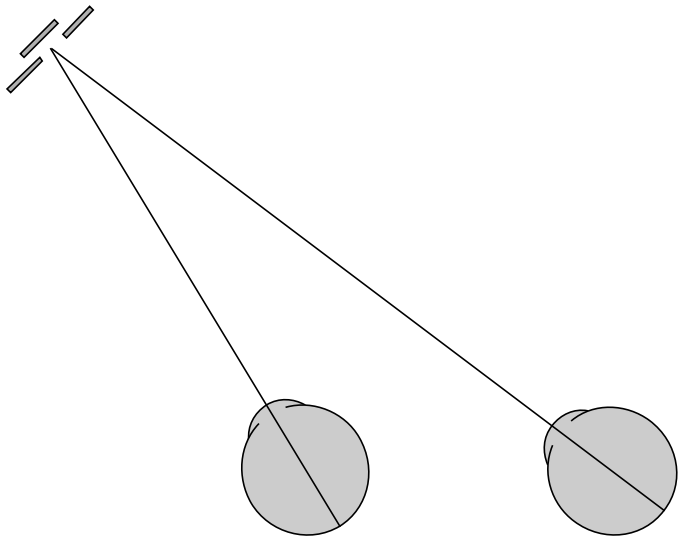


Figure 9.4: The stimuli for the final form of the bootstrapping experiment were constructed to contain disparities consistent with binocular gaze directed to the left or right, while being presented in a center position.

with the global stimuli in Section 9.3 was that my inconsistent stereograms were after all approximately consistent with certain eye positions. I needed stereograms that were not even roughly compatible with any possible eye positions. To build these, I abandoned cyclodisparities and considered stereograms that simulated what the eyes would see on extreme rightward or leftward gaze.

In my simulations of epipolar lines (see Appendix E) I had found that the pattern of epipolar lines created by eccentric horizontal viewing positions did not resemble any other viewing geometry, enabling us to create stimuli that really were inconsistent with any viewing geometry whatsoever.

The pattern of disparities in the stimuli I used (see Figure 9.4) corresponded to a simulated viewing situation where a collection of dots is presented to the subject at an eccentric gaze angle. As before, these stimuli contained zero, one or two disparity-encoded discs. The discs were located at the top and the bottom of the stimulus. In the inconsistent stimuli either the top half of the stimulus was consistent with a gaze position looking left and the bottom half consistent with gaze pointing right or the other way around.

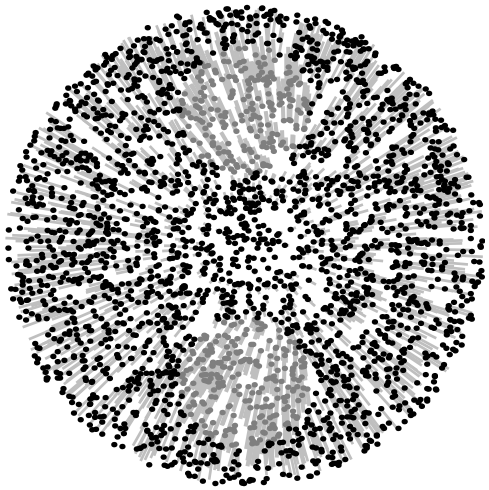


Figure 9.5: The disparity field for the stimulus configuration shown in Figure 9.4. The dots mark random dotpositions in the right eye's image, while the lines give the binocular disparity for each dot. The expanding pattern is due to the fact that the left eye is closer to the target.

The stimuli were presented on a computer screen, using LCD shutter glasses, eliminating the need for vergence and, more importantly, removing the restriction on my subject pool that had enabled only free fusers to participate in my experiments.

When I tested my subjects on these stimuli, using the same procedure described above, I found the consistent stimuli to be significantly easier to perceive than the inconsistent ones ($p<0.01$). See Figure 9.6 for the results.

These results suggest that the global bootstrapping hypothesis is correct: the stereoptic system computes a global eye position estimate from the retinal distribution of disparities, and uses this estimate to facilitate stereo matching, effectively incorporating the epipolar constraint.

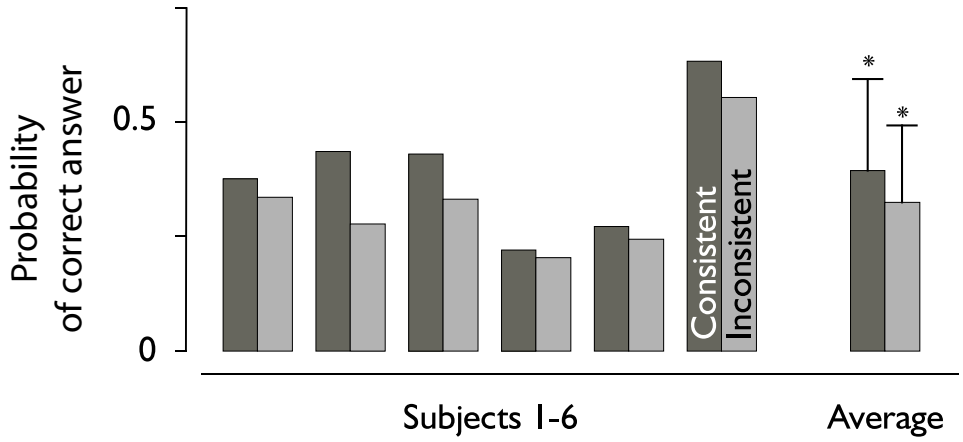


Figure 9.6: Experimental results for the stimuli with horizontal search line geometry.

Chapter 10

Summary

10.1 Conclusions

In this work, I have investigated the interaction between the geometry of stereo-matching and eye movements using both psychophysical experiments on human subjects and mathematical simulations.

1. Stereopsis is a sensorimotor process. The pattern of eye movements can not be explained without taking into account the demands of the visual system. Conversely, a full description of the problem the visual system has to include the effect of eye movements on stereo correspondence.

2. The search regions for stereo-matching are retina-fixed. Specifically they are not adjusted to coincide with the locations of the moving epipolar lines for changing eye positions.

3. The retina-fixed search regions have to be large enough to cover the area the epipolar lines can slide through. Therefore stereo-matching is an inherently two-dimensional problem.

4. The motor program humans use during near vision, the binocular extension of Listing's law, can be described as a compromise between the program used for far vision, Listing's law, and the motor program minimizing the size of the retinal search zones.

5. Visual information about eye position and the location of the epipolar lines is used to simplify matching in a process I call visual bootstrapping.

6. The theoretical horopter, the region in space where objects are seen single rather than double, hitherto characterized only incompletely, is here extended into a full two-dimensional surface using a mathematical device called the Moore-Penrose pseudoinverse to compute the contributions of Panum's fusional area.

10.2 Contributions

1. The interactive processes in the stereo-motor system provide a good and convincing example of the general claim that analysis of subsystems of the brain can be in danger of overlooking important interactions between systems.

2. The study described in Chapter 6 and published in Nature (Schreiber *et al.*, 2001) was the first to show that stereo-matching is retina-fixed and the epipolar constraint therefore is not employed. These results have now been confirmed by Raymond van Ee (Personal communication).

3. The deviation from Listing's law known as L2 was first described about 10 years ago. This disseratation offers the first clear demonstration of a functional advantage for L2 that can convincingly explain its use by the oculomotor system.

4. The extension of the horopter into a full two-dimensional surface shifts the focus of binocular correspondence from pointwise matching toward the matching of fusional areas. Also, the predictions stemming from the particular shape of the horopter can be tested experimentally and shed light on the principles governing the formation of binocular correspondence between retinal regions in the two eyes.

5. I have written simulations that allow the visualization of three- dimensional rotations and epipolar geometry. In addition to the direct purposes served by the simulations this is of great value for the teaching of binocular vision and communication of its basic principles and methods.

10.3 Publications

The results of this dissertation have been or will be published in the form of refereed journal articles:

1. The results of the psychophysical experiments showing that the search zones are retina-fixed and an overview on the results of the simulations of the retinal search regions for different motor programs have been published in Nature (Schreiber *et al.*, 2001).

2. An article elaborating the principles of sensorimotor interaction and the findings in the stereo-motor system has been submitted to the journal Strabismus and is currently under review.

3. The details of the simulations of the retinal search regions will be the subject of a longer version of (Schreiber *et al.*, 2001)

4. The results and conclusions of the bootstrapping experiments.

5. A theoretical study of three-dimensional rotational geometry and the mathematics of binocular correspondence. This article will try to unify different approaches to the definition of vertical disparity.

6. The description and formulas of the extended theoretical horopter.

10.4 Perspectives

It has become clear from this work that the stereo-motor system, that is, the system that drives the two eyes in order to achieve stereopsis, cannot be fully understood from an analysis of the two subsystems it most naturally breaks down into: the stereo-visual and the oculomotor system. How the eyes move influences the problem vision has to solve, and the demands of vision in turn shape the details of motor control.

10.4.1 Subject Pool and Stimulus Specifics

The subjects in our first study were selected for their ability to perceive a random-dot stereogram free fusion. Since this is only a small subset of

the population with stereopsis, we used LCD shutter glasses in the subsequent studies to overcome this limitation, obtaining similar results. More importantly, we obtained the same qualitative results of differences between conditions independent of the subjects' ability to perceive the stereograms.

But even in these studies, my subject pool was restricted to young healthy persons with normal vision. It would be interesting to investigate the role of age and a variety of visual disorders on the performance in these viewing conditions.

