J Math Imaging Vis (2008) 31: 171–187
DOI 10.1007/s10851-008-0076-3

The Structure of Visual Spaces

Jan Koenderink · Andrea van Doorn

Published online: 12 April 2008

Abstract The "visual space" of an optical observer situated at a single, fixed viewpoint is necessarily very ambiguous. Although the structure of the "visual field" (the lateral dimensions, *i.e.*, the "image") is well defined, the "depth" dimension has to be inferred from the image on the basis of "monocular depth cues" such as occlusion, shading, *etc.* Such cues are in no way "given", but are guesses on the basis of prior knowledge about the generic structure of the world and the laws of optics. Thus such a guess is like a hallucination that is used to tentatively interpret image structures as depth cues. The guesses are successful if they lead to a coherent interpretation. Such "controlled hallucination" (in psychological terminology) is similar to the "analysis by synthesis" of computer vision. Although highly ambiguous, visual spaces do have geometrical structure. The group of ambiguities left open by the cues (*e.g.*, the well known bas-relief ambiguity in the case of shape from shading) may be interpreted as the group of congruences (proper motions) of the space. The general structure of visual spaces for different visual fields is explored in the paper. Applications include improved viewing systems for optical man-machine interfaces.

Keywords Visual space · Human perception · Isotropic metric · Visual field · Parallactic optical structure · Panoramic vision · Projective structure · Riemann metrics · Ground plane · Ambiguity groups · Pictorial space · Stereopsis

J. Koenderink (✉) · A. van Doorn
Buys-Ballot Laboratory, Princetonplein 5, Utrecht 3584 CC, Netherlands
e-mail: j.j.koenderink@phys.uu.nl

1 Introduction

With "structure" we intend "geometrical structure", to be understood as the conventional "projective", "affine", "metrical", *etc.*, type of structures, much in the sense of Felix Klein's [40] *Erlangen Programm.* "Visual spaces" is used in the plural because there indeed exist categorically distinct entities often denoted as "visual space". In this paper we (at least) distinguish entities alternately known as "optic array" (or "viewing sphere"), "visual field", "visual space" proper (the space of optical objects experienced with your eyes open in normal conditions) and "pictorial space". We focus the discussion on biological vision since most of the relevant literature derives from that, but there is little doubt that similar issues are bound to arise in computer vision and—especially—in image based human interfaces.

The term "optic array" is of a physical nature. Physicists simply speak of "radiance" (since it is the standard term [10] in the exact sciences this is much to be recommended), though in artificial intelligence one speaks of the "plenoptic function [2]". It describes the photon number flux per solid angle (perhaps also per temporal and per spectral interval) for any direction as seen from a given vantage point (or perhaps as seen from any vantage point within a certain volume, on a certain surface, on a certain curve, or on a number of distinct points). "Viewing sphere [5]" sometimes simply stands for the space of visual directions ($\mathbb{S}^2$), sometimes for radiance. The term "optic array" was introduced by Gibson [24], who wrestled with its definition for years. In the final instance, Gibson's optic array became—at least in our interpretation—effectively the radiance.

The "visual field" denotes the optical structure as perceived in any direction from a given vantage point, either with or without eye movements. It describes a thread of experience in the presence of certain optical stimulation (read:

radiance). The visual field has a geometrical structure, that is at least a topological structure, but is pre-categorical, thus somewhat of an unreachable fringe of experience. As a visual artist in the realist tradition one learns to attend to experiences in terms of the *visual field*, this is often known as "learning to see [36]". Most people find this difficult or even impossible. The visual field is limited (though its bounds are not visually apparent), of limited resolution (though this is rarely experienced), and has a topological structure that (for the human observer) amounts to the two dimensional disk $\mathbb{D}^2$. In most observers there is also a projective and even a metrical structure [27]. The topology is known as "local sign [52]" and is an ill-understood though very fundamental property of vision. That the existence of a topology is not trivial (as it is usually taken to be) is evident from the existence of atypical observers that lack a fully developed topology (an infrequent but not exceedingly rare condition termed *tarachopia* [32]). The topology can only be due to the correlation structure of optic nerve activity [30, 43], which derives from the overlap of retinal receptive fields. This is what Gibson [25] most likely meant with his "nested solid angles", though he never attempted a formal theory. In computer vision local sign is *a priori* settled as the structure of the pixel array.

"Visual space" is the geometrical structure of the scene in front of you as revealed through optically induced experience. Apart from being seen in a certain direction, things are seen at a certain "depth". You perceive an articulated three-dimensional space, with various degrees of translucency and luminosity, filled with generally opaque objects [28]. Notice that such terms as "luminous", "translucent", "opaque" are phenomenological descriptive terms, not physical entities or properties. Although you see only the frontal sides of opaque objects, you experience them as solids "in the round". You not only perceive geometrical properties, but also material properties, causal relations, agenthood, and so forth. Much of your perceptions are dominated by prior multimodal experiences of generally self induced or initiated interactions ("transactions" may be a better term) with your biotope. From this derive generic notions (immediately perceived) as "rigid body", Euclidean movement (congruence), and so forth. Thus "visual space" is only a limited facet of your actual experiences, already somewhat of an abstraction [59, 64]. Nevertheless, there exist theories of the structure of visual space and a continuing stream of empirical studies.

Current science is far from a formal or even phenomenological understanding of visual space. In artificial intelligence the issue still has to be addressed, in machine vision it is simply skipped. Of course the "depth" dimension of the visual space for a single perspective center in a static world is highly ambiguous. One may not expect the structure of visual spaces to be like Euclidean space, for instance, the depth and lateral ("image") dimensions are categorically different and may not be compared or mixed as a Euclidean rotation would allow one to do.

In this paper we develop *a priori* formal theories of visual space from scratch, trying to clarify their ontologies. We discuss their descriptive and unifying value through the review of empirical studies (no novel data are presented). In many cases we interpret these data in novel ways, not necessarily in accord with the context in which the data were originally presented.

2 Basic Theory

In this section we discuss certain fundamental concepts. They deal with optical sampling, the structure of visual fields of general optically guided agents, "cues" and their related ambiguity groups, and the role of the "beholder's share" in the autogenous generation of optically controlled "presentations".

2.1 Parallactic Optical Structure

Consider a purely "optical" sensor, without absolute directional references (*e.g.*, the sagittal plane, or the "vertical" and "frontal" directions). Let the sensor sample all directions with the same optical resolution and sampling density. Without an absolute reference you have only parallactic structure, that is to say, spatial structure referenced to itself. This is another instance where Gibson's [25] "nested solid angles" are relevant, it was Gibson's way to get rid of an absolute reference system, doing away with "space as a container", a thoroughly Leibnitzian view [4].

Next consider the groups of similarities about the vantage point and of the rotations about the vantage point. We will say that either type of transformation conserves "parallactic structure". Since arbitrary rotation-dilations about the vantage point conserve the parallactic structure they are visually irrelevant to the "parallactic observer". Notice that binocularity or translation of the vantage point induce transformations that violate the parallactic structure.

The metric of such a space is

$$\begin{aligned}
\mathrm{d}s^2 &= \frac{\mathrm{d}x^2+\mathrm{d}y^2+\mathrm{d}z^2}{x^2+y^2+z^2} \\
&= \frac{\mathrm{d}\rho^2}{\rho^2}+\mathrm{d}\vartheta^2+\sin^2\vartheta\,\mathrm{d}\varphi^2 \\
&= (\mathrm{d}\log\rho)^2+\mathrm{d}\vartheta^2+\sin^2\vartheta\,\mathrm{d}\varphi^2,
\end{aligned}$$

where $\{x, y, z\}$ denote Cartesian coordinates of $\mathbb{R}^3$, and $\{\rho, \vartheta, \varphi\}$ polar coordinates of $\mathbb{E}^3 - \{\mathbf{o}\}$ centered at the vantage point $\mathbf{o}$. Spherical symmetry and invariance with respect to scaling are obviously implemented. For a spherical shell about the vantage point you obtain the standard spherical geometry, for a visual direction you find that the half-ray

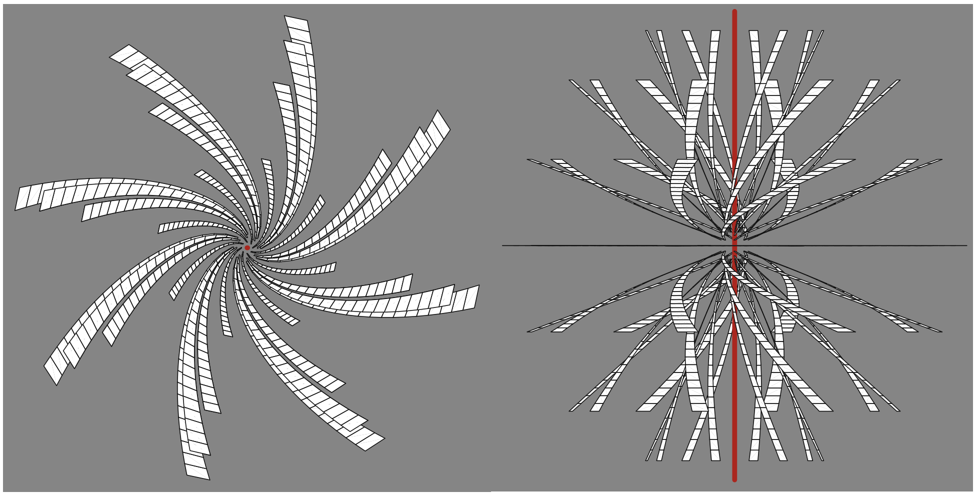


Fig. 1 Example of a Killing field (zero Lie derivative with respect to the metric; *i.e.*, the infinitesimal generators of the isometries), seen from two different directions. Here we have taken short segments in the φ-direction on the unit sphere and have "pushed them along" with the Killing field in order to form ribbons. This yields a visually intuitive representation of the field. The line is the "polar-axis", that is the $\vartheta = k\pi$ $(k = 0, 1)$ direction. This is also the axis of rotation in this example

at the vantage point is mapped on the full affine line (the origin being arbitrary) via a logarithmic transformation. Thus a radial distance between two points depends (only) on the ratio of their Euclidean distances. This is at it should be, there being no absolute scale of distance available [38].

