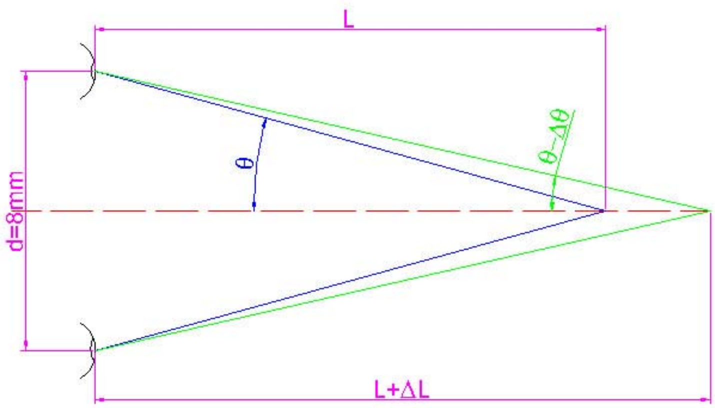
3D-vision and 3D-visualization

The human perception of distance is based on several stimuli from the vision. The primary one is however the convergence, i.e. how much the eyes need to be angled (squinted), so that both eyes will generate as identical images as possible. The angular resolution for a normal-sighted eye (Visus = 1) is barely θ = 0.4 mrad, which we will use in the calculations below on the distance resolution of the vision. This kind of estimation of distance obviously presupposes two reasonably equally performing eyes, i.e. the sight of one eye must not be noticeably better than the sight of the other one, and the image on the retina has to be approximately of the same size for both eyes.
A secondary method of distance perception is accomodation (adjustment of the strength of the eye). Furthermore, the brain utilizes different types of experiences e.g. of sizes, appearances of foregrounds etc. which will not be covered by this text.

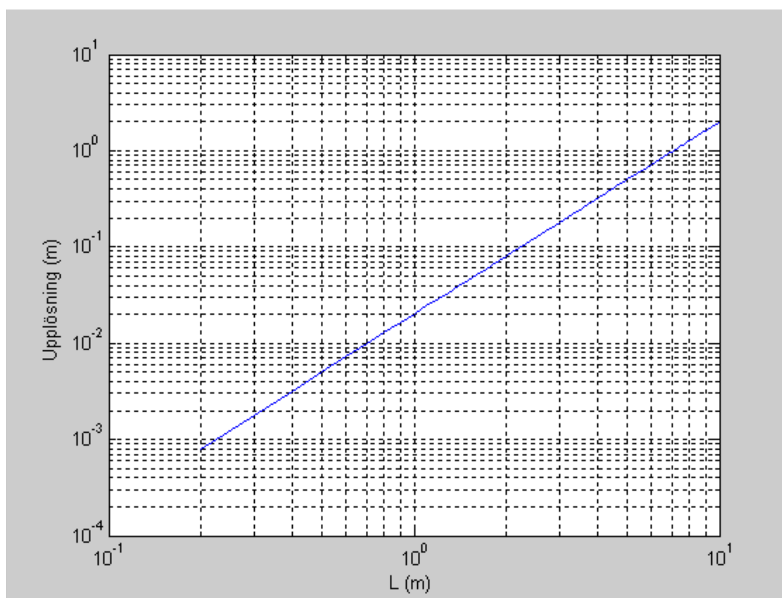
Convergence



The figure above describes the convergence of the eyes at two different object distances (the blue case and the green case). The geometry gives:

$$\frac{d/2}{L} = \tan\theta \approx \theta \quad \Rightarrow \quad L = \frac{d}{2\theta} \quad \Rightarrow \quad \Delta L = \frac{-d}{2\theta^2}\Delta\theta = \frac{2L^2}{d}\Delta\theta$$

Plotting the resolution as a function of object distance gives us the curve besides (observe the logarithmic scale)
The convergence works best if one has two objects close to each other in the field of vision with slightly different distances. For distances larger than 10 – 20 m, the resolution is the same as the distance, i.e. there is no upper limit for how far the object can be located.