Secondly, I wanted to make an important general point about the processing of visual information by the stereoptic system. By eliciting statistically significant response differences for one stimulus configuration (number and size of random dots, luminance profile of dots, etc.), I have been able to demonstrate some general principles of information processing. It would be interesting to study the dependence of these principles on stimulus parameters such as dot density, spatial frequency, color, luminance, contrast, etc., to specify which neural pathways most likely are involved.

10.4.2 Plasticity

It has been shown in several studies that the motor pattern of L2 can be adapted using binocular stimuli (Schor *et al.*, 2001) (Schor & McCandless, 2002) (Takagi *et al.*, 2001). There is also evidence for changes in motor control in direct and immediate response to stereograms (Kapoula *et al.*, 1999). On the other hand children that have oculomotor deficits or strabismus can develop what is known as anomalous correspondence, where regions in different parts of the two eyes' retinas are connected by binocular cells. This could indicate that during development binocular vision and oculomotor control interact to develop both the binocular connectivity in Panum's fusional areas and the oculomotor program that enables the system to have the best stereovision at the least metabolic cost.

This interaction would entail several predictions: if improving the binocular register of the two retinas during eye movements is indeed the purpose of L2, and if this is learned during childhood, patients with early childhood strabismus should show deviations from the motor pattern of L2. Also, since binocular vision has a sudden onset at about 4 to 5 months of age, it would be useful to look for changes in oculomotor control that accompany this developmental step.

A second very interesting possibility is that the visual system retains some

of the plasticity into adulthood, making it possible to increase the size of Panum's fusional areas. There is some indication that the tolerable range of disparities increases within the first 10 to 15 minutes of stereogram viewing, and there is also strong indication that the ability to perceive binocular stimuli can be trained over a longer period of time—I was hardly able to cross fuse stereograms when I started on this dissertation project and have thresholds at about double the usual subject's average today.

10.4.3 Full Geometry

During all of the experiments the subjects' heads were fixed in place by chin rests. Epipolar geometry depends only on the relative orientation of the two eyes, so head position in itself does not change the rules of retinal correspondence. However, any movement of the head influences eye movements, through vestibular and postural reflexes and by simply changing the relative eye orientations required to fixate a space-fixed target.

Misslisch *et al.* reported in 2001 that the amount of ocular counterroll, the torsional rotation of the eyes in response to torsional rotations of the head, is reduced for near vision and large vergence angles (Misslisch *et al.*, 2001). This may imply that the problem of stereo-matching has implications for the kinematics of eye-head movements as well and is an interesting field for further research.

Appendix A

Means of Stereoscopy

There are several different techniques for controlling binocular input to the two eyes in order to create stereoscopic depth. While some of these are used mainly for entertainment purposes, others have been used extensively in stereovision research. In the studies contained in this thesis I have used both free fusion techniques and shutter glasses to present the random-dot stimuli to my subjects.

A.1 Free Fusion

The simplest approach, as far as the technology employed is concerned, is the presentation of the two retinal images side by side. Spectators then must converge or diverge their eyes so that each eye sees the appropriate image (see Figure A.1).

This techniques has the enormous advantage that it can be used anywhere, with no equipment at all. All that is necessary is a piece of paper with the two images.

The setbacks are: Image size is limited by the interocular distance for divergent fusion. For convergent fusion, while the images can be in principle as large as desired, the limits of convergence and of display distance will restrict this somewhat.

But more importantly, free fusion is not easy to achieve. The eyes are fixating on one distance, but need to accommodate to another distance. The coupling between accommodation and vergence can in principle be over-

come[1], but not all persons can do it with ease, and some cannot at all[2]. Most of the experiments described in this thesis were done using the free fusion technique, and it was quite difficult to find subjects who could reliably see the stimuli.

Also, since free fusing stereograms can be seen by only a minority of people, it might be the case that experiments done on free fusers do not actually probe the stereoptic system in general but rather the stereoptic systems of free fusers. Given the smooth distribution of the ability to perceive stereograms it seems very unlikely that there exists a fundamental difference between the visual systems of people who can free fuse and the visual systems of people who can't, but the possibility remains that effects shown for free fusion do not apply to normal binocular vision.

The second concern for free fusion is that the accommodation signal and the vergence signal carry different information about the depth of the viewed object. If these two sources of depth information are integrated and used in the processing of stereo information, the conflict of information in free fusion alone might lead to a change in the system's behavior.

A.2 Stereoscopes

The first stereoscope was invented by Wheatstone and presented in his 1838 report to the Royal Society (Wheatstone, 1838) (see Section 2.1.5 for a description).

The basic idea of the stereoscope is to present different images to the two eyes by diverting the lines of sight with the use of optical instruments, such as prisms, mirrors and lenses. The enormous advantage over free fusion techniques is that by properly arranging the devices used, one can make vergence angle and accommodation distance agree, thus eliminating both the conflicting input to the visual system and the accommodation problems subjects face.

[1] See for example (Carpenter, 1991), p. 160ff

[2] The frustrated viewer trying to see an autostereogram in free fusion, but failing, has even found entry into popular culture in the movie Mallrats where a person trying to see a beautiful sailboat in a stereograms displayed in a Mall eventually gives up and destroys the post the stereogram stands on.

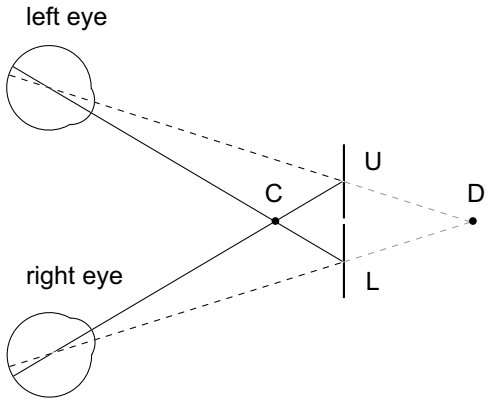


Figure A.1: The free fusion techniques of convergent and divergent fusion. When the eyes converge on C, the left eye sees the lower image L and the right eye sees the upper image U. When the eyes diverge to focus on D, the left eye sees U and the right eye sees L.

A.3 Anaglyphs and Polarized Images

The former two methods worked by spatially separating the left and the right eye's view, providing each eye with an image that could be manipulated independently. The two devices in this section separate the two eyes' images that are presented in overlay by physically filtering out the unwanted image for each eye.

Wilhelm Rollmann in 1853 published a description of his stereoscopic method, that now is called the anaglyph method[3]. In its original form, the two images are presented in different colours but in the same spatial location, say a green line drawing for the left eye and a red drawing for the right eye. The spectator then wears red-green-coloured glasses. The left eye, having a red lens in front of it, will only see the green image, while the right eye, through the green lens, will only see the red image. This technique originally allowed only black and white images to be shown, but has later been expanded to enable several colours through the use of specially designed colour filters.

A similar technique exploiting a different physical property of light is the polarization technique. Light rays in general are electromagnetic waves, and can, through the use of polarization filters, be made to oscillate in a

[3] Anaglyph meaning "like a relief".

plane perpendicular to their direction of propagation (Feynman, 1963-65). When polarized light passes through a polarization filter, the intensity being transferred is $I = I_0 \cos(\alpha)$, with α being the angle between the transmission direction of the polarization filter and the polarization plane of the beam of light.

The images destined for the two eyes are both polarized, with the polarization planes oriented so that they are perpendicular to each other. Two polarization filters worn in front of the eyes, their transmission directions aligned with the respective polarization direction of one of the two images, will let that image pass (α being 0, $I = I_0$) while eliminating the other image ($\alpha = 90°$, $I = 0$). This technique works in full colour.

A.4 LCD Displays and Shutter Glasses

More recently the miniaturisation of LCD displays has made two more techniques for producing disparate images available. The technology of head mounted displays uses two small LCD screens directly in front of the eyes, to create large field 3D-stimuli in full colour. Usually the head mounted apparatus also contains gyrosocopes for measuring of head position and velocity, to accurately correct the scene displayed according to the new head position. Due to its low resolution and image distortion this technique is used mainly in the gaming industry.

The shutter glasses also use LCDs. The transparent glasses on the front of the goggles can go opaque at high frequencies and independently of each other. The full display apparatus contains a stereo enabled graphics card providing a synchronization signal, an infrared emitter sending the signal to the glasses and the shutter glasses themselves. A high frequency monitor is desirable.

The software has to create one full-resolution full-colour image for each eye. It then uses OpenGL to send these two images to the left and right eyes' buffers on the graphics card. The card will display the two buffers alternatingly at half the monitors' refresh rate, i.e. a monitor working at 100 Hz will have sequence of left eye's image and right eye's image 50 times a second. The shutter glasses through the emitter are synchronized with the display, so that the left eye's LCD is opaque while the right eye's image is on the screen and transparent while the left eye's image is on the screen. The converse happens for the right eye's LCD.

A.5 Holograms

Another means of creating a depth percept is holography, invented in 1948 by Denis Gabor as an improvement to electron microscopy (Gabor, 1948).[4]

To describe the state of a light waves at a point in space, its amplitude and phase are needed. Standard photography only records the amplitude of the wave. This is the deep reason for photographs being two-dimensional.

In holography, a special interference technique using coherent light from a laser source is used to record both amplitude and phase of light in the surface plane of the image. When this image later is viewed, ideally with a point light source, the full physical properties of the light recorded are recreated, leading to a full 3D percept of the recorded object.