For the metric tensor [66, 72]

$$g = \begin{pmatrix} \frac{1}{\rho^2} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \sin^2\vartheta \end{pmatrix},$$

the Christoffel symbols become

$$\Gamma^1_{11} = -\frac{1}{\rho}, \qquad \Gamma^2_{33} = -\sin\vartheta\cos\vartheta, \qquad \Gamma^3_{32} = \cos\vartheta,$$

(the Christoffel symbols are symmetric under interchange of the last two indices, only the independent components are displayed) and the Riemann tensor $R^\lambda_{\mu\nu\sigma}$

$$R^2_{332} = -\sin^2\vartheta, \qquad R^3_{232} = 1$$

(where one obtains R^1_{231} from R^1_{213} using the antisymmetry of the Riemann tensor under exchange of the last two indices). The scalar curvature is $+2$, thus constant and positive, the space is "elliptic" [33]. The geodesic equations are

$$\frac{\mathrm{d}u^1}{\mathrm{d}\tau} = \frac{u^1}{\rho}, \qquad \frac{\mathrm{d}u^2}{\mathrm{d}\tau} = \sin\vartheta\cos\vartheta\,(u^3)^2,$$

$$\frac{\mathrm{d}u^3}{\mathrm{d}\tau} = -2\cot\vartheta\, u^2 u^3,$$

where $\mathbf{u}(\tau) = u^1(\tau)\mathbf{e}_\varrho + u^2(\tau)\mathbf{e}_\vartheta + u^3(\tau)\mathbf{e}_\varphi$ is the "velocity". These equations are easily integrated. (A simpler procedure is to see that the structure of the metric forces the geodesics to be planar, restrict the metric to the equatorial plane and conclude that $\mathrm{d}s^2 = (\mathrm{d}\log\rho)^2 + (\mathrm{d}\vartheta)^2$ is the Euclidean line element.) The geodesics are planar logarithmic spirals [16] in planes through the origin. The squared geodesic distance between two points is the sum of squares of their angular separation and the logarithm of the ratio of their egocentric distances, thus there exists a global distance function.

Although the structure of this metric looks very simple (Fig. 1), the space generated by its geodesics fails to generate a global projective structure. It is not hard to show that the Pasch axiom [61] is violated: in a triangle *abc* the geodesics that connect vertex *a* with a point on the geodesic arc *bc* and vertex *b* to a point on the geodesic arc *ac* typically fail to intersect. Thus one cannot define "planes". It is hard to say whether this rules out the metric, the existing data are simply insufficient for that. We know of only a single experiment [23] that attempts to address the issue. The result is indecisive, but an empirical check appears to be just about feasible: For an equilateral spherical triangle with sides $\pi/2$, vertices at depths in the ratios $1:2:3$, the depths at the direction of the intersection of the arcs connecting the vertices to the midpoints of the opposite sides are in the ratios $1:1.07:1.15$, probably just enough to make an observable difference.

The space is almost certainly too general to describe the case of the stationary, monocular observer in a static world. The reason is that the parallactic structure still assumes that ratios of egocentric distances are valid data items. This cannot be the case for a "purely optical observer" though, because it assumes an independent depth gauge (optical range finder (binocularity), acoustic echo device, radar, *etc.*). "Depth" for this purely optical observer is only a "mental" (inferred rather than observed) entity, quite distinct from separation in the visual field, which is a physical entity, *e.g.*, reflected in separate retinal stimulations. In the next section we discuss formal methods to handle this.

2.2 Visual Fields

2.2.1 *Boundless, Panoramic Visual Field*

Consider an organism with a perfect panoramic eye, that is to say, all visual directions are sampled with the same optical resolution and sampling density. Next consider the groups of similarities about the vantage point and of the rotations about the vantage point. We submit that either type of transformation is irrelevant—with respect to probing for information—for vision. This is perhaps most obvious in the former case, because dilations and contractions about the vantage point not even affect the available optical structure. This is generally understood and played with in "Gulliver's Travels [77]" and "Alice in Wonderland [13]". Lilliput and Brobdingnag are optically the same until the appearance of Gulliver introduces an absolute length unit. Though perhaps less obvious in the case of rotations, a rotation about the vantage point can be cancelled by an eye movement or a change of (spherical) coordinates. Thus the only change induced by such a rotation is a rigid movement of the visual field with respect to an absolute reference. Such an absolute reference has to be of extra-optical origin (*e.g.*, the bilateral symmetry and frontal orientation of the body and the vestibular system [65]). In discussions of vision *per se* we may safely ignore such transformations. Arbitrary rotation-dilations about the vantage point are thus irrelevant. We are especially interested in such transformation that are of a merely mental (imagined) nature.

When transformations are merely mental, then the geometrical structure of the visual field must reflect the structure of the optic array. "Transformation" merely implies the adjustment of depth values along each visual ray individually. The points of visual space are like beads on a string in the sense that they cannot pass each other or break loose from the string, but are freely moveable along the strings. Here the "strings" are the visual rays, that are the points of the visual field or optic array (Fig. 3). This severely limits the possible transformations. It forces the depth dimension to be *isotropic*, because isotropic dimensions are conserved as a family under general similarities and individually under similarities that appear as identities in the visual field. (We will return to this issue in a later section.)

Thus we rewrite the metric as

$$\mathrm{d}s^2 = (\mathrm{d}\vartheta^2 + \sin^2\vartheta\,\mathrm{d}\varphi^2) + \varepsilon^2(\mathrm{d}\log\rho)^2,$$

where the nilpotent "dual number" $\varepsilon \neq 0$ with $\varepsilon^2 = 0$ forces the depth dimension to be isotropic [7, 79, 80]. This has important consequences discussed below. Notice that the violation of the Pasch axiom has been avoided in the sense that the "gap" is imaginary.

The mentally applied central rotation-dilations are to be regarded as the "movements" or "congruences" of "Visual Space $\mathbb{V}^3$". Following Felix Klein's [40] *Erlangen Programm* they define the geometrical structure of visual space.

2.2.2 *Panoramic, Horizontally Centered Visual Fields*

A horizontally centered visual field is organized about a great circle in the unit sphere of visual directions, the "horizon". The horizon corresponds to the horizontal plane at eye level. The extent is panoramic, that is to say, (almost) the full circle, though it typically makes sense to distinguish a "forward direction". The extent orthogonal to the horizon is considered limited, here we assume it to be infinitesimal. (If so desired the model can be extended to the finite elevation domain through Mercator projection [57] (the map $\{\vartheta, \varphi\} \mapsto \{\xi, \zeta\}$, with $\xi = \varphi, \zeta = \log(\tan(\frac{\pi}{2} - \frac{\vartheta}{2})))$.) Such a model applies to animals with a well developed "visual streak [21]", who effectively live in "Flatland" [1]. (See Fig. 2.)

Since we assume the vertical extent to be infinitesimal, we will concentrate on the horizon itself. Although the vertical extent does not figure in the formalism, it is required in order to establish well defined (traceable) "features" or "landmarks" on the horizon.

There are various contexts in which the human observer is approximately described as such an observer [45, 47–49].

Physical space is a plane $\mathbb{E}^2 - \{\mathbf{o}\}$, with the vantage point singled out. We delete the vantage point (origin) because "the eye cannot see itself" [78]. The metric discussed above simplifies to (but see below)

$$\mathrm{d}s^2 = (\mathrm{d}\log\rho)^2 + \mathrm{d}\vartheta^2,$$

that is the Euclidean metric in the $\{\vartheta, \log\rho\}$-plane, which is the log-polar transform of the punctured plane. The geodesics are the straight lines of the $\{\vartheta, \log\rho\}$-plane, which are logarithmic spirals in the punctured plane. The logarithmic spirals centered on the vantage point are shifted within themselves by central rotation-dilations. This metric does not (yet) take the isotropic nature of the depth dimension into account though (see below).

The angular dimension spans the full horizon. In order to arrive at a projective structure we will assume $\vartheta \in (-\infty, +\infty)$, that is to say, we consider the points $\varphi + n2\pi$, $n = \ldots, -1, 0, +1, \ldots$ to be distinct. This is perhaps less strange if one thinks of a material object encircling the observer: Then these directions are assumed at different moments in time (Fig. 4).

We follow Berkeley [8] in holding that *depth* as such is not optically specified. The observer is free to assign depth values *ad libitum*, except that the assigned depths should respect the "cues". Cues are aspects of optical structure that the observer considers to be constraints on the field of depth values. Thus depths are generated by the observer, but checked against the available optical structure, they are

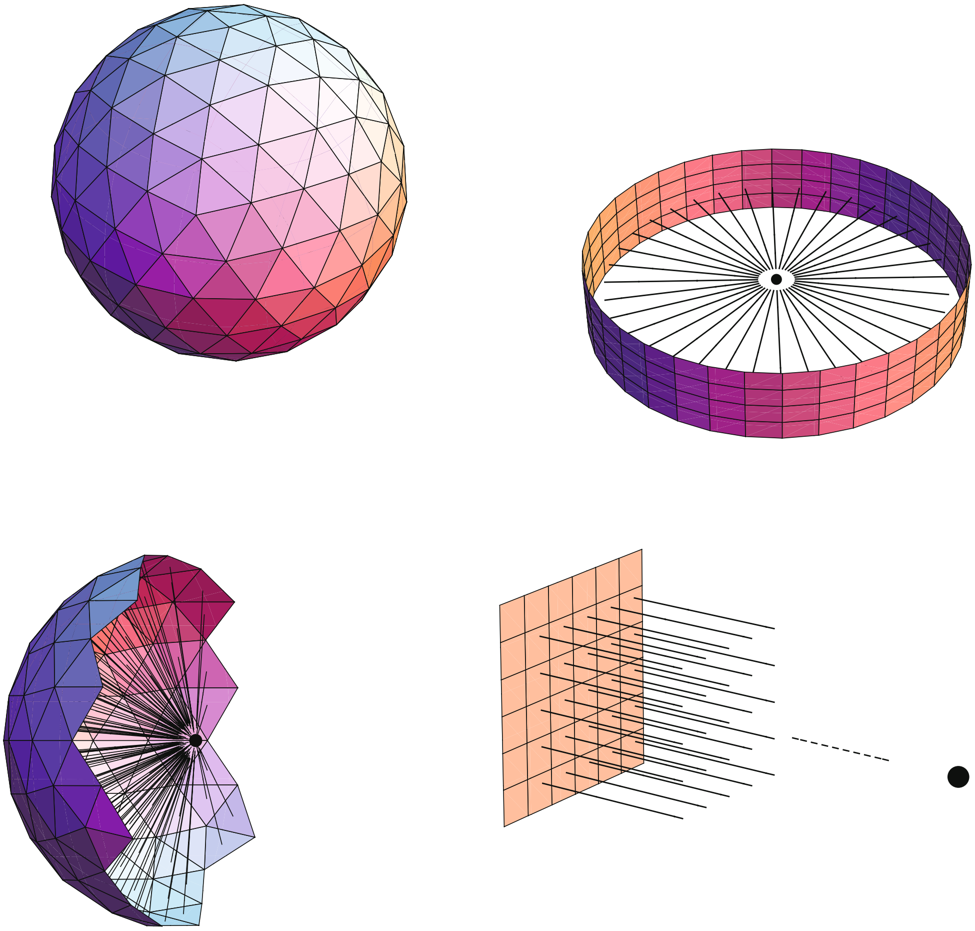


Fig. 2 At *top-left* a full panoramic observer, similar to some insect's eyes, its visual field encompasses all directions. At *top-right* an observer that observes only a narrow strip along the horizon of the panoramic observer, similar to the visual field of animals with a "visual streak". At *bottom-left* the visual field of an observer with forward viewing visual field, in this example about a hemispherical field. At *bottom-right* an observer with a narrow visual field near the forward direction, similar to animals with a well developed fovea. The human observer is a mixture of such a foveal observer with one that observes a hemispherical field (at much lower resolution)

