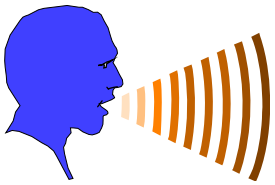





Video Processing & Communications

Stereo and 3D Video

Yao Wang
Polytechnic Institute of New York University
Brooklyn, NY11201
(with significant contribution from Amy Reibman)

Ranges of applications for 3D

- Understanding the world around us, using vision systems
  - Acquisition and interpretation of 3D
- Communicating, depicting the world around us
  - Stereo, free viewpoint TV
  - Telling a more compelling story

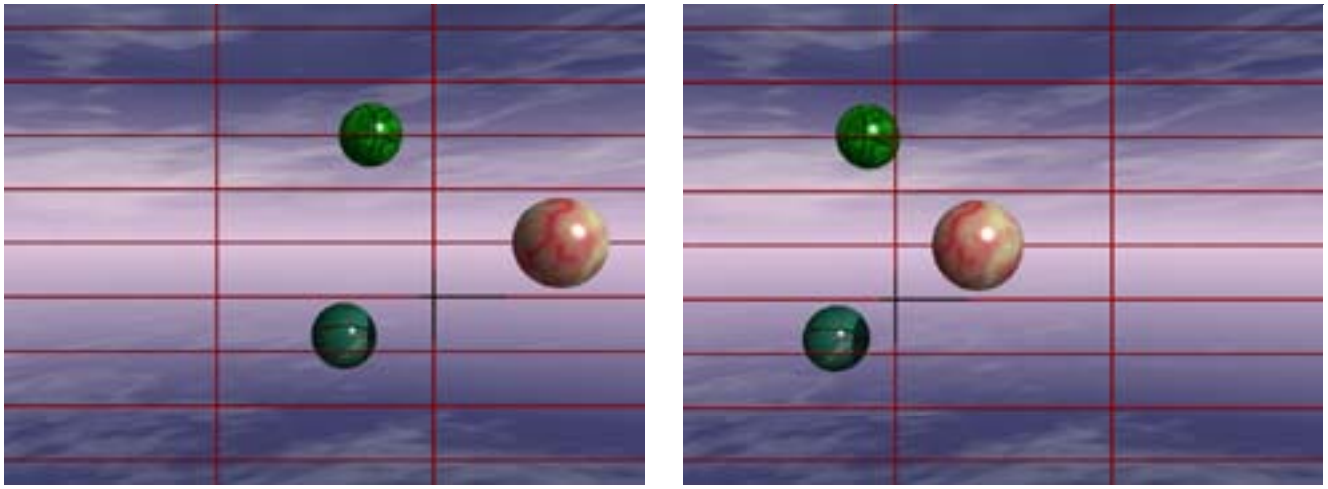
Outline

- Understanding the 3D world: Geometry
  - Acquisition: geometry of parallel and converged cameras
  - Depth from disparity
  - Finding correspondences; disparity estimation
  - Rectification
  - Intermediate view synthesis
- Human depth perception
  - Depth cues
  - Fusion
  - What causes eye fatigues?
  - Other related phenomena
- Displays: stereo, autostereoscopic, volumetric
- Communication: Stereo and multiview sequence coding
  - MPEG-2 Temporal scalability and H.264 Multiview coding

Depth Perception by Stereopsis

- Human eye perceives depth by having two eyes with slightly shifted views
  - The shift is called “disparity”
  - Perceived depth depends on the the “disparity”
- There are other monocular cues for depth perception
  - E.g. far away objects look smaller

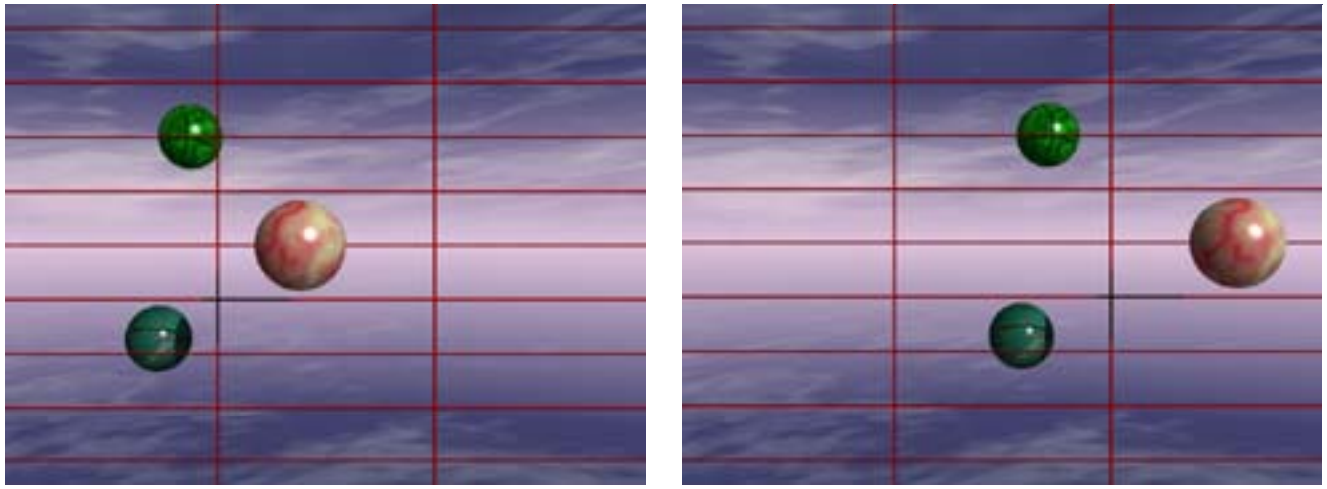
A Visual Experiment



Try to look at the left and right images using your left and right eyes separately while try to merge the two images into one. Can you tell which ball is closer?

Pictures generated by ray-tracing. Courtesy of Rushang Wang

A Visual Experiment



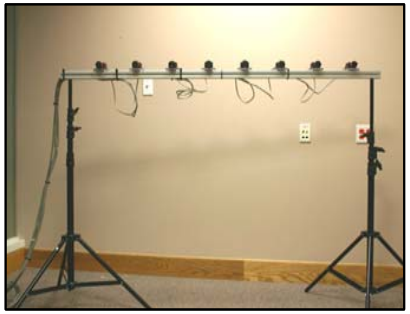
Try to look at the left and right images by crossing your left and right eyes while trying to merge the two images into one. Can you tell which ball is closer?

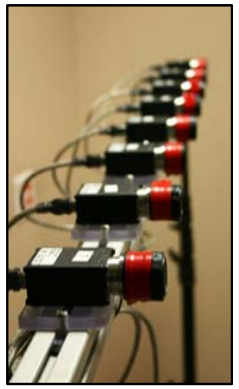
Different people merge left and right views differently: wall or cross

Pictures generated by ray-tracing. Courtesy of Rushang Wang

Stereo Imaging Configuration: Acquisition

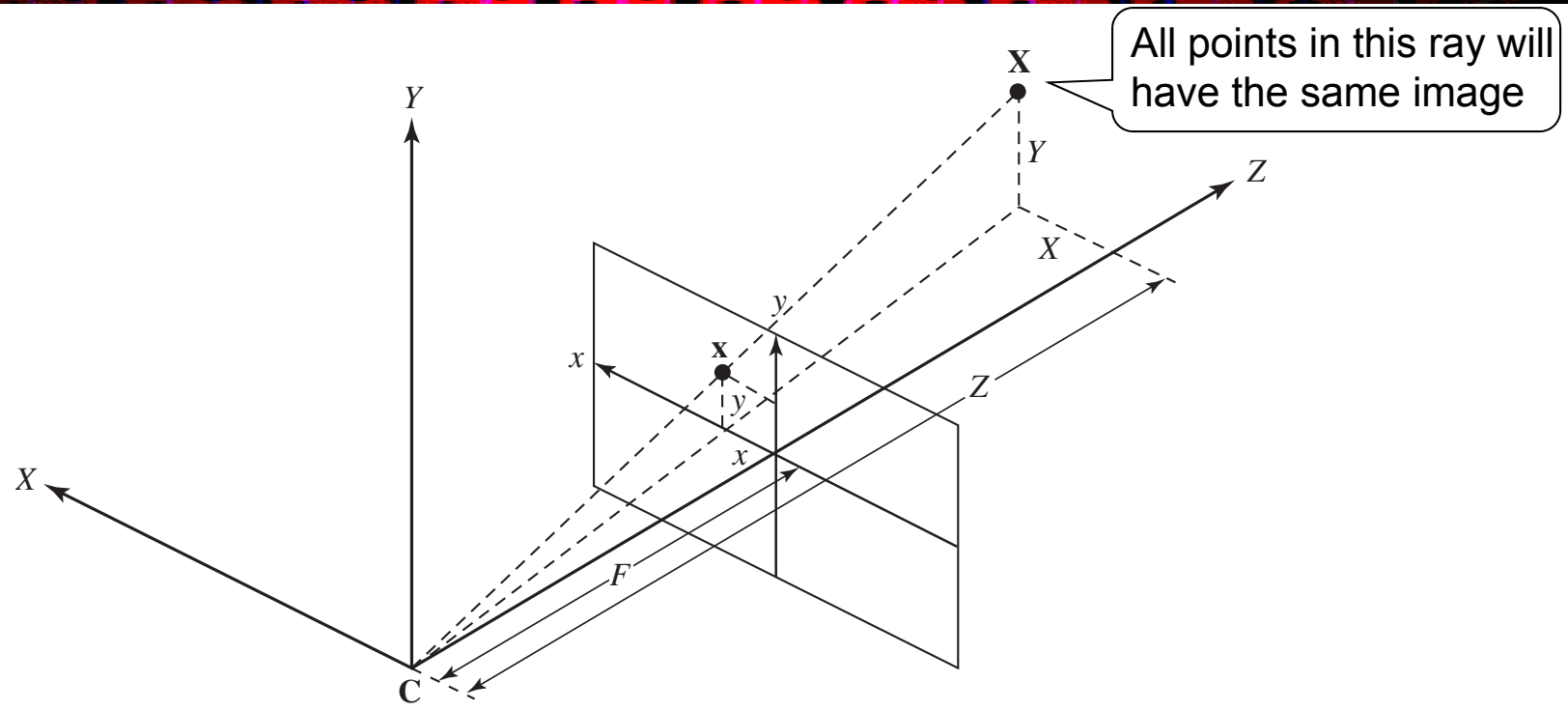
- Live action
  - Parallel configuration
  - Converging





- Animation
  - Computer generated animation uses native 3D modeling
  - Additional view(s) can easily be rendered

Perspective Projection Revisited

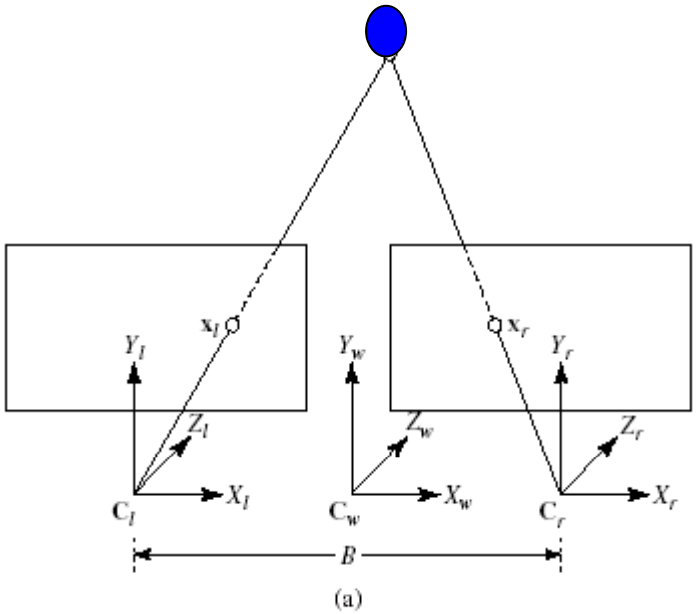


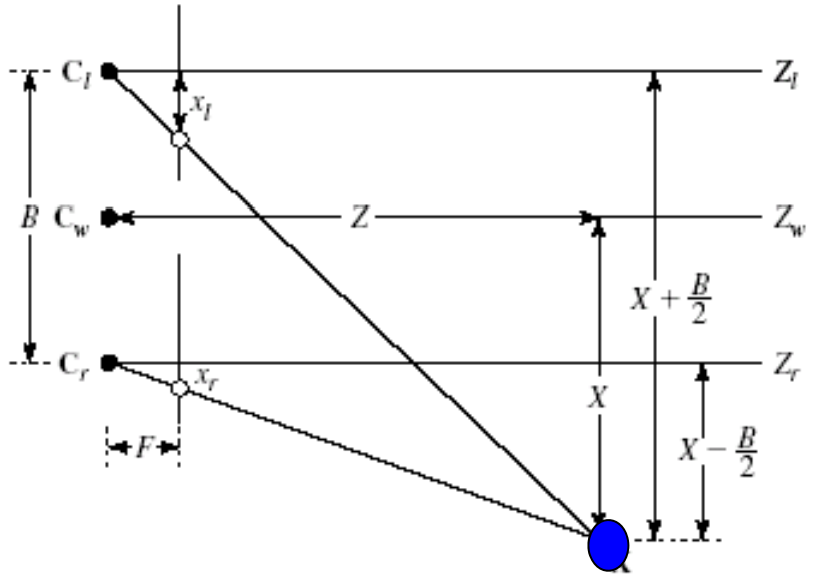
(a)

$$\frac{x}{F}=\frac{X}{Z},\frac{y}{F}=\frac{Y}{Z}\Rightarrow x=F\frac{X}{Z},y=F\frac{Y}{Z}$$

x, y are inversely related to Z

Parallel Camera Configuration





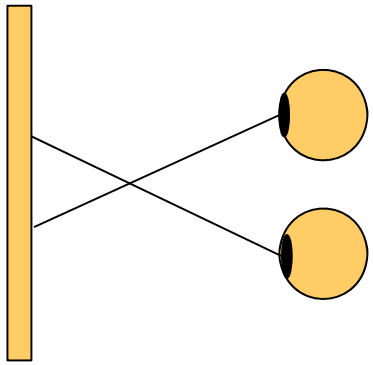
$$X_l = X + \frac{B}{2}, \quad X_r = X - \frac{B}{2}, \quad Y_l = Y_r = Y, \quad Z_l = Z_r = Z;$$

$$x_l = F\frac{X + B/2}{Z}, \quad x_r = F\frac{X - B/2}{Z}, \quad y_l = y_r = y = F\frac{Y}{Z}.$$

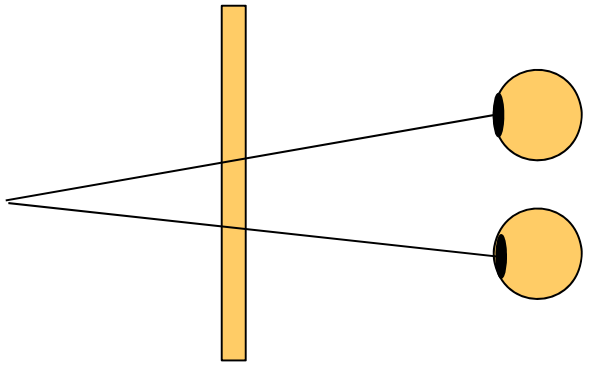
$$d_x = x_l - x_r = \frac{FB}{Z}.$$

i) **Only horizontal disparity**
ii) **Disparity is inversely proportional to Z**
iii) **Range of disparity increases with B**