A.6 Lenticular Images

Lenticular images were developed from parallax barrier methods, which were patented in the US by Frederick E. Ives in 1903. Parallax barrier stereograms interleave the two eyes' views and use a barrier to block each eye from seeing the other eye's image (see Figure A.2. Due to their geometry parallax barrier stereograms are only visible from exactly the right viewing distance. Another setback is that they can only be seen through the barrier, occluding half of the field of view.

Lenticular images, proposed by Ives, use essentially the same technique, but instead of occluding the stripes intended for the other eye, a lenticular sheet is placed before the image. This sheet consists of many long cylindrical lenses, placed side by side. Their optical properties and distance to the image are chosen so that again a selected subset of stripes is projected through the lenses to each eye.

An improved version, both of parallax stereograms and lenticular images, has several images for each eye, which through the arrangement of the barrier or the lenticular sheet can be seen from different viewing angles. Thus it is possible to create the impression of motion.

Lenticular images enjoy great popularity, both in the form of little images changing their appearance when they or their observers move, or as lenticular

[4]The word *hologram* being derived from greek *holos* and *gramma*, meaning "All the message".

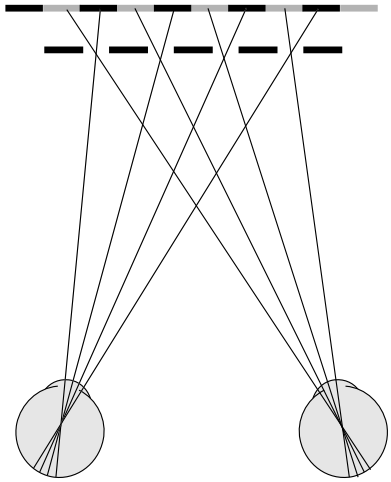


Figure A.2: The geometry of parallax barrier images. In front of the image surface a barrier, commonly an array of wires, is placed. The image surface carries both eyes' images in alternating stripes. The arrangement is such that each eye sees only the stripes from the image intended for it through the barrier. This leads to the impression of a 3D scene behind the occluders.

stereograms creating the impression of depth. These stereograms are commonly used in advertising and can be bought on postcards in almost every tourist cardshop.

Appendix B

On Pseudoscopes

In his original 1838 paper on stereo vision, Wheatstone reported

> A very singular effect is produced when the drawing originally intended to be seen by the right eye is placed at the left hand side of the stereoscope and that designed to be seen by the left eye is placed on its right hand side. A figure of three dimensions, as bold in relief as before, is perceived, but it has a different form from that which is seen when the drawings are in their proper places. There is a certain relationship between the proper figure and this, which I shall call its *converse* figure. Those points which are nearest the observer in the proper figure are the most remote from him in the converse figure, and *vice versa*, so that the figure is, as it were, inverted; but it is not an exact inversion, for the near parts of the converse figure appear smaller, and the remote parts larger than the same parts before the inversion.[1]

In 1852, when he published his continuation report on stereo vision, he described an instrument he called a pseudoscope, because it "conveys to the mind false perceptions of all external objects"[2].

This device, shown in Figure B.1, consisted of two adjustable triangular prisms. The optical arrangement of the apparatus had each eye see the image it would have normally seen, but reflected horizontally. Since this reverses the horizontal direction of disparities, it also reverses depth impression.

[1] (Wheatstone, 1838), p. 377
[2] (Wheatstone, 1852), page 11

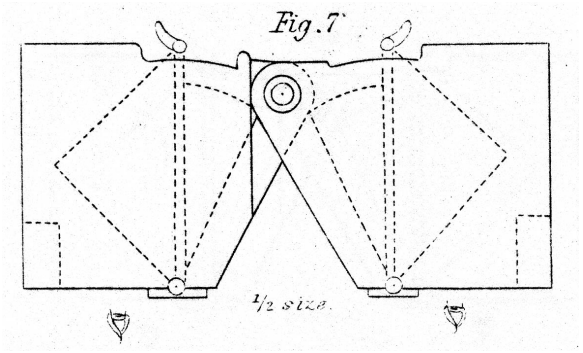


Figure B.1: Wheatstone's pseudoscope. Two triangular prisms are used to invert the image in the horizontal dimension and thus reverse the perceived depth. From (Wheatstone, 1852).

In 1906 Ewald and Gross published a paper on pseudoscopes, where they discussed Wheatstone's design of the apparatus and presented their own, simpler design (Ewald & Gross, 1906). Their pseudoscope, shown in Figure B.2, consisted of four small mirrors, arranged to present the left eye with the image the right eye would have seen and vice versa. See Figure B.2 for the optical geometry of the apparatus.

For demonstration purposes I have built an Ewald pseudoscope using first surface mirrors.

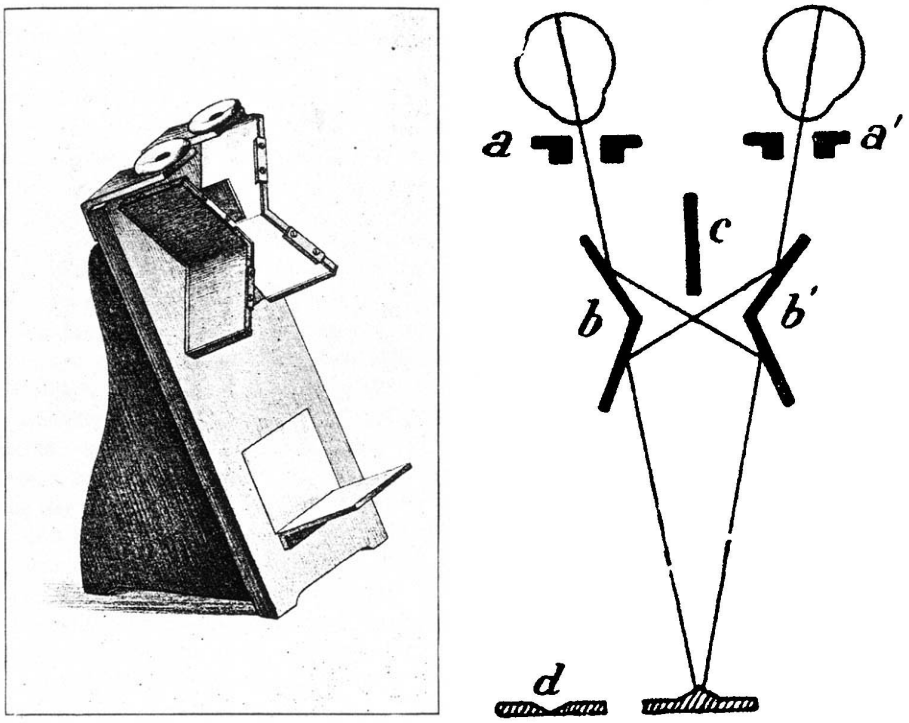


Figure B.2: Ewald's pseudoscope. Two triangular prisms are used to invert the image in the horizontal dimension and thus reverse the perceived depth. The figure on the left shows the lines of sight for both eyes, being reflected twice of the mirrors, effectively swapping the two eyes' images. From (Ewald & Gross, 1906).

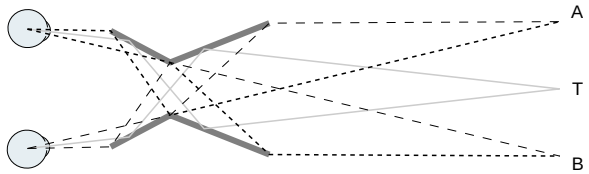


Figure B.3: The actual geometry used for my own version of the pseudoscope. The mirrors are placed so that the left and right eyes view are swapped, and the field of view afforded by the mirrors (between points A and B) has full binocular overlap. Any target placed between A and B is then seen in inverse depth.

Appendix C

The Ghost Theatre

For U of T day on the 16th of October 1999, I created a display showing the double nail illusion and demonstrating Marr's and Poggio's cooperative algorithm.

Figure C.1: The ghost theatre (front view).

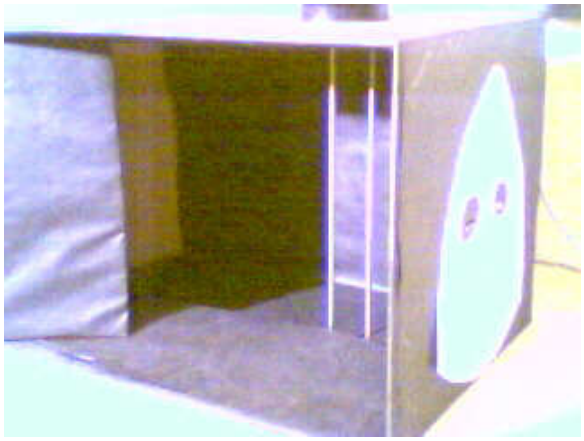


Figure C.2: The ghost theatre (side view).

Appendix D

Epipolar Geometry: VRML-Simulation

The change of the epipolar geometry described in Section 3.4 can be viewed in any browser equipped with the VRML plugin using this model I have written in the virtual reality markup language, VRML. The advantage of VRML for visualizations of this kind is that objects defined can be easily rotated by just rotating their frames of reference.