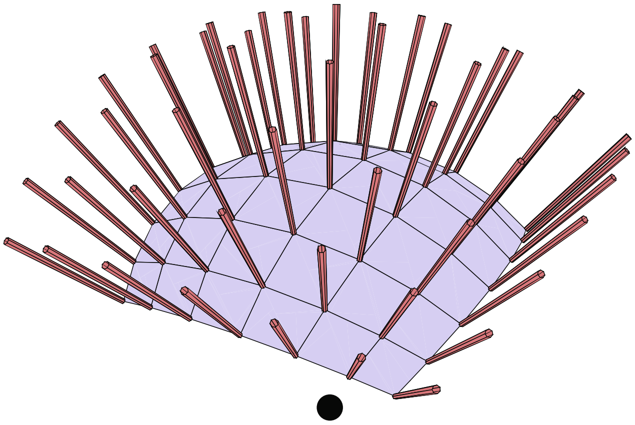


Fig. 3 The visual field as a trivial fiber bundle with as base space the visual field $\mathbb{S}^2$ and fibers the "visual rays". The visual rays are half-lines in physical space, but are to be treated as (full) affine lines of $\log\rho$, rather than depth ρ. Since it should be impossible to "mix" the transverse and depth dimensions, the visual rays are most conveniently treated as isotropic lines

"controlled hallucinations". In machine vision one would speak of "analysis by synthesis" [83, 84].

Please note that we have used "depth" for the sake of simple exposition, but one might substitute higher order derivatives of depth, such as "surface attitude" (first order, depth proper being zeroth order), "surface curvature" (second order), and so forth. There are good reasons to believe that human observers primarily assign second order properties. In many cases higher order spatial variations of depth are better presented than absolute depth, or depth on a point by point basis. It is possible to have curvature (shape) without absolute depth, this is even typical for pictorial perception for instance. There are good reasons for this because many

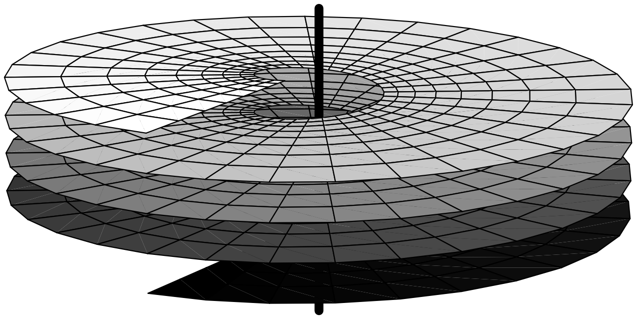


Fig. 4 Riemann surface for the visual streak model. The infinitely sheeted Riemann surface maps on the isotropic plane, logarithmic spirals (encircling the origin infinitely many times) mapping on straight lines

so called "depth cues" actually don't specify depth, but various derivatives of depth. (A good example is the shading cue.) We continue to speak simply of "depth values", so the reader should remain aware that the intention is more general.

That depth values can (in principle) be assigned independently at different locations is important. It prohibits the "mix" of the visual field and the depth dimensions. Such a mix occurs for instance in Euclidean space, where rotations can bring any dimension in coincidence with any other. This should be impossible in visual space $\mathbb{V}^2$. Intuitively, if you see the front of an object you cannot apply a rotation to see the back, at least not in *visual space*, where such a rotation would have to be a mental one. A good example is a painting, of any painted object only the front is painted at all, the back is nonexistent. You will never see the backside of a painted object, in visual space you cannot "turn around" in depth. This forces the depth dimension to be isotropic.

We construct "visual space $\mathbb{V}^2$" through the log-polar transform of $\mathbb{E}^2 - \{\mathbf{0}\}$:

$$\{u, v\} = \left\{\vartheta, \log \frac{\rho}{\rho_0}\right\},$$

where ρ denotes the distance from the vantage point, ρ_0 an arbitrary reference, and ϑ is the angular distance from the forward direction. We define the metric as

$$\mathrm{d}s^2 = \mathrm{d}u^2 + \varepsilon^2 \mathrm{d}v^2,$$

with $\varepsilon^2 = 0$, $\varepsilon \neq 0$, that is to say, the v-dimension is isotropic. In this space the visual directions become parallel lines and the equidistance circles the orthogonal family of parallel lines. Logarithmic spirals in $\mathbb{E}^2$ centered on the vantage point become straight lines. (See Fig. 5.)

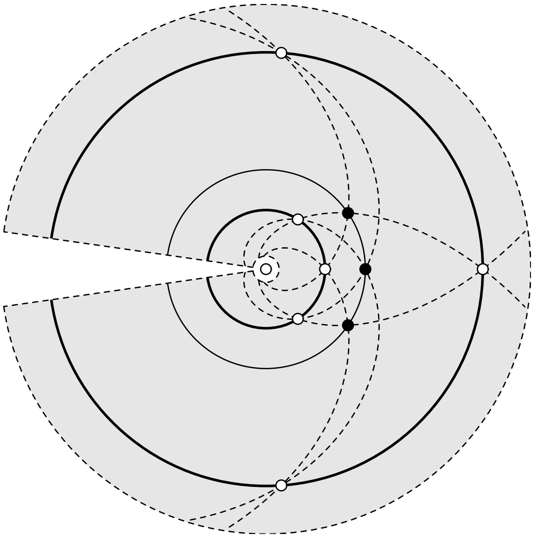


Fig. 5 The Pappus configuration. The Pappus structure is conserved over arbitrary (isotropic) similarities

The full group of orientation preserving similarities is [67, 68, 73–76]

$$u' = \alpha u + \beta,$$

$$v' = \gamma u + \lambda v + \mu,$$

where β and μ denote translations (they correspond to the aforementioned central rotation-dilations), γ a (non-Euclidean) rotation, and α and λ two moduli of similarities. (See Fig. 6.) This is a 5-parameter group, whereas the group of orientation preserving similarities of $\mathbb{E}^2$ is merely a 4-parameter group. This is due to the *hyperbolic angle metric* in $\mathbb{V}^2$, thus both angles and distances can be scaled, whereas you can only scale distances in $\mathbb{E}^2$ due to the fact that the angle measure is elliptic (periodic). To see the analogy to the Euclidean case notice that in the case of a congruency ($\alpha = \lambda = 1$) you have ($\mathbf{R}$ a rotation matrix, $\mathbf{t}$ a translation vector)

$$\mathbf{R}\begin{pmatrix} u \\ \varepsilon v \end{pmatrix} + \mathbf{t} = \begin{pmatrix} \cos \varepsilon\gamma & -\sin \varepsilon\gamma \\ \sin \varepsilon\gamma & \cos \varepsilon\gamma \end{pmatrix}\begin{pmatrix} u \\ \varepsilon v \end{pmatrix} + \begin{pmatrix} \beta \\ \varepsilon\mu \end{pmatrix}$$

$$= \begin{pmatrix} u + \beta \\ \varepsilon(\gamma u + v + \mu) \end{pmatrix} = \begin{pmatrix} u' \\ \varepsilon v' \end{pmatrix}.$$

Thus you regain the transformations introduced above. Apparently the familiar "Euclidian" form of the rotation matrix automatically changes into a (Euclidean) "shear" (which is indeed an isotropic rotation!) when you consequently carry through the algebraic consequences of the nilpotency of ε.

2.2.3 *Limitation to the Ground Plane*

Consider an observer with eye height h standing on an infinitely extended ground plane. The spherical, isotropic metric *limited to the ground plane* becomes

$$\mathrm{d}s^2 = \frac{\mathrm{d}x^2 + \mathrm{d}y^2}{h^2 + x^2 + y^2},$$

where $\{x, y\}$ are Cartesian coordinates in the ground plane, with the location of the observer's feet as origin. This is an important special case that applies to many daily life situations as well as many experimental settings.

The elements of the Riemann tensor (again, these are subject to the usual symmetries) are $R^1_{221} = -2r^2/(1+r^2)^2$ and $R^2_{121} = 2/(1+r^2)^2$, where $r = \sqrt{x^2 + y^2}$. The scalar curvature is $4/(1 + r^2)$, thus $+4$ at the location of the observer to zero near the horizon.

The geodesics are easily found through numerical integration. We find results that are similar to empirical results obtained for human observers in near space.

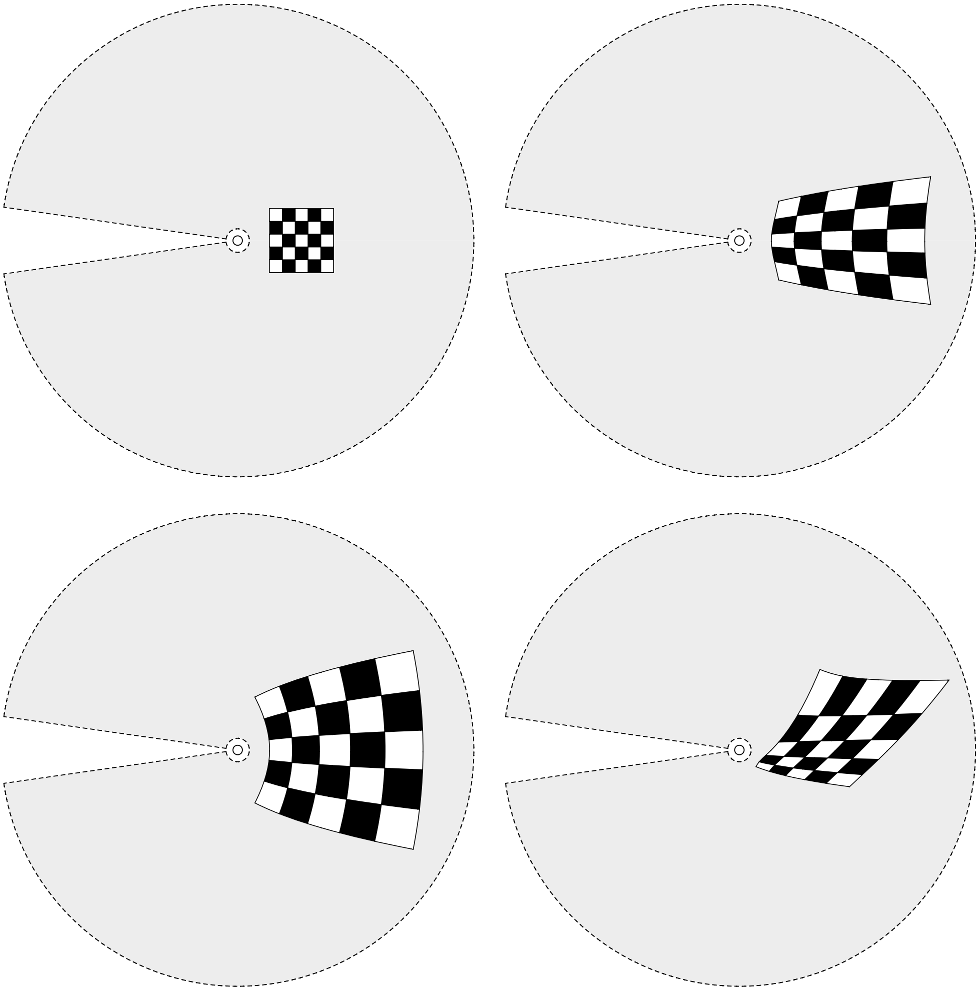


Fig. 6 A regular square (*top-left*) and the effect of some similarities (*top-right*, *bottom-left*) and a rotation (*bottom-right*)

2.2.4 *Frontally Centered Visual Fields*

A "frontally centered visual field" has a well defined "primary visual direction" and is organized concentrically about this direction, for instance, the visual resolution typically decreases monotonically with distance from the primary direction [21]. At the primary direction you may find a specialized "fovea". The extent of the visual field is typically the half-space in front of the observer, but may extend all the way up to the occipital pole. Due to its central organization the system depends on a system of eye muscles (or head muscles, as in the case of the barn owl [55], the nature of the "platform" is not important) that let the observer "fixate" targets of momentary interest. Here we concentrate on the human observer.

