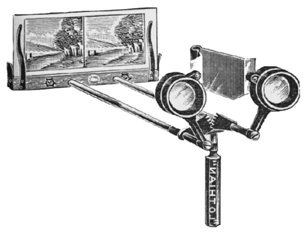
Stereoscopy: Theory and Practice

Paul Bourke
Peter Morse



1895 1995 2005

Outline

Paul Bourke

- Introduction to depth cues, parallax, and how stereo pairs are captured or generated.
- How create effective stereo, what to avoid.
- Technical details of projection and viewer systems, concentrating on mainstream digital systems.
- Special topics and demonstrations..

Peter Morse

- Brief history of stereoscopic imaging.
- Restoration & usage of old stereo pairs.
- Contemporary digital approaches - Practical issues.
- Projects.

Stereoscopy relates to the depth perception that arises because we have two eyes that are separated horizontally and therefore our visual cortex is presented with two slightly different views of the world.

Depth cues: 2D

By 2D ... they can be observed in a drawing, photograph or painting.
Apply even if you had only one eye.

- Occlusion. An object that blocks another is closer.
- Perspective. Objects reduce in size with distance.
  Parallel lines converge with distance.
- Expectation: We have an expectation of the size of some objects, for example, a plane may appear to be the same size as a bird so it must be further away.
- Motion. In a group of similarly moving objects, the faster moving ones are closer.
- Lighting. Diffuse reflection on a surface gives curvature information.
- Shadows. Depth relationships are implied (with respect to the light source) if one object casts a shadow on another.
- Detail. More detail is observed in closer objects.

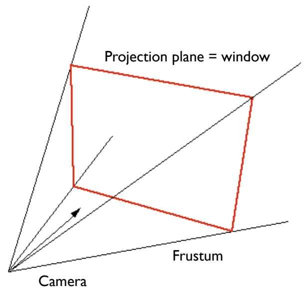
Depth cues: 3D

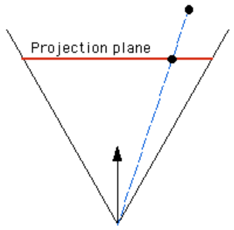
- Accommodation. The muscle tension needed to change the focal length of the eye lens in order to focus at a particular depth.
- Convergence. The muscle tension required to rotate each eye so they converge on the object of interest.
- Binocular disparity. The difference in the images projected onto the back the eye (and then onto the visual cortex) because the eyes are separated horizontally by the interocular distance.

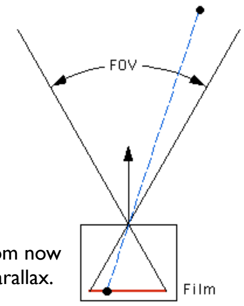
- It is this last depth cue that we are concerned with simulating in stereographics. If we can artificially create and present two (correctly formed) images to the visual system then they will induce the same sense of depth we enjoy in real life.
- Important for the other depth cues to be consistent!
- Note that accommodation cue is almost always in conflict. (More later)

Basics of perspective: Pinhole camera

- Pinhole (perspective) camera model.
- Determine where a point is drawn on the projection plane (film) by drawing a straight line from the point, through the camera position (pinhole), and see where it intersects the projection plane (film).
- Projection plane can be located anywhere perpendicular to the view direction. The image is horizontally mirrored if the projection plane is located behind the lens.
- View frustum is the rectangular cone with edges from the camera through each corner of the projection plane. The view frustum defines those parts of the scene that can be "seen" from a particular position.



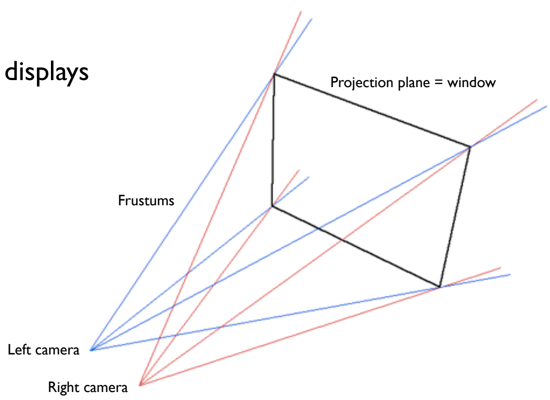




Will use views from the top from now on since there is no vertical parallax.

View through a window

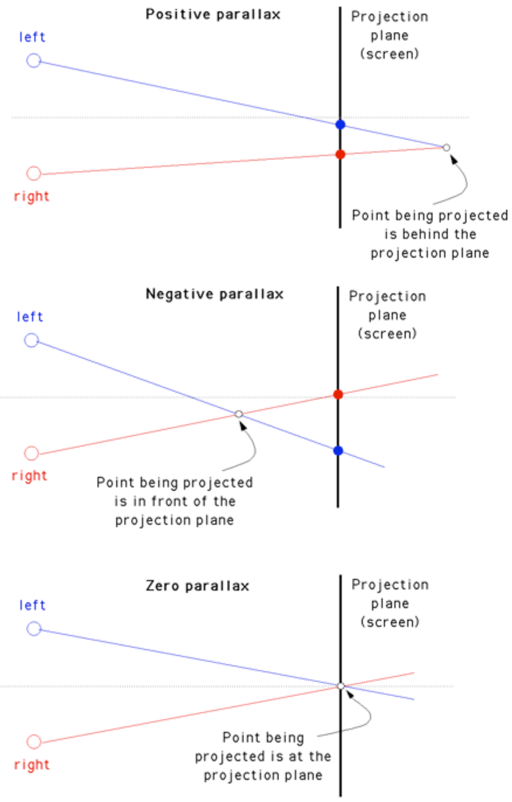
- Consider two cameras (eyes) and a single projection plane: correct model to think about this is a window through which the world (virtual or real) is viewed.
- Note that if the camera/eye/observer moves then view changes. There are some parts of the world than can be seen from one position and not other, and objects in the scene are viewed from different angles.
- This is not accounted for in most stereoscopic projection systems and explains the "shearing" effect one experiences when one moves.
- This can be accounted for in realtime stereoscopic projection by tracking the head position and view direction of the observer, with the sacrifice that the stereo pairs generated are only strictly correct for one observer.
- More on this later when discussing multiple wall displays

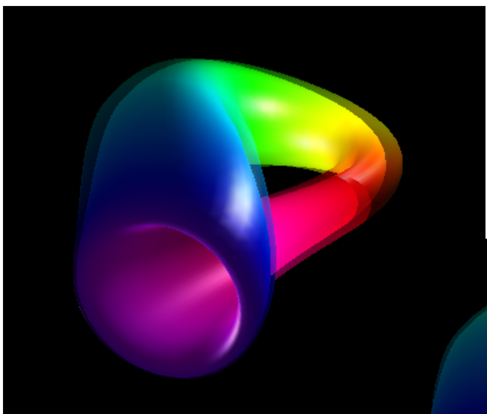


Example: Window shear (Socrates)