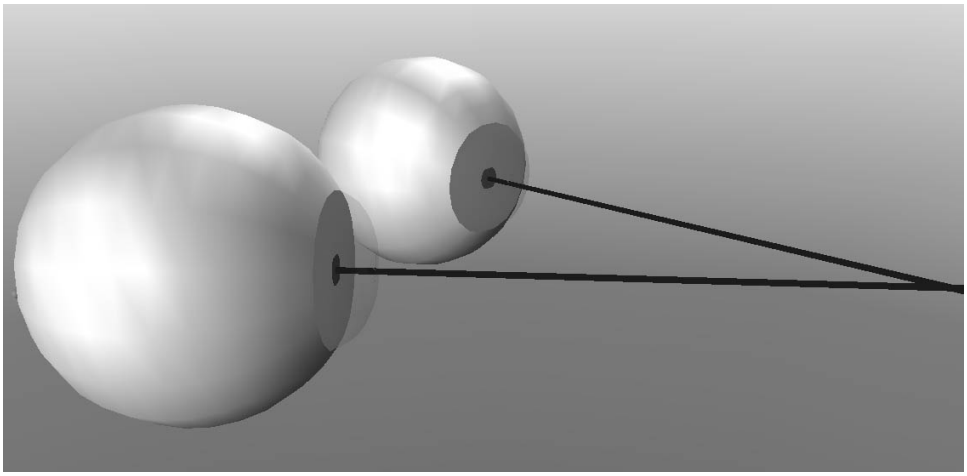


Figure D.1: Screenshot of the VRML simulation. The two eyes will converge and diverge in the horizontal plane and change vertical position while converged, following Listing's law. The viewpoint can be changed to behind three different retinal locations on each eye, and the projection of these locations into space can be switched on, showing the movement of the projection from the other eye (the epipolar line) on the retina. See Section 3.4 for details.

Appendix E

Epipolar geometry: retinal search line simulations

It is quite difficult to conceptualize the geometry of retinal disparities for any eye position other than when gaze is straight ahead and parallel. For this reason I have written a simulation program for epipolar geometry.

This program allows for the easy manipulation of all angles of eye position and immediately displays the resulting changes in epipolar geometry. Figure E.1 shows the programs' interface, consisting of the parameter window, the eye position display window and the epipolar display window.

E.1 Use of the Program

In the parameter window, both the version and vergence angles for horizontal, vertical and torsional eye position can be specified either by entering them into the appropriate fields or by sliding the controls. While torsion can be selected manually, it is also possible to select either Listing's law or L2 ($\mu = 0.5$) for automatic setting of torsion. This makes it possible to observe the change in epipolar geometry for both of these motor programs.

The program can display the epipolar lines themselves, evenly spaced dots on the lines and disparity traces connecting points on the left and right eye's epipolar line. It is also capable of exporting the lines into Matlab readable files for generation of illustrations.

The epipolar lines are displayed in retinal $\lambda\mu$ coordinates. It should be noted that these coordinates have poles (the left and right side of the display

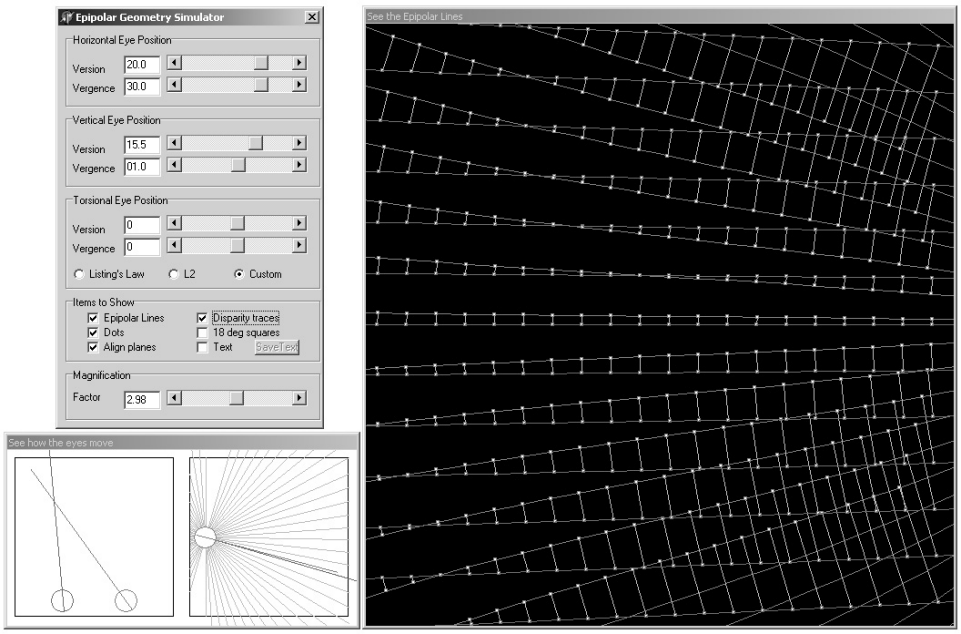


Figure E.1: Program interface of the epipolar simulator. See text for details.

are in fact retinal points) and that at the top and bottom sides traces get reflected due to the geometry of the coordinate system.

Figures E.2 to E.5 show a selection of epipolar line geometries for different gaze configurations.

The epipolar lines are displayed at a distance of 5°of elevation angle, one full set for each eye.

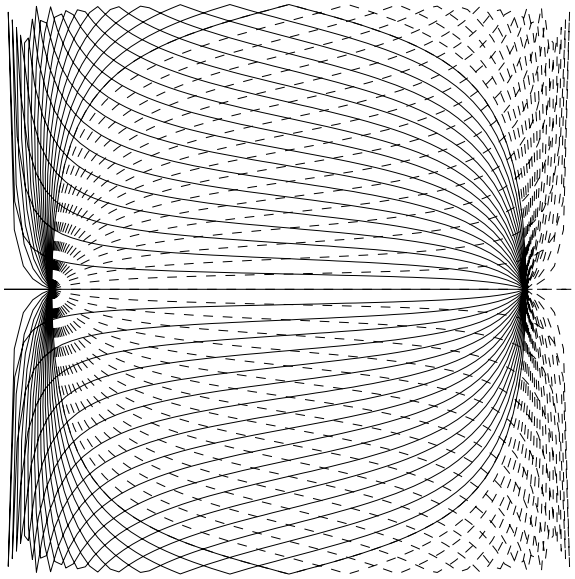


Figure E.2: Epipolar lines for 30° of horizontal vergence.

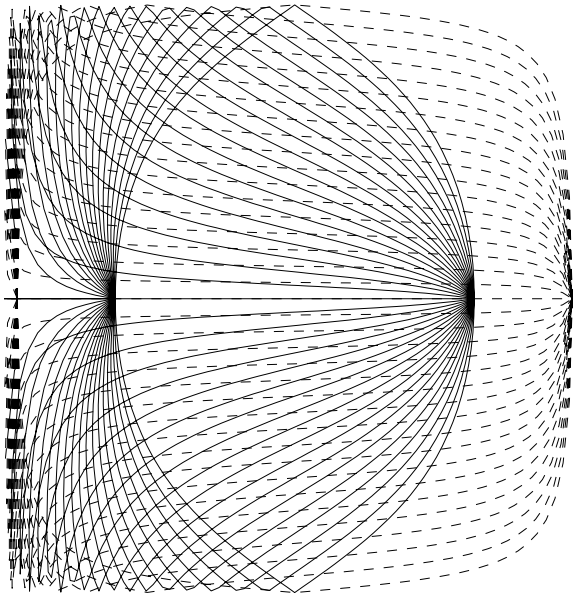


Figure E.3: Epipolar lines for 30° of horizontal vergence and 20° of horizontal version, corresponding to the viewing situation in the experiments described in section 9.5.

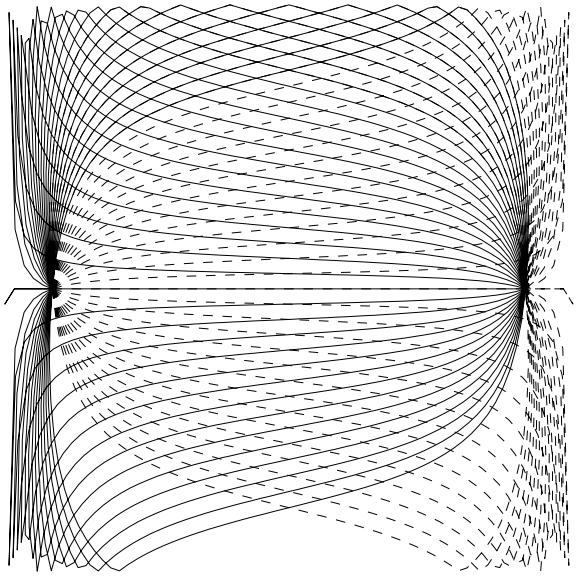


Figure E.4: Epipolar lines for 30° of horizontal vergence and 20 degrees of vertical version with L2. Note the torsional tilt of the epipolar lines.

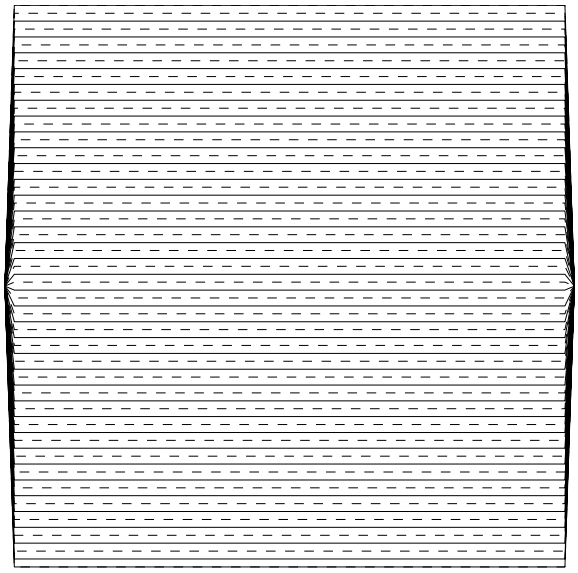


Figure E.5: Epipolar lines for 2.5° of vertical vergence. This

Appendix F

Simulations of Panum's Areas

To visualize the change of the expected size and shape of Panum's fusional areas with different motor programs described in Chapter 7, I wrote a simulation package that allows the quick and easy computation of these zones.

