

Introduction to Stereoscopy

2017 Fall

National Cheng Kung University

Instructor: Min-Chun Hu 胡敏君



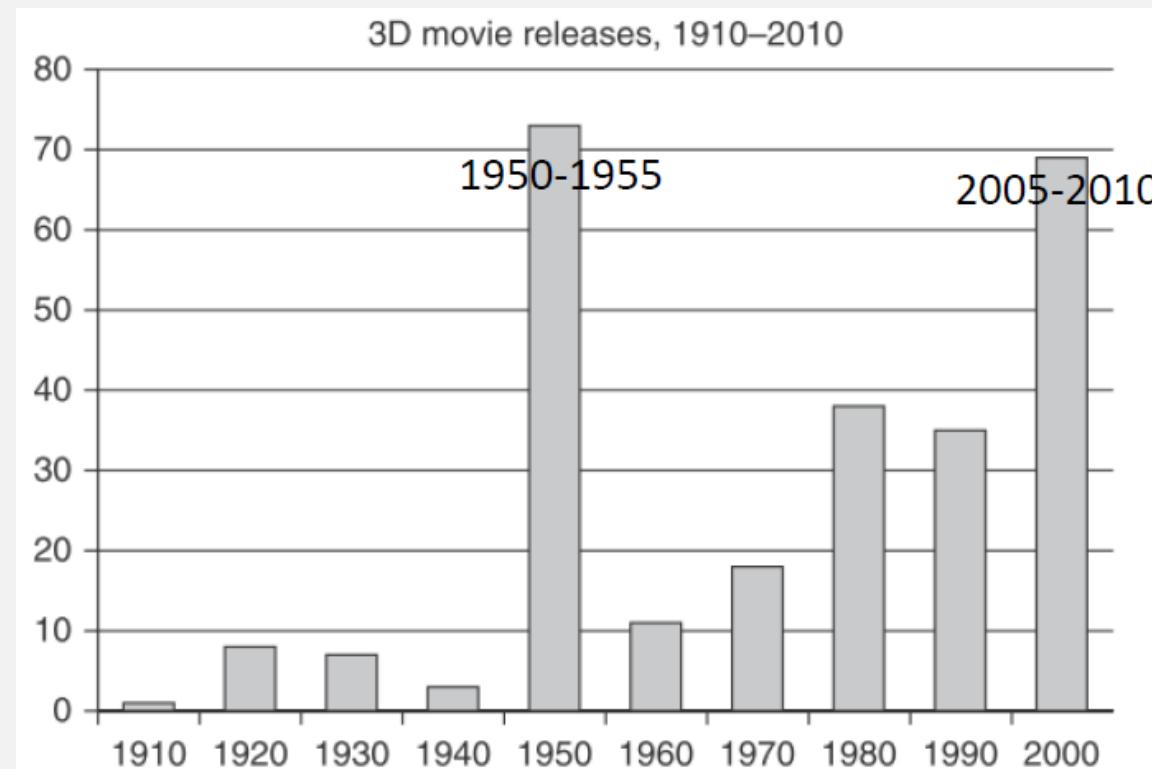
3D has a long history

- 1830s, stereoscope
- 1920s, first 3D film
 - *The Power of Love*
 - projected dual-strip in the red/green
 - anaglyph format
- 1920s, televue system



3D has a long history

- 1950s, the "golden era" of 3-D
- The attempts failed because immature technology results in viewer discomfort.
- 1980s, rebirth of 3D, IMAX



The Earliest 3D Image



3D is hot today



3D TVs



3D Computers



3D Notebooks



3D Game Consoles



3DS

HTC EVO 3D



dcfever.com

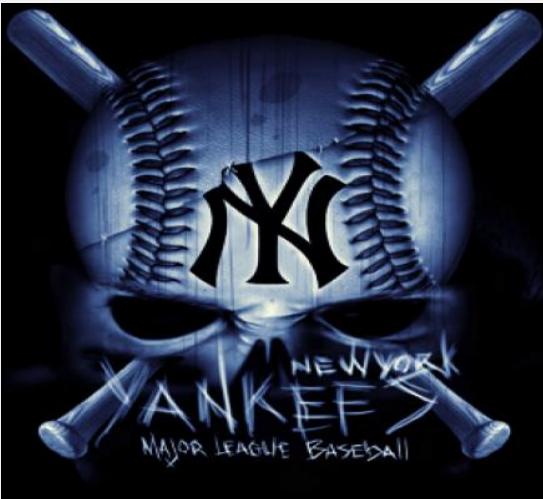
3D Contents (Games)