Accomodation

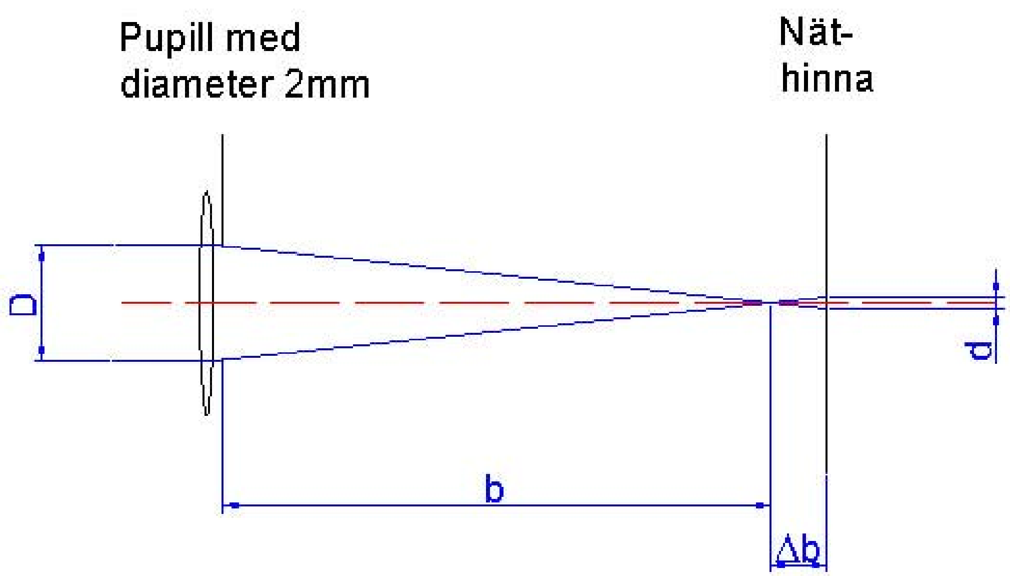
In order to describe the physical effect of accomodation, we need an eye model. There are different models to choose among, with different degrees of complexity, but for the purpose of demonstration (and of a surprisingly high degree of quantification) of the phenomena we are looking at, we can describe an eye as a thin lens with a focal width of 20 mm in the relaxed state (Strength 50 D), the distance between the lens and the retina being 20 mm.

When the eye accomodates itself to vision at a small distance, the strength of the lens increases (the focal length diminishes). This increase of strength is called accomodation. The faculty of accomodation decreases with age, a 20-year old person can accommodate 10 D, while a middle-aged person can accommodate 5 D and an old person 1-2 D.

In order to understand the effect more easily, one can split the lens in two, one representing the relaxed eye of 50 D, and the other one variable between 0 and the maximal accomodation. The lenses are placed at a negligible distance from each other, which means that their order is arbitrary. Place the accomodation lens first. Its task will then be to deliver parallel light to the 50 D lens.

It will fulfil the task if the object is placed in its first focus. Maximal accomodation corresponds to the shortest focal length of the accomodation lens, i.e. a person capable of accomodating from 0 to 5 D has a sharp vision between infinity and 0.20 m, while an old person accomodating between 0 and 2 D will see sharply between infinity and 0.50 m.

How small accomodation errors is the eye capable of detecting? We start again with a figure.