Consider Helmholtz's [31] theory of the *Richtkreisen*, which is relevant for the understanding of the structure of the visual field. A "*Richtkreis*" is defined on the unit sphere of visual directions $\mathbb{S}^2$. The "straight ahead", or "primary", direction is taken as the "north pole" and the occipital direction as the "south pole". In a polar coordinate system we use the "polar distance ϑ and the "azimuth φ" (reckoned from the vertically upwards direction) as coordinate angles. Eye movements in accordance with Donders' and Listing's Laws shift certain curves within themselves. These curves turn out to be small circles through the south pole, these are Helmholtz's "*Richtkreisen*". Since the eye movements, as constrained by the Donders–Listing Laws [19, 51], form a group that is isomorphic to the group of translations of the Euclidean plane $\mathbb{E}^2$, it is natural to consider the *Richtkreisen* as the analogues of the straight lines of $\mathbb{E}^2$. In this interpretation the space of visual directions $\mathbb{S}^2$ can be fitted out with a metric such as to become the "visual field $\mathbb{V}^2$". As Helmholtz notes this is most conveniently done by using the Cartesian plane $\mathbb{R}^2$ with the Euclidean metric as a model of $\mathbb{V}^2$, taking the stereographic projection from the south pole of $\mathbb{S}^2$ to the tangent plane at the north pole as the relation $\mathbb{V}^2 \leftrightarrow \mathbb{S}^2$. The south pole is mapped to a single point at infinity. Helmholtz shows through psychophysical experiments that this model of the visual field is indeed an apt one.

Formally, when ϑ denotes the angular distance from the frontal direction, φ the azimuth, then the Helmholtz representation is given by $x + \mathrm{i}y = \cot(\vartheta/2)\exp(\mathrm{i}\varphi)$, where x is the direction to the right and y the vertically upwards direction. (As a convenience we use the complex line $\mathbb{C}^1$ as a model for the Euclidean plane $\mathbb{E}^2$.) Conversely, if $\{X, Y, Z\}$ are the Cartesian coordinates of the unit vector on the viewing sphere, then

$$X + \mathrm{i}Y = \frac{2(x+\mathrm{i}y)}{1+x^2+y^2}, \quad \text{and} \quad Z = \frac{|x+\mathrm{i}y|^2-1}{|x+\mathrm{i}y|^2+1}.$$

The straight lines $\{aX + bY + c = 0\}$ correspond to the *Richtkreisen*.

We refer to the translations of $\mathbb{V}^2$ as "shifts". The shifts are transformations of the visual field $\mathbb{V}^2$. We also consider "blowups", that are scalings of the "depths". The "depth" is the "egocentric distance" in the "visual world" $\mathbb{V}^3$, it is the subjective correlate of the distance from the vantage point. We now consider the group of shift-blowups. These transformations should leave all relations in $\mathbb{V}^3$ invariant, we consider them to be *congruences*, the group of *translations* of $\mathbb{V}^3$.

If we assume $\mathbb{V}^3$ to be a *homogeneous* space, that is a space that looks the same as seen from any one of its points, then its structure is fixed. (Notice that homogeneity was the major assumption in Luneburg's [9, 53, 54] theory. It is implicitly assumed in many treatments.) It has to be one of the 27 Cayley–Klein spaces [14, 41, 42, 79], the one with a single isotropic dimension. The metric becomes

$$\mathrm{d}s^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \varepsilon^2 \mathrm{d}w^2,$$

where $w = \log z/z_0$ and $\varepsilon^2 = 0$, $\varepsilon \neq 0$.

2.2.5 *Infinitesimal Visual Fields (Pseudo-Perspective)*

For a very restricted ("infinitesimal") visual field, it is not necessary to distinguish between the different structures of $\mathbb{V}^2$ and $\mathbb{S}^2$, both can be treated as infinitesimal pieces of $\mathbb{E}^2$. In terms of physical space one talks of "pseudo-perspective" projection [20]. This is the case that we will use to describe "pictorial space". The visual space for the infinitesimal visual field is simply a narrow part around the primary visual direction of visual space. All visual directions are effectively parallel to each other. The space is thus the three-dimensional Cayley–Klein space with a single isotropic direction. We define "Frontal Space $\mathbb{F}^{2+1}$" as

$$\{u, v, w\} = \left\{\vartheta \cos\varphi, \vartheta \sin\varphi, \varepsilon \log \frac{z}{z_0}\right\},$$

where $\{u, v\}$ are Cartesian coordinates in the tangent space to the visual field at the forward direction, z is "depth", and z_0 some reference depth. We assume $\vartheta \ll 1$, but the forward model can be extended by the introduction of the Riemann normal coordinates $u = \vartheta \cos\varphi$, $v = \vartheta \sin\varphi$. Riemann normal coordinates are also known as "Postel projection" [18] and have been applied up to $\vartheta = 90°$ by Barre and Flocon. At $\vartheta = 90°$ the deformation of the Postel projection is appreciable, small circles map to ellipses with $\pi/2 \approx 1.57$ aspect ratio.

"Pictorial space" $\mathbb{P}^{2+1}$ is very similar to the frontal space $\mathbb{F}^{2+1}$. Empirically one finds that human observers treat pictures in pseudo-perspective mode, even if the actual field of view occupied by the picture is not small [15, 29, 62]. Instead of the parameters $\{u, v\}$ in the visual field near the forward direction we take the picture plane coordinates $\{x, y\}$. Notice that $\{x, y\}$ can be taken as physical measures if the observer perceives the relations in the visual field veridically (as we shall assume), whereas the depth z is a mental one, there being no physical analog. There simply is no third physical measure corresponding to depth since there is no depth in the picture. Pictures are planar distributions of pigments in some simultaneous pattern [17]. The picture is a flat physical object, whereas pictorial space is a "thick" (extended in depth) mental object. A similar reasoning applies to data structures involving "depth" inferred from an image in machine vision, though "mental" is not an apt term then. The formalism of the isotropic depth domain should be useful in machine vision (as it is in human vision), yet we haven't seen it used.

Since the case of the infinitesimal frontal field is formally identical with that of pictorial space, we discuss it in detail later under this heading.

The "Panoramic, Horizontally Centered Visual Field" and the "Frontally Centered Visual Field" models agree in an infinitesimal environment of the forward direction. Thus we need no additional formalism to describe pictorial space.

2.3 The Ambiguity Groups of the Cues

A "cue" is a substructure of the available optical structure that serves in constraining perception of a given observer in some particular way in the presence of that optical structure [65]. A simple example is a "T-junction", which is an image structure interpreted in terms of the occlusion of one opaque object by another in front of a background. Notice that "cues" exist only relative to observers. The observer selects optical data and (given situational awareness and current interests) promotes these to "cue" status. Most of this is preconscious activity, for "perceptions" are "presentations" in the sense that they simply happen to the observer, like sneezing. They are due to autogeneous processes [12] akin to "hallucinations". That perceptions often effectively subserve transactions with the world is due to the fact that they are constrained such as not to contradict the (selected!) cues. This renders the hallucinations "true (because efficacious) knowledge". Perceptions are "constrained hallucinations" in the same sense as scientific theories are "freely invented" but are constantly confronted with (thus constrained by) observed phenomena. In machine vision "cues" are usually inserted in the process by an external intelligence (the designer of the system). The identification of what are effective "cues" in a general context is a difficult problem that remains largely unsolved.

Such a view is distinct from the Marrian [56, 60] idea that perceptions are the result of "inverse optics [63]" calculations on the basis of the available optical structure. It seems unlikely that such a scheme could be made to work at all and it is not clear how such "perceptions" could be intentional [11, 35] (*about* the world). However, we will not enter into this discussion here, but will merely consider how either scheme deals with ambiguity.

No single cue suffices to determine (part of) the scene in front of you. The inverse optics calculation typically yields an infinity of "solutions". The various solutions may be transformed into each other through a member of the group of ambiguities characteristic for the calculation. Only for a few cues do we possess a well developed formal theory [6, 44]. Moreover, even in the well researched cases the theory depends on a large number of prior assumptions. Some of these are stated as part of the theory, but many others are willy-nilly assumed. The problem is that it is not possible to fully enumerate the prior assumptions in any case (they are like Searle's [70] "background"). It is always possible to think of others, because any exposition has to start from *something* and that something can always be questioned in numerous ways. Thus it is not possible to fully describe the ambiguity groups. The inverse optics calculations thus yield an ill defined set of possible interpretations of the scene in front of the observer. In real life the set cannot be exhaustively described. Since the inverse optics calculations exhaust the available information, the observer has to select a member of the set of solutions as the present interpretation of the scene in front of the eye. Such a selection is guided by prior knowledge of the way the world is likely to be. Modern Marrians use Bayesian estimation techniques to select a unique answer [22]. A problem is that the required prior probabilities are hard to get, especially since the full extent of the set of possible interpretations is very ill defined. The scheme is at least a challenging one if one prefers a cheerful view.