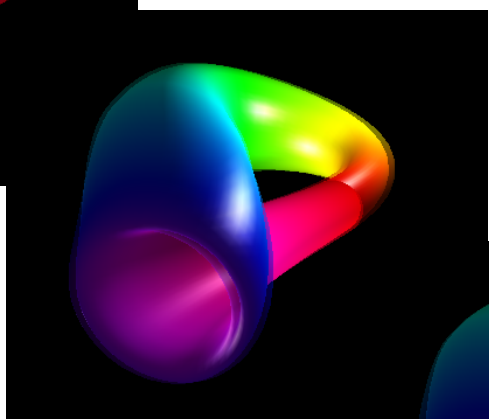
Parallax, +'ve, 0, and -'ve

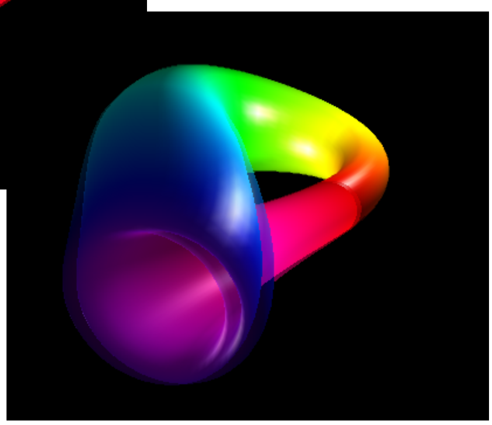
- There are three relationships that a point in space can have with the projection plane and the camera.
- They are classified depending on the relative position of the projection of an object onto the plane compared to the camera positions.
  - negative parallax = objects are closer than projection plane
  - positive parallax = objects are behind the projection plane
- The greatest possible positive parallax occurs when an object is at infinity, the maximum parallax separation is the same as the eye separation.
- It is the distance at which objects appear to be at zero parallax that forms the link between the virtual (or real world) and the environment into which the images are viewed. Objects at zero parallax will appear to be at the screen depth.





Zero parallax at front surface

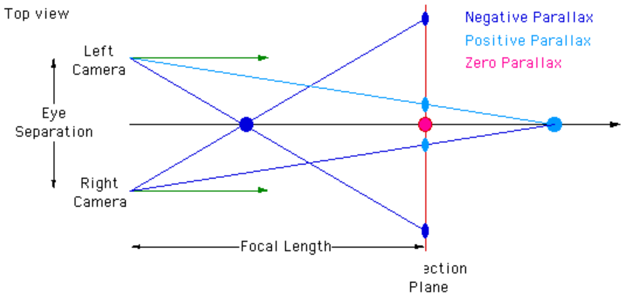




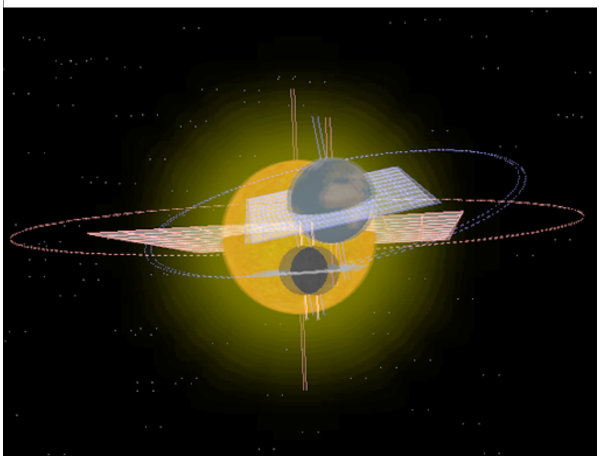
Example

Zero parallax at back surface

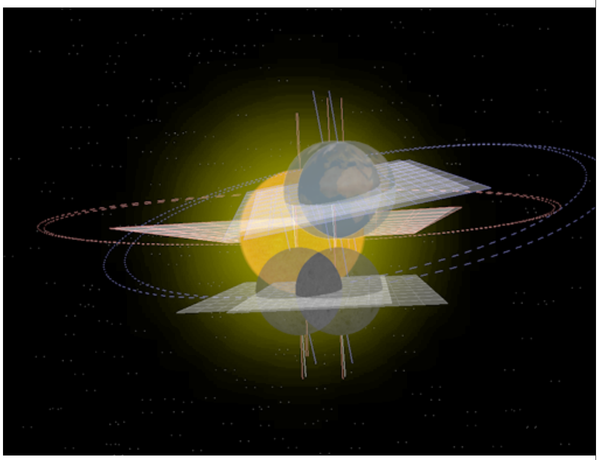
Negative parallax, the danger zone



- As an object comes closer to the camera the negative parallax goes to infinity.
- The degree to which an observers visual system will fuse large negative parallax depends on the quality of the projection system (degree of ghosting) .... but it will always eventually fail.
- High values of negative parallax is a key contributor to eyestrain.
- When the point in question is half way between the cameras and the projection plane, the negative parallax is the same as the camera separation. This is generally considered to be a good limit to "safe" negative parallax.
- It is the limiting negative parallax distance that plays a key role in the design of stereoscopic content, at least when attempting to achieve comfortable viewing. For realtime content one can use front clipping planes to avoid problems with close objects, other solutions can be used for film or movie style content.



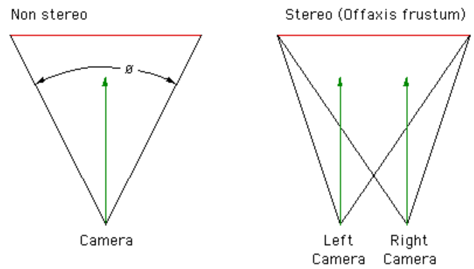
Modest negative parallax



Increasingly extreme negative parallax as the moon gets closer to the camera

Example

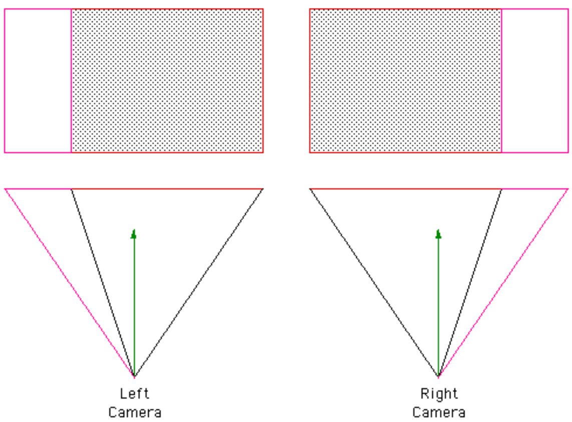
Symmetric and asymmetric frustums: Offaxis frustum



- We are used to thinking about view frustums that are symmetric about the view direction.
- Consider the view through a window, each camera (eye) position has a slightly different view of the world.
- Frustums for stereo pairs are not symmetric, see diagram top right.
  Often called "offaxis frustums".
- The vast majority of real cameras only support symmetric frustums although there are some specialist cameras that have offaxis lens/film configurations.
- The majority of computer rendering/animation software also only support a symmetric frustum.
- OpenGL (main API for realtime/interactive stereoscopic pair generation) does support asymmetric frustums.