Disparity and Depth

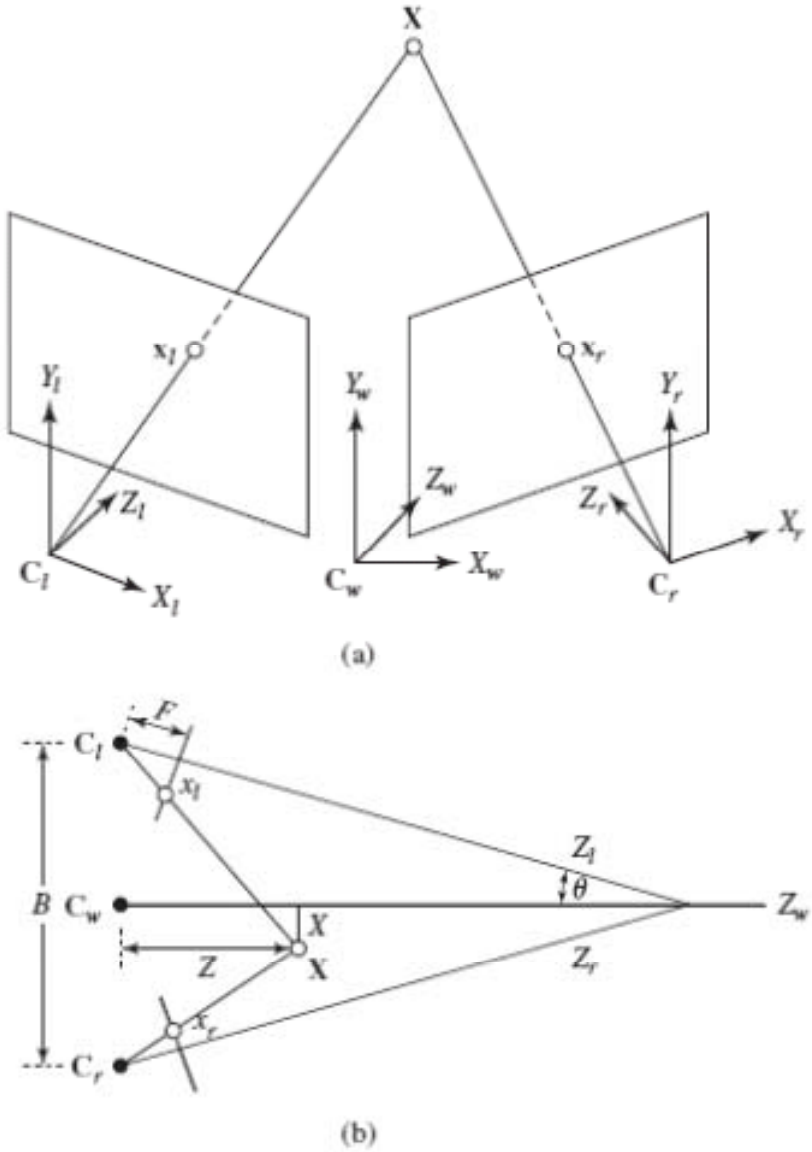


Negative disparity:
Object in front of the screen



Positive disparity
Object behind the screen

Convergent Camera Configuration



$$x_l = F\frac{\cos\theta(X + B/2) - \sin\theta Z}{\sin\theta(X + B/2) + \cos\theta Z},$$

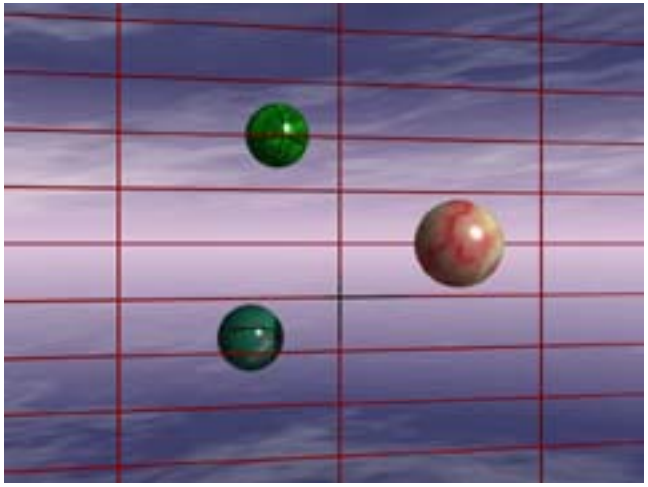
$$x_r = F\frac{\cos\theta(X - B/2) + \sin\theta Z}{-\sin\theta(X - B/2) + \cos\theta Z},$$

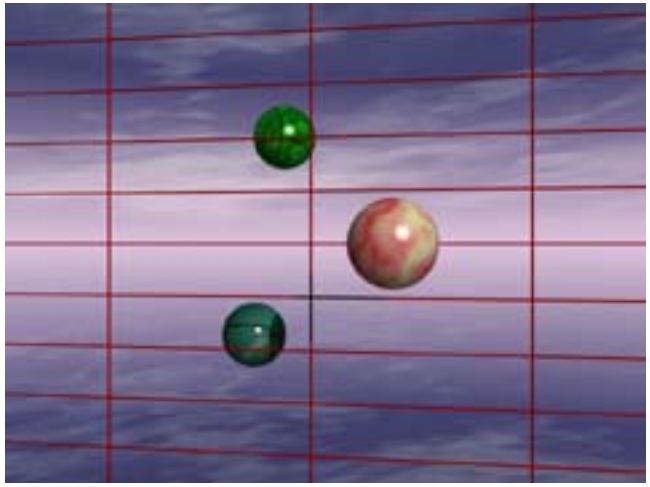
$$y_l = F\frac{Y}{\sin\theta(X + B/2) + \cos\theta Z},$$

$$y_r = F\frac{Y}{-\sin\theta(X - B/2) + \cos\theta Z}.$$

both horizontal and vertical disparity

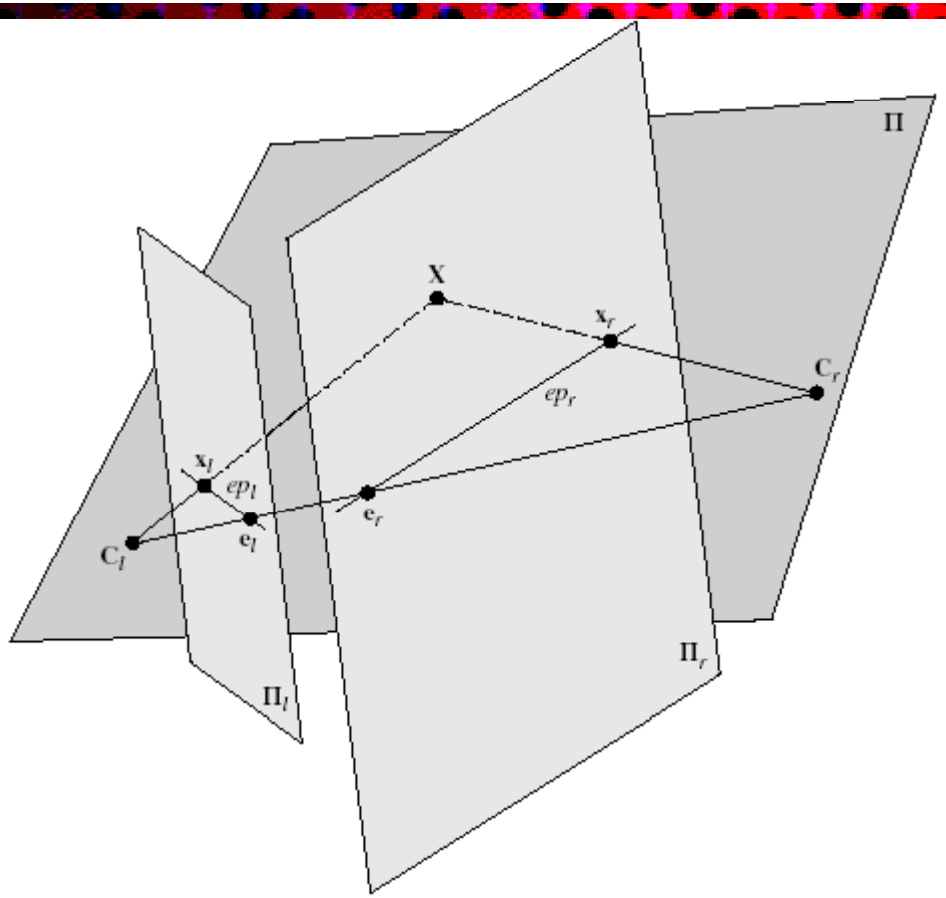
Example Images (Converging Camera)





Notice the keystone effect
Can get a better depth perception of objects closer to the camera than with the parallel set up
But when displayed using a parallel projection system and viewed by the human eye, the vertical disparity causes perceptual discomfort. Geometric correction is needed before displaying these images.

Epipolar Geometry: Arbitrary Case



Π: Epipolar plane

ep $_{l,r}$: Epipolar lines

$$\tilde{\mathbf{x}}_r^T [\mathbf{F}] \tilde{\mathbf{x}}_l = 0, \quad \tilde{\mathbf{x}}_l^T [\mathbf{F}]^T \tilde{\mathbf{x}}_r = 0.$$

[F]: fundamental matrix, Depends on camera setup

Epipolar constraint: For any point that is on the left epipolar line in the left image, its corresponding point in the right image must be on the epipolar line, and vice versa.

Epipolar Geometry: Parallel Case

- Epipolar constraint: the corresponding left and right image points should be on the same horizontal line (only horizontal disparity exists)

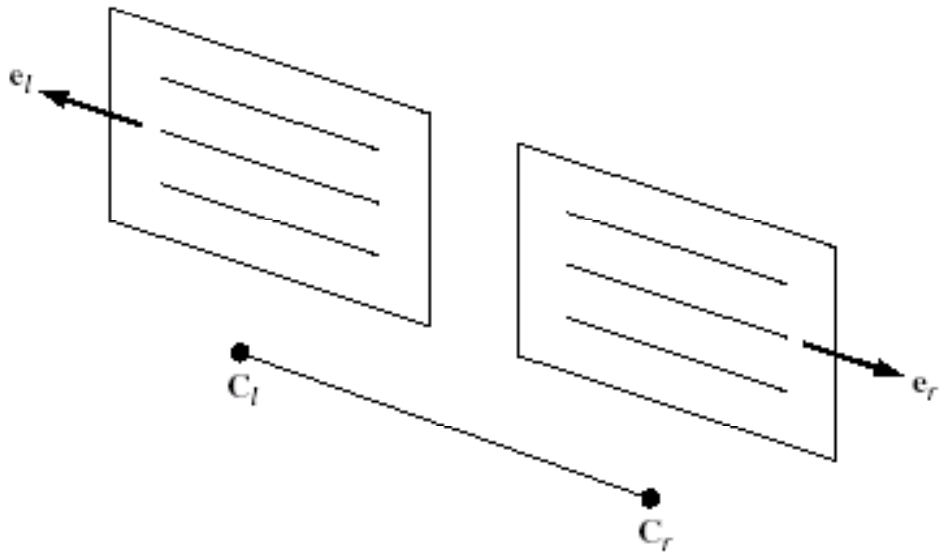


Figure 12.8 Epipolar geometry for a parallel camera configuration: epipoles are at infinity, and epipolar lines are parallel. Adapted from O. Faugeras, *Three-Dimensional Computer Vision—A Geometric Viewpoint*, Cambridge, MA: MIT Press, 1993. Copyright 1993 MIT Press.

Rectification: creation of images as if acquired from parallel cameras

Disparity Estimation

- Disparity Estimation Problem:
  - For each point in one (anchor) image, find its corresponding point in the other (target) image – Similar to motion estimation problem
  - Parallel configuration: only horizontal disparity needs to be estimated.
  - Difference from motion estimation: Has more physical constraints; disparity range may be very large for objects nearby (up to 40-50 pixels)
  - Problem set-up:
    - Determine disparity at every pixel to minimize Disparity Compensated Prediction (DCP) error
- Constraints for disparity estimation
- Block-based disparity estimation
- Mesh-based disparity estimation
- 3D structure estimation

Constraints for Disparity Estimation

- Epipolar constraints:
  - For parallel set up: two corresponding points are in the same line, only horizontal disparity need to be searched
  - For an arbitrary camera set up: given x_r, possible x_l sits on a line (epipolar line)
- Ordering constraint:
  - If two points in the right image are such that $x_{r,1} < x_{r,2}$,
  - Then the corresponding two points in the left image satisfy $x_{l,1} < x_{l,2}$.
- Models for disparity functions

Models for Disparity Functions

- Affine model for plane surface, parallel set-up:
  - If an imaged object has a plane surface

$$Z(X, Y) = aX + bY + c.$$

then the disparity function for points on the surface satisfies affine model:

$$d_x(x_r, y_r) = \frac{1}{c/B + a/2}(F - ax_r - by_r).$$

- For an arbitrary scene, we can divide the reference (right) image into small blocks so that the object surface corresponding to each block is approximately flat. Then the disparity function over each block can be modeled as affine.
- Using similar approach, can derive models for curved surfaces (higher order polynomial)

Block-Based Disparity Estimation

- Following the method for block-based motion estimation
  - Divide the anchor image into regular blocks
  - Assume disparity function over each block is constant or an affine function
  - Determine the constant or the affine parameters
  - For parallel set up: Only1 horizontal disparity or 3 affine parameters
- Difference from motion estimation
  - Constant disparity model is less effective than constant motion model even over small blocks
  - Affine model is quite good
  - Need a large search range to account for disparities in objects nearby

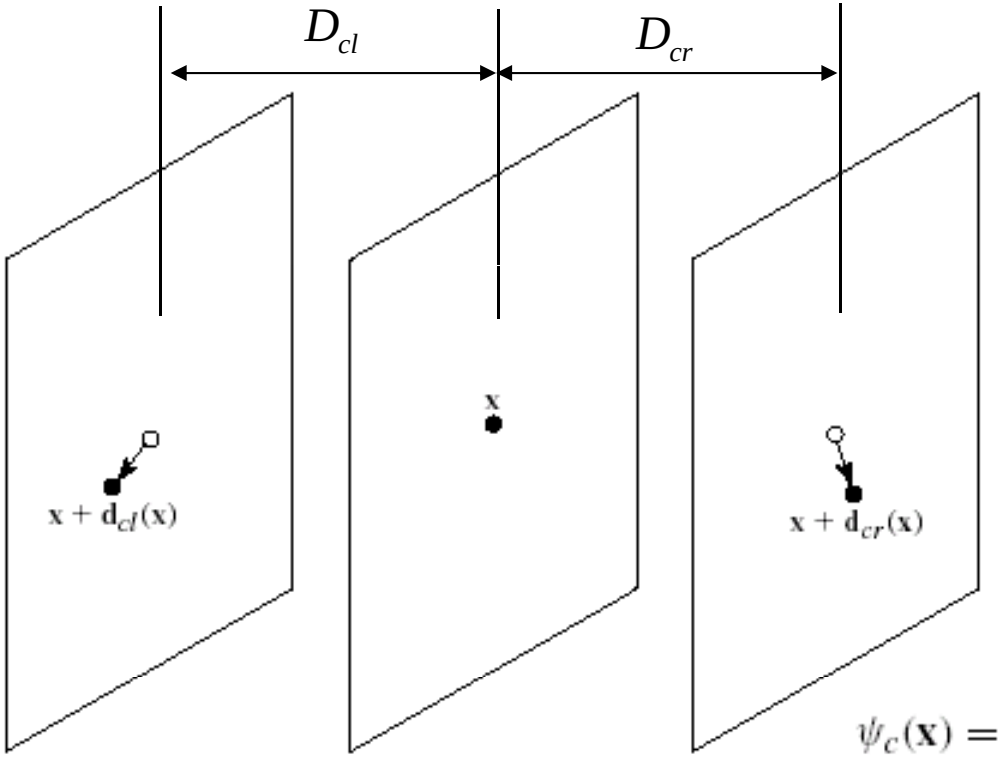
Intermediate View Synthesis

- Problem:
  - Interpolate intermediate views from given views
  - Necessary for virtual reality applications
- Linear interpolation: can lead to blurred images

$$\psi_c(\mathbf{x}) = w_l(\mathbf{x})\psi_l(\mathbf{x}) + w_r(\mathbf{x})\psi_r(\mathbf{x}).$$

- Disparity-compensated interpolation

Disparity Compensated View Synthesis



Baseline distances:
Dcl and Dcr

$$\psi_c(\mathbf{x}) = w_l(\mathbf{x})\psi_l(\mathbf{x} + \mathbf{d}_{cl}(\mathbf{x})) + w_r(\mathbf{x})\psi_r(\mathbf{x} + \mathbf{d}_{cr}(\mathbf{x})).$$

Figure 12.13 Disparity-compensated interpolation: x is interpolated from $\mathbf{x} + \mathbf{d}_{cl}(\mathbf{x})$ in the left view and $\mathbf{x} + \mathbf{d}_{cr}(\mathbf{x})$ in the right view.

$$w_l(\mathbf{x}) = \begin{cases} \frac{D_{cr}}{D_{cl}+D_{cr}}, & \text{if } \mathbf{x} \text{ is visible in both views,} \\ 1, & \text{if } \mathbf{x} \text{ is visible only in the left view,} \\ 0, & \text{if } \mathbf{x} \text{ is visible only in the right view.} \end{cases}$$

Problem

- How to determine disparity from the central (unknown) view?
- One approach:
  - First determine disparity between left and right for every pixel in the left d_lr(x_l)
  - Then determine disparity between left and central based on distance, d_lc(x_l)=D_cl/(D_cl+D_cr)
- For every point x_l in left, find corresponding point in central x_c=x_l+d_lc(x_l) and right x_r=x_l+d_lr(x_l), interpolate central point
- But the central point may not be an integer pixel!
- When using block-based method for estimating d_lr, there may be uncovered points in the central view or multiple-covered points; a dense depth field is better

Perception of Depth

- Monocular cues:
  - Shape/size
  - Occlusion (one object blocks another)
  - Shading and texture
  - Linear perspective (think railroad tracks)
  - Relative height (with respect to the horizon)
  - Motion parallax
  - Aerial haze (blueness on the horizon)
- Motion cues
  - motion parallax
- Binocular cue: Stereopsis
  - The use of two images (or their disparity) to form a sense of depth

Monocular Depth Cues

- Interposition
  - An object that occludes another is closer
- Shading
  - Shape info. Shadows are included here
- Size
  - Usually, the larger object is closer
- Linear Perspective
  - parallel lines converge at a single point
- Surface Texture Gradient
  - more detail for closer objects
- Height in the visual field
  - Higher the object is (vertically), the further it is
- Atmospheric effects
  - further away objects are blurrier
- Brightness
  - further away objects are dimmer

From Seeing 3D from 2D Images

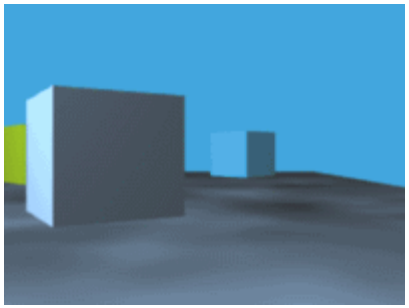
Monocular Cues

- What can you tell from each image?