The "vision as controlled hallucination" view neatly avoids such problems because the "hallucinations" are specific to begin with, thus the selection problem doesn't occur. Here the problem is rather to come up with viable hallucinations, *i.e.*, those that are likely to survive the confrontation with the cues. Being able to see becomes a problem of coming up with likely guesses as to the nature of the scene in front of you, the ability to recognize and use cues, and even to create possible cue violations through probing actions. The formalism behind the "inverse optics" algorithms is still useful, but in a very different way. In many cases they may be replaced with forward algorithms, "running simulations" as it were. The idea behind the Bayesian prior probabilities remains useful too. After all it is only experience that allows the observer to come up with viable hallucinations. Good hallucinations require creativity tempered by experience. In machine vision the "hallucinations" are part of the algorithm, they are imposed by the designer, an external intelligence.

2.4 The "Beholder's Share"

Perceptions are presentations, that are creative constructs in accord with the cues. The cues are again creative constructs. Generically, the cues package aspects of prior experience. Specifically, certain optical structures have to be designated "cue". Thus even a correct cue may easily be misapplied. A cue is "correct" if it usually works, *i.e.*, if the perceptions usually lead to efficacious actions. Thus perceptions are far from inverse optics. They rely heavily on the observer. The part of the perceptions not due to optical structure is the "beholder's share" (a convenient term due to Gombrich [26]).

The beholder's share is constrained by the ambiguity group of the cues. This is a slightly different perspective from that discussed above that will turn out to be fruitful. We must expect the structure of visual space to be codetermined by the ambiguity groups, the ambiguities necessarily being "congruences" of the space.

The groups of congruences identified above are indeed likely to agree with the intersection of the ambiguity groups of all visual cues. This is a very important topic that will doubtless reward close attention. At present the formal understanding of even the most common cues (a few dozen) is virtually nonexistent.

3 The Structure of Pictorial Space

3.1 Formal Development

Of all structures discussed so far that of "pictorial space" is the simplest instance. We use it to illustrate a number of important issues. The generalization to the more complicated structures is immediate.

In looking at pictures observers typically discount the fact that their "visual rays" diverge from the pupil of the eye. Rather, everything is referred to the normals to the picture plane. This is why pictorial space can be treated like the "Frontally Centered Visual Field Model", even though the angular subtend under which the picture is seen may be far from "infinitesimal" (Fig. 7).

Apparently $\mathbb{P}^{2+1} = \mathbb{E}^2 \times \mathbb{A}$, where $\mathbb{A}$ denotes the affine line. The metric is simply the metric of the Euclidean plane, the depth dimension being isotropic. Notice that two points need not be identical if their mutual distance vanishes, for

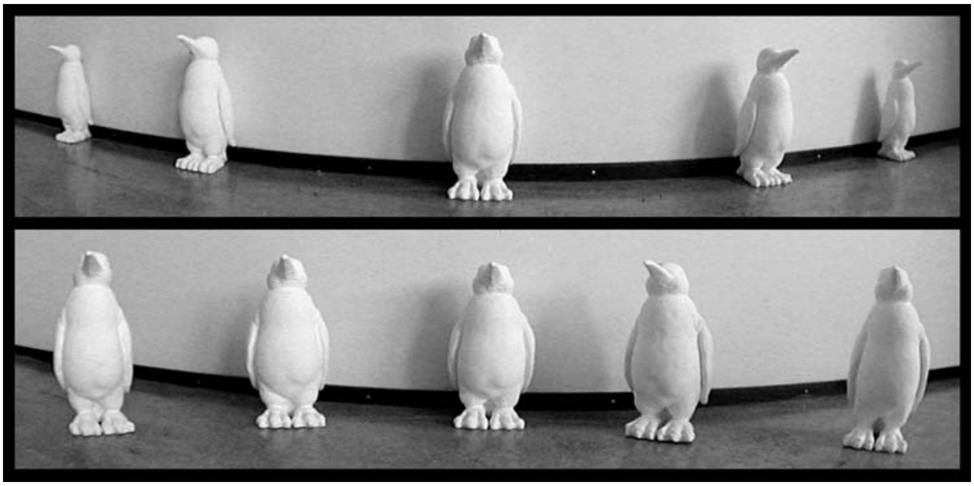


Fig. 7 Two photographs taken with a fish-eye lens (implementing Postel projection, *i.e.*, Riemann normal coordinates [50]) measuring 120° along the width of the picture. The (identical) puppets were placed in military formation (in a frontoparallel, equidistant, facing frontally) in the top scene, and placed on a circle centered at the camera anterior nodal point, equal angular spacing, facing the camera, in the bottom scene. Notice that the photograph at *bottom* looks much more like a military order than the *top* one

they may still differ in depth and thus be distinct. Such points are denoted *parallel.* In that case (and *only* in that case) they may be assigned a "special" distance, namely their separation in log-depth. Something very similar applies to planes. As the "distance" of two planes we may use the angle subtended by them (measured in the isotropic angle measure, see below). If the angle between two planes vanishes they may still differ, one calls them parallel. The log-depth separation can then (and only then) be used as a "special" angle.

A plane $ax + by + c = z$ is called "regular" if $\sqrt{a^2+b^2} < \infty$. All regular planes are metrically the same as the plane $z = 0$, thus $\mathbb{E}^2$. Planes $ax + by = 0$ ($\sqrt{a^2+b^2} > 0$) are special, for they contain isotropic lines. Such planes are not Euclidean, their metric being degenerated. We discuss the geometry of the special planes through the example of the plane $y = 0$, but all special planes have the same metrical structure so the example is generic.

The formalism becomes especially simple if you identify the special plane $y = 0$ with the dual number plane $\mathbb{D}$. The dual numbers [80] are complex numbers $u = x + \varepsilon z$, where the nilpotent imaginary unit $\varepsilon \neq 0$ satisfies $\varepsilon^2 = 0$. They have formally similar properties as the familiar complex numbers with imaginary unit $\mathrm{i}^2 = -1$. For instance, they can be written in polar form $\rho \exp \varepsilon\varphi$, where the modulus $\rho = \sqrt{u\bar{u}} = \sqrt{(x+\varepsilon z)(x-\varepsilon z)} = x$ and the argument $\varphi = \arctan z/x = z/x$. This is easily verified through direct calculation, applying $\varepsilon^2 = 0$ wherever applicable.

The linear transformation $u' = au + b$ (with complex coefficients) can be written explicitly as

$$x' = \alpha x + \lambda,$$

$$z' = \beta x + \alpha z + \mu,$$

where $a = \alpha + \varepsilon\beta$ and $b = \lambda + \varepsilon\mu$.

Geometrically, you have an isotropic scaling $\{x', z'\} = \alpha\{x, z\}$, a translation $\{x', z'\} = \{x, z\} + \{\lambda, \mu\}$, and a shear $\{x', z'\} = \{x, \beta x + z\}$. This "shear" changes the argument by a constant amount $\varphi' = z'/x' = \beta + z/x = \beta + \varphi$. Thus what looks like a shear to the Euclidean eye is actually an isotropic rotation over an angle β.

Notice that rotations may change the argument of the vector $\exp \varepsilon\varphi$ over the range $(-\infty, +\infty)$, that is from the direction towards the observer to the direction away from the observer. The vector cannot "turn around", you can see only the "front" of any visual object.

The difference of two points $x_{1,2} + \varepsilon z_{1,2}$ transforms as $\alpha(x_1 - x_2) + \varepsilon(\beta(x_1 - x_2) + \alpha(z_1 - z_2))$. Thus for $\alpha = 1$ we have that $x_1 - x_2$ is invariant, thus the signed difference $x_1 - x_2$ can be defined as a distance, and the linear transformations with $\alpha = 1$ are congruences, with $\alpha \neq 1$ they are similarities. In case $x_1 = x_2$ you have that $z_1 - z_2$ is invariant under congruences. Hence the definition: The distance between the points equals the regular distance $x_1 - x_2$, in case the regular distance vanishes, the distance equals the special distance $\varepsilon(z_1 - z_2)$. We define the metric as $\mathrm{d}s^2 = \mathrm{d}x^2 \pm \varepsilon^2 \mathrm{d}z^2$. Notice that the second term is zero ($\varepsilon^2 \mathrm{d}z^2 = 0$), we write it because it reveals the isotropic metric as a limiting case of either the Euclidean [39] or the Minkowski plane. Regarded as limit of the Euclidean plane we see that the isotropic plane is like an infinitesimally thick strip along the x-axis. Regarded as limit of the Minkowski [58] plane, the "light cones" are degenerated into the isotropic lines. Distinct isotropic lines are in the "elsewhere" region of each other, thus the isotropic lines are causally isolated from each other. This implements Berkeley's [8] notions fairly closely.

The unit circle centered at the origin is $x = \pm 1$. Because $\varphi = z$ for points on the circumference, we see that angles are measured as arc lengths along the unit circle, exactly as in the familiar Euclidean case.

Consider a curve $\mathbf{p}(x) = x + \varepsilon z(x)$. Differentiating once we find $\mathbf{t}(x) = 1 + \varepsilon z_x(x)$, with $\|\mathbf{t}\| = 1$, thus $\mathbf{t}$ is the unit tangent and the curve is parameterized by arc length. Differentiating once more we obtain $\kappa(x)\mathbf{n}(x) = \varepsilon z_{xx}(x)$. Thus the normal to the curve is ε (of unit special length, in the isotropic direction) and the curvature is $z_{xx}(x)$. It is perhaps disconcerting that the normal doesn't vary along the curve, but because $|\mathbf{n} \cdot \mathbf{t}| = 0$, the normal is apparently properly orthogonal to the tangent as should be. The expression for the curvature $\kappa(x) = z_{xx}(x)$ is much simpler than the Euclidean equivalent $\kappa = z_{xx}/(1 + z_x^2)^{3/2}$, a pleasant surprise. It is a true differential invariant. Thus differential geometry is really simple in the isotropic planes, much simpler than the familiar Euclidean differential geometry.

A curve of constant unit curvature satisfies $z_{xx} = 1$. Integrating twice we obtain $z(x) = x^2/2 + C_1 x + C_2$ for arbitrary (real) constants $C_{1,2}$. Consider $C_{1,2} = 0$. The curve $z(x) = x^2/2$ looks like a parabola to the Euclidean eye, but is evidently a unit circle centered at the origin. It is denoted "unit circle of the second kind", whereas the previously defined unit circle $x = \pm 1$ is denoted "unit circle of the first kind". These circles hold many properties in common, though they are clearly different. All well known properties of the Euclidean circle can be demonstrated on both the circles of the first and of the second kind. Both types have many uses.

Because the angle metric is hyperbolic, you can scale angles in the isotropic plane, just as you can scale distances. Thus the general similarity is

$$x' = \alpha_1 x + \lambda,$$

$$z' = \beta x + \alpha_2 z + \mu,$$

where α_1 is the modulus of the "similarity of the first kind", and α_2 the modulus of the "similarity of the second kind". A general similarity involves the moduli of both kinds.

Remember that z stands for the logarithm of the depth Z (say). Thus $z' = \alpha_2 z + \mu$ implies $Z' = \exp^{\mu} Z^{\alpha_2}$. This gives some intuitive meaning to the constants α_2 and μ. The translation μ causes a linear scaling, whereas the modulus α_2 causes a nonlinear scaling (generally known as "gamma transformation" in the intensity domain [37]).