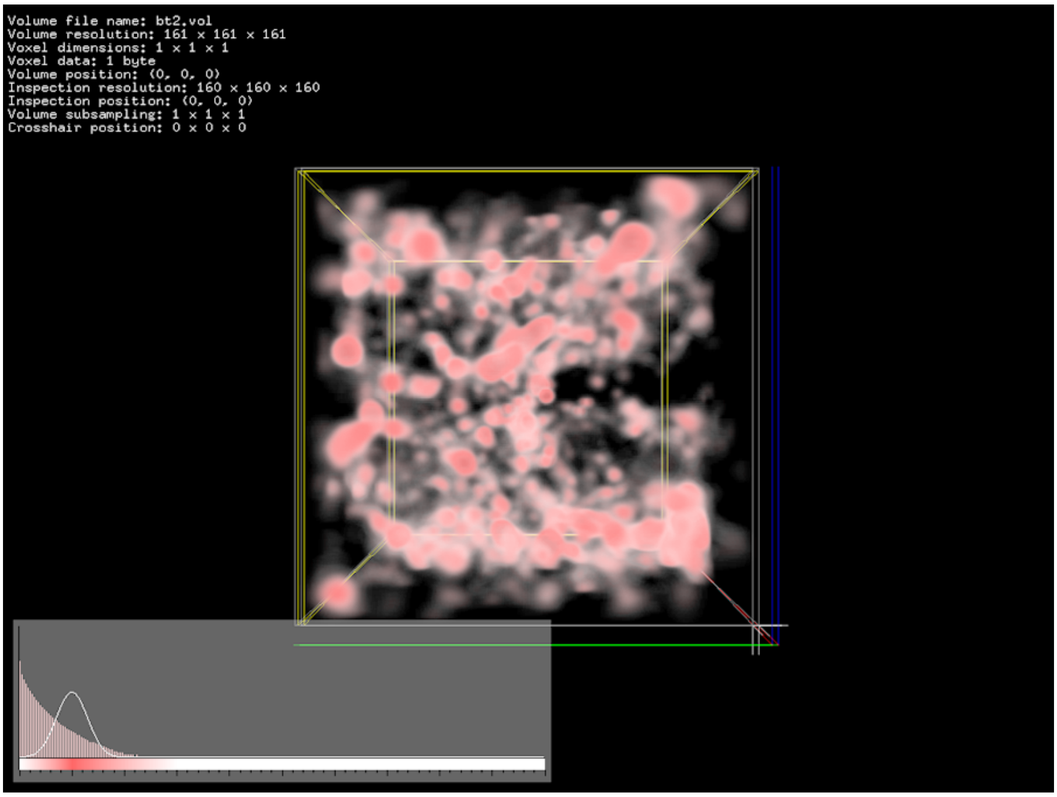
Asymmetric projection frustum from a symmetric frustum

- An asymmetric frustum can be achieved by taking photos or rendering with two parallel cameras and removing columns from each image in order to create the desired projections. This is equivalent to sliding the images horizontally with respect to each other and cropping.
- The exact amount to remove can be calculated given the desired eye separation and distance to zero parallax.
- "w" is the desired image width, "e" the eye separation, "fo" the intended distance to zero parallax, "a" is the desired horizontal field of view. Then the amount the trim "delta" from each image and the new aperture "a`" is given below.



$$\text{delta} = \frac{e\,w}{2\,f_o\,\tan(a/2)}$$

$$a' = 2\,\text{atan}\frac{(w + \text{delta})\tan(a/2)}{w}$$

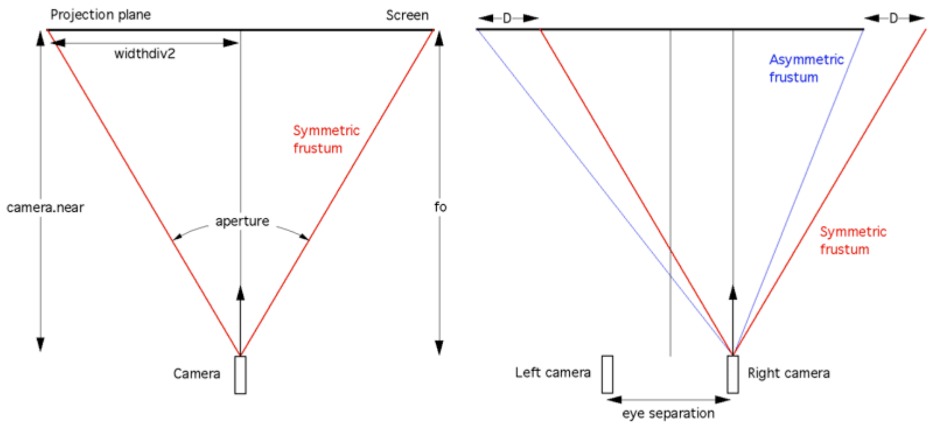


Example: Local group

Example

Asymmetric frustums: OpenGL

- OpenGL supports a sufficiently powerful frustum definition to directly create arbitrary asymmetric frustums. For some more exotic environments (eg: multiple wall displays) these frustums can be very much more asymmetric even for monoscopic content, see later.
- Need to ensure orthonormal view direction and up vector for correct calculation of the “right” vector.
- Distinguish between horizontal and vertical field of view. Horizontal used here.



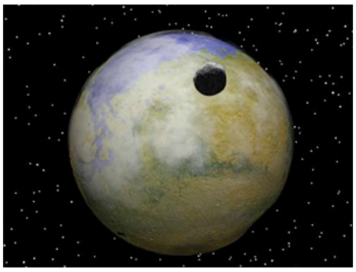

```
aspectratio = windowwidth / (double)windowheight;         // Divide by 2 for side-by-side stereo
widthdiv2   = camera.neardist * tan(camera.aperture / 2); // aperture in radians
cameraright = crossproduct(camera.dir,camera.up);         // Each unit vectors
right.x *= camera.eyesep / 2.0;
right.y *= camera.eyesep / 2.0;
right.z *= camera.eyesep / 2.0;

   // Right eye
   glMatrixMode(GL_PROJECTION);
   glLoadIdentity();
   // For frame sequential, earlier use glDrawBuffer(GL_BACK_RIGHT);
   glViewport(0,0,windowwidth,windowheight);
   // For side by side stereo
   //glViewport(windowwidth/2,0,windowwidth/2,windowheight);
   top    =   widthdiv2;
   bottom = - widthdiv2;
   left   = - aspectratio * widthdiv2 - 0.5 * camera.eyesep * camera.near / camera.fo;
   right  =   aspectratio * widthdiv2 - 0.5 * camera.eyesep * camera.near / camera.fo;
   glFrustum(left,right,bottom,top,camera.neardist,camera.fardist);
   glMatrixMode(GL_MODELVIEW);
   glLoadIdentity();
   gluLookAt(camera.pos.x + right.x,camera.pos.y + right.y,camera.pos.z + right.z,
             camera.pos.x + right.x + camera.dir.x,camera.pos.y + right.y + camera.dir.y,camera.pos.z + right.z + camera.dir.z,
             camera.up.x,camera.up.y,camera.up.z);
   // Create geometry here in convenient model coordinates

   // Left eye
   glMatrixMode(GL_PROJECTION);
   glLoadIdentity();
   // For frame sequential, earlier use glDrawBuffer(GL_BACK_LEFT);
   glViewport(0,0,windowwidth,windowheight);
   // For side by side stereo
   //glViewport(0,0,windowidth/2,windowheight);
   top    =   widthdiv2;
   bottom = - widthdiv2;
   left   = - aspectratio * widthdiv2 + 0.5 * camera.eyesep * camera.near / camera.fo;
   right  =   aspectratio * widthdiv2 + 0.5 * camera.eyesep * camera.near / camera.fo;
   glFrustum(left,right,bottom,top,camera.neardist,camera.fardist);
   glMatrixMode(GL_MODELVIEW);
   glLoadIdentity();
   gluLookAt(camera.pos.x - right.x,camera.pos.y - right.y,camera.pos.z - right.z,
             camera.pos.x - right.x + camera.dir.x,camera.pos.y - right.y + camera.dir.y,camera.pos.z - right.z + camera.dir.z,
             camera.up.x,camera.up.y,camera.up.z);
   // Create geometry here in convenient model coordinates
```