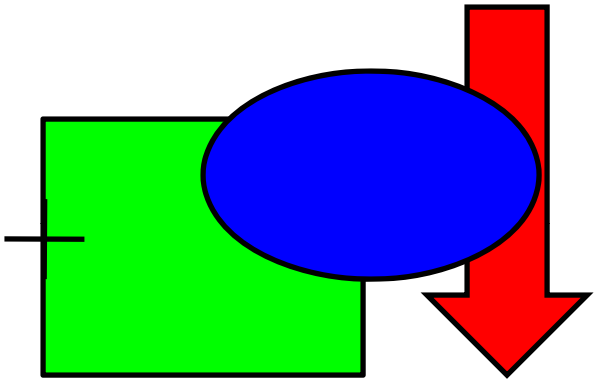
Motion Parallax

- Motion parallax: As we move, objects that are closer to us move farther across our field of view than do objects that are in the distance. The animation below attempts to demonstrate how motion parallax works for driving along the road
  - http://psych.hanover.edu/krantz/motionparallax/motionparallax.html

- Another example :
  - http://en.wikipedia.org/wiki/Parallax

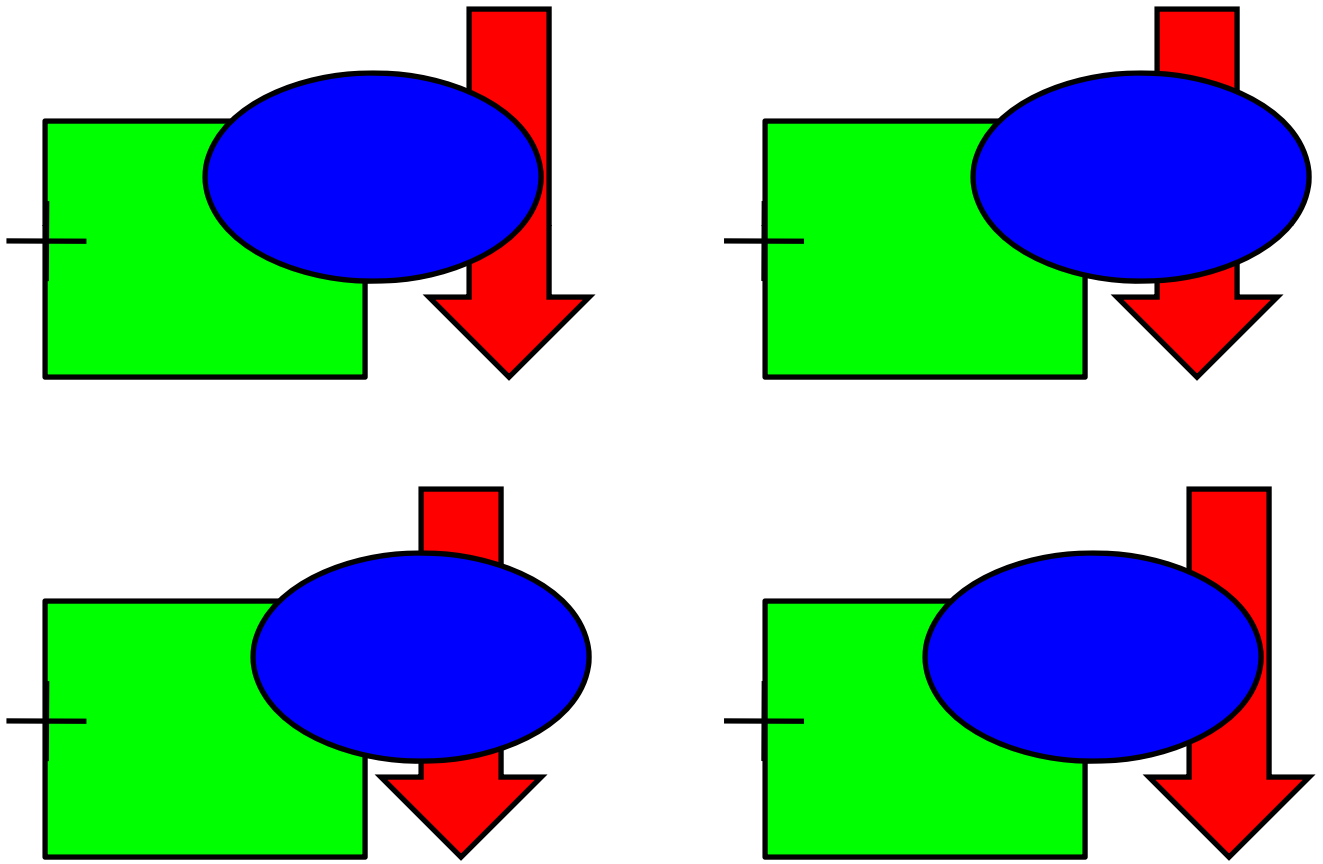


Occlusion cues only



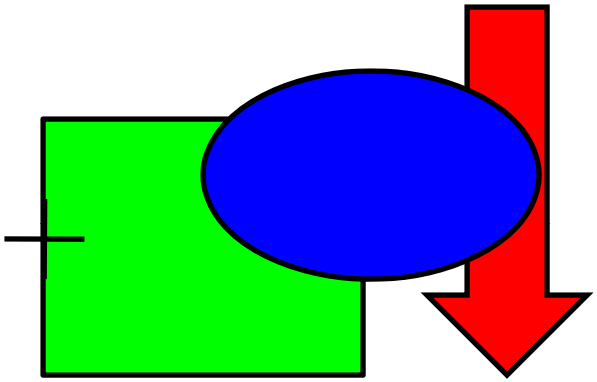
The occlusion of blue over red and green tells us that blue is closer.
But what about green vs. red?

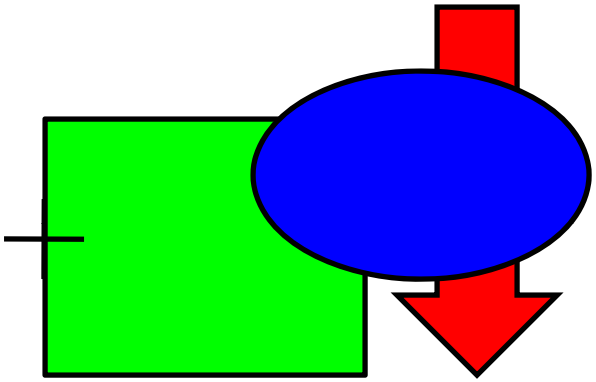
Stereo and occlusion cues



Stereo and occlusion cues

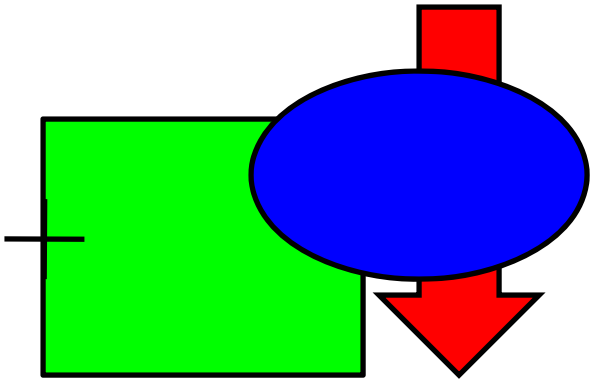
Cross:
consistent

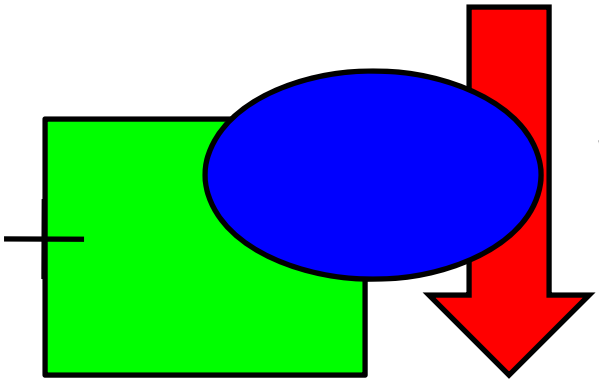




Wall:
conflicting

Cross:
conflicting





Wall:
consistent

Stereopsis

- Stereopsis
  - The use of two images to form a sense of depth
- About 10% of viewers cannot achieve stereopsis
- Stereopsis is particularly important inside “personal space” (actual depth less than 1.5 meters)
- Stereoscopic depth discrimination is of an order of magnitude finer than visual acuity
- Motion parallax much stronger than stereopsis*
- M. Changizi hypothesizes the evolutionary reason for two eyes is NOT stereopsis, but being able to see around leaves

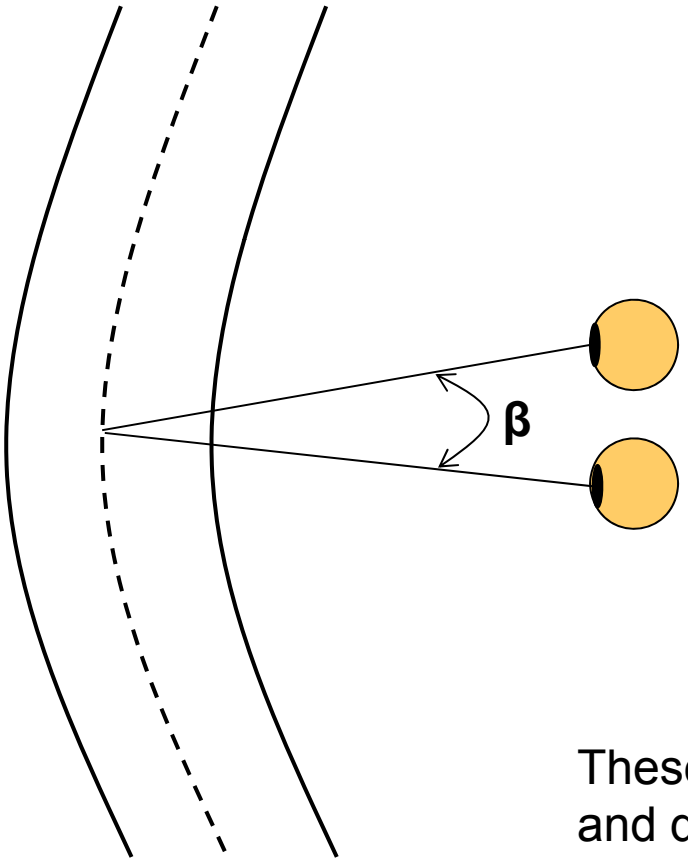
Vergence and Accommodation

- Fusion
  - Fusing Left and Right images into a single image
- Vergence: simultaneous movement of both eyes in opposite directions to obtain or achieve fusion
  - Convergence: two eyes rotate towards each other to look at a close-by object
  - Divergence: two eyes rotate away from each other to look at a far-away object
- Accommodation: change of focus as the object moves
- Changing the focus of the eyes to look at an object at a different distance will cause both vergence and accommodation.
  - Our eyes are trained to accommodate and converge in unison

Fusion and fusing image pairs

- Fusion occurs where L and R images “match” sufficiently
  - Fusion may not occur at all pixels
- Fusion depends on vergence
  - Hold up a finger and look in distance…
  - Humans are constantly adapting their two views (verging and accommodating)
- Fusing objects can be tiring, even in real life
  - If they have rapidly changing depth
  - If they are very near

Horopter and Panum's fusional area



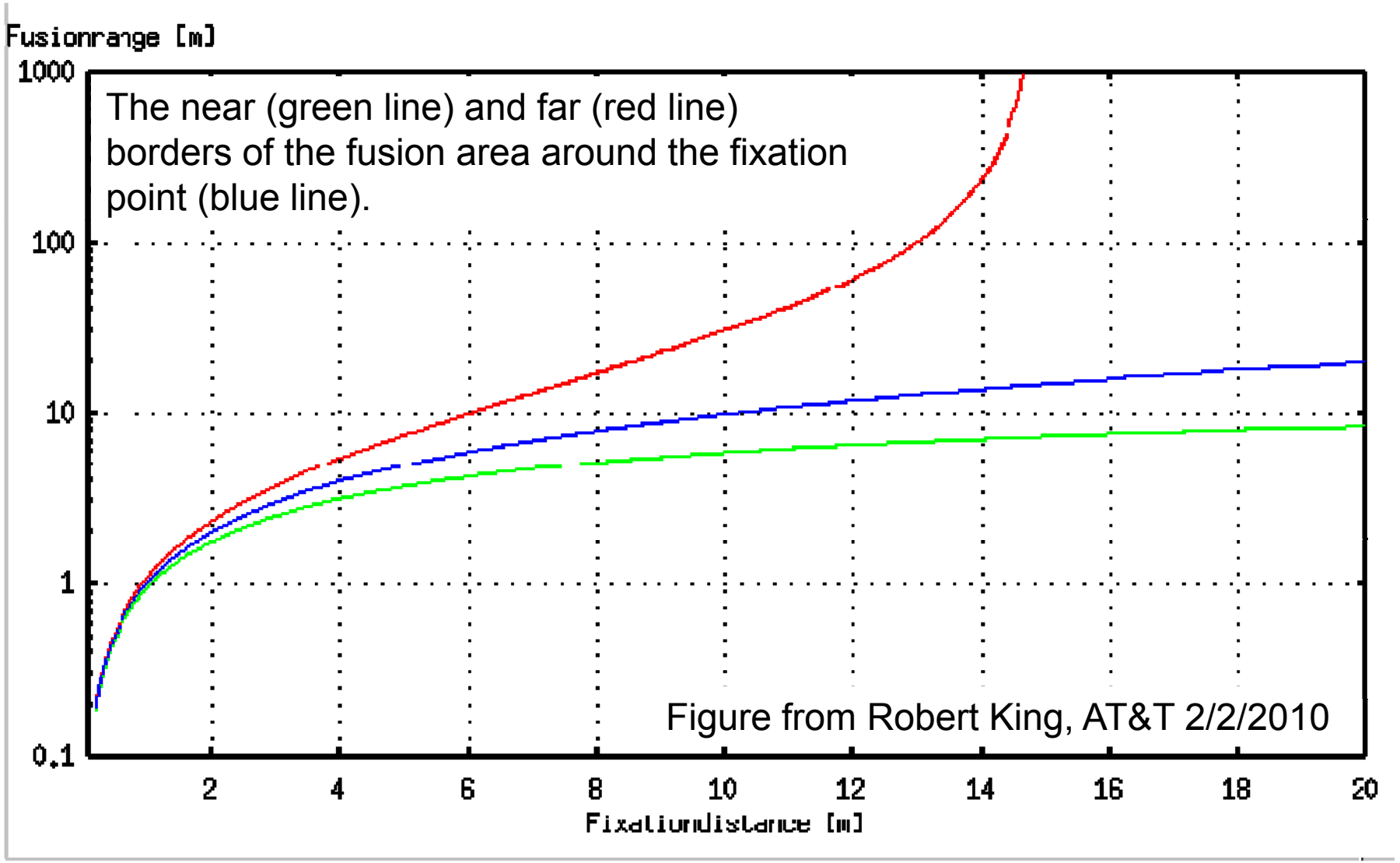
The horopter are the points in 3D space which project to corresponding images points on the retina, for a given convergence angle, β.