3.2 Application of the Formalism

First consider transformations that leave the picture plane intact, thus $x' = x$. The (log-)depth dimension is scaled according to $z' = \beta x + \alpha_2 z + \mu$. The "rotation" $z' = \beta x + z$ is perhaps the most unexpected component. It skews the apparent frontoparallel over an (isotropic) angle β. The coefficient α_2 describes a scaling of angles (similarity of the second kind), thus a *proper movement* has $\alpha_2 = 1$, *i.e.*, $z' = z + \beta x + \mu$.

Do the proper movements indeed describe the phenomenology and how should they be understood? The latter question is a pressing one, for the proper movements leave (by design) the visual field fully unchanged. They merely affect the depth. Looking at a picture, you look at a planar surface covered with pigments in some simultaneous pattern. Indeed, looking *at* a picture presents you with the perception of such a planar surface. However, you (most of the time) can also look *into* the picture. Then you are presented with pictorial space, pictorial objects and pictorial relief. These entities are not physically present, thus purely *mental* objects. The proper movements apply to these entities and are thus themselves purely mental entities. They are not spatial objects, but transformations and deserve to be called "mental movements" or "mental changes of view". Such mental changes of view are apparently applied by the autogenous process that produces the presentations, often in (as far as we know) idiosyncratic manner. It is part of the "beholder's share".

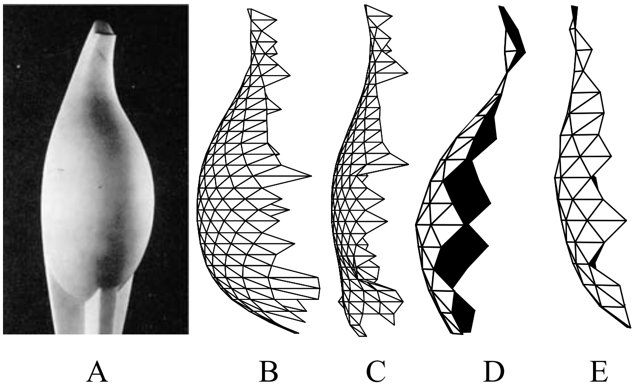


Fig. 8 A picture (**A**) and four psychophysically obtained "pictorial reliefs" (**B**–**E**). The reliefs are plotted with the depth dimension from *left* to *right*. The reliefs were obtained from different observers. Notice the differences in depth and apparent orientation: Apparently the observers apply isotropic rotations and depth scalings (isotropic similarities of the second kind) idiosyncratically

We find empirically (using novel methods of psychophysics [46]) that for the same observer and the same picture the pictorial reliefs observed at different occasions are typically related by (apparently) random proper movements. (See Fig. 8.) The pictorial reliefs observed for different observers confronted with the same picture often differ greatly, coefficients of variation for depth at corresponding location being at chance level, whereas a regression *modulo* proper movements will often change this to coefficients of variation near unity.

3.2.1 The Cue Structure

Of course the pictorial relief depends on the cues available in the structure of pixel values (in case of a pixelated picture, although irrelevant this is a convenient thought model). The cues are created by the observer (certain simultaneous patterns of pixel values being assigned cue status) and used to constrain the presentation. Some observers may use cues that others don't and perhaps assign cues to pixel value distributions in different ways. However, for observers (like human members of a certain culture) that are of identical evolutionary origin and have been raised in essentially identical biotopes, it is likely that the repertoire of available cues and the constraint induced by the cues are very similar. This is to be expected because the cues are due to efficacious interaction with the physical environment, they are threads of generic causal dependencies that occur over and over again and are comparatively independent of the specific total flux of events. Observers dealing with the same generic environment (biotope) are indeed likely to arrive at a similar bag of tricks. Thus one need not be surprised at the very high coefficients of variation for pictorial relief modulo proper movements found for different observers, at least when a rich bouquet of cues is made available by the structure of the picture.

This is typically the case for a straight, detailed and sharp and full scale photograph of a generic scene or a painting in a realistic style. Most people feel assured that you can safely discuss the contents of such pictures under the assumption that we all "see the same thing". If this were not the case one could only discuss the structure of the simultaneous pattern of pigments in the picture plane, as is indeed common in the case of Jackson Pollock's action paintings, and so forth.

In order to observe idiosyncrasies in the presentations of different observers one needs to weaken the constraints imposed by the cues. This can be done by deteriorating the retinal image through blurring, locally scrambling, diminishing contrast, changing the type of rendering (photograph, drawing, cartoon-style drawing, mere silhouette) or weakening possible relations to generic physical scenes (varieties of abstract painting). Observers will typically experience articulated presentations in any case, but they are likely to be idiosyncratic. For instance, Leonardo mentions the fact that observers have phantasmic presentations when looking at a rough, dirty wall (nowadays a random picture on a computer screen) that change from moment to moment. Here the only loosely constrained autogenous process is indeed close to an uncontrolled hallucination.

3.2.2 *Familiarity Cues*

In a great many cases of practical interest the mental views taken by the observer are readily understood in terms of certain global expectancies that (of course) derive from frequent and efficacious experiences. In that respect these constraints are similar to the cues, except for the fact that they apply at more global and semantically "higher" levels. For instance, it is much more likely to (erroneously) confuse a low-relief sculpture with sculpture "in the round" than *vice versa.* The reason is obviously that it is infinitely much more likely to meet a normal proportioned human being (in the case of a human effigy) than a "flattened" one. Similar observations apply to "skewed" sculptural reliefs as are common in the side panels of medieval triptychs or paintings viewed eccentrically.

"Skewed" sculptural reliefs, or pictures viewed (or painted) obliquely often lead to pictorial reliefs that are more frontal than expected on the basis of a (naive) reasoning on the assumption that vision is necessarily "veridical". In such cases the observers turn out to use the (isotropic) rotation that is part of the proper motions, that is to say, the parameter β. We have observed changes of the apparent frontoparallel plane by (Euclidean) angles as large as 70°. Such changes are encountered with photographs of objects that have (in the physical world) a well developed bilateral symmetry (thus "canonical views" like the human torso or head) but are photographed obliquely such as to destroy the symmetry in the picture. Here observers sometimes apply mental views that promote a "canonical view".

The German sculptor Hildebrand [34] wrote an influential treatise (in 1893) in which he describes how the depth dimension is highly volatile as compared with the dimensions of the visual field. He identifies the phenomenology of the proper movements in a lucid manner. His understanding is close to the theory presented here, except for a lack of formalism.

3.2.3 *Viewing Modes*

Other cases of practical interest in which the mental views taken by the observer are readily understood are those where the observer uses certain viewing methods, either with or without the use of optical instruments. Early observations are due to Leonardo who observes that in order to experience good depth in a painting the observer should close one eye and assume the proper vantage point at the center of perspective. This is good advice, for violating these conditions will favor the presentation of the picture as a physical (planar) object, rather than that of pictorial space. Empirically, we find that the presentations typically strike a balance, some linear combination, or weighted average, of either. Thus closing one eye easily (depending upon the binocularity of the observer) increases the depth range by a factor of two or more. Similar effects are found for oblique viewing. Optical instruments that promote high pictorial relief are constructed on these principles, that is to say, they influence the balance between seeing *into* and *at* the picture.

An effective way to suppress the presentation of the picture plane and favor the presentation of pictorial space is to use a view box. The view box works due to two effects, firstly it presents the picture in isolation, thus suppressing the presentation of a physical (planar) object and, secondly, it enforces monocular viewing from a well defined vantage

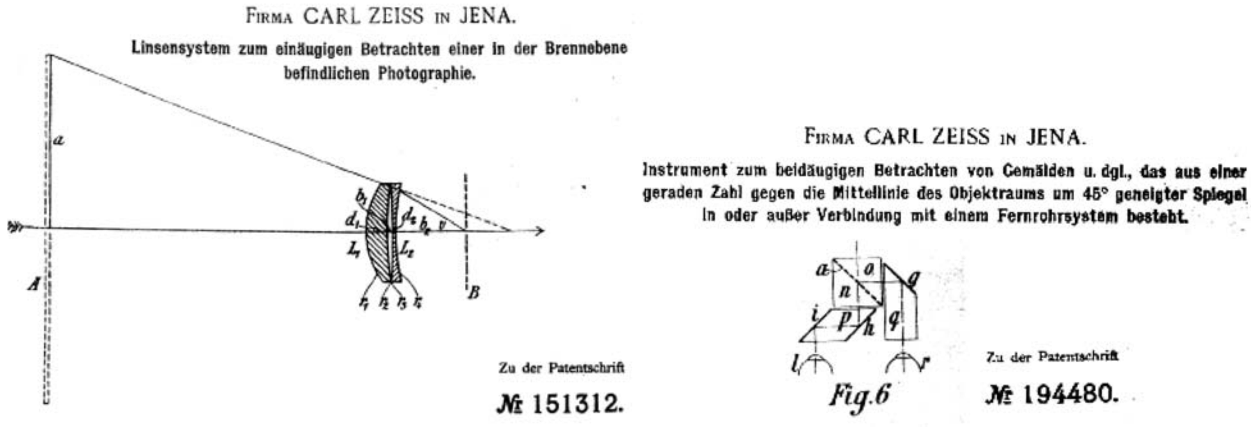

FIRMA CARL ZEISS IN JENA.
Linsensystem zum einäugigen Betrachten einer in der Brennebene befindlichen Photographie.
a
A
b1
d1
L1
L2
d2
b2
v
B
l1 l2 l3 l4
Zu der Patentschrift
№ 151312.
FIRMA CARL ZEISS IN JENA.
Instrument zum beidäugigen Betrachten von Gemälden u. dgl., das aus einer geraden Zahl gegen die Mittellinie des Objektraums um 45° geneigter Spiegel in oder außer Verbindung mit einem Fernrohrsystem besteht.
a
o
n
p
h
q
i
l
r
Fig.6
Zu der Patentschrift
№ 194480.