F.1 Retinal Locations

Retinal locations can be added to the display by simply clicking on the screen, or by choosing from a preset set of configurations, including horizontal and vertical locations and a circle of locations (shown in Figure F.1). The radius of the circle can be set in the command panel on the right.

The retinal locations can also be dragged around the retina using the middle mouse button, allowing for the exploration of the change of the predicted search zones.

F.2 Motor Programs

In the "Torsion Motor Program" section of the command panel the motor program used for the simulations can be specified either as a mixture between Listing's law and L2 (see Chapter 7 for details) or as a custom motor program. The significance of each of the ten parameters is noted above the line showing their current value.

F.3 Simulation Details

The section labelled "Movement Range and Steps" specifies what the extreme gaze angles for the simulated range of gaze should be and how many intermediate angles should be calculated. This can be selected separately for the vertical Helmholtz coordinate of both eyes (vertical vergence is not a part of this simulation), the conjugate horizontal coordinate and the vergence coordinate. Both eyes' torsion at each of these simulated gaze positions is of course specified by the selections made in the "Torsion Motor Program" section of the panel.

It is also possible to change the properties of the epipolar segment computed for each eye position. The four parameters are the relative length of the segment as a percentage of its distance from the eye, the number of points on the segment (more points of course yielding a more accurate fit to the actual curvature of the segment), the number of points linearly interpolated between the ones actually simulated and the size of individual points when printed on the display. Since the computations are quite extensive, for exploration of the zones it is best to compute few points and have the point size relatively large. This will blur the outline of the search zones somewhat, but will greatly increase the speed of computations.

F.4 Exporting

The program also has the facility to export the overall size of the search zones for the set of retinal locations chosen, for a segment of parameter space. This can either be a one-dimensional cut along a parameter axis, specified by the base and a direction parameter vector, of a two-dimensional cut, specified by the base and two vectors. One-dimensional cuts will be displayed online and can optionally be saved to an external file, two-dimensional cuts have to be exported.

To make a two-dimensional cut through parameter space, including Listing's law and L2, the base should be left all zero, the first vector should have -0.5 in the second to last position, corresponding to Listing's law, and the second vector should additionally have 0.5 in the position before that, corresponding to L2.

The area is computed by putting a fine two-dimensional grid over the retina and counting the number of grid elements containing a dot.

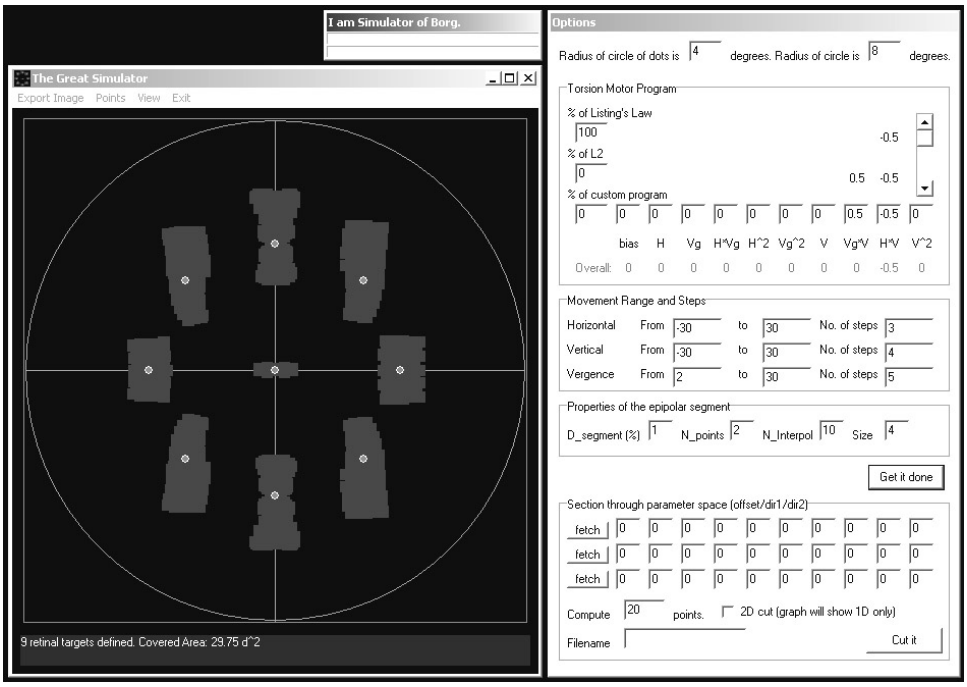


Figure F.1: Interface of the retinal search zone simulation program. In the panel on the right the program's parameters can be set. The window on the left will display the search zones.

The simulations in Chapter 7 were not done using this program, but with a Matlab routine described in the chapter.

Acknowledgments

I have received lots of help and advice during the work on this dissertation, for which I am very grateful.

For help with the experiments involving coil measurements I want to thank Dr. J. Douglas Crawford, Denise Henriques, Dr. James Sharpe and Hongying Wang.

I thank Dr. Matthias Niemeier for inspiring discussions on the geometric problems of binocular disparity.

The members of my advisory committee, Dr. Peter Hallett, Dr. Hon Kwan, Dr. R. David Tomlinson and Dr. James Winslow have been a great source of advise and feedback, bringing to my attention problems that would have gone unnoticed otherwise and guiding the progress of this work.

My greatest debt is to my supervisor Dr. Douglas Tweed, who has invited me to do this exciting project in his lab, and who's style of work and scientific insight have inspired and energized me. Dr. Tweed is an excellent teacher, an outstanding scientist and a pleasure to interact with. He has made this dissertation project and my stay in Canada a great experience both professionally and personally.

I would also like to thank my wife, Caroline Härdter, for bearing with me and backing me up in my strange adventures in the land of binocular vision.

Index

References

Adams, W, Frisby, JP, Buckley, D, Garding, J, Hippisley-Cox, SD, & Porrill, J. 1996. Pooling of vertical disparities by the human visual system. *Perception*, **25**(2), 165–76.

Aguilon, Francois D'. 1613. *Opticorum libri sex.* Widows and Sons of Jan Moretus, Antwerp.

Alhazen. 1989. Book of Optics. *In:* Sabra, A. I. (ed), *The optics of Ibn Al-Haytham.* Warburg Institute, University of London.

Allen, MJ. 1954. The dependence of cyclophoria on convergence, elevation and the system of axes. *Am J Optom*, **31**, 297–307.

Allen, NPL. 1998. *The Turin shroud and the crystal lens. Testament to a Lost Technology.* Empowerment technologies, South Africa.

Anzai, A, Ohzawa, I, & Freeman, RD. 1999a. Neural mechanisms for encoding binocular disparity: receptive field position versus phase. *J Neurophysiol*, **82**(2), 874–90.

Anzai, A, Ohzawa, I, & Freeman, RD. 1999b. Neural mechanisms for processing binocular information I. Simple cells. *J Neurophysiol*, **82**(2), 891–908.

Anzai, A, Ohzawa, I, & Freeman, RD. 1999c. Neural mechanisms for processing binocular information II. Complex cells. *J Neurophysiol*, **82**(2), 909–24.

Arditi, A. 1982. The dependence of the induced effect on orientation and a hypothesis concerning disparity computations in general. *Vision Res*, **22**(2), 247–56.

Arditi, A, Kaufman, L, & Movshon, JA. 1981. A simple explanation of the induced size effect. *Vision Res*, **21**(6), 755–64.

Aschenbrenner, Claus M. 1954. Problems in getting information into and out of air photographs. *Photogrammetric Engineering*, **20**, 398–401.

Backus, BT, Banks, MS, van Ee, R, & Crowell, JA. 1999. Horizontal and vertical disparity, eye position, and stereoscopic slant perception. *Vision Research*, **39**(6), 1143–70.

Backus, BT, Fleet, DJ, Parker, AJ, & Heeger, DJ. 2001. Human cortical activity correlates with stereoscopic depth perception. *J Neurophysiol*, **86**(4), 2054–68.

Baere, JI. 1931. Parva naturalia. De Somniis. *Pages 461b–462a of:* Ross, W. D. (ed), *The works of Aristotle*. Oxford University Press, London.

Banks, MS, Backus, BT, & Banks, RS. 2002. Is vertical disparity used to determine azimuth? *Vis Res*, **42**, 801–7.

Barlow, HB, Blakemore, C, & Pettigrew, JD. 1967. The neural mechanism of binocular depth discrimination. *J Physiol (Lond)*, **193**(2), 327–42.

Beare, JI. 1906. *Greek theories of elementary congition from Alcmaeon to Aristotle*. Clarendon Press, Oxford.

Bergua, A, & Skrandies, W. 2000. An early antecedent to modern random dot stereograms –'the secret stereoscopic writing' of Ramon y Cajal. *Int J Psychophysiol*, **36**(1), 69–72.

Birch, EE, Gwiazda, J, & Held, R. 1982. Stereoacuity development for crossed and uncrossed disparities in human infants. *Vision Res*, **22**(5), 507–13.

Braddick, O, Atkinson, J, Julesz, B, & Kropfl, W. 1980. Cortical binocularity in infants. *Nature*, **288**, 363–65.