When our eyes are converged, we can also fuse image points not on the horopter.
Points outside Panum's fusional area are not fused and create double images.

These regions are determined experimentally and depend on the convergence angle/distance.

When creating and displaying a stereo video, the design should avoid objects outside the Panum's area, to reduce eye fatigue!

Horopter and Panum's fusional area



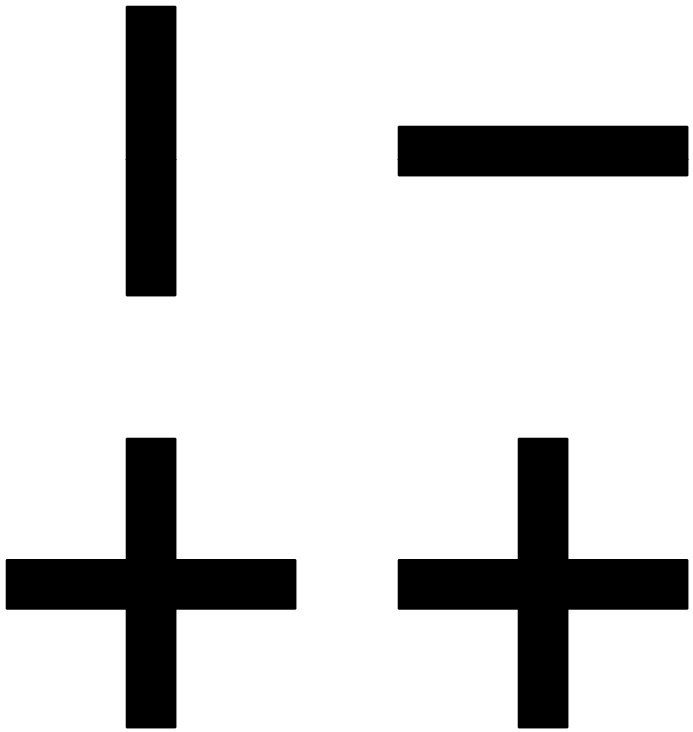
More on binocular fusion

- Lots of psychology literature on fusion
  - Hubel and Wiesel discovered binocular cells in the primary visual cortex in 1959
  - Fusion is based on information in independent spatial-frequency channels
  - Wheatstone: can fuse two objects of different sizes!!
  - Julesz: can fuse random-dot stereograms (no monocular cues)
- Factors that limit fuse-ability
  - Blurriness, spatial impairments, spatial aliasing, brightness
  - Conflicting fusion (multiple possible vergences)
    - Regular HF pattern over a bland LF background
    - A nearly-see-thru pattern on glass
    - Ambiguous figures & optical illusions

Some Related Perception Phenomena

- Binocular rivalry and suppression
- Wallpaper stereo
- Eye dominance

Binocular rivalry and suppression

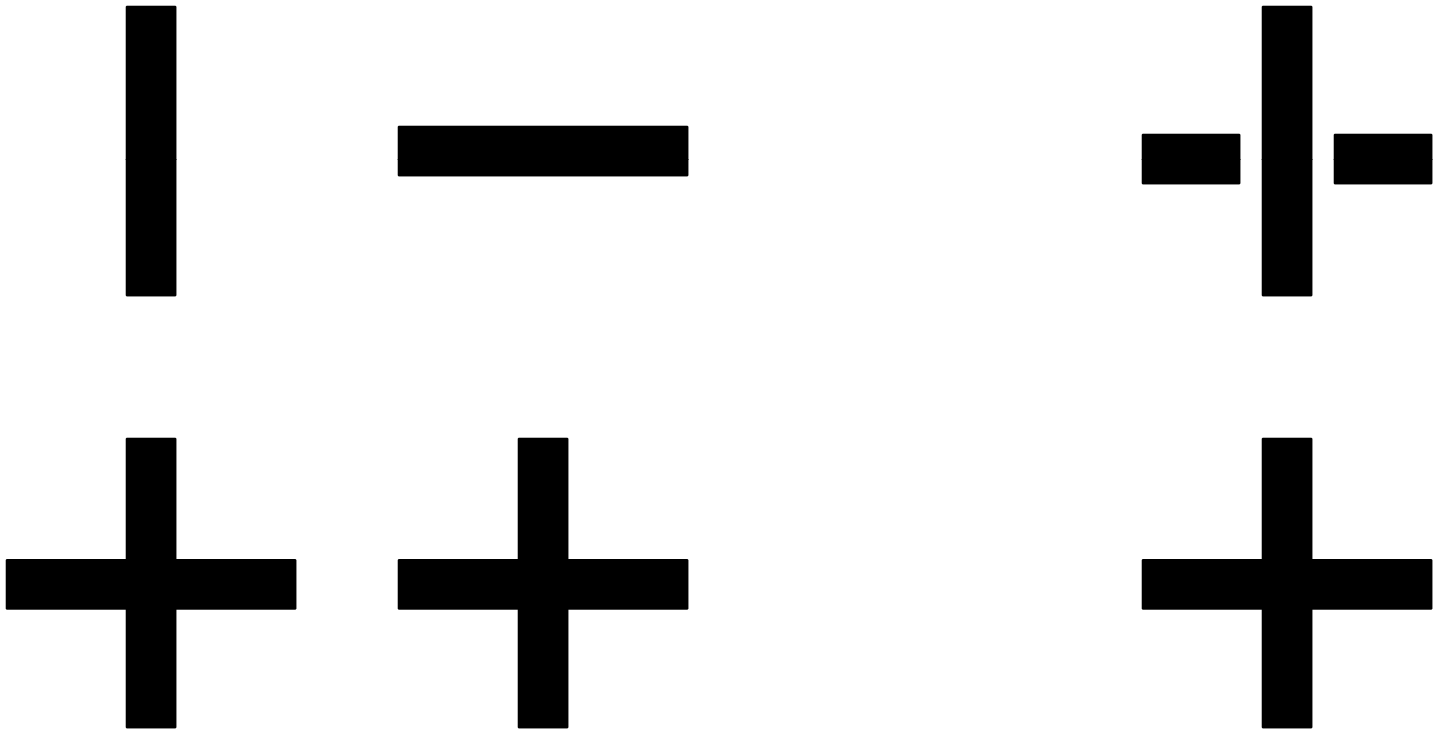


Binocular rivalry
- **Brain swaps between each eye's images**

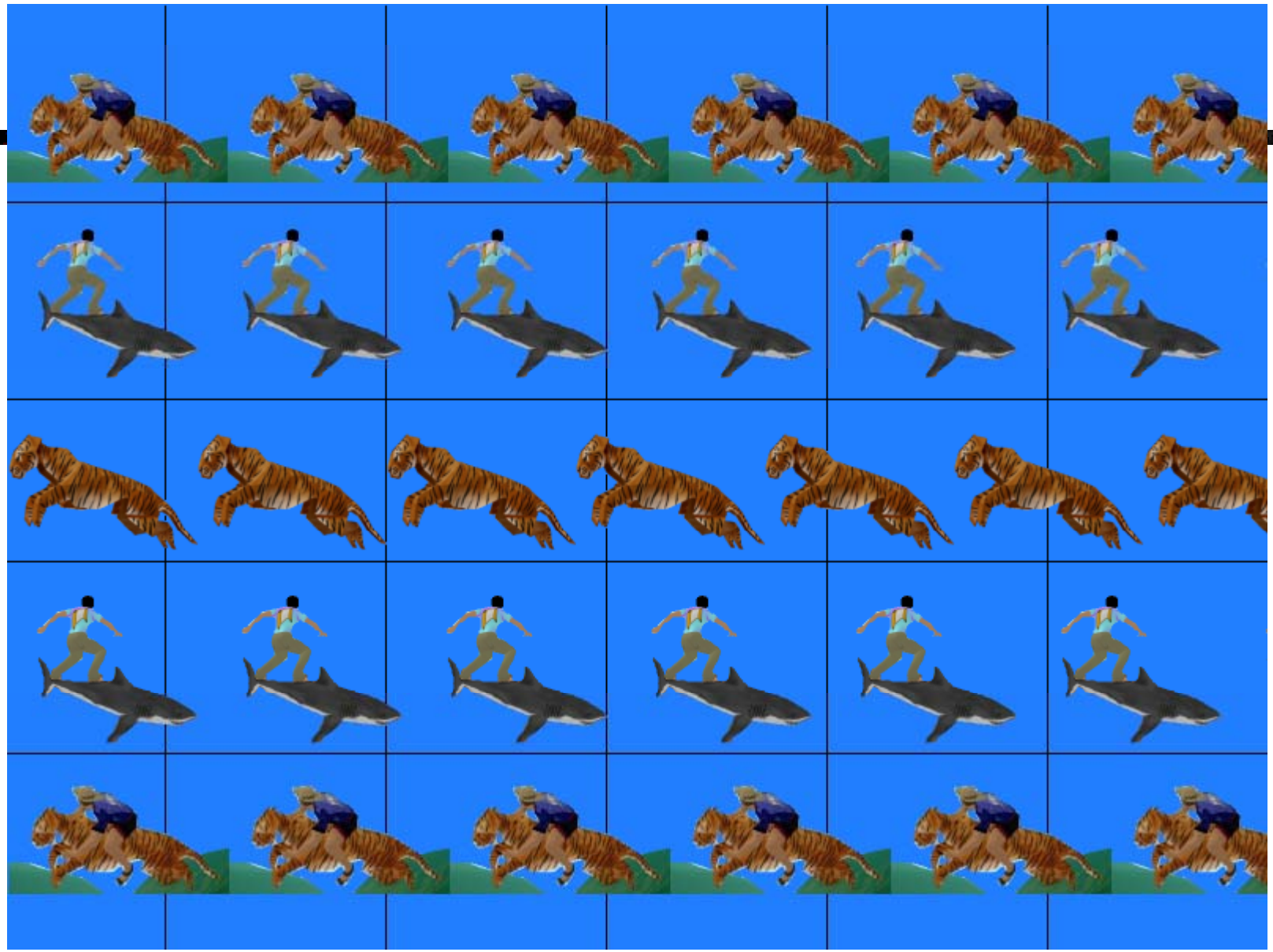
Binocular suppression
- **Brain sees only one of the two eye's images**

What do you see when you try to fuse the two images in each row together?

Binocular rivalry and suppression



Wallpaper Autostereogram



From: http://en.wikipedia.org/wiki/Autostereogram

Wallpaper Stereo

- When viewed with proper vergence, the repeating patterns in a wall paper appear to float above or below the background.
- When the brain is presented with a repeating pattern like wallpaper, it has difficulty matching the two eyes' views accurately. By looking at a horizontally repeating pattern, but converging the two eyes at a point behind the pattern, it is possible to trick the brain into matching one element of the pattern, as seen by the left eye, with another (similar looking) element, beside the first, as seen by the right eye. With the typical wall-eyed viewing, this gives the illusion of a plane bearing the same pattern but located behind the real wall. The distance at which this plane lies behind the wall depends only on the spacing between identical elements.

http://en.wikipedia.org/wiki/Autostereogram

Eye Dominance

- Eye dominance: the tendency to prefer visual input from one eye to the other.

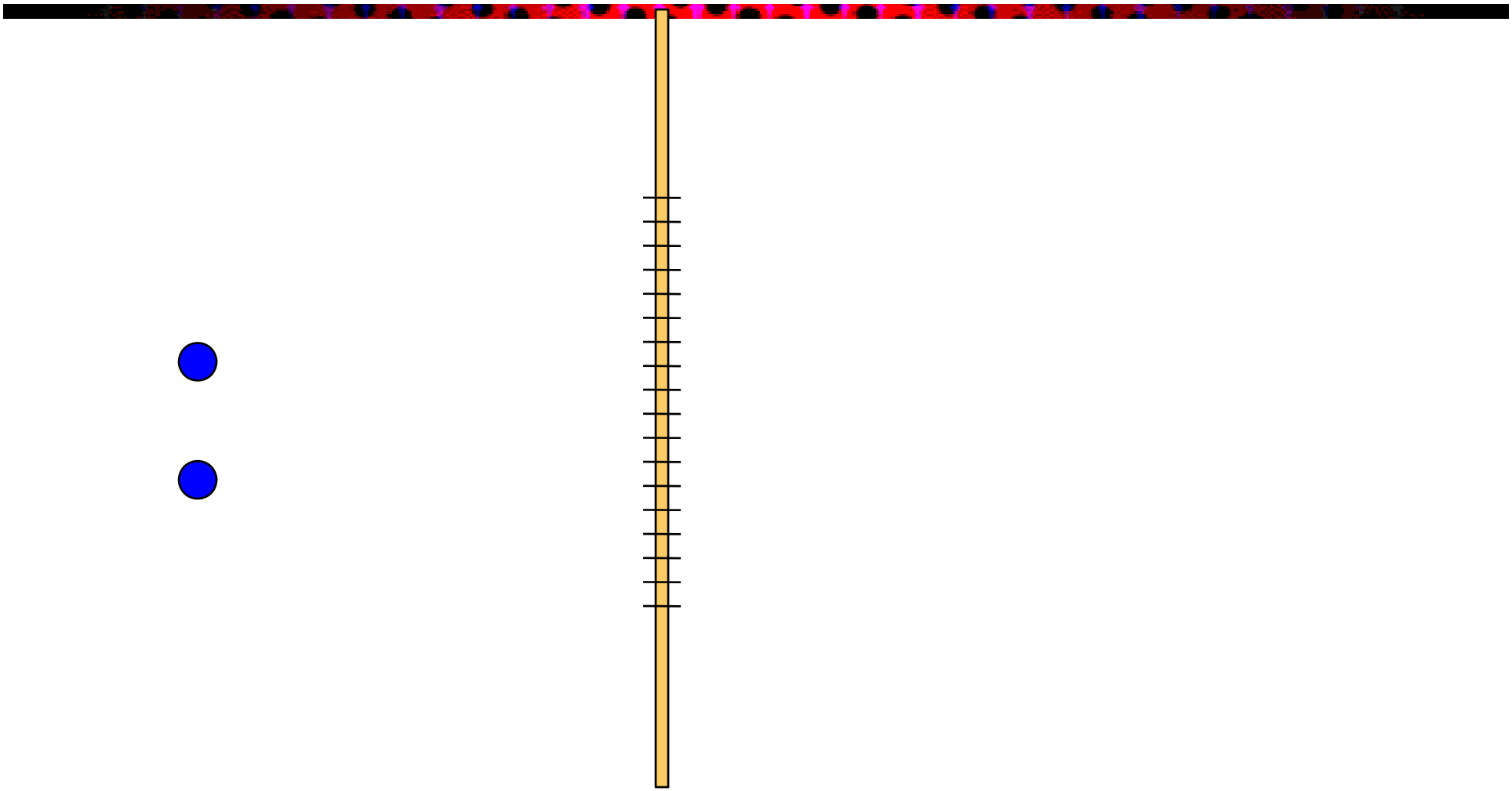
- The Miles test. The observer extends both arms, brings both hands together to create a small opening, then with both eyes open views a distant object through the opening. The observer then alternates closing the eyes or slowly draws opening back to the head to determine which eye is viewing the object (i.e. the dominant eye)
- Approximately two-thirds of the population is right-eye dominant and one-third left-eye dominant; however in a small portion of the population neither eye is dominant.