Fig. 9 *Left* the Verant, *right* (one of) the Synopter design(s). The Verant is a loupe with the center of rotation of the eye at the exit pupil, the object at the left focal plane. The system has a flat field and very low distortion. Thus eye movements do not induce monocular parallax and accommodation is fixed, effectively eliminating the physiological cues as to the flatness of the picture plane. The synopter design illustrated here is superior to the one actually marketed, which had unequal magnification for the two eyes. In the illustrated design the optical path lengths are identical, placing the apparent eye positions both in the midpoint of the interocular segment, thus rendering the observer effectively *cyclopic*. This removes binocular disparity cues that might reveal the flatness of the picture surface

point. Such instruments were very popular in Europe for several centuries. The ultimate view box was devised by von Rohr and marketed by the Zeiss company in the early twentieth century, the "Verant" [81] (Fig. 9 left). The verant is designed to view photographs taken with a single lens camera, thus a well defined center of perspective. One eye being occluded, the other eye views the photograph through a (triplet) lens. The lens is chromatically corrected and has a flat field, the photograph being placed in the focal plane. The center of rotation of the eye is placed in the exit pupil, thus nullifying the parallactic effect of eye movements. A specially shaped ocular cup is provided to constrain the positioning of the eye. Thus the instrument removes the conflicting cues due to binocular disparity, monocular parallax and accommodation. The picture is seen in perfect perspective, the anterior nodal point being the focal length of the camera removed from the picture (at the time the pictures were contact prints, not enlargements, and were not cropped). This instrument was described as a "monocular stereoscope" at the time and was very popular because it really delivered.

Modern viewers for slides, *etc.*, are generally inferior to the Verant because these principles are not well understood. Of course the "principles" are largely of a psychological nature, the geometrical optics and mechanical design being derived from them. Designing a quality "loupe" on the basis of geometrical optics is simple enough, but is likely to violate one or more of the psychological objectives. Inferior designs are common in modern head mounted displays, *etc.*, and have unfortunate effects on the resulting "visual space", especially for wide angle designs.

3.2.4 *Monocular Versus Binocular Stereopsis*

The invention of the (binocular) stereoscope in the early nineteenth century unfortunately greatly impeded the understanding of pictorial perception. Even today "stereopsis" is taken to be synonymous with "binocular stereopsis". (The Oxford American Dictionaries on our computer have "**stereopsis:** *the perception of depth produced by the reception in the brain of visual stimuli from both eyes in combination; binocular vision*." for instance.) Binocular disparity is generally taken to "explain" depth vision (stereopsis) and the very notion of "monocular stereopsis" is experienced as paradoxical [3, 69]. Indeed, in the second half of the nineteenth century we find papers with titles as "Paradoxical Monocular Stereopsis" in minor scientific journals, such contributions apparently being refused by the leading journals. In one such paper it is described how the pictorial space obtained by fusing (using a standard household stereoscope) two *identical* photographs beats that of a proper stereo pair in the following respect: In the latter case objects appear as flat cut-outs (coulisses) staggered in depth, whereas in the former case they have well developed pictorial reliefs. In modern developments one often purposely decreases the degree of binocular disparity in order to let monocular stereopsis contribute its share, with substantial advantages in realistic rendering in depth [71].

The device of using two identical images produces a zero disparity field (like when viewing objects at infinite distance), whereas viewing a single image with two eyes produces a nonzero disparity field due to the planar picture surface, evidently a "conflicting cue" from the perspective of monocular stereopsis. Early instruments that allowed binocular viewing of pictures without the conflicting binocular disparity cue included the "Zograscope". In this instrument one uses a large positive lens (large enough to look through it with two eyes) to view pictures in the focal plane. The instrument was typically used to view engravings, *etc.*, of which only a single copy was available. The instrument works to some degree, but the ones we examined suffer from inferior optical quality. A number of much improved designs are (again) due to von Rohr, and one (the "Synopter") was actually patented by the Zeiss company in the early twentieth century [82]. The highlighted design in the patent uses two mirrors, one semi-silvered to present the eyes with identical perspectives. It suffers from a magnification difference between the eyes. We have not been able to examine an actual specimen, but a reconstruction proved to work remarkably well. The best designs by von Rohr create a synthetic "cyclopean eye" located at the midpoint of the interocular segment (Fig. 9 right). A prototype turns out to work very well except for the fact that the visual field is limited (due to vignetting inevitably forced by the design) to about 27°(one effectively looks through a glass "tunnel" that vignets the view; the shortening of the optical path due to the refractive index of the glass helps somewhat though). In an experiment we found the instrument to function as advertised, for one observer the gain in depth range over binocular vision exceeded a factor of four.

The synopter principle is likely to be very effective in head mounted displays, but has—to the best of our knowledge—not been exploited in recent times. Neither the Verant, nor the Synopter designs are to be found in modern texts on display technology.

3.2.5 Ecologically Valid Cue Variations

The instruments described above leave the monocular cues invariant, they merely change the amount of conflicting cues. (*E.g.*, the cues that serve to reveal the flatness of the picture surface but contain no information concerning the *content* of the picture.) Thus it is no surprise that we encounter only changes in the depth range, described by the parameter μ of the proper movements. Changes of mental view are also observed in circumstances in which the cues are actually changed. Examples include photographs of a single physical scene under different lighting conditions. In such cases one might expect deviations from the proper movements. In one experiment we photographed an object several times from exactly the same vantage point, thus the resulting pictures were geometrically identical. We varied the direction of illumination from picture to picture, using a directional, though diffuse beam. The resulting pictures are very different as pixel values are concerned, though they look very similar as pictorial scenes. We find that the pictorial reliefs are very similar and that the (very significant) differences in depth distributions can be largely accounted for by proper movements. In such cases the pixel values identified as "cues" are apparently processed with very similar results (of course that is exactly the definition of "cue") and the changes that are observed remain within the group of cue ambiguities, which again coincides with the proper movements.

4 Discussion

This paper is about the data structures involving scene inferences (especially "depth") from single views. Most of the discussion was in terms of human perception, this makes sense because there exists hardly any comparable material of this nature in the literature of machine vision or artificial intelligence. On the other hand, the literature on human perception is mainly of an empirical nature with relatively little general discussion of a formal kind. Here we have tried to bridge the gap, this seems useful because the topic is bound to come up in machine vision/artificial intelligence and especially in the context of image based man-machine interfaces.

Many of the problems familiar from biological vision have (yet) not surfaced in machine vision because so much can simply be taken for granted. By way of an example consider the issue of "local sign". The problem was first faced squarely by the philosopher Lotze in the mid nineteenth century. Roughly speaking the problem is: How does the brain "know" where the wires in the optic nerve derive from? The optic nerve is a bundle of several million wires, each carrying pulse coded signals with a maximum frequency of about a kilohertz. The brain somehow derives the topological structure of the two-dimensional visual field from this input. One cannot assume that the mapping of "which wire" to "which pixel" is available (hardwired lookup table) at birth, because cases are known in which the local sign fails to develop ("tarachopia" or "scrambled visual field"). Since visual acuity increases in human infant (at least) up to the age of fifteen, it would seem that the topology of the visual field is established from coarse to fine over a long time span. In machine vision the problem is skipped because the pixel array of the camera is assumed known. However, it is not hard to conceive of more general cases in which a machine might notice the insertion of a connector with a few million contacts without any precognition as to the topology (dimension, layout of pixels, and so forth). Many of the issues familiar from human vision are bound to arise in machine vision at some time or other.

Similar issues apply to the "visual cues". Cues are often understood as image structures (for instance, T-junctions, specularities, and so forth). However, this only applies to very limited contexts. In general, a cue exists only in terms of a potential interpretation of the image. Given a theory of occlusion of opaque objects, certain image structures may (tentatively) be denoted "T-junction" or "edge". Such a tentative interpretation is like a "hallucination" that needs to be corroborated by the success of the resulting inference. General vision systems may take very little for granted, perhaps only the laws of optics and the generic statistics of the environment. Such systems presently hardly exist, almost all successful machine vision systems today are more or less dedicated.

A system that builds a spatial interpretation of a "scene in front of it" on the basis of a mere static image has to rely heavily upon prior knowledge. The resulting datastructure is necessarily a "controlled hallucination". The data-structure will have a spatial framework that may be denoted "visual space" or "pictorial space" that need not even exist as a "physical reality" (for instance in the case the system looks at a painting done from fantasy). This spatial framework is necessarily ambiguous because all static monocular cues are inherently infinitely ambiguous. (For instance, a painting on any arbitrarily shaped surface is as valid an interpretation as any.) The group of ambiguities left by the cues defines the freedom of the observer. Different spatial frameworks that each fully exhaust the constraints of the combined cues are related by what may be understood as a "congruence of visual space". Thus an understanding of the group of cue ambiguities serves to define the structure of visual space through its congruences or "proper movements".

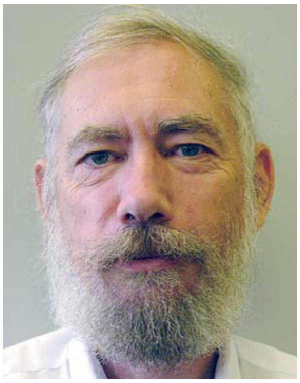
In this paper we have explored ways to describe possible structures of visual spaces in a formal, geometrical sense. The key features of the structure of pictorial space have been identified in the psychophysics of spatial perception of human observers. Human observers indeed apply the proper movements as defined by the cues as they see fit.

Several applications of these ideas to machine vision are immediate. For instance, the 3D shape parameters derived from 2D images have to be differential invariants of the group of congruences in visual space, rather than differential invariants of the Euclidean group in physical space. A systematic development of the concepts developed here is thus likely to be rewarding.