Brenner, E, Smeets, JB, & Landy, MS. 2001. How vertical disparities assist judgements of distance. *Vision Res*, **41**(25-26), 3455–65.

Brewster, D. 1844a. On the knowledge of distance given by binocular vision. *Transactions of the Royal Society of Edinburgh*, **15**, 663–74.

Brewster, D. 1844b. On the law of visible position in single and binocular vision, and on the representation of solid figures by the union of dissimilar plane pictures in the retina. *Transactions of the Royal Society of Edinburgh*, **15**, 349–68.

Buchert, M, Greenlee, MW, Rutschmann, RM, Kraemer, FM, Luo, F, & Hennig, J. 2002. Functional magnetic resonance imaging evidence for binocular interactions in human visual cortex. *Exp Brain Res*, **145**(3), 334–9.

Bürger, Gottfried August. 1786. *Wunderbare Reisen zu Wasser und Lande, Feldzüge und Lustige Abenteuer des Freiherrn von Münchhausen.* Dietrich, Göttingen.

Busettini, C, Fitzgibbon, EJ, & Miles, FA. 2001. Short-latency disparity vergence in humans. *J Neurophysiol*, **85**(3), 1129–52.

Carpenter, RHS (ed). 1991. *Eye movements.* Vision and visual dysfunction, vol. 8. CRC Press Inc., Boca Raton, Ann Arbor, Boston.

Crone, RA, & Everhard-Halm, Y. 1975. Optically induced eye torsion. *Albrecht Von Graefes Arch Klin Exp Ophthalmol*, **195**(4), 231–9.

Cumming, BG. 2002. An unexpected specialization for horizontal disparity in primate primary visual cortex. *Nature*, **418**(6898), 633–6.

Cumming, BG, & DeAngelis, GC. 2001. The physiology of stereopsis. *Annu Rev Neurosci*, **24**, 203–38.

Cumming, BG, & Parker, AJ. 1997. Responses of primary visual cortical neurons to binocular disparity without depth perception. *Nature*, **389**(6648), 280–3.

Cumming, BG, & Parker, AJ. 1999. Binocular neurons in V1 of awake monkeys are selective for absolute, not relative, disparity. *J Neurosci*, **19**(13), 5602–18.

Cumming, BG, & Parker, AJ. 2000. Local disparity not perceived depth is signaled by binocular neurons in cortical area V1 of the Macaque. *J Neurosci*, **20**(12), 4758–67.

Cumming, BG, Johnston, EB, & Parker, AJ. 1991. Vertical disparities and perception of three-dimensional shape. *Nature*, **349**(6308), 411–3.

Demer, JL. 2002. The orbital pulley system: a revolution in concepts of orbital anatomy. *Ann N Y Acad Sci*, **956**(Apr), 17–32.

Demer, JL, Miller, JM, Poukens, V, Vinters, HV, & Glasgow, BJ. 1995. Evidence for fibromuscular pulleys of the recti extraocular muscles. *Invest Ophthalmol Vis Sci*, **36**(6), 1125–36.

Demer, JL, Oh, SY, & Poukens, V. 2000. Evidence for active control of rectus extraocular muscle pulleys. *Invest Ophthalmol Vis Sci*, **41**(6), 1280–90.

Derrington, AM, & Lennie, P. 1984. Spatial and temporal contrast sensitivities of neurones in lateral geniculate nucleus of macaque. *J Physiol*, **357**(Dec), 219–40.

Dobson, Jessie. 1969. *John Hunter*. E & S Livingstone, Edinburgh and London.

Erkelens, CJ, & Collewijn, H. 1985. Motion perception during dichoptic viewing of moving random-dot stereograms. *Vision Res*, **25**(4), 583–8.

Erkelens, CJ, & Collewijn, H. 1991. Control of vergence: gating among disparity inputs by voluntary target selection. *Exp Brain Res*, **87**(3), 671–8.

Erkelens, CJ, & van Ee, R. 1998. A computational model of depth perception based on headcentric disparity. *Vision Res*, **38**(19), 2999–3018.

Erwin, E, & Miller, KD. 1999. The subregion correspondence model of binocular simple cells. *J Neurosci*, **19**(16), 7212–29.

Euclid. 300 BC/1945. Optics. *J Opt Soc Am*, **35**, 357–72. Translated by H. E. Burton.

Ewald, JR, & Gross, O. 1906. Über Stereoskopie und Pseudoskopie. *Pflügers Archiv*, **115**, 514–21.

Farell, B. 1998. Two-dimensional matches from one-dimensional stimulus components in human stereopsis. *Nature*, **395**(6703), 689–93.

Feynman, RP. 1963-65. *Feynman lectures in physics*. Addison-Wesley Pub. Co., Reading, Mass.

Fick, Adolf. 1854. Die Bewegungen des menschlichen Augapfels. *Zeitschrift für rationelle Medizin*, **4**, 109–28.

Finger, S. 1994. *Origins of Neuroscience*. Oxford University Press, New York.

Fisher, MH. 1922. Beiträge und kritische Studien zur Heterophoriefrage auf Grund systematischer Untersuchungen. *Alb v Graefes Arch f Ophthalmol*, **108**, 251–84.

Fry, Glenn A, & Hill, W W. 1962. The center of rotation of the eye. *American Journal of Optometry*, **39**(11), 581–95.

Gabor, D. 1948. A new microscopic principle. *Nature*, **161**, 777–8.

Galen. 1968. *De usu partium corporis humani*. Cornell University Press, Ithaca, NY. Translated by M. T. May.

Gårding, J, Porrill, J, Mayhew, JE, & Frisby, JP. 1995. Stereopsis, vertical disparity and relief transformations. *Vision Research*, **35**(5), 703–22.

Gillam, B, & Lawergren, B. 1983. The induced effect, vertical disparity, and stereoscopic theory. *Percept Psychophys*, **34**(2), 121–30.

Goethe, Johann Wolfgang von. 1952. *Faust: parts one and two. Translated by George Madison Priest*. Encyclopaedia Britannica, Chicago.

Gonzalez, F, & Perez, R. 1998. Neural mechanisms underlying stereoscopic vision. *Prog Neurobiol*, **55**(3), 191–224.

Granrud, CE. 1986. Binocular vision and spatial perception in 4- and 5-month-old infants. *J Exp Psychol Hum Percept Perform*, **12**(1), 36–49.

Graves, Robert P. 1882. *Life of Sir William Rowan Hamilton*. Hodges, Figgis and Co., Dublin.

Gwiazda, J, Bauer, J, Thorn, F, & Held, R. 1989. Stereoacuity development for crossed and uncrossed disparities in children. *Invest Ophthalmol Vis Sci Suppl*, **30**, 313.

Hammond, JH. 1981. *The camera obscura. A chronicle.* Adam Hilger Ltd, Bristol.

Harris, Joseph. 1775. *A treatise of Optics.* B. White, London.

Held, Richard. 1991. Development of Binocular Vision and Stereopsis. *Chap. 9, pages 170–8 of:* Regan, D (ed), *Binocular Vision.* Vision and visual dysfunction, vol. 9. CRC Press Inc., Boca Raton.

Helmholtz, Hermann von. 1863. Über die normalen Bewegungen des menschlichen Auges. *Arch. Ophtalmol.*, **IX**, 153–214.

Helmholtz, Hermann von. 1867. *Handbuch der physiologischen Optik.* Voss, Hamburg.

Hering, KEK. 1868. *Die Lehre vom binokularen Sehen.* Engelmann, Leipzip.

Hofmann, F.B., & Bielschowsky, A. 1900. Über die der Willkür entzogenen Fusionsbewegungen der Augen. *Arch. f. d. ges. Physiol.*, **80**, 1–40.

Hooge, IT, & van den Berg, AV. 2000. Visually evoked cyclovergence and extended Listing's Law. *J Neurophysiol*, **83**(5), 2757–75.

Hooten, K, Myers, E, Worrall, R, & Stark, L. 1979. Cyclovergence: the motor response to cyclodisparity. *Albrecht Von Graefes Arch Klin Exp Ophthalmol*, **210**(1), 65–8.

Horn, Berthold K.P. 1990. Relative Orientation. *International Journal of Computer Vision*, **4**, 59–78.

Howard, IP, & Rogers, BJ. 1995. *Binocular vision and stereopsis.* Oxford University Press, Oxford.

Howard, IP, Sun, L, & Shen, X. 1994. Cycloversion and cyclovergence: the effects of the area and position of the visual display. *Exp Brain Res*, **100**(3), 509–14.

Howard, IP, Allison, RS, & Zacher, JE. 1997. The dynamics of vertical vergence. *Exp Brain Res*, **116**(1), 153–9.

Howard, IP, Fang, X, Allison, RS, & Zacher, JE. 2000. Effects of stimulus size and eccentricity on horizontal and vertical vergence. *Exp Brain Res*, **130**(2), 124–32.

Hubel, DH, & Wiesel, TN. 1959. Receptive fields of single neurones in the cat's visual cortex. *Journal of Physiology*, **148**, 574–91.

Hubel, DH, & Wiesel, TN. 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, **160**, 106–54.

Hueck, Alexander Friedrich von. 1838. *Die Achsendrehungen des Auges.* Kluge, Dorpat.

Hunter, John. 1786. *Observations on certain parts of the Animal Oeconomy.* London.

Iwami, T, Nishida, Y, Hayashi, O, Kimura, M, Sakai, M, Kani, K, Ito, R, Shiino, A, & Suzuki, M. 2002. Common neural processing regions for dynamic and static stereopsis in human parieto-occipital cortices. *Neurosci Lett*, **327**(1), 29–32.