3D Contents (Films)



3D Contents (Broadcasting)



3D Cameras



Fuji Real3D W1 and W3



Sony HDR-TD10E



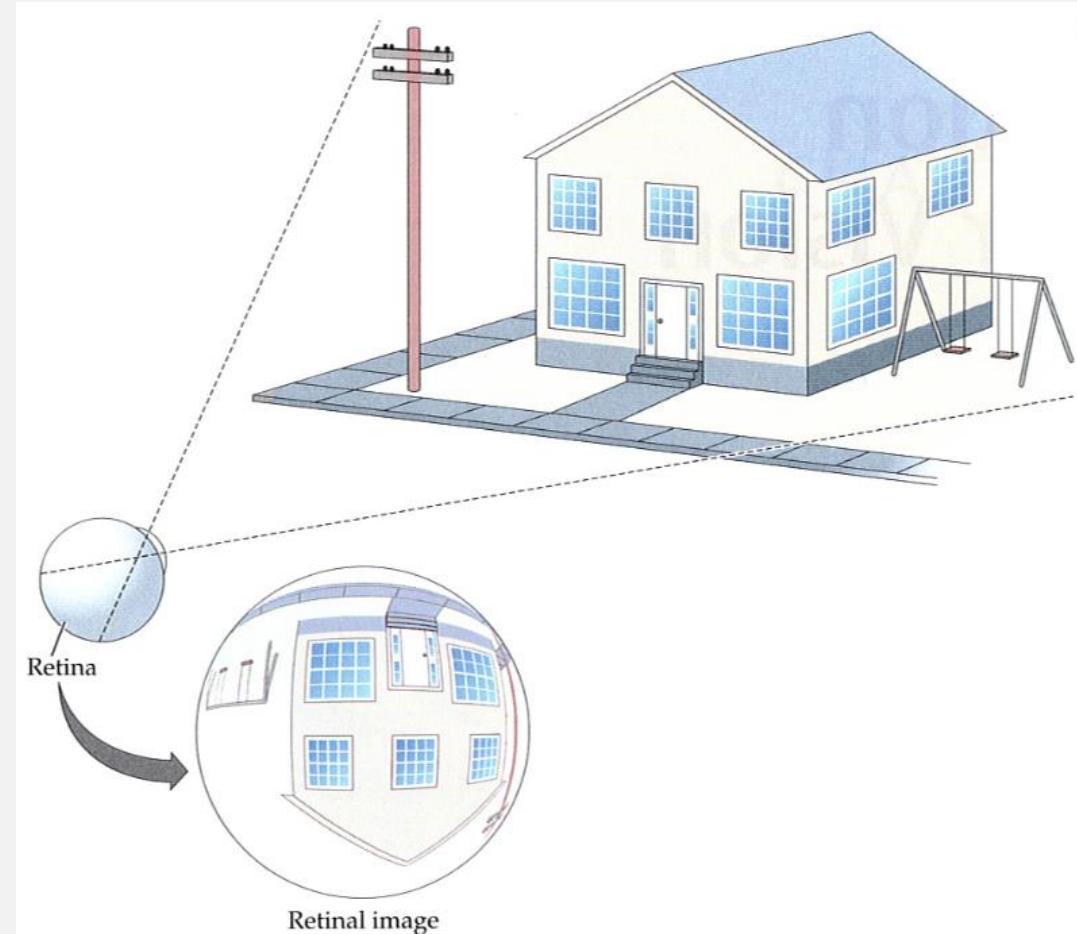
Outline

- Human depth perception
- 3D displays
- 3D cinematography
- Stereoscopic media postprocessing

Human Depth Perception