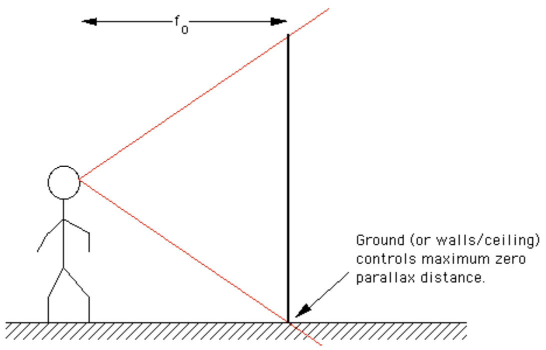
Camera separation - rule of thumb

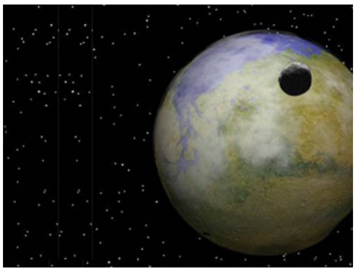
- There are two options for choosing camera (eye) separation:
  1. Use human eye separation.
  2. Use a separation relevant to the scale of the objects being captured.
- (1) is the only approach that can give a correct "real world" sense of scale and distance. Perfectly appropriate for (most) photography or rendering of objects on our scale. Can result in little depth perception for distant objects, as in real life.
- (2) is required in order to give a sense of depth for objects that are at very different scales, for example very small or very large.
- For abstract scaleless geometry one can either employ (2) or impose a human scale to the geometry and use (1).
- The rule of thumb for "safe stereo" is to choose the object and therefore distance that you want zero parallax to occur (the distance at which the object will appear to be at the screen depth) and choose a camera separation no greater than 1/30 of that zero parallax distance

Creating effective stereo content, part 1

- Objects that have negative parallax but cut the image border. These in general will not appear 3D since our visual system clearly knows the frame border is at 0 parallax.
- Positive parallax objects that cut the frames are fine.
- This often leads to the ground or walls limiting the zero parallax distance and therefore the eye separation.



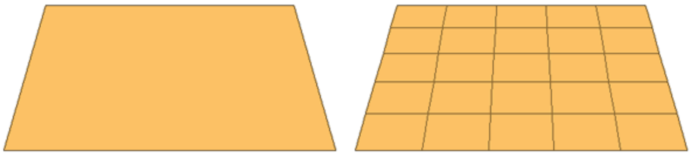


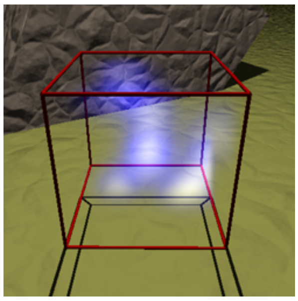


Example: Tritorus

Creating effective stereo, part 2

- Motion cues are perhaps the next most important depth cue.
  Adding motion (eg: camera panning or rotation) can greatly enhance the sense of depth.
- Parallax is preserved on reflection by a planar mirror.
  This is not the case for curved mirror surfaces.
- Vertical structure.
  Parallax information is carried in vertical structure.
  For example, a horizontal line only presents parallax differences at the two ends. This can be improved by using dashed lines.
- Gaseous objects, clouds, smoke, and so on don't convey strong parallax cues.

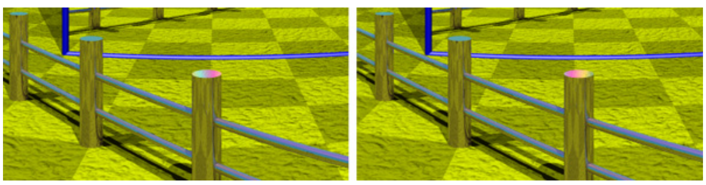




Example: Motion cues (6dF) and gaseous objects (Furnace)

Adverse effects, part 1

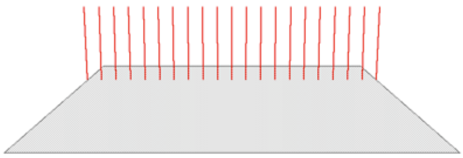
- Our visual system is very sensitive to differences in brightness and/or colour levels between the two views.
  While usually largely of concern for photography rather than CG.
  It can arise however between unmatched projectors or displays.
- Low resolution stereo pairs may not provide enough pixels to represent a range of parallax distances. This results in images that appear to be made up of layers of depth rather than continuous depth range. Occurs when the range of parallax within an object is on the order of 1 pixel width.
- Avoid frequent and abrupt changes in zero parallax distances and eye separation.
- Specular highlights. Specularity (eg: shiny metal objects) depends on the relative position of the camera to the light sources, thus a specular highlight can occur in camera (eye) and not the other.

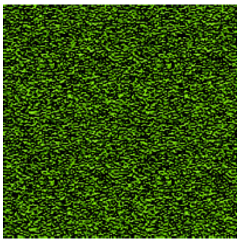


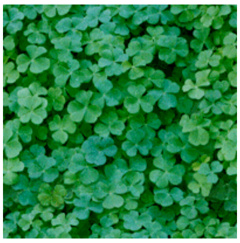
Example: Boolardy caravan

Adverse effects, part 2

- Noisy textures.
  At a sufficiently high frequency there will not be visually identifiable structure to qualify as stereo pairs. A form of aliasing that might not be noticed in monoscopic images.
- Parallax interference.
  Geometric or texture structure that is the same spacing as the parallax can result in ambiguous depth perception.
- Large transitions between light and dark regions will aggravate any ghosting inherent in the projection/presentation system. Not uncommon to use grey rather than black backgrounds for visualisation applications. This is dependent on characteristics of the projection hardware, see later.
- There are various other techniques used for 3D rendering that can be problematic for stereo pairs. eg: Processes that may have different random number generators between runs (cameras).

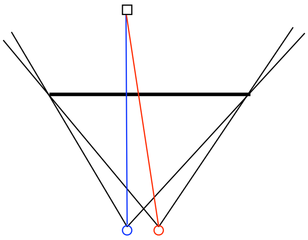




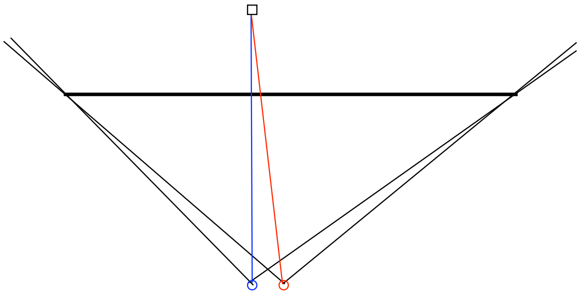


Adverse effects, part 3

- Fast moving objects and slight differences in frame synchronisation in capture and/or presentation hardware. Two effects can occur, parallax between the left and right images can change significantly between sequential time steps, and the visual system may not have time to maintain convergence and focus.
- Divergence. Objects an infinite distance should be separated by the eye separation, say 6.5cm. Wider than that is asking the eyes to diverge, they are only designed to converge! This places a limit on the amount of horizontal sliding that can be achieved when setting zero parallax. This is a common problem when content is designed for a small screen stereo projection system is moved to a larger screen. For interactive content this can be avoided by simply changing to the appropriate field of view. For precomputed content the only solution is the render a number of versions, each with a different field of view. and sweet spot position.