- Eye dominance does not affect stereopsis!

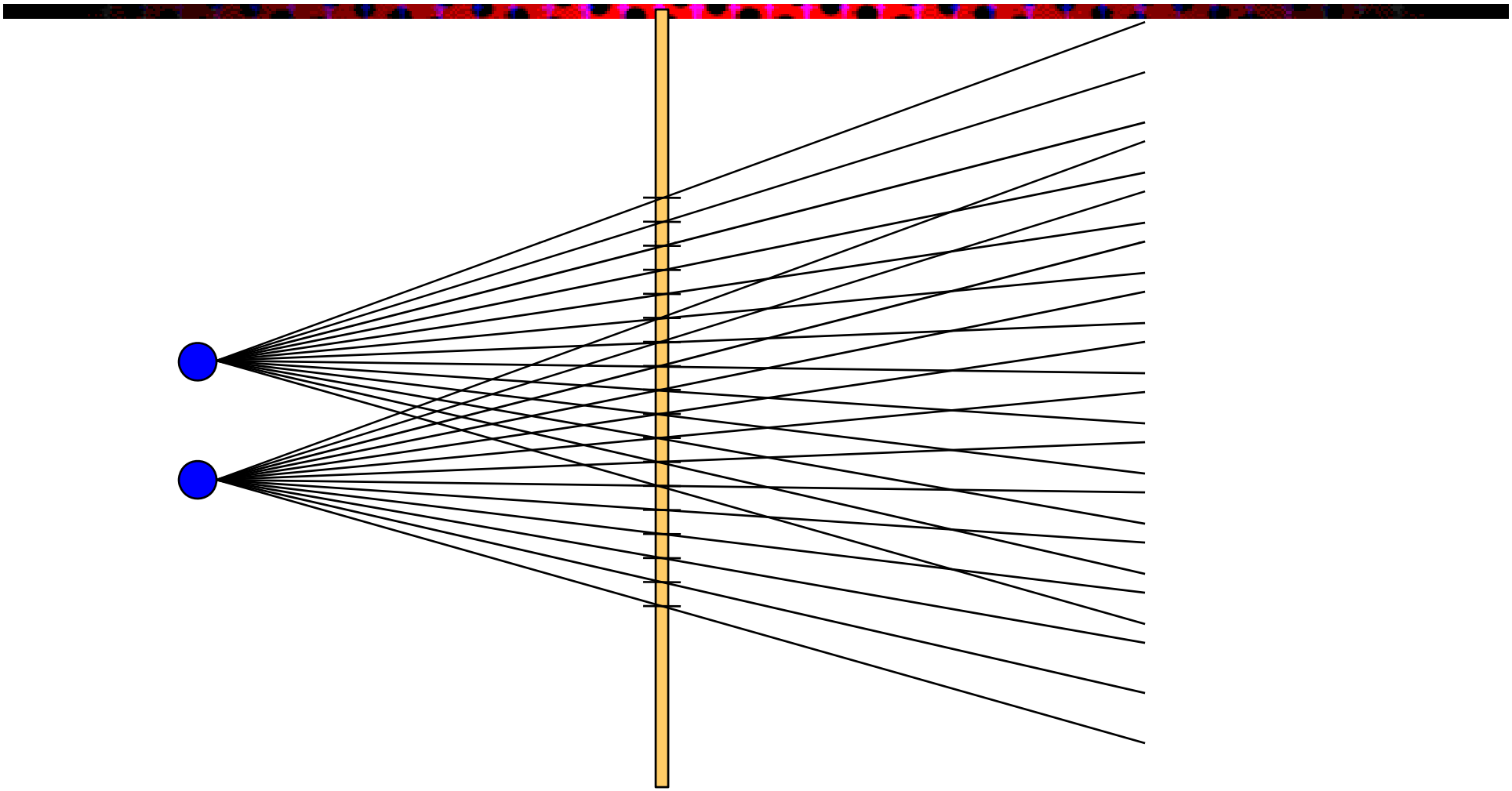
Problems in Stereo Display

- Depth plane quantization
- Impact of viewer position
- Impact of screen size

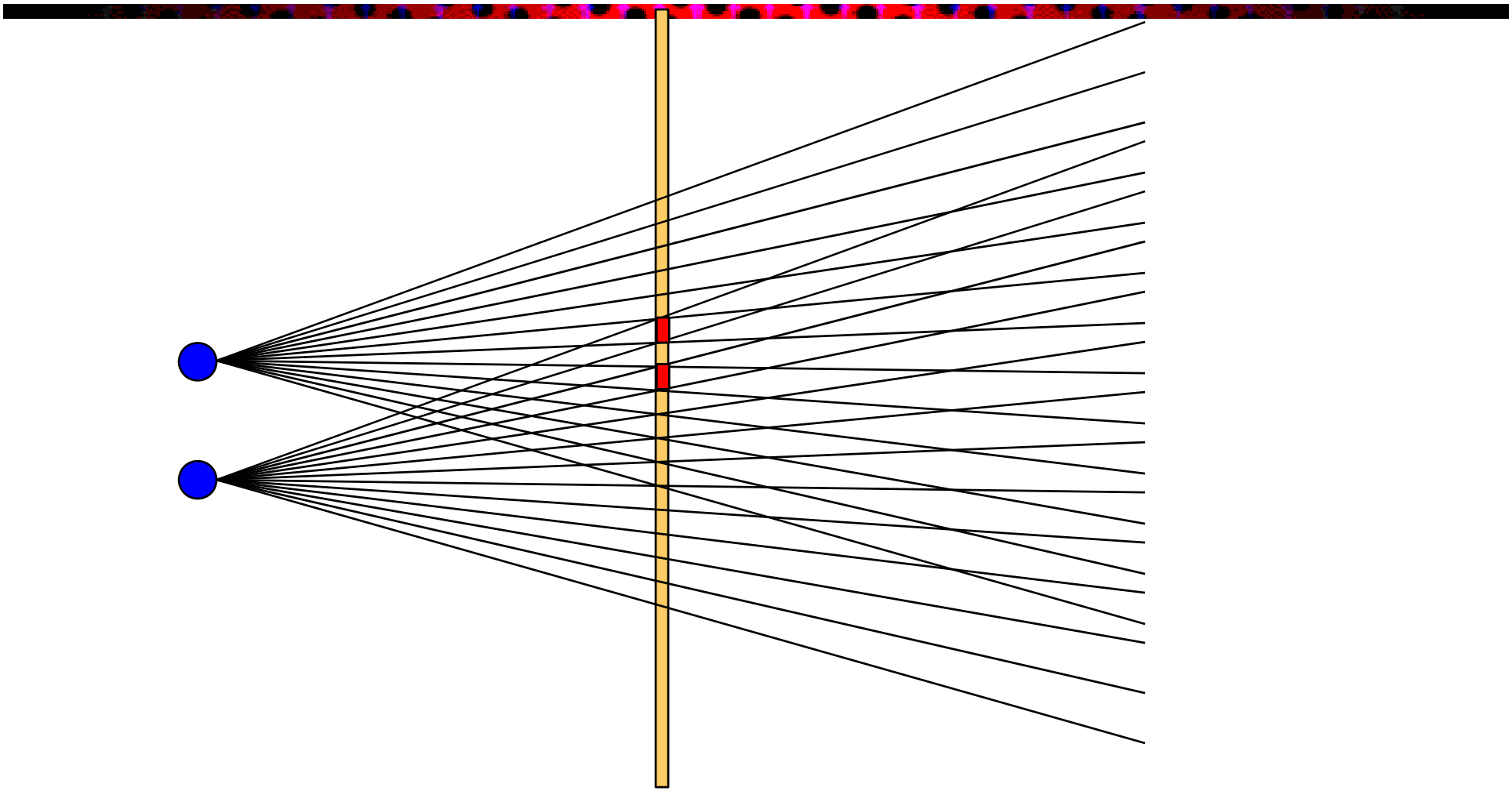
Depth-plane Quantization Due to Pixelization