Judge, Stuart. 1991. Vergence. *Chap. 7, pages 157–72 of:* Carpenter, RHS (ed), *Eye movements.* Vision and visual dysfunction, vol. 8. CRC Press Inc, Boca Raton.

Julesz, Bela. 1960. Binocular depth perception of computer-generated patterns. *Bell Syst Tech J*, **34**(5), 1125–63.

Julesz, Bela. 1964. Binocular depth perception without familiarity cues. *Science*, **145**, 356–62.

Kapoula, Z, Bernotas, M, & Haslwanter, T. 1999. Listing's plane rotation with convergence: role of disparity, accommodation, and depth perception. *Exp Brain Res*, **126**(2), 175–86.

Kepler, Johannes. 1604. *Ad Vitellionem Paralipomena.* C. Marnium, Frankfurt.

Kepler, Johannes. 1611. *Dioptrice.* Augsburg.

Kertesz, AE. 1972. The effect of stimulus complexity on human cyclofusional responses. *Vision Research*, **12**, 699–704.

Kertesz, AE, & Jones, RW. 1970. Human cyclofusional response. *Vision Research*, **10**(9), 891–6.

Koenderink, JJ. 1986. Optic flow. *Vision Res*, **26**(1), 161–79.

Kontsevich, LL, & Tyler, CW. 1999. Bayesian adaptive estimation of psychometric slope and threshold. *Vision Research*, **39**(16), 2729–37.

Kontsevich, LL, & Tyler, CW. 2000. Relative contributions of sustained and transient pathways to human stereoprocessing [In Process Citation]. *Vision Research*, **40**(23), 3245–55.

Krol, JD, & van de Grind, WA. 1980. The double-nail illusion: experiments on binocular vision with nails, needles, and pins. *Perception*, **9**(6), 651–69.

Kuhn, Thomas S. 1996. *The structure of scientific revolutions*. University of Chicago Press.

Land, MF. 1999. Motion and vision: why animals move their eyes. *J Comp Physiol [A]*, **185**(4), 341–52.

LeGrand, Yves, & ElHage, Sami G. 1980. *Physiological optics*. Berlin: Springer.

Leight, RJ, & Zee, DS. 1991. *The neurology of eye movements*. 2nd edn. F.A. Davis Company, Philadelphia.

Lindberg, DC. 1976. *Theories of vision from Al-Kindi To Kepler*. University of Chicago Press, Chicago.

Listing, JB. 1845. *Beitrag zur physiologischen Optik*. Vandenhoeck und Ruprecht, Göttingen.

Liu, L, Stevenson, SB, & Schor, CM. 1994. A polar coordinate system for describing binocular disparity. *Vision Res*, **34**(9), 1205–22.

Livingstone, M, & Hubel, D. 1988. Segregation of form, color, movement, and depth: anatomy, physiology, and perception. *Science*, **240**(4853), 740–9.

Livingstone, MS, & Hubel, DH. 1987. Psychophysical evidence for separate channels for the perception of form, color, movement, and depth. *J Neurosci*, **7**(11), 3416–68.

Marr, D, & Poggio, T. 1979. A computational theory of human stereo vision. *Proc R Soc Lond B Biol Sci*, **204**(1156), 301–28.

Mayhew, JE, & Longuet-Higgins, HC. 1982. A computational model of binocular depth perception. *Nature*, **297**(5865), 376–8.

Mayhew, John E.W., & Frisy, John P. 1981. Psychophysical and Computational Studies towards a Theory of Human Stereopsis. *Artificial Intelligence*, **17**, 349–85.

Minken, AW, & van Gisbergen, JA. 1994. A three-dimensional analysis of vergence movements at various levels of elevation. *Exp Brain Res*, **101**(2), 331–45.

Misslisch, H, Tweed, D, & Hess, BJ. 2001. Stereopsis outweighs gravity in the control of the eyes. *J Neurosci*, **21**(3), RC126(1–5).

Mok, D, Ro, A, Cadera, W, Crawford, JD, & Vilis, T. 1992. Rotation of Listing's plane during vergence. *Vision Res*, **32**(11), 2055–64.

Müller, Johannes. 1826. *Zur vergleichenden Physiologie des Gesichtssinnes des Menschen und der Thiere*. Leipzig.

Nagel, A. 1868. Über das Vorkommen von wahren Rollungen des Auges um die Gesichtslinie. *Arch. f. Ophth.*, **14**, 228–46. pt. 2.

Nikara, T, Bishop, PO, & Pettigrew, JD. 1968. Analysis of retinal correspondence by studying receptive fields of binocular single units in cat striate cortex. *Exp Brain Res*, **6**(4), 353–72.

Nunes, Pedro. 1542. *De Crepusculis liber unus*. Rodrigues, Lissabon.

Ogle, Kenneth N. 1950. *Researches in Binocular Vision*. Hafner.

Ogle, Kennth N., & Ellerbrock, Vincent J. 1946. Cyclofusional movements. *Archives of Ophthalmology*, **36**, 700–35.

Ogle, KN. 1938. Induced size effect. I A new phenomenon on binocular space perception associated with the images of the two eyes. *Arch Ophthalmol*, **20**, 604–23.

Ogle, KN. 1939. Induced size effect. II An experimental study of the phenomenon with restricted fusion contours. *Arch Ophthalmol*, **21**, 604–25.

Ohzawa, I. 1998. Mechanisms of stereoscopic vision: the disparity energy model. *Curr Opin Neurobiol*, **8**(4), 509–15.

Ohzawa, I, & Freeman, RD. 1986a. The binocular organization of complex cells in the cat's visual cortex. *J Neurophysiol*, **56**(1), 243–59.

Ohzawa, I, & Freeman, RD. 1986b. The binocular organization of simple cells in the cat's visual cortex. *J Neurophysiol*, **56**(1), 221–42.

Ohzawa, I, DeAngelis, GC, & Freeman, RD. 1990. Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors. *Science*, **249**(4972), 1037–41.

Palmisano, S, Allison, RS, & Howard, IP. 2001. Effects of horizontal and vertical additive disparity noise on stereoscopic corrugation detection. *Vision Res*, **41**(24), 3133–43.

Panum, PL. 1858. *Physiologische Untersuchungen "uber das Sehen mit zwei Augen*. Schwers, Kiel.

Park, R S, & Park, G E. 1933. The center of ocular rotation in the horizontal plane. *American Journal of Physiology*, **104**, 545–51.

Platter, Felix. 1583. *De corporis humani structura et usu*. Basel.

Poggio, GE. 1995. Mechanisms of stereopsis in monkey visual cortex. *Cereb Cortex*, **5**(3), 193–204.

Popper, Karl R. 1935. *Logik der Forschung*. Julius Springer Verlag, Vienna.

Porrill, J, Frisby, JP, Adams, WJ, & Buckley, D. 1999a. Robust and optimal use of information in stereo vision. *Nature*, **397**(6714), 63–6.

Porrill, J, Ivins, JP, & Frisby, JP. 1999b. The variation of torsion with vergence and elevation. *Vision Res*, **39**(23), 3934–50.

Prince, SJ, Pointon, AD, Cumming, BG, & Parker, AJ. 2000. The Precision of Single Neuron Responses in Cortical Area V1 during Stereoscopic Depth Judgments. *J Neurosci*, **20**(9), 3387–3400.

Quaia, Christian, & Optican, Lance M. 1998. Commutative saccadic generator is sufficient to control a 3-D ocular plant with pulleys. *Journal of neurophysiology*, **79**(6), 3197–215.

Ramon y Cajal, Santiago. 1901. Recreaciones estereoscopicas y binoculares. *La Fotografia*, **27**, 41–8.

Ramon y Cajal, Santiago de. 1995. *Histology of the nervous system of man and vertebrates*. Oxford University Press, New York.

Raphan, T. 1997. Modeling control of eye orientation in three dimensions. *Pages 359–374 of:* Fetter, M, Haslwanter, T, Misslisch, H, & Tweed, D (eds), *Three dimensional Kinematics of Eye, Head and Limb movement*. Harwood Academic, The Netherlands.

Raphan, Theodore. 1998. Modeling control of eye orientation in three dimensions. I. Role of muscle pulleys in determining saccadic trajectory. *Journal of neurophysiology*, **79**(5), 2653–67.

Rashbass, C, & Westheimer, G. 1961. Disjunctive eye movements. *J Physiol (Lond)*, **159**, 339–360.

Regan, D, Erkelens, CJ, & Collewijn, H. 1986. Necessary conditions for the perception of motion in depth. *Invest Ophthalmol Vis Sci*, **27**(4), 584–97.

Robinson, David A. 1963. A method of measuring eye movement using a scleral search coil in a magnetic field. *IEEE Transactions on Biomedical Electronics*, **10**, 137–45.

Rogers, B, & Koenderink, J. 1986. Monocular aniseikonia: a motion parallax analogue of the disparity-induced effect. *Nature*, **322**(6074), 62–3.

Rogers, BJ, & Bradshaw, MF. 1993. Vertical disparities, differential perspective and binocular stereopsis. *Nature*, **361**(6409), 253–5.

Rogers, BJ, & Bradshaw, MF. 1996. Does the visual system use the epipolar constraint for matching binocular images? *Investigative Ophthalmology & Visual Sciences (Supplement)*, **37**(3), 684.

Ruete, CGT. 1846; 2nd edition, Vol. 1, 1853. *Lehrbuch der Ophthalmologie*. Braunschweig.