The figure shows an eye that has accomodated more than what is needed in order to see a given object. The image is formed at a distance b behind the lens and at a distance Δb before the retina. If d is the size of the diffuse spot, and θ_{sudd} is the angle the spot is taking up on the retina, we obtain

$$\frac{d}{D} = \frac{\Delta b}{b} \approx \frac{\Delta f}{f} \approx \frac{\Delta P}{P}$$

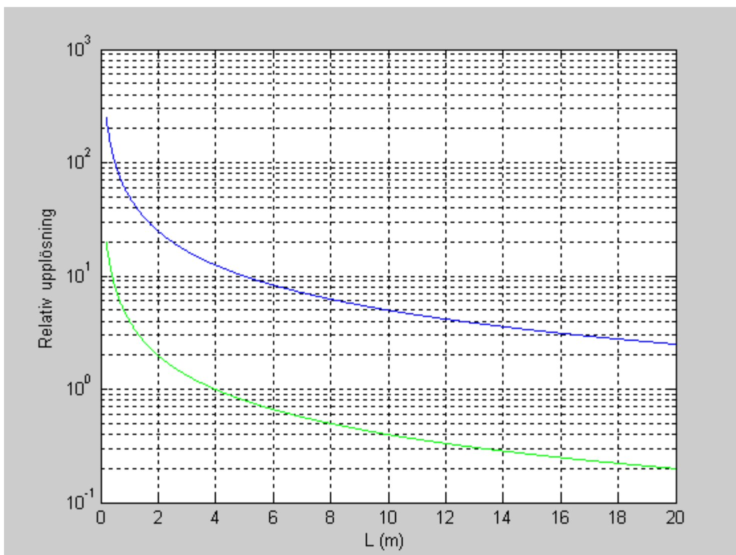
The last equality is due to the fact that a relative change in strength P equals the relative change in focal width (this is shown through differentiation).
The angle the diffuse spot is taking up from the lens is the same as the angle of diffuseness in reality, i.e. the diffuseness (in radians) is obtained through:

$$\theta_{\text{diffuseness}} = \frac{d}{f} = \frac{D\Delta P}{fP} = D\Delta P$$

For this diffuseness to be noticeable, the angle must be larger than the resolution of the eye, i.e. 0.4 mrad (see above). With a 2 mm pupil, the smallest possible detectable change in accomodation is about 0.2 D, which gives us:

$$\Delta L = L^2 \Delta P = 0.2 m^{-1} L^2$$

Plotting the relative resolution (= L/ΔL) for the accomodation and convergence, respectively, as methods of distance determination, we obtain



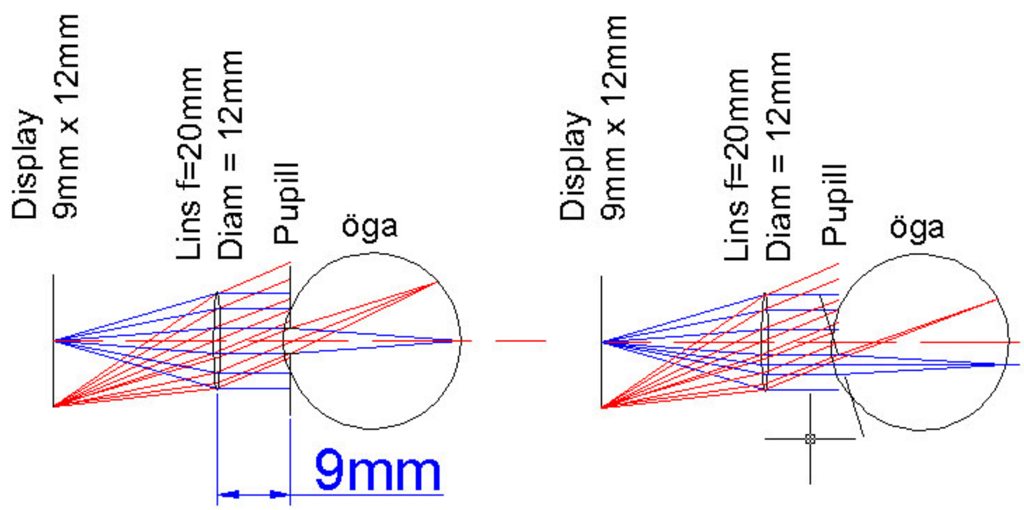
The blue curve is for the convergence and the green curve is for the accomodation. When the curve approaches 1 (= 10^0) the method becomes unusable ("the error" = 100%) and we see that for distances above 3-4 m the accomodation plays no role and equipment destined to give 3D-impression only needs to address the convergence.
For distances below 2-3 m accommodation errors will be noticeable and an observer will get contradictory visual impressions if the simulated 3D-effects are not addressing both convergence and accomodation with the same distance information. The effect of this varies between different persons, but can include indisposition, dizziness and headache. As we will see below, the accomodation is much harder to simulate, which is why most methods should be applied at distances above 2-4 m, but below 100 m, since at this distance the 3D-effects will not be noticed.