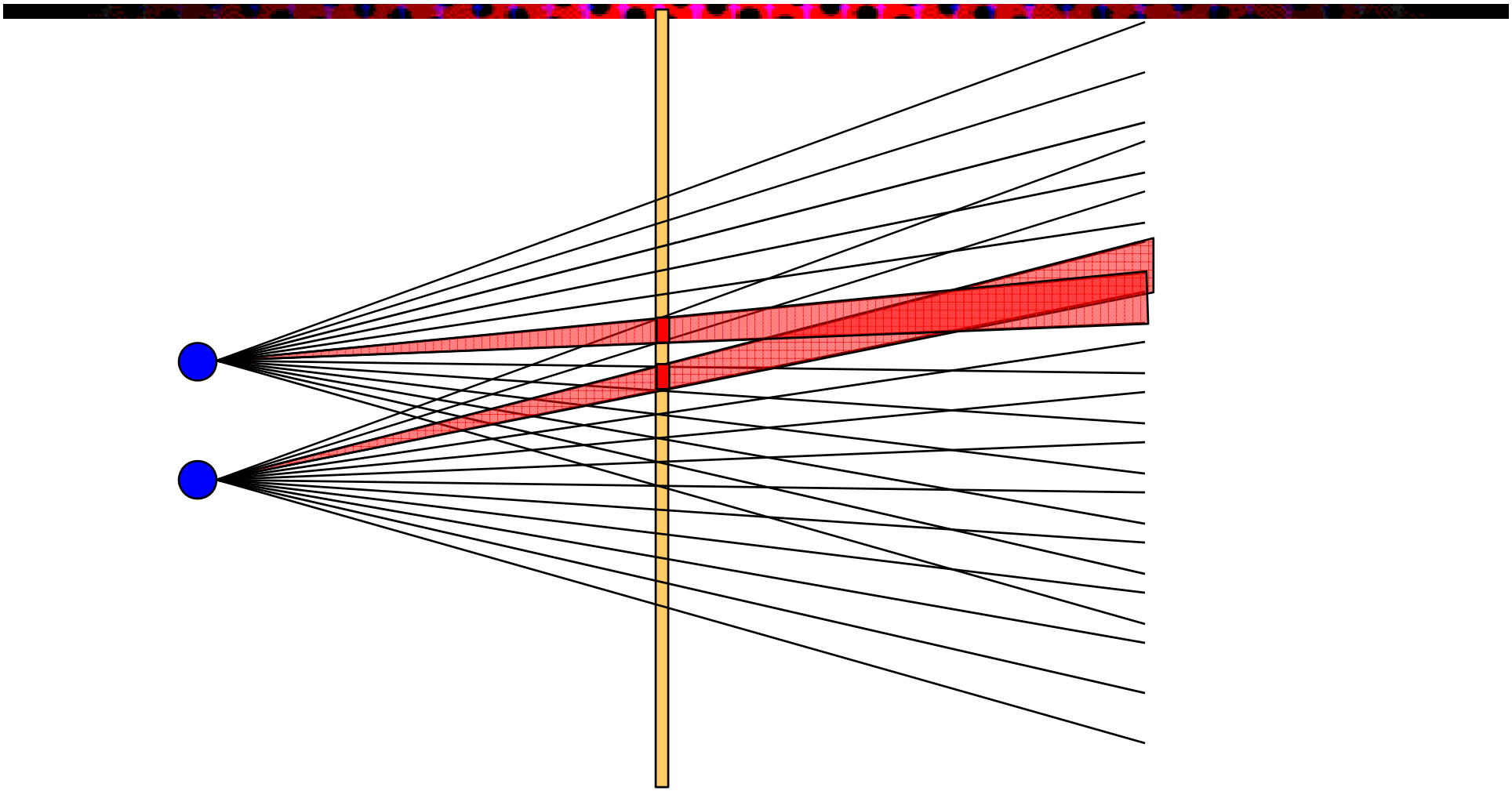
Depth-plane quantization



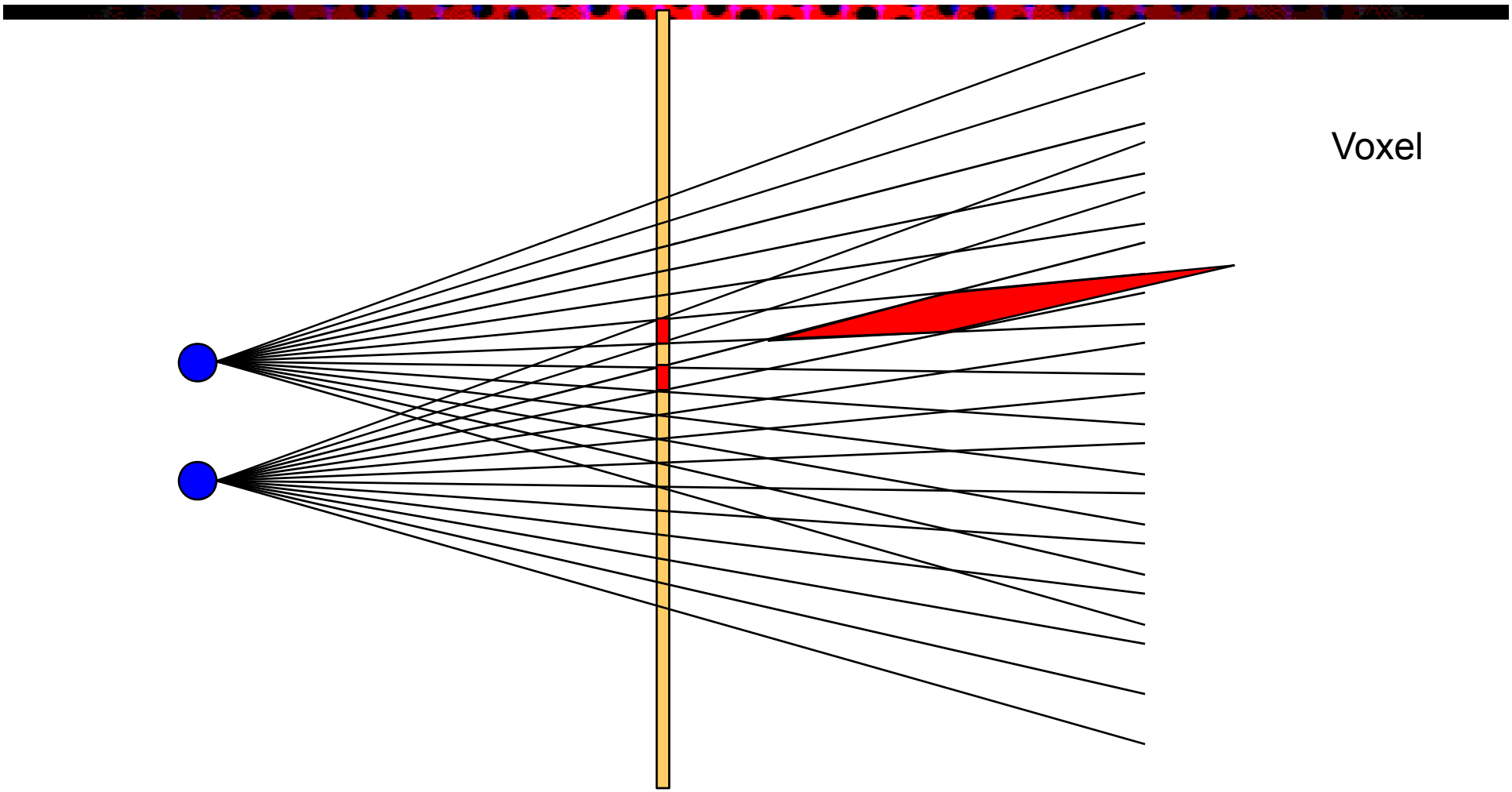
Depth-plane quantization



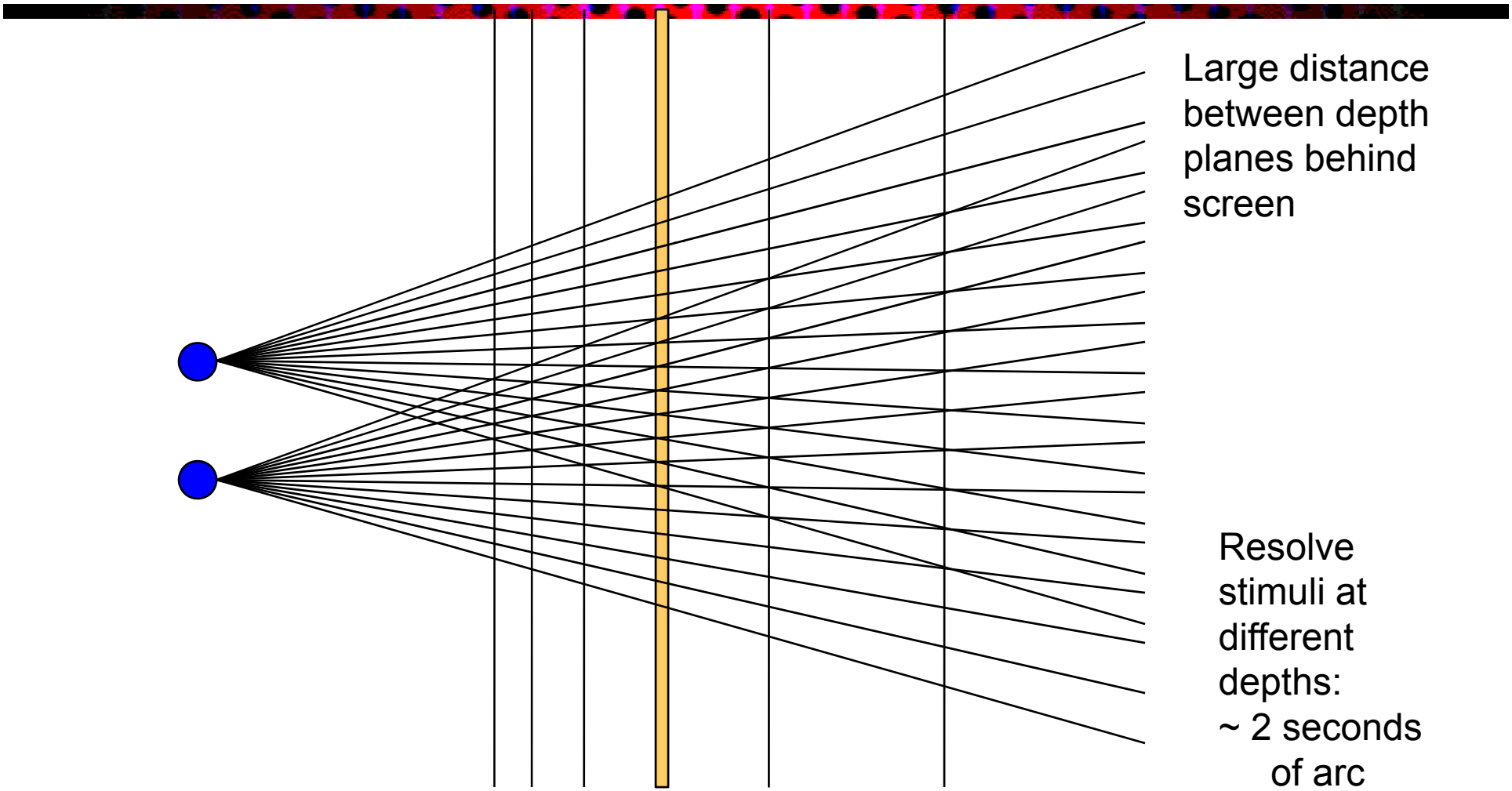
Depth-plane quantization



Depth-plane quantization



Depth-plane quantization

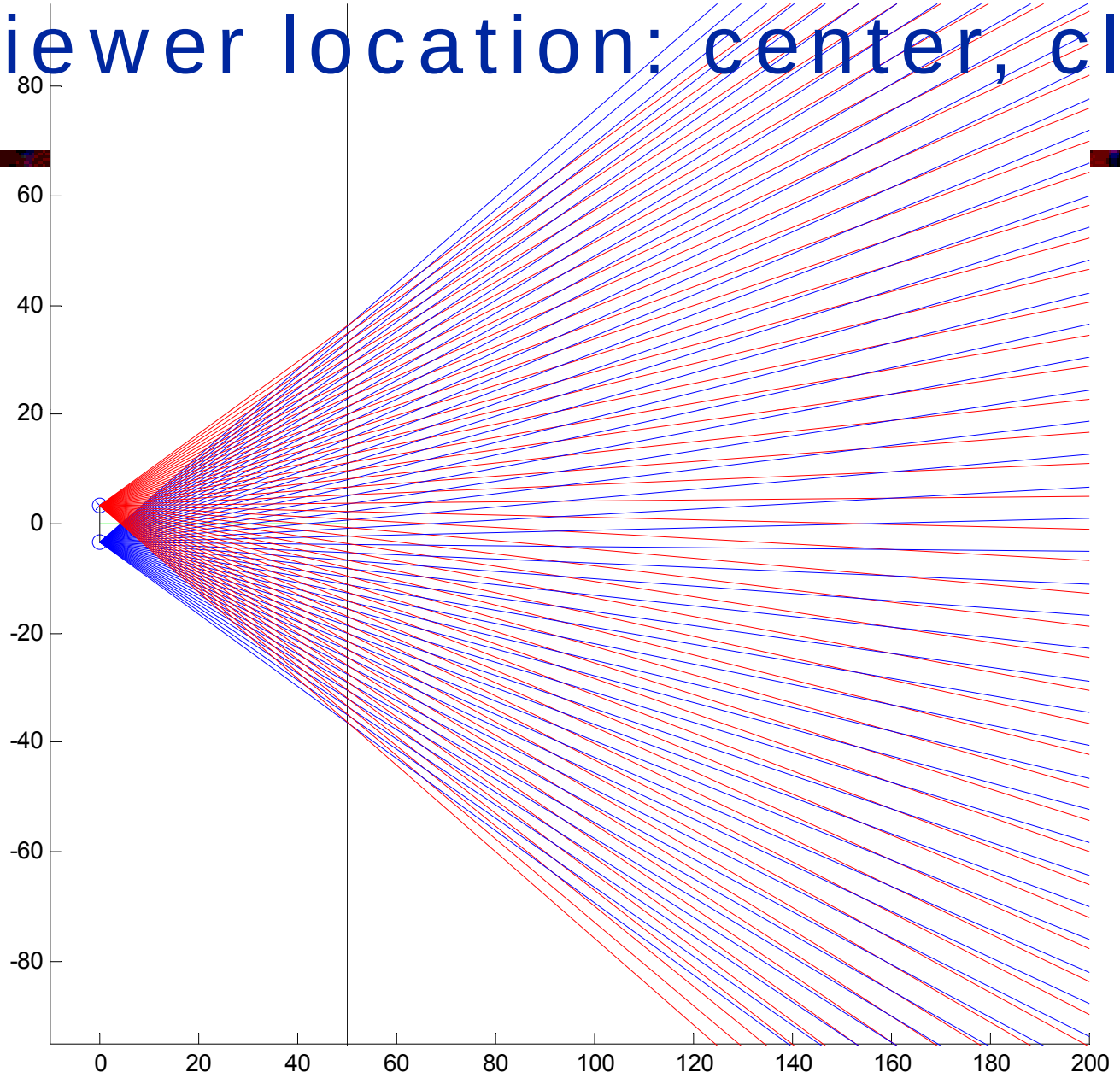


Large distance between depth planes behind screen

Resolve stimuli at different depths:
~ 2 seconds of arc

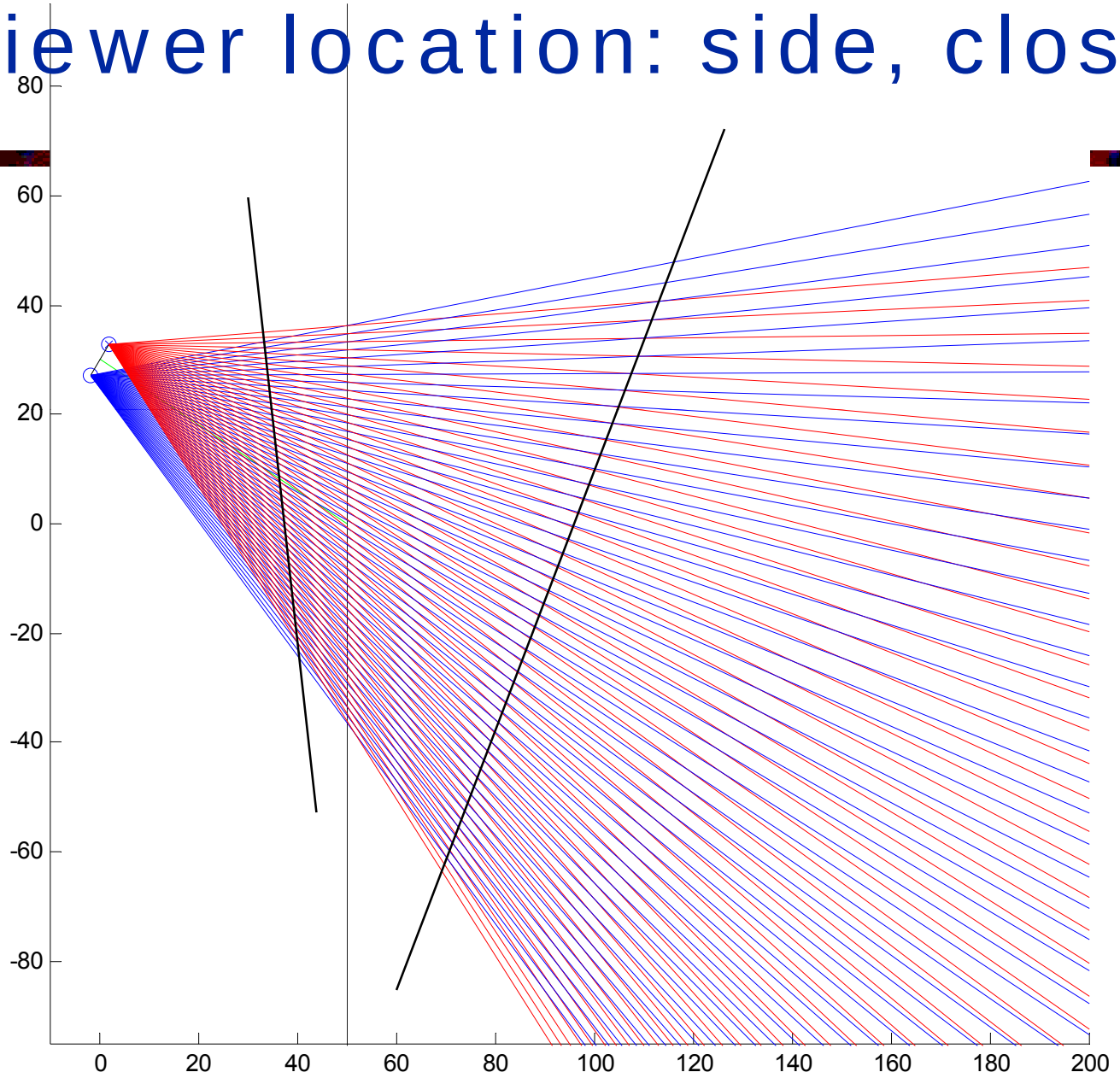
Impact of viewer location

Viewer location: center, closer



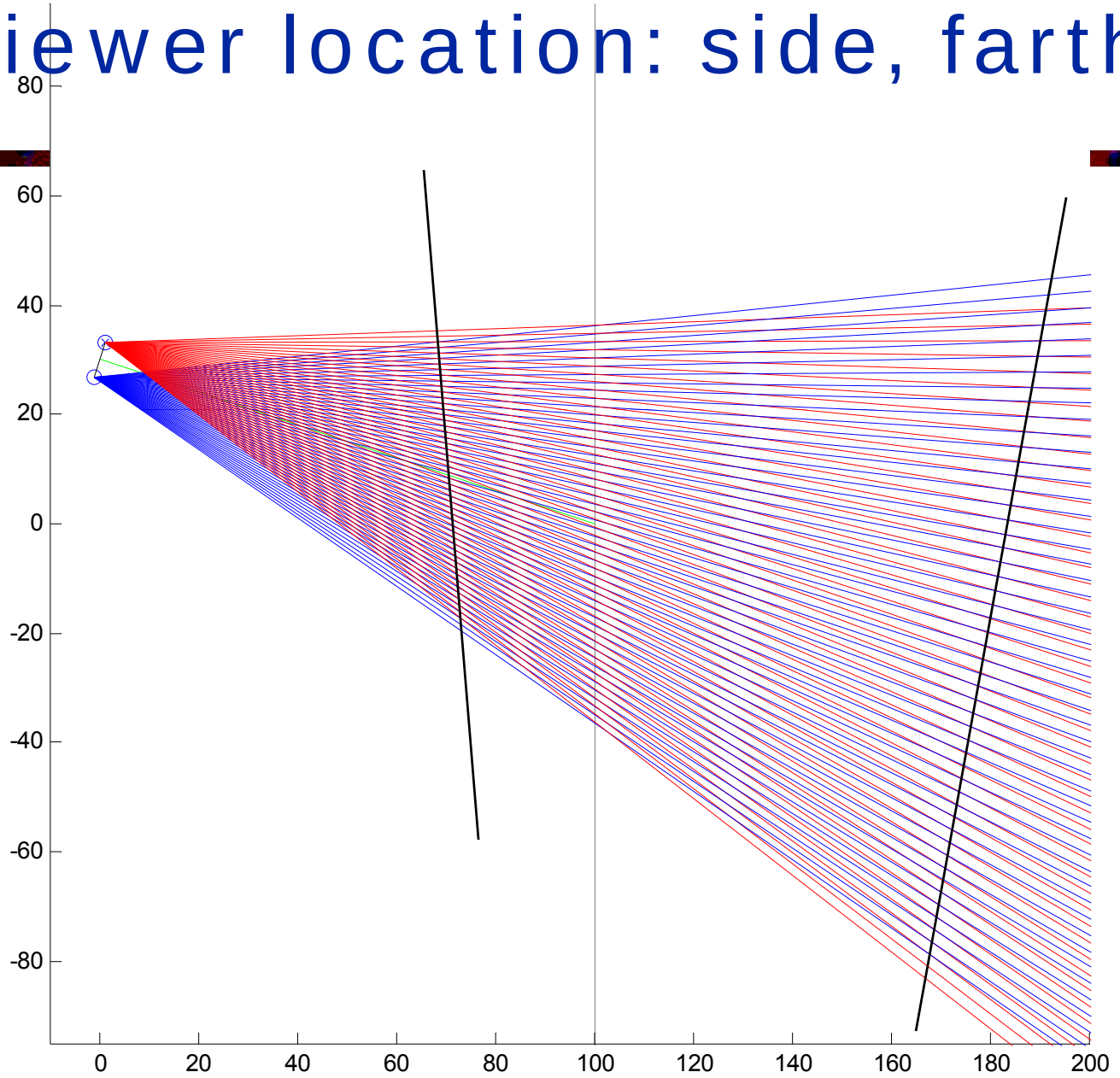
Screen: 50
Center view

Viewer location: side, closer



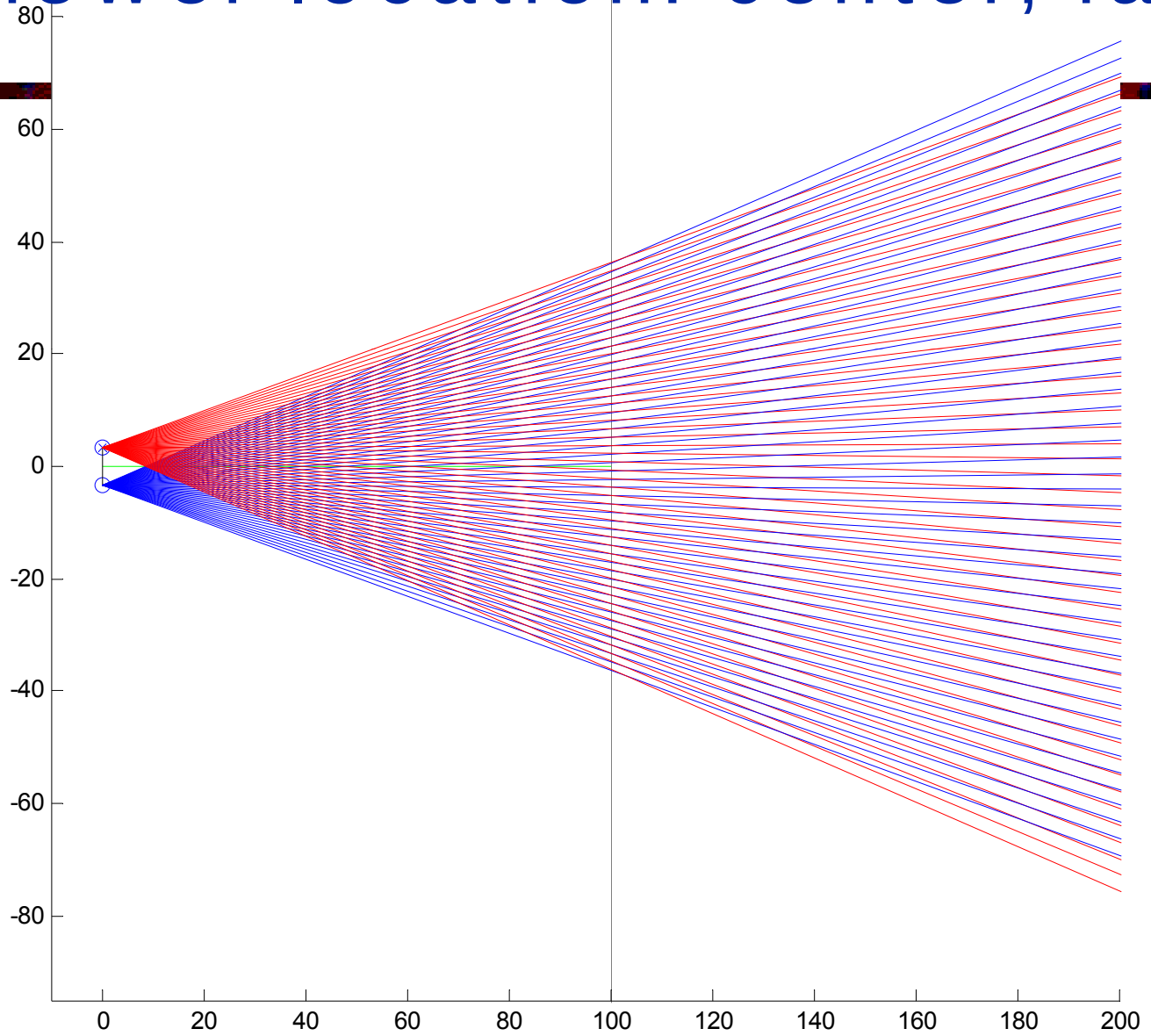
Screen: 50
Side view

Viewer location: side, farther



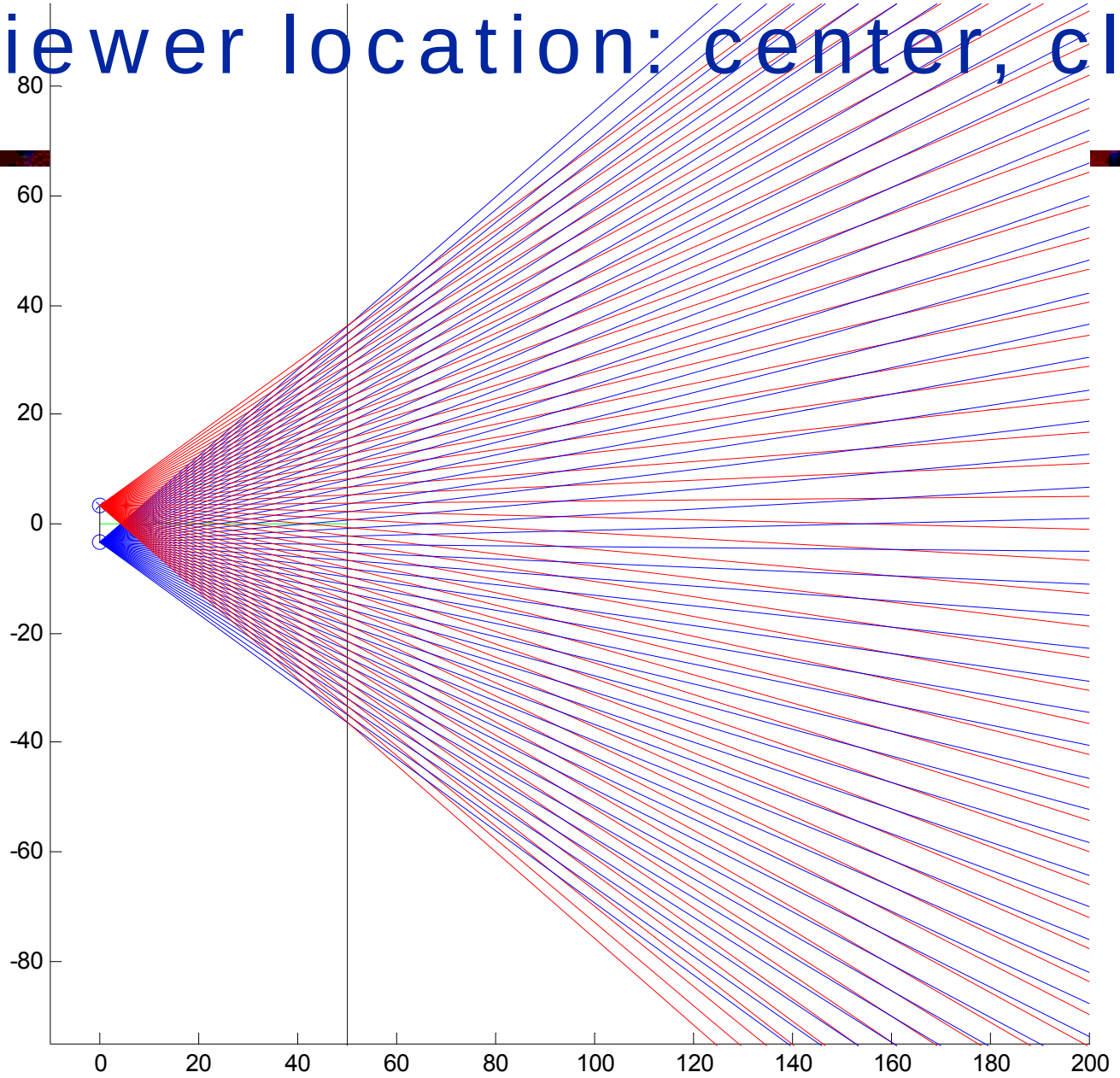
Screen: 100
Side view

Viewer location: center, farther



Screen: 100
Center view

Viewer location: center, closer



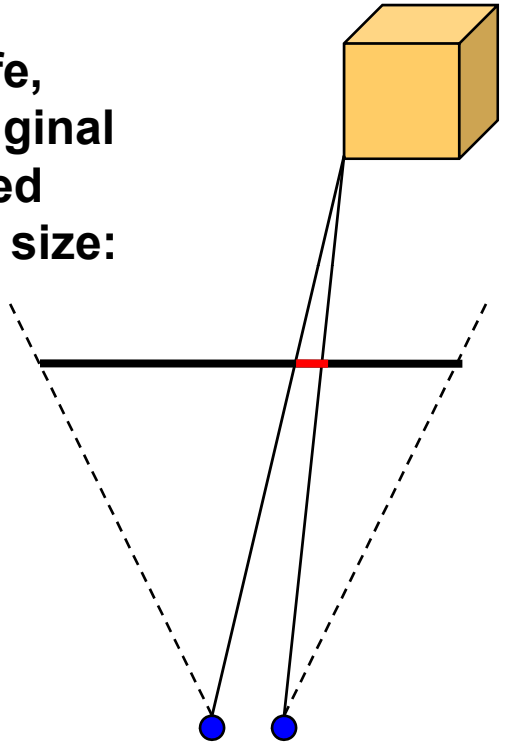
Screen: 50
Center view

Screen geometry

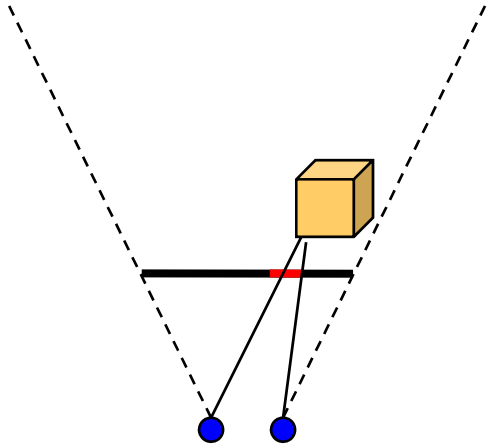