Small screen



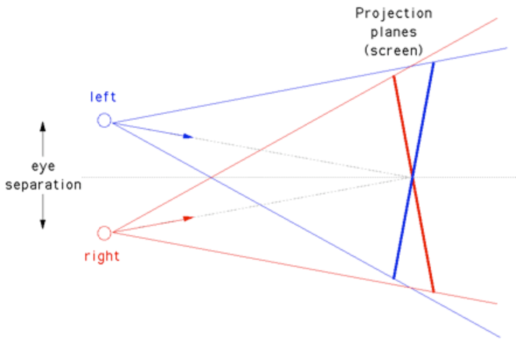
Larger screen, different FOV and zero parallax distance
Very different result by scaling the small screen content.

Creating effective stereo, summary (Take home message!)

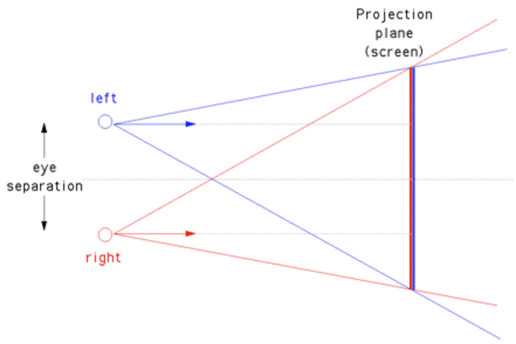
- In general one should be conservative with eye separation and the degree to which extreme negative parallax is employed. This is particularly important for stereo content that will be viewed for extended periods.
- While negative parallax is "sexy", positive parallax is much easier to look at for extended periods than negative parallax.
- For the best results design the content for the geometry for the intended display system. That is, match the camera field of view to match the field of view given the screen width and intended sweet spot for the viewer.
- Creating stereoscopic content that places minimal strain on the visual system is difficult! This is the skill or art for stereoscopic producers.

Alternative approach: Toe-In

- In the photography and filming industry (less so in the computer generated stereo pairs) it has been common practice historically to not use parallel cameras.
- The cameras are turned inwards such that they converge at the distance of zero parallax.
- Toe-In introduces vertical parallax towards the corners increasing eyestrain.
- Toe-In has advantages for low resolution capture (eg: video) since pixels are not discarded horizontally.
- Toe-In has simpler processing requirements, no trimming so images are ready immediately after capture.
- Parallel cameras have a simpler alignment procedure, zero parallax distance can be adjusted over a wider range in post production.



Toe-in cameras (above) vs parallel cameras (below)

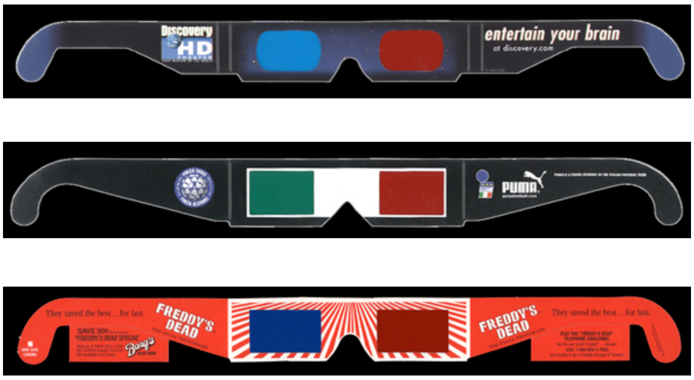


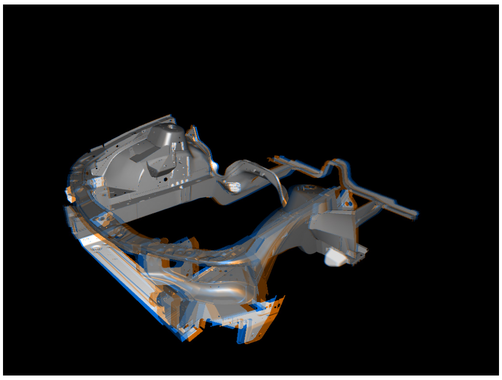
Projection/viewing technology

- Goal is always the same, to independently present a left and right eye image to the corresponding eye. Common requirement to all presentation systems!
- Will discuss the more established technologies within each category, there are variations in the exact technology within each category.
- Will only consider technologies capable of delivering digital stereo pairs in real time. Excludes single image based viewing, holography, laser/mirror based systems, and some of the more exotic systems.
- Goal is to give a sense of each technology, how they might be evaluated, and in what circumstances is one more appropriate than another.
- Important for content developers to have an understanding of the limitations of the projection hardware they will use. This can affect how the content is created.

Anaglyph

- This is a rather old approach but illustrates the basic principle of delivering images independently to each eye.
- The left and right eye are encoded as two colours on a single image, for example red and cyan. Matching glasses presents the appropriate stereo pairs to each eye.
- Generally very high cross talk (ghosting).
- One of the few ways to present stereo images without special hardware and in print.
- Doesn't represent colour faithfully.
- Main options are red-cyan, red-blue, and red-green.





Mirror based viewers

- Modern version of the Wheatstone mirror stereoscope (circa 1838).
- Generally limited to a single person.
- Largely a gimmick, rarely used on an ongoing basis except perhaps for aerial mapping and geology.

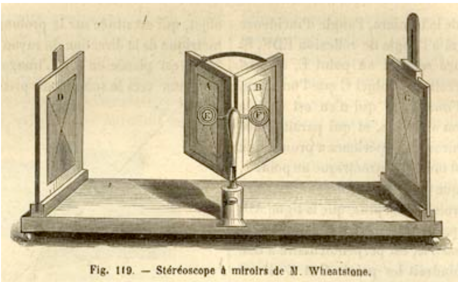
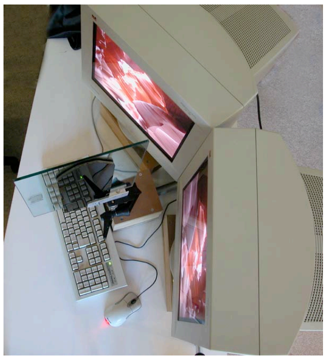
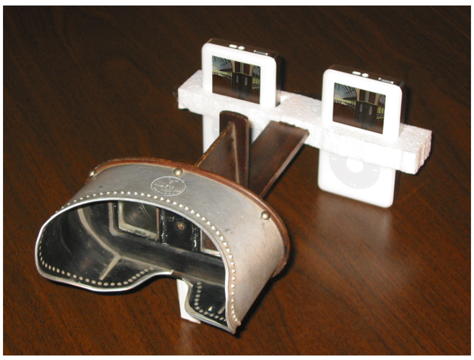


Fig. 119. — Stéréoscope à miroirs de M. Wheatstone.