Methods to give 3D-illusion

Of all the methods incorporating screens or otherwise plane image surfaces, either large or small, one or two will be able to affect the convergence, but not the accomodation. In order to affect also the accomodation, it is required that a wave be generated which really diverges from points located at different depths. In order to obtain a projection with a depth distribution, holography or a direct projection on the retina can be used. These methods will be treated last.

One screen for each eye

This is the method used in most so-called virtual-reality helmets. The screens are small (usually about 1”) and are observed through a magnifying glass where the display is placed in the first focus of the lens. The angle taken up by the screen is then arctan(image width/focal length), which for a 1” screen (width 16 mm) and a focal width of 20 mm gives 38.7^{o} or 0.7 mrad per pixel (which is reasonable according to above, almost too good). Since the image is placed at infinity, the accomodation will never be affected, the whole effect deals with convergence. A problem in this context is making the lens large enough. When the eye is rotated in order to see an object at the edge of the field of vision, the pupil is being moved, and if the lens is too small and/or placed at too long a distance from the eye, image loss will occur.



In the leftmost case above the eye is looking straight ahead and is observing the starting point of the blue rays in the middle of the field of vision and the starting point of the read rays in the peripheral field of vision. But to the right, when the eye is rotated (the center of rotation lies about 10 mm inside the eye) in order to make the starting point for the red rays appear in the middle of the field of vision (one is simply looking at that direction) one can notice that the red rays are almost completely missing their target. Had the lens been smaller, the lens further away from the eye or the display larger, the outer point would not have been visible at all.

Since one must be able to look this obliquely through the lens, rather stringent requirements are put on it. It is above all necessary to keep the oblique astigmatism under control.

One screen, info alternately to each eye

The integration time of the eye is about 10 – 20 ms, depending on the method of calculation. This means that one gets the same visual impression from a picture sending out light continuously as from a picture lit under 2 ms and dark under 8 ms, but with a 5 times higher luminance. Now, if the observer is wearing glasses with a shutter (e.g. based on liquid crystals) alternating between transmission for each eye synchronized with the picture on the screen alternating between right-eye and left-eye pictures, a proper convergence can be generated for the eyes.
A variation on this is to send the right-eye and left-eye pictures with different polarizations and letting the observer wear polaroid glasses with corresponding different polarizations. This will of course restrict movements like ”shaking the head”.
The drawback in both cases is the requirement of having to wear eyeglasses.