- Each technology has its own screen size and resolution
- IMAX
  - 48-foot screen; 2048x1080: aspect ratio 1.4
  - Typically all seats are within one screen height
- Real-D XLS:
  - 20-foot screen; 2048x858 per view; aspect ratio 1.85
  - Typically seats are within {single digit} screen heights
- Home TV
  - Typically 8 feet viewing distance
- Screen parallax (i.e. disparity) is affected by the size of the display screen

Mismatch in screen sizes and viewing distances (movie theater vs. home)

Real-life, and original intended screen size:



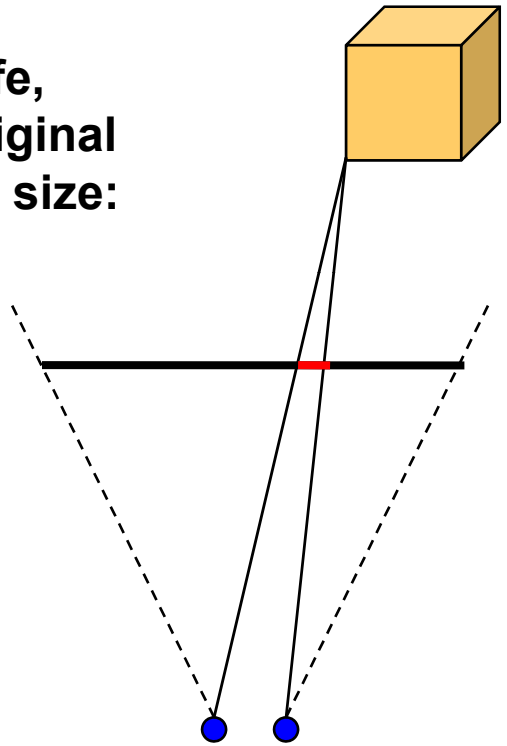
Same screen angle, smaller disparity.
Result: different object size and distance



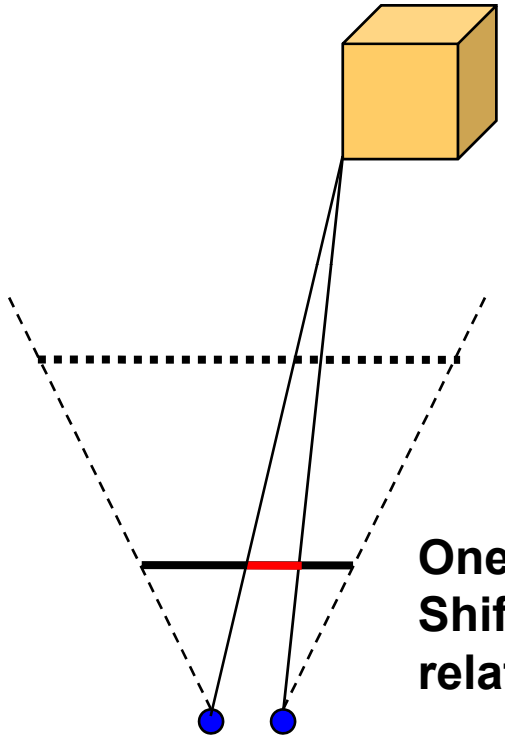
Objects appear to be in a puppet theater: small and close together

Adaptation of disparity for different screen sizes

Real-life, and original screen size:



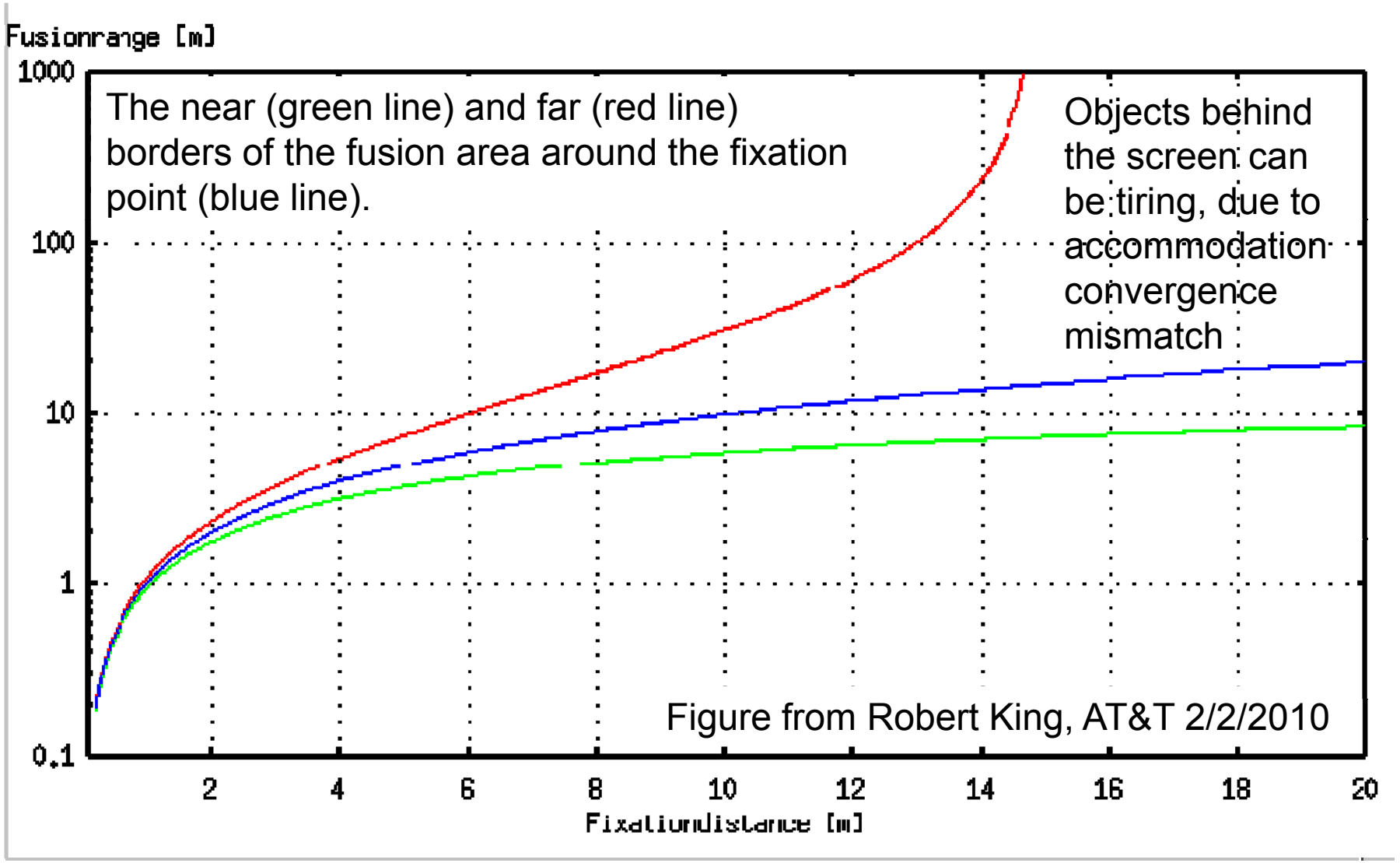
Same screen angle, shifted disparity.
Result: same object size