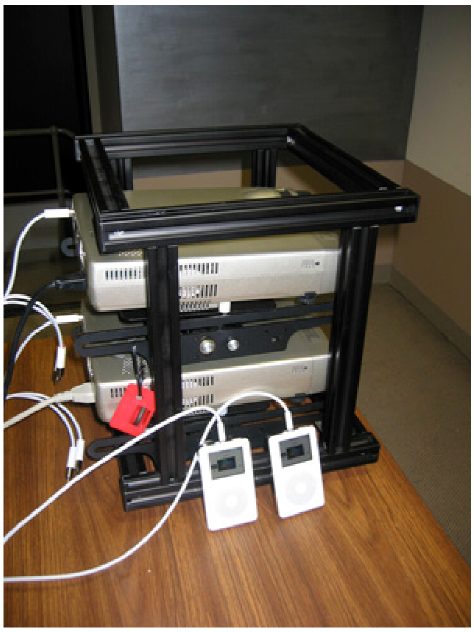


iPod Stereoscope

1905



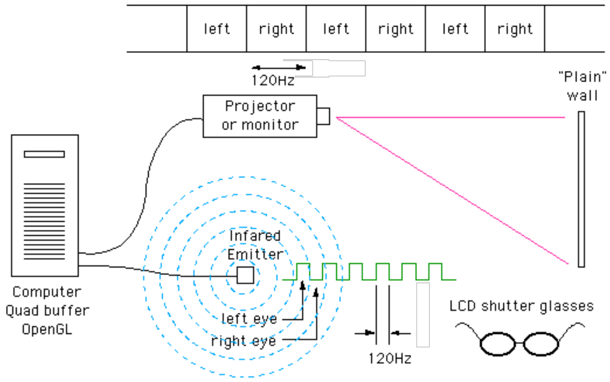
2005



Worlds smallest stereoscopic playback system

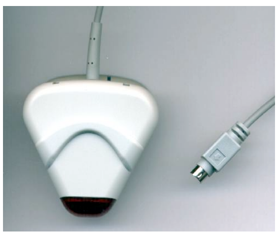
Frame sequential (also known as quad buffer)

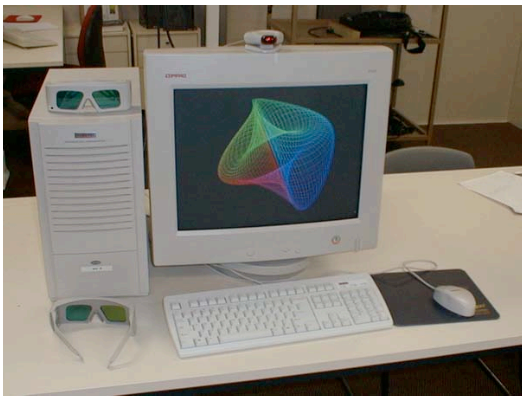
- This has been the main stereoscopic presentation method over the last 20 years.
- Left and right eye images multiplexed in time, typically at 120Hz.
- Typically employs shutter glasses, synchronised directly or through an infared emitter by a signal from an additional port on the graphics card.
- Suitable for monitors (CRT only) or CRT projectors (there are however now some digital displays and high end projectors capable of the required refresh rates).
- Eye strain from flicker (blanking signal in CRT technology), imperfect LCD extinction, and phosphor decay (varies with phosphor colour).



Shutter glasses and desktop arrangement

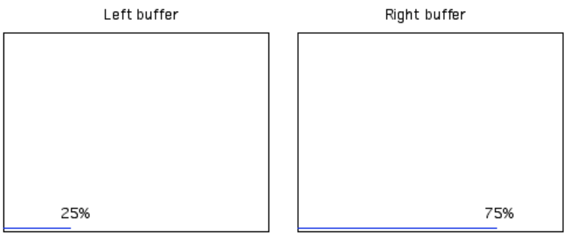
- Still the most prevalent stereoscopic presentation system for desktop displays. Note that LCD/Plasma displays generally only switch at or around 60Hz.
- Requires particular graphics cards, namely those with sync port for the glasses or emitter. These cards also have left and right buffers that can be written to by OpenGL applications before a swapbuffer is called.
- left&right back buffer, left&right display = quad buffer.

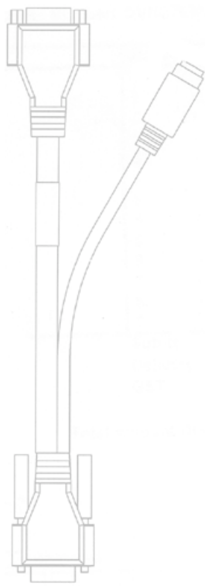




Frame sequential: Blue line

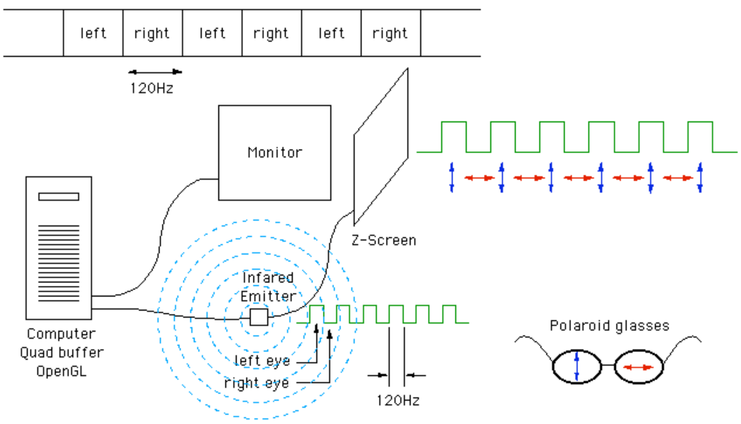
- Low cost alternative for cards that might support high refresh rates but don’t support a synchronisation port for the glasses.
- Pretty much limited to custom software solutions.
- Not recommended, not fashionable .... only included for completeness.





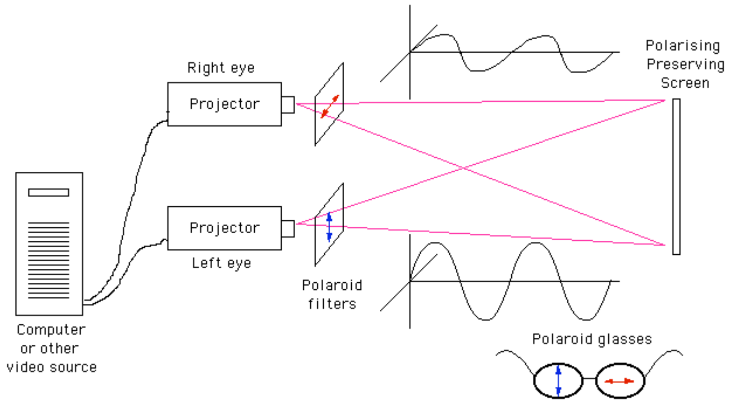
Z-Screen

- A transparent cover that is placed in front of the monitor or projector.
- Alternates the circular polarisation state of light.
- Suffers from flicker in the same way as the active shutter glasses for frame sequential stereo.
- Computer drives frame sequential images, user wears circular polaroid glasses.
- In the early days this was quite popular for monitor based stereo viewing, now mostly only used in digital stereoscopic capable theatres.