References

1. Abbot, E., Flatland, E.: A Romance of Many Dimensions. Buccaneer Books, Cutchogue (1976) (first edition 1884)
2. Adelson, E.H., Bergen, J.R.: The plenoptic function and the elements of early vision. In: Landy, M., Movshon, J.A. (eds.) Computational Models of Visual Processing, pp. 3–20. MIT Press, Cambridge (1991)
3. Ames, A. Jr.: The illusion of depth from single pictures. J. Opt. Soc. Am. **10**, 137–148 (1925)
4. Alexander, H.G. (ed.): The Leibniz-Clarke Correspondence. Manchester University Press, Manchester (1956)
5. Barre, A., Flocon, A.: La Perspective Curviligne, de l'Espace Visuel à Limage Construite. Flammarion, Paris (1968)
6. Belhumeur, P.N., Kriegman, D.J., Yuille, A.L.: The bas-relief ambiguity. Int. J. Comput. Vis. **35**, 33–44 (1999)
7. Bell, J.L.A.: Primer of Infinitesimal Analysis. Cambridge University Press, Cambridge (1998)
8. Berkeley, G.: An essay towards a new theory of vision (1709). In: Theory of Vision and Other Writing by Bishop Berkeley. Dent and Sons, New York (1925)
9. Blank, A.A.: The Luneburg theory of visual space. J. Opt. Soc. Am. **43**, 717–727 (1953)
10. Born, M., Wolf, E.: Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light. Cambridge University Press, Cambridge (1999)
11. Brentano, F.: Psychology from Empirical Standpoint. Routledge, London (1995) (original 1874)
12. Brown, J.W.: Fundamentals of process neuropsychology. Brain Cogn. **38**, 234 (1998)
13. Carrol, L.: Alice's Adventures in Wonderland. Macmillan, London (1865)
14. Cayley, A.: Sixth memoir upon the quantics. Philos. Trans. Rcmillan, R. Soc. Lond. **149**, 61–70 (1859)
15. Cornish, V.: Scenery and the Sense of Sight. Cambridge University Press, London (1935)
16. Coxeter, H.S.M.: Introduction to Geometry. Wiley Classics Library Series, New York (1989)
17. Denis, M.: Manifesto of symbolism. Revue Art et Critique (1890)
18. Destombes, M.: Guillaume Postel cartographe. In: AAVV, Guillaume Postel 1581–1981. Actes du Colloque International d'Avranches, pp. 361–371. Guy Trédaniel, Paris (1985)
19. Donders, F.C.: Anomalies of Accommodation and Refraction. New Sydenham Society, London (1864)
20. Dubery, F., Willats, J.: Perspective and Other Drawing Systems. Van Nostrand Reinhold, New York (1983)
21. Duke-Elder, S.: The Eye in Evolution. System of Opthalmology, vol. I. Kimpton, London (1958)
22. Edwards, W., Lindman, H., Savage, L.J.: Bayesian statistical inference for psychological research. Psychol. Rev. **70**, 193–242 (1963)
23. Foley, J.M.: The size-distance relation and intrinsic geometry of visual space: implications for processing. Vis. Res. **12**, 323–332 (1972)
24. Gibson, J.J.: The Perception of the Visual World. Houghton-Mifflin, Boston (1950)
25. Gibson, J.J.: The Senses Considered as Perceptual Systems. Houghton-Mifflin, Boston (1966)
26. Gombridge, E.: Art and Illusion: A Study in the Psychology of Pictorial Representation. Phaedon, London (1969)
27. Graham, C.H.: Vision and Visual Perception. Wiley, New York (1966)
28. Hale, N.C.: Abstraction in Art and Nature. Watson-Guptill, New York (1972)
29. Hauck, G.: Die subjektive Perspektive und die Horizontalen Curvaturen des Dorischen Styls. Wittwer, Stuttgart (1875)
30. von Helmholtz, H.: Die Tatsachen in der Wahrnehmung. Hirschwald, Berlin (1878)
31. von Helmholtz, H.: Handbuch der Physiologischen Optik. Voss, Leipzig (1860)
32. Hess, R.F.: Developmental sensory impairment: amblyopia or tarachopia? Hum. Neurobiol. **1**, 17–29 (1982)
33. Hilbert, D., Cohn-Vossen, S.: Geometry and the Imagination. Dover, New York (1944) (Orig. (G.) Anschauliche Geometrie, 1932)
34. Hildebrand, A.: The Problem of Form in Painting and Sculpture (transl. by M. Meyer and R.M. Ogden). Stechert, New York (1945) (First, German edition, Das Problem der Form, 1893)
35. Husserl, E.: Logische Untersuchungen. Zweite Teil: Untersuchungen zur Phaenomenologie und Theorie der Erkenntnis (1901)

36. Jacobs, T.S.: Drawing with an Open Mind. Watson-Guptil, New York (1986)
37. Jäne, B.: Practical Handbook on Image Processing for Scientific and Technical Applications. CRC Press, Boca Raton (2004)
38. Jaynes, E.T.: Prior probabilities. IEEE Trans. Syst. Sci. Cybern. **4**, 227–241 (1968)
39. Kheirandish, E.: The Arabic Version of Euclid's Optics. Springer, New York (1998)
40. Klein, C.F.: Vergleichende Betrachtungen über neuere geometrische Forschungen. Andreas Deichert, Erlangen (1872)
41. Klein, F.: Über die sogenannte nicht-Euklidische Geometrie. Math. Ann. **6**, 112–145 (1871)
42. Klein, F.: Vorlesungen über nicht-Euklidische Geometrie. Springer, Berlin (1928)
43. Koenderink, J.J.: The concept of local sign. In: van Doorn, A.J., van de Grind, W.A., Koenderink, J.J. (eds.) Limits in Perception. VNU Science Press, Utrecht (1984)
44. Koenderink, J.J., van Doorn, A.J.: The generic bilinear calibration-estimation problem. Int. J. Comput. Vis. **23**, 217–234 (1997)
45. Koenderink, J.J., van Doorn, A.J.: Exocentric pointing. In: Harris, L.R., Jenkin, M. (eds.) Vision and Action, pp. 295–313. Cambridge University Press, Cambridge (1998)
46. Koenderink, J.J., van Doorn, A.J., Kappers, A.M.L., Todd, J.T.: Ambiguity and the "mental eye" in pictorial relief. Perception **30**, 431–448 (2001)
47. Koenderink, J.J., van Doorn, A.J., Lappin, J.S.: Direct measurement of the curvature of visual space. Perception **29**, 69–79 (2000)
48. Koenderink, J.J., van Doorn, A.J., Kappers, A.M.L., Todd, J.T.: Pappus in optical space. Percept. Psychophys. **64**, 380–391 (2002)
49. Koenderink, J.J., van Doorn, A.J., Lappin, J.S.: Exocentric pointing to opposite targets. Acta Psychol. **112**, 71–87 (2003)
50. Kumler, J.J., Bauer, M.: Fisheye lens designs and their relative performance. Proc. SPIE **4093**, 360–369 (2000)
51. Listing, J.B.: Beitrag zur physiologischen Optik, Göttinger Studien. Vandenhoeck and Ruprecht, Göttingen (1845)
52. Lotze, H.: Mikrokosmos. Ideen zur Naturgeschichte und Geschichte der Menschheit. Versuch einer Anthropologie. Hirzel, Leipzig (1876)
53. Luneburg, R.K.: Mathematical Analysis of Binocular Vision. Princeton University Press, Princeton (1947)
54. Luneburg, R.K.: The metric of binocular visual space. J. Opt. Soc. Am. **40**, 627–642 (1950)
55. Martin, G.: An owl's eye; schematic optics and visual performance in Strix aluco L. J. Comput. Physiol. **145**, 341–349 (1982)
56. Marr, D.: Vision. Freeman, San Francisco (1982)
57. Mercator, G.: Atlas sive Cosmographicae Meditationes de Fabrica Mundi et Fabricati Figura, Duisburg, 1595. Lessing J. Rosenwald Collection, Library of Congress
58. Naber, G.L.: The Geometry of Minkowski Spacetime. Springer, New York (1992)
59. Nicod, J.: La géométric dans le monde sensible. Thèse, University of Paris, Paris (1923)
60. Palmer, S.E.: Vision Science: Photons to Phenomenology. MIT Press, Cambridge (1999)
61. Pasch, M.: Vorlesungen über neuere Geometrie von M. Pasch. Teubner, Leipzig (1912)
62. Pirenne, M.H.: Optics, Painting and Photography. Cambridge University Press, Cambridge (1970)
63. Poggio, T.: Low-level vision as inverse optics. In: Rauk, M. (ed.) Proceedings of Symposium on Computational Models of Hearing and Vision, pp. 123–127. Academy of Sciences of the Estonian S.S.R. (1984)
64. Poincaré, H.: Science et la Méthode. Flammarion, Paris (1908)
65. Riedl, R.: Die Ordnung des Lebendigen. Systembedingungen der Evolution. Paul Parey, Hamburg (1975)
66. Riemann, B.: Ueber die Hypothesen, welche der Geometrie zu Grunde liegen (Aus dem dreizehnten Bande der Abhandlungen der Königlichen Gesellschaft der Wissenschaften zu Göttingen), 10 June 1854
67. Sachs, H.: Ebene isotrope Geometrie. Friedrich Vieweg, Braunschweig (1987)
68. Sachs, H.: Isotrope Geometrie des Raumes. Friedrich Vieweg, Braunschweig (1990)
69. Schlosberg, H.: Stereoscopic depth from single pictures. Am. J. Psychol. **54**, 601–605 (1941)
70. Searle, J.: The Rediscovery of the Mind. MIT Press, Cambridge (1992)
71. Siegel, M., Tobinaga, Y., Akiya, T.: Kinder gentler stereo. IS&T/SPIE'98 1-10 (1998)
72. Spivak, M.: Comprehensive Introduction to Differential Geometry. Publish or Perish, Berkeley (1990)
73. Strubecker, K.: Differentialgeometrie des isotropen Raumes I. Sitzungsber. Akad. Wiss. Wien **150**, 1–43 (1941)
74. Strubecker, K.: Differentialgeometrie des isotropen Raumes II. Math. Z. **47**, 743–777 (1942)
75. Strubecker, K.: Differentialgeometrie des isotropen Raumes III. Math. Z. **48**, 369–427 (1943)
76. Strubecker, K.: Differentialgeometrie des isotropen Raumes IV. Math. Z. **50**, 1–92 (1945)
77. Swift, J.: Gulliver's Travels and Other Works. Routledge, London (1906) (orig. 1726)
78. Witgenstein, L.: Tractatus Logico-Philosophicus. Routledge and Kegan Paul, London (1922) (German text with an English translation on regard by C.K. Ogden; with an introduction by Bertrand Russell)
79. Yaglom, I.M.: A Simple Non-Euclidean Geometry and Its Physical Basis: An Elementary Account of Galilean Geometry and the Galilean Principle of Relativity (transl. by A. Shenitzer). Springer, New York (1979)
80. Yaglom, I.M.: Complex Numbers in Geometry (transl. by E. Primrose from 1963 Russian original). Academic Press, New York (1968)
81. Zeiss, C., von Rohr, M.: Linsensystem zum einaugigen Betrachten einer in der Brennebene befindlichen Photographie. Kaiserliches Patentamt Patentschrift Nr. 151312 Klasse 42h. (1904)
82. Zeiss, C., von Rohr, M.: Instrument zum beidäugigen Betrachten von Gemälden. Kaiserliches Patentamt Patentschrift Nr. 194480 Klasse 42h Gruppe 34. (1907)
83. Yuille, A., Kersten, D.: Vision as Bayesian inference: analysis by synthesis? Trends Cogn. Sci. **20**, 1–7 (2006).
84. Zucker, S.W.: The emerging paradigm of computational vision. Ann. Rev. Comput. Sci. **2**, 69–89 (1987)



Jan Koenderink graduated in Physics and Mathematics in 1967 at the Universiteit Utrecht. He has been associate professor in Experimental Psychology at the Universiteit Groningen, in 1974 returned to the Universiteit Utrecht where he held a chair in the Department of Physics and Astronomy. After retiring in 2008 he is connected to the Man-Machine-Interaction group of Delft University of Technology and as professor emeritus to the Universiteit Utrecht. He has received an honorific degree (D.Sc.) in Medicine from the University of Leuven and is a member of the Royal Netherlands Academy of Arts and Sciences.

Andrea van Doorn graduated in Physics and Mathematics in 1971 at the Universiteit Utrecht. She has participated in research on vision at Groningen University and is now connected with the Helmholtz Instituut of the Universiteit Utrecht where she works on various topics in visual psychophysics and modelling of visual functions in humans. She is associate professor at the Department of Industrial Design of Delft University of Technology.