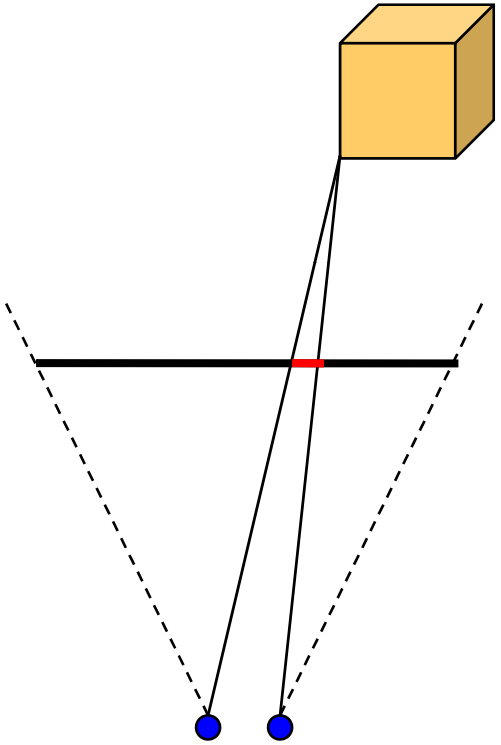
One way to "fix": Shift one image relative to the other

Now, the object appears the correct size. However, objects are almost always behind the screen. Causing conflict of converging and accommodation

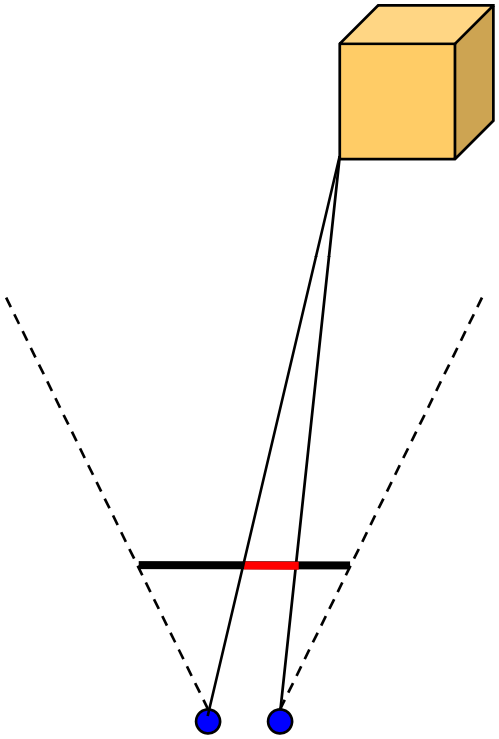
Horopter and Panum's fusional area



Different screen sizes



Same screen angle,
shifted disparity.
Result: same object size



BUT, this isn't valid with
a different screen distance
or an offset viewer. Need to use
Intermediate View Synthesis.

Depth-based adaptation of 3D content

- Screen parallax is affected by the size of the screen

- To display for different screens, adjust stereo disparity
- Limit maximum disparity to avoid too much eye strain

- Shifting/offsetting one image has only limited success

- Ideally, for a given viewer distance and viewer location, generate an intermediate view for that viewer: Intermediate view synthesis

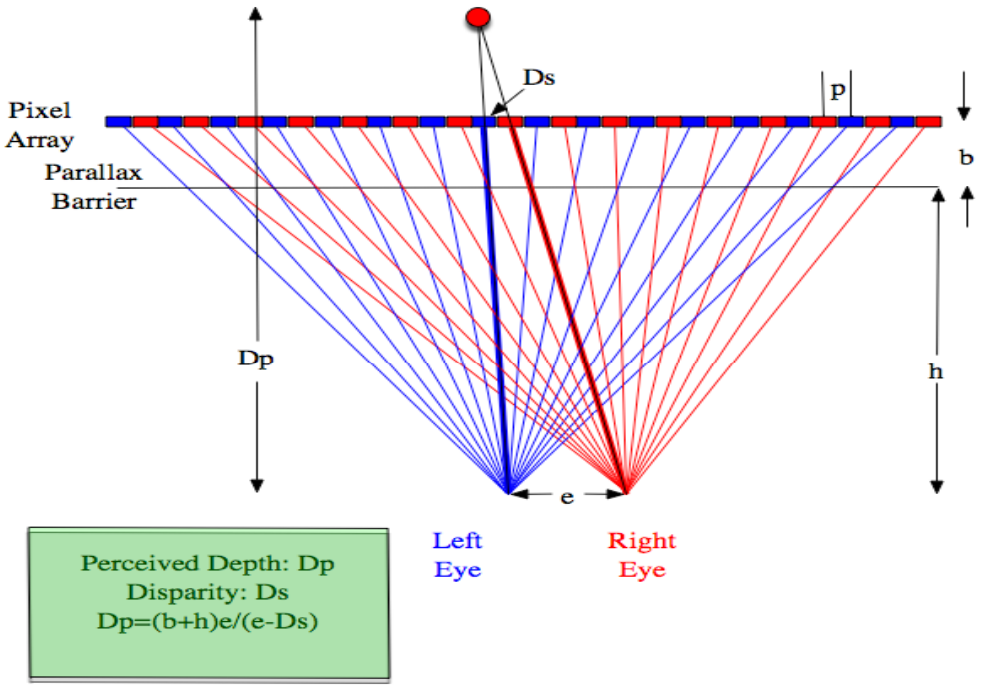
Display of Stereo Images/Sequences

- Principle:
  - Project images for the left and right eyes simultaneously in a way that the two images can be received separately by left and right eyes
- Separation mechanism in stereoscopic display
  - Color filter (Cannot be used for display color stereo images)
  - Polorization
  - Interlace in time the left and right views (Stereographics, Inc.)
  - Viewers need to wear special glasses
- Auto-stereoscopic display
  - Present two or multiple views on the same screen simultaneously
  - A viewer sees different view when looking from different angle
  - Viewers do not need to wear glasses
  - Autostereoscopic lenticular screens

Using glasses

- Anaglyph. Two-color separation of left/right view. Poor color rendition.
- Polarized. For viewing stereo pairs projected through suitable polarizing filters. Better image quality.
- Shutter glasses. Liquid crystal. Expensive. Require high refresh rate. Require synchronization of display and glasses

Autostereoscopic display principle



- Blue pixels in pixel array display left view; Red display right view.
- Parallax barrier or lens array allows left eye to see `left pixels' and right eye to see `right pixels'.
- Thus separate views are presented to left and right eye creating illusion of depth.

3D-ready consumer TVs

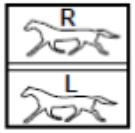
- Display stereo pairs in time-sequential manner
- Active shutter glasses

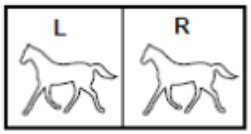
- Options
  - 3D DLP technology from TI (Samsung & Mitsubishi)
  - 3D plasma (Samsung)

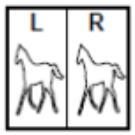
3D Video Delivery Formats

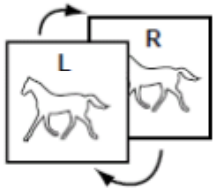
① Anaglyph  L=Red R=Blue

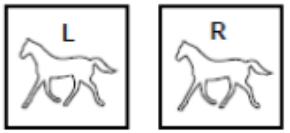
② Planar  **+ Depth map**

③ Above/Below 

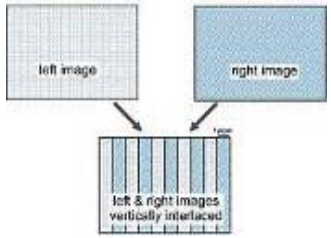
④ Side by Side 

⑤ Side by Side Squashed 

⑥ Page Flipping 

⑦ Stereo Pair 

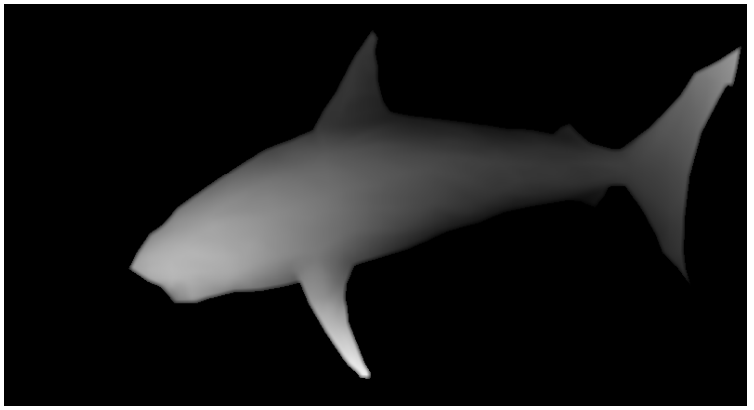
⑧ Interline 

⑨ 

JBoyce, 2011

Examples of depth maps

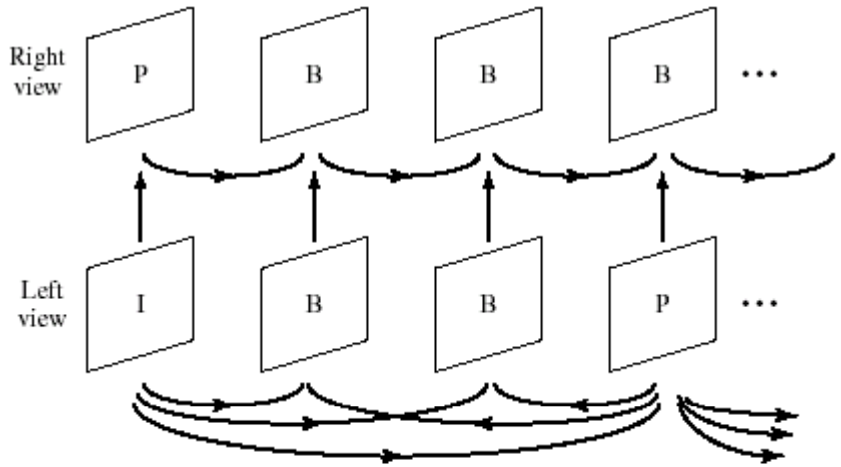
- **Brighter objects closer, dark objects further away**



Coding of Stereo Sequences

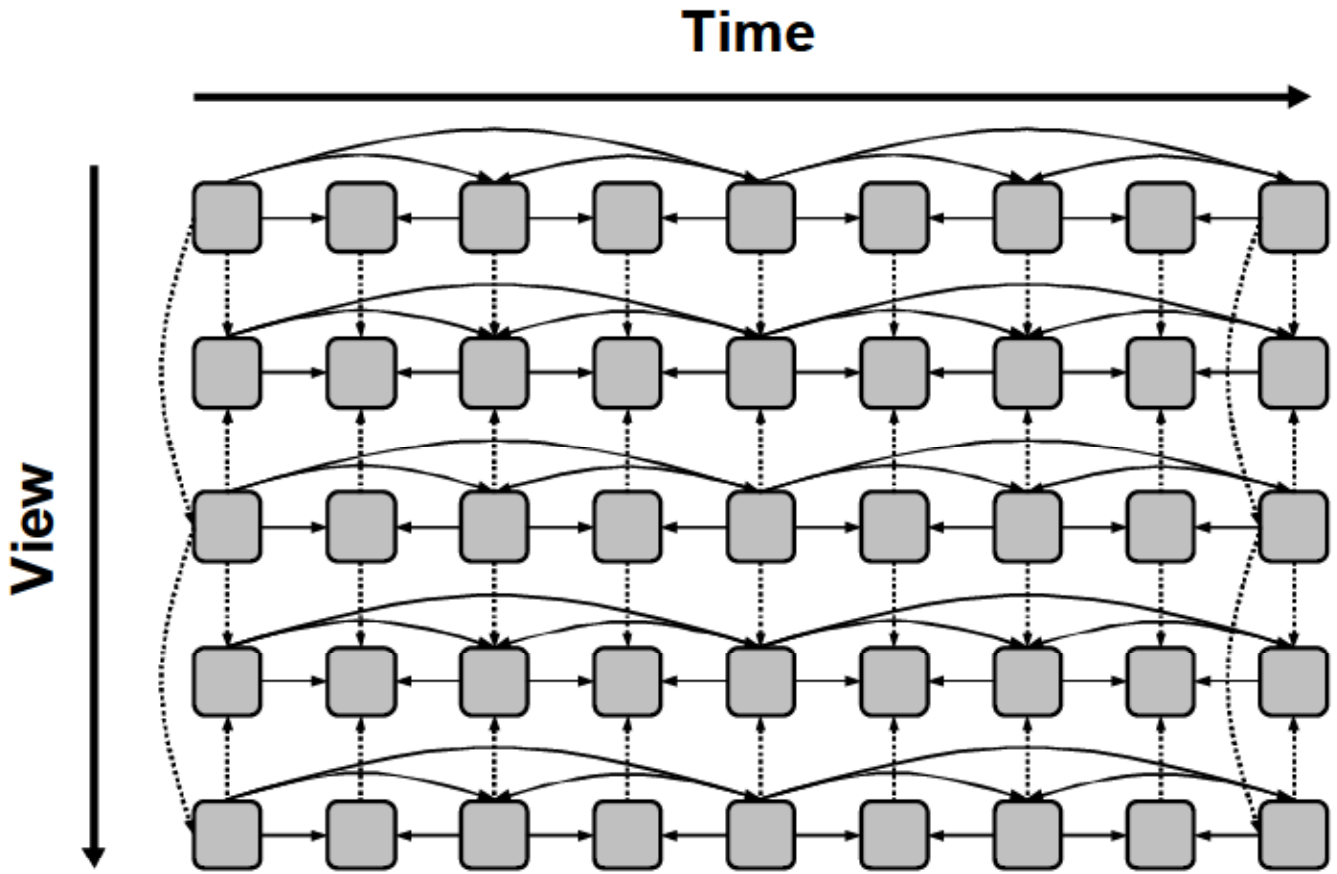
- Simulcast:
  - Code each view independently
- Extension from block-based hybrid code:
  - Code one view using a standard video coder, using MCP between successive frames in this view
  - Code the other view by using both DCP between views at the same frame and MCP in the same view
  - MPEG-2 Multiview profile
- Mixed resolution
  - Code one view at the desired spatial/temporal resolution
  - Code another view with reduced spatial/temporal resolution
- More advanced, object-based approach

MPEG-2 Multiview Profile



- Left view: MCP only
- Right view: combination of MCP and DCP, using the bi-directional prediction mode
- Can be implemented using the temporal scalability tool:
  - left view treated as the base-layer, right view treated as the enhancement layer
- Only limited gain over simulcast
  - MCP is typically more effective than DCP
  - Ineffectiveness of DCP due to inaccuracy of block-based constant disparity model

H.264 Multiview prediction structures



Barriers to mass market

- Data and delivery format
- Quality of 3D video production
  - Content creators must be aware of 3D videography
  - Re-purposing of 3D content from cinema into the homes
- Human factors
  - Stereoscopic glasses are no fun
  - Auto-stereoscopic has its own issues
  - Avoid objectionable 3D effects

Additional perceptual issues

- Both too *much* depth and too many *fast changes* of depth cause visual fatigue

- Conflicting depth information causes visual fatigue
  - Accommodation and vergence are linked when scanning the scene (but can be decoupled over time)
  - Compression, aliasing, other impairments (like keystoning) can make fusing more difficult
  - Screen or glasses scratch or dust

- Cross-talk
  - Left eye sees some of what Right eye should see
  - Stronger in high-contrast and large-disparity areas
  - (But fusing is easier in high-contrast areas)

Summary

- Human perception of depth
- Principle in stereo imaging:
  - Relation between depth and disparity for parallel set-up and other more general camera set-ups.
  - Epipolar constraint for an arbitrary set-up
- Disparity estimation:
  - Formulation as an optimization problem similar to motion estimation
  - Block-based approach
  - Mesh-based approach: regular mesh vs. adaptive mesh
  - Dynamic programming: not required
  - Joint motion and structure estimation: not required
- Intermediate view synthesis
- Stereo sequence coding
  - Extension from standard video coder: Joint MCP and DCP
  - Simulcast with mixed resolution
- Stereo image/video display

Homework

- Reading assignment: Chap. 12
- Written assignment
  - Prob. 12.1
  - Prob. 12.2

Additional references

- A. Woods, T. Docherty, and R. Koch, “Image distortions in stereoscopic video systems”, Proceedings SPIE Stereoscopic Displays and Applications IV, vol. 1915, San Jose, CA, Feb. 1993.