Salisbury, John of. 1955. *Metalogicon*. University of California Press. Translated with an introduction and notes by Daniel D. McGarry.

Scheiner, Christoph. 1619. *Oculus, hoc est fundamentum opticum*. Innsbruck.

Schiller, PH, Logothetis, NK, & Charles, ER. 1991. Parallel pathways in the visual system: their role in perception at isoluminance. *Neuropsychologia*, **29**(6), 433–41.

Schor, CM, & McCandless, J. 2002. Adaptive Control of Vergence in Humans. *Ann N Y Acad Sci*, **956**(Apr), 297–305.

Schor, CM, Maxwell, JS, & Graf, EW. 2001. Plasticity of convergence-dependent variations of cyclovergence with vertical gaze. *Vision Res*, **41**(25-26), 3353–69.

Schreiber, K, Crawford, JD, Fetter, M, & Tweed, D. 2001. The motor side of depth vision. *Nature*, **410**(6830), 819–22.

Schubert, G. 1927. Studien über das Listingsche Bewegungsgesetz am Auge. *Pflügers Arch Physiol*, **215**, 553–87.

Siderov, J, Harwerth, RS, & Bedell, HE. 1999. Stereopsis, cyclovergence and the backwards tilt of the vertical horopter. *Vision Res*, **39**(7), 1347–57.

Simonsz, HJ, & den Tonkelaar, I. 1990. 19th century mechanical models of eye movements, Donder' law, Listing's law and Helmholtz' direction circles. *Documenta Ophthalomologica*, **74**, 95–112.

Steffen, H, Walker, MF, & Zee, DS. 2000. Rotation of Listing's plane with convergence: independence from eye position. *Invest Ophthalmol Vis Sci*, **41**(3), 715–21.

Steffen, H, Walker, M, & Zee, DS. 2002. Changes in Listing's plane after sustained vertical fusion. *Invest Ophthalmol Vis Sci*, **43**(3), 668–72.

Stenton, SP, Frisby, JP, & Mayhew, JE. 1984. Vertical disparity pooling and the induced effect. *Nature*, **309**(5969), 622–3.

Sullivan, MJ, & Kertesz, AE. 1978. Peripheral stimulation and human cyclofusional response. *Invest. Ophthalmol. Vis. Sci.*, **18**, 1287–91.

Takagi, M, Trillenberg, P, & Zee, DS. 2001. Adaptive control of pursuit, vergence and eye torsion in humans: basic and clinical implications. *Vision Res*, **41**(25-26), 3331–44.

Theophrastus. 1917. On the senses. *In:* Stratton, G. M. (ed), *Theophrastus and Greek Physiological Psychology before Aristotle*. Macmillan, New York.

Thomas, OM, Cumming, BG, & Parker, AJ. 2002. A specialization for relative disparity in V2. *Nat Neurosci*, **5**(5), 472–8.

Tweed, D. 1997a. Visual-motor optimization in binocular control. *Vision Research*, **37**(14), 1939–51.

Tweed, DB, Haslwanter, TP, Happe, V, & Fetter, M. 1999. Non-commutativity in the brain. *Nature*, **399**(6733), 261–3.

Tweed, Douglas. 1997b. A three-dimensional model of the human eye-head saccadic system. *Journal of neurophysiology*, **77**, 654–66.

Tweed, Douglas, & Vilis, Tutis. 1987. Implications of rotational kinematics for the oculomotor system in three dimensions. *Journal of neurophysiology*, **58**, 832–49.

Tweed, Douglas, Haslwanter, Thomas, & Fetter, Michael. 1998. Optimizing gaze control in three dimensions. *Science*, **281**, 1363–6.

Tyler, Christopher. 1991. The Horopter and Binocular Fusion. *Chap. 2, pages 19–37 of:* Regan, D. (ed), *Binocular Vision*. Vision and Visual Dysfunction, vol. 9. CRC Press Inc., Boca Raton, Ann Arbor, Boston.

Tyler, CW, & Clarke, MB. 1990. The autostereogram. *Pages 182–97 of:* Fisher, SS, & Merritt, JO (eds), *Stereoscopic Displays and Applications*, vol. 1256. International Society for Optical Engineering.

van Rijn, LJ, & van den Berg, AV. 1993. Binocular eye orientation during fixations: Listing's law extended to include eye vergence. *Vision Res*, **33**(5-6), 691–708.

Vieth, GUA. 1818. Über die Richtung der Augen. *Annalen der Physik*, **28**(3), 233–53.

Volkmann, Alfred Wilhelm. 1836. *Neue Beiträge zur Physiologie des Gesichtssinnes*. Leipzig.

Volkmann, AW. 1859. Die stereoskopischen Erscheinungen in ihrer Beziehung zu der Lehre von den identischen Netzhautpunkten. *Albrecht v. Graefes Arch Ophthal*, **5**(2), 1–100.

von Gudden, JBA. 1874. Ueber die Kreuzung der Fasern in Chiasma nervorum optici. *Archiv für Ophtalmologie*, **20**(2. Abth.), 249–68.

Wade, Nicholas. 1998. *A Natural History of Vision*. MIT Press, Cambridge, Massachusetts.

Westheimer, G. 1957. Kinematics of the eye. *J Opt Soc Am*, **47**, 967–74.

Westheimer, G. 1978. Vertical disparity detection: is there an induced size effect? *Invest Ophthalmol Vis Sci*, **17**(6), 545–51.

Westheimer, G. 1979. Cooperative neural processes involved in stereoscopic acuity. *Exp Brain Res*, **36**(3), 585–97.

Wheatstone, C. 1838. On some remarkable, and hitherto unobserved, Phenomena of Binocular Vision. *Phil Trans R Soc A*, 371–94.

Wheatstone, C. 1852. On some remarkable, and hitherto unobserved, Phenomena of Binocular Vision (continued). *Phil Trans Roy Soc*, **142**, 1–17.

Wigner, Euegne. 1960. The unreasonable effectiveness of mathematics in the natural sciences. *Communications in Pure and Applied Mathematics*, **13**(1), 1–14.

Zee, David S, & Leigh, R John. 1983. *The Neurology of Eye Movements*. Contemporary neurology series, vol. 23. Philadelphia: F.A. Davis Company.

Stereopsis is the reconstruction of depth from slight differences in the two retinal images. The first step in stereopsis is stereo-matching, the locating of corresponding visual features in the two eyes' retinal images. Stereo-matching is a demanding task, but it could be simplified by exploiting a geometric constraint: matching features are constrained to retinal bands called epipolar lines.

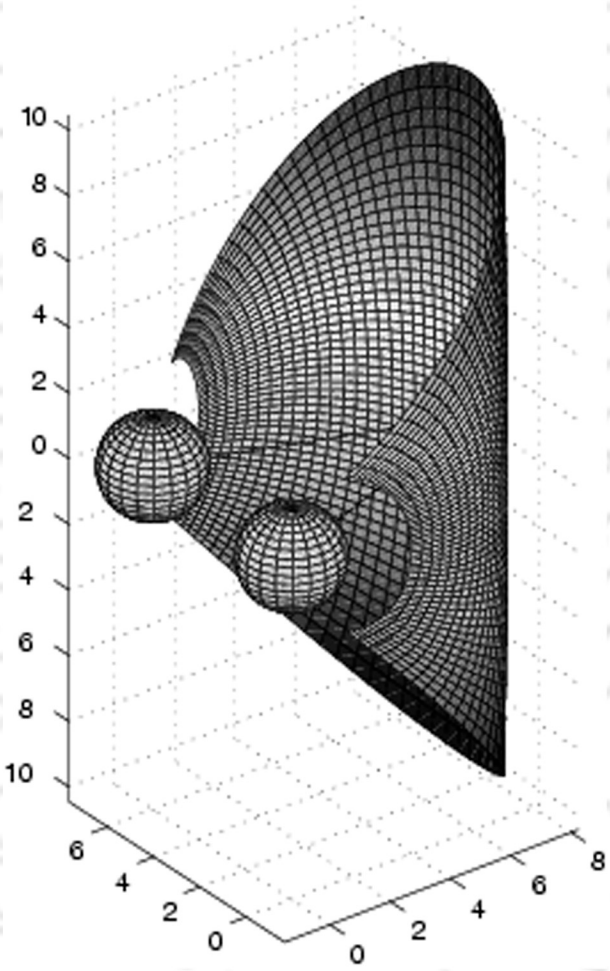
The epipolar lines slide on the retina when the eyes move, so eye position information is necessary if the epipolar constraint is to be used. Using random-dot stereograms it is shown that human stereo-matching does not shift retinal search regions along with the moving epipolar lines, but searches large retina-fixed two-dimensional retinal patches instead.

The size of these patches depends on the pattern of eye movements. In simulations of eye movement effects on the search zone size it is demonstrated that human eye movement follows a movement pattern that reduces the size of these search zones. This movement pattern is a compromise between motor advantages and the demands of stereopsis.

It is also shown that the visual system could use the epipolar constraint with retina-fixed search regions by using visual information alone to compute the location of the epipolar lines. This process is called visual bootstrapping. Psychophysical data are presented showing that the visual system does use visual bootstrapping.

The fusional horopter is the region of space where objects are seen single. A mathematical extension of the classical theoretical point-horopter is presented that allows the computation of the point-horopter as a full two-dimensional surface.

This dissertation shows that it is necessary to study the interaction of sensory and motor systems in order to fully understand the purpose of each. It uses the interaction of stereopsis and the oculomotor system as an example.



ISBN 3-930171-28-7
Genista
Tübingen - Toronto