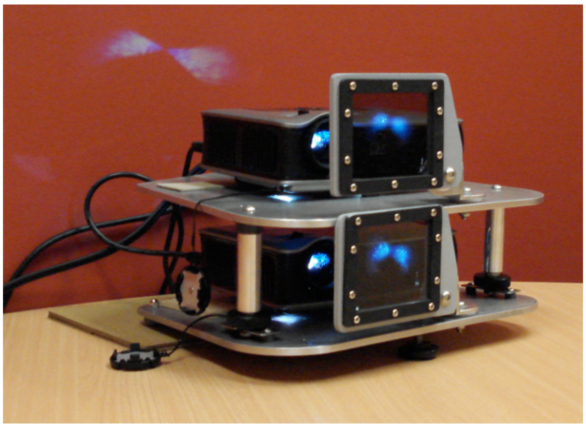


Polaroid filters (also known as passive stereo)

- Perhaps now the most commonly used stereoscopic projection method.
- Lowest cost option employing orthogonally aligned filters on the projectors and in the glasses.
- Only (normally) used for projected stereo, typically using 2 projectors, but desktop units do exist.
- Two options, linear and circular polaroid. Linear limits head tilt, circular offers higher ghosting levels due primarily to wavelength dependence of 1/4 wave plate layer on the filter.
- Requires a special screen surface that preserves polarisation on reflection (eg: silver screens for front projection) or transmission.



Hardware examples



Side-by-side stereoscopic projection

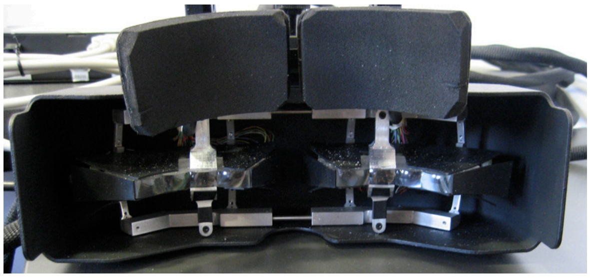
- Commonly used for passive stereo projection systems.
- Doesn't require a card with high refresh support or emitters.
- Does require either two video outputs, either one card with two display outputs, or more recently the Matrox dual-head-to-go splitters (Used for this workshop).
- OpenGL software model is straightforward, one simply draws the left and right images to the corresponding halves of the combined displays.
- Advantages over frame sequential is that some things are easier: slideshows and movies can be played with non-stereo specific software, eg: QuickTime.
- Disadvantage compared to frame sequential is the operating system cannot (generally) be viewed at the same time.
- Disadvantage that most existing stereoscopic software only frame sequential approach. nVidia have a special driver mode that allows one to use frame sequential software. The driver spits the left/back and right/back to two output ports. There are also some (expensive) splitter units that do the same thing in hardware.

Infitec (Interference filter technology)

- One of the more recent technologies, based upon sharp cut-off RGB filters in front of the projectors and in the glasses. Only been on the market for perhaps 3 years but was used in house at DaimlerChrysler Research much earlier.
- Has the key advantage of very low ghosting levels compared to all other techniques.
- Does not require a special screen surface.
- Main disadvantage is a careful colour calibration is required on a regular basis over the life of a projector globe.
- The Infitec filters are now offered internal with a few projector suppliers.
- Glasses are not "commodity", priced at about the same levels as active shutter glasses.
- Still employs 2 projectors so software is either frame sequential based with splitter or the nVidia option, or side-by-side display.

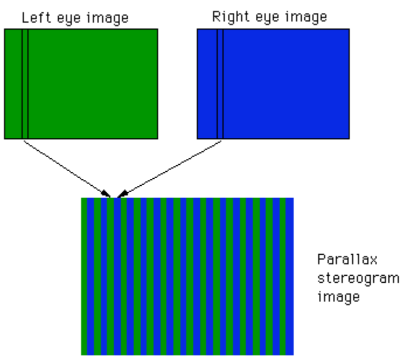
HMD: Head Mounted Displays

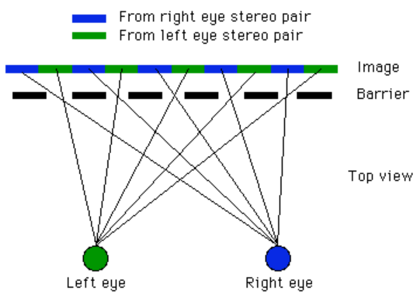
- Small, typically LCD panels in front of each eye along with focusing lens.
- Commodity priced HMDs have historically been low resolution, not the case now with 1280x1024 per eye quite common.
- Usually feels like tunnel vision due to a narrow field of view.
- Additional panels to support wide field of view exist, at a price.
- Often contains head tracking unit so the view can be naturally adjusted as the wearer moves their head.



Autostereoscopic (Glasses free)

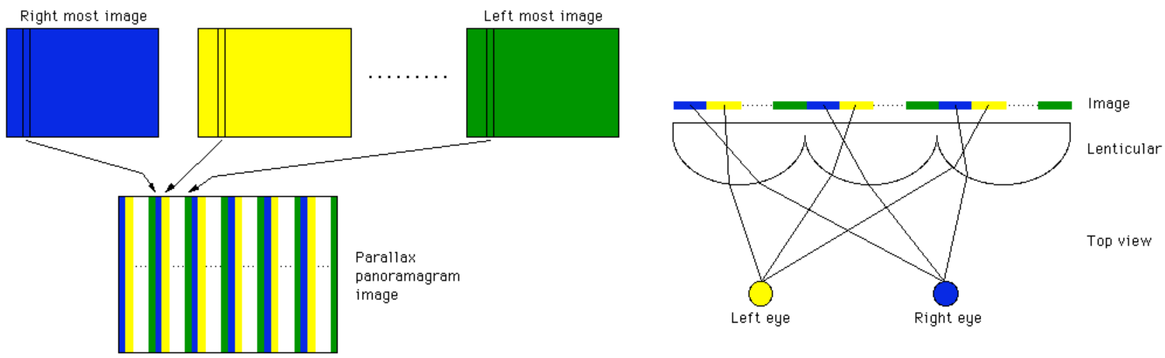
- Lots of options but basic idea is grasped by considering barrier strip displays.
- Limited viewing angle, a function of how many images are used.
- Limited resolution, inversely a function of the number of images used.
- Often camera head tracking is used to align the images correctly depending on the observers head.
- Can additionally be used for autostereoscopic prints.





Autostereoscopic: Lenticular

- Lenticular systems give a wider viewing angle, but there are still dead viewing zones in between the transition between the leftmost and rightmost image.
- Can be viewed by a number of people at once, although each needs to be in a correct zone.
- Incorporates more than just two images.
- Very limited resolution, vertical pixilation is usually very obvious.



	Z-Screen	Frame sequential	Passive	Infitec	HMD	Auto-stereoscopic
Ghosting	Yes	Yes	Yes	Essentially none	None	NA
Flicker	Yes	Yes	None	None	None	None
Glasses	Circular polaroid, low cost	LCD, active electronics	Polaroid, low cost	Optical elements	Yes	None
Resolution	Average	Average	Range	Range	Average	Generally low
Special requirements	CRT technology (mostly)	CRT technology (mostly)	Special projection screen surface	Proprietary	Single person	Limited viewing positions
Cost	Medium	Range	Range	High	Range	Range
Suitability for public exhibit	Low	Low	Highest	Low	Low	High
Suitability for research	Average	Average	Average	High	Medium	Low
Suitability for cinema	High	Average/low	High	High	Low	Low
Suitability for gaming	Low	Most suitable for monitor	Average	Low	Average	Average

Sources and discussion of eye strain in projection technologies

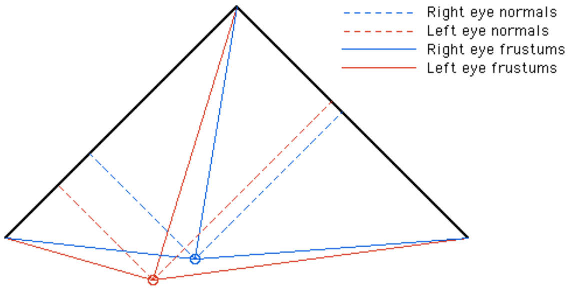
- Ghosting.
  Common to almost all technologies except Infitec.
- Flicker.
  Mainly relevant to frame sequential systems.
  Not present in current digital projectors without blanking signal of CRTs.
- Lack of consistent accommodation cues.
  Common to all stereoscopic systems.
- Discrete time steps.
  25 or 30 fps is often not fast enough to capture temporal information.
  Current digital projection typically runs at 60Hz.
- Resolution.
  Limits parallax dynamic range as well as impacting on image quality.
  Primarily an issue with autostereoscopic displays.
- Viewing position error.
  Non tracked systems and autostereoscopic displays.
- Non colour or brightness matched displays/projectors.