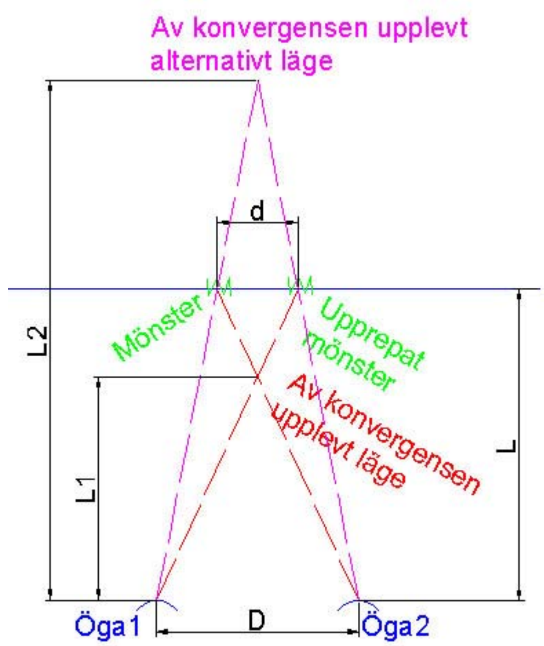
One screen, directed info

Another variant is to make every pixel double so that one part of it will direct light to one eye and the other part to the other one. This is done through putting a lens with the required directionality above each TFT-pixel.
The drawback is that the head must be kept fixed.

Autostereogram

The dot patterns (or other patterns) one sometimes sees in journals and books, having the property that after a moment of ”fixing” parts of the pattern are seen as a relief in front of the paper, are based on repeating the pattern so that the eyes can obtain the same image without converging to the same point on the paper.

Cf. Fig. to the right
The repetition distance, d, is often chosen to be smaller than the distance between the eyes, D, which has the effect that there are two convergences for which one obtains repetition. One of them leads to a picture in front of the paper (at a distance L1 from the observer) and the other one behind the



paper (at distance L2).
For the first position the geometry gives

$$\frac{D}{L1} = \frac{d}{L - L1} \quad \Rightarrow \quad L1 = \frac{D}{d + D} L$$

This position is always possible to find, but one must not hold the paper so close that L1 comes inside the near point.
The back position is given by

$$\frac{D}{L2} = \frac{d}{L2 - L} \quad \Rightarrow \quad L2 = \frac{D}{D - d} L$$

One can see that this position does not exist if d > D, which can be useful in order to avoid the ambiguity.
Autostereogram should be seen more as a method of creating illusions rather than a true means of imaging. There have been propositions on how to make real images based on this principle, but these are of dubious value.
The method obviously requires that one converges to one distance and accomodates to another one. This is an “unnatural” position for the eyes, which explains that everybody cannot see this effect.

Affecting the accomodation

It is not difficult to influence the accomodation, e.g. in a VR-helmet, through replacing the lens with a zoomable objective. The problem is that one gets the same accomodation for the entire picture, which is not what is required for the feeling of depth.
In order to affect both accomodation and convergence, the wavefronts must appear as coming from different objects.

In projection, this can only happen if the light scattering points are really located at different distances from the observer, the realization of this having been tried in many more or less unsuccessful ways. For instance with thin gauze weave, smokescreens or ionizing the air in each pixel at the very moment it is scattering light. Another variant is to steer laser beams so that they are really drawing a real image of the object in the air and not scatter the light subsequently, but rather look at the unscattered rays.

The only method that really does the job is the holography, which is not realistic for live images in a foreseeable future. The reason for this is that one needs a resolution of at least a few hundreds of lines/mm on the screen used, i.e. about 100 000 pixels/mm^2, which corresponds to 10^{10} pixels on a reasonably large screen. If this is to be updated 50 times per second with 24 bits per pixel (3 colors) one needs 10^{13} bits/s, which is clearly beyond the hopes of the most optimistic enthousiasts today. Normal compression algorithms like jpg and mpg cannot be used either, since it is precisely a very dense line pattern one wants to have. Anybody inventing a compression algorithm capable of dealing with this will not be asked to do much more than so in his/her active professional life.

Retinal display

A conceivable method to modify the concept “one screen for each eye” is, instead of a display and optics, draw directly on the retina using a scanning laser beam (or three scanning laser beams if one wants to have colors). If it is possible to vary the divergence of the laser beam for each pixel, one can in this way create a true 3D-image really addressing both convergence and accomodation.

This does not sound risk-free, but laser drawing on the retina (with carefully adjusted laser power) has been used during decennia e.g. in flight simulators.

This is the author’s guess on how 3D will be created in a couple of decennia.