Special topics - Some of my current related projects

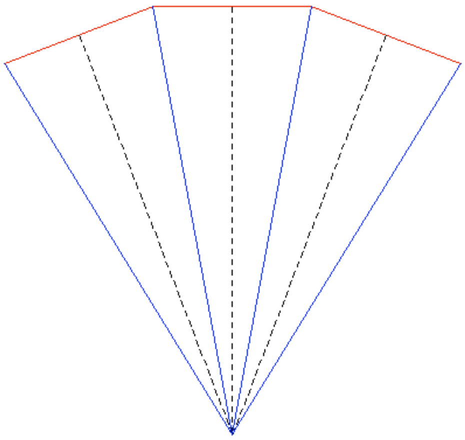
- Multiple screens, implications for projection frustums.
- Stereoscopic panoramic pairs, example by Peter Morse from Antarctica.
- Tiled stereoscopic displays.
- Reconstruction from stereo pairs, examples from SiroVision.

Multiple stereoscopic screens: immersive environments

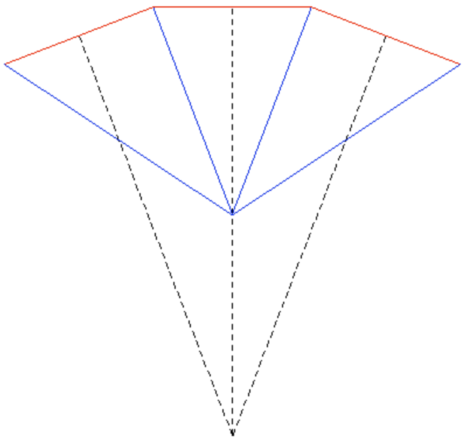
- Extreme case of offaxis frustums. Head tracking critical for a non-distorted view, precludes multiple participants (at least if they wish to look around independently with an undistorted view).
- Issue with polaroid based systems: loss of polarisation preservation with incident angle.
- Issues with projection technologies: light bleed and reflections between screens at an angle.
- Can the seams be made invisible?



Three wall example (Frustums for only one eye shown)

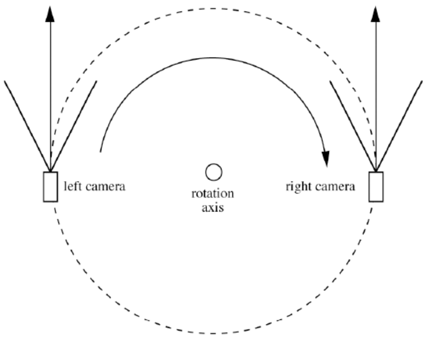


There is only one position where symmetric frustums can be used.

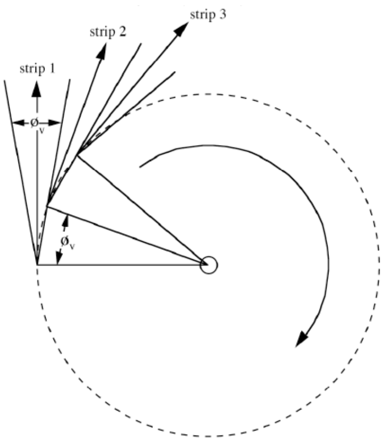


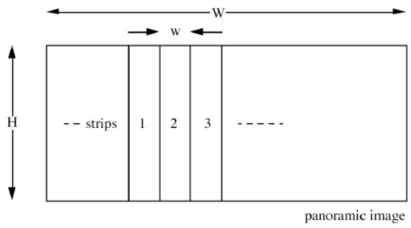
If the intention is for multiple viewers then a "sweet" spot is usually chosen for three asymmetric frustums (per eye).

Stereoscopic panoramic pairs



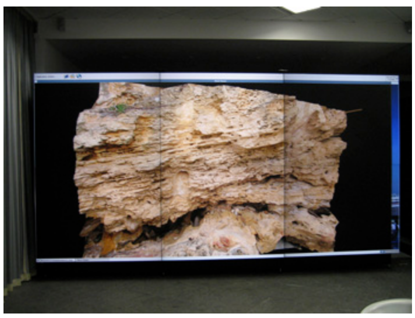
- Two cameras rotating about their common center.
- Ideally capture as slit cameras.
- Alternatively stitch the two sets of images to form the two panoramic images.
- Care needs to be taken over exposure and lighting.
- Obviously issues of movement in the scene.





Tiled stereoscopic displays

- Even very modest seams anchor the images to the 2D display depth.
- Acceptable for positive parallax, generally unacceptable for negative parallax.
- Still requires offaxis frustums but now they are offaxis vertically as well as horizontally.



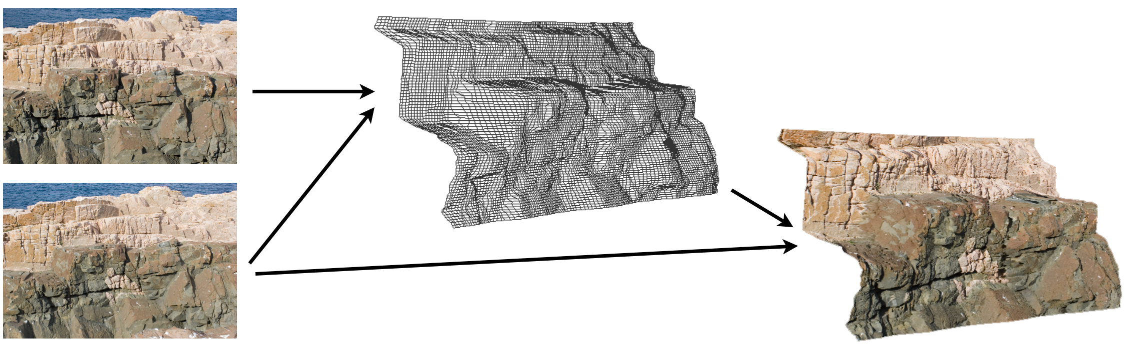
4K x 2K pixel display at IHPC, Singapore. 3x2 tile

Example: Boolardy (Peter Morse)

Left eye

Right eye

Reconstruction from stereo pairs



- Use the parallax information in a stereo pair to reconstruct the geometry of the terrain, one image can then be draped over the mesh to give a realistic 3D model. Which in turn can then be viewed in stereo.
- Representative of a whole range of technologies in topography and machine vision.
- In general, stereo pairs captured for this purpose are not suitable for viewing! The main difference is the camera separation required for successful surface reconstruction is much wider than is suitable for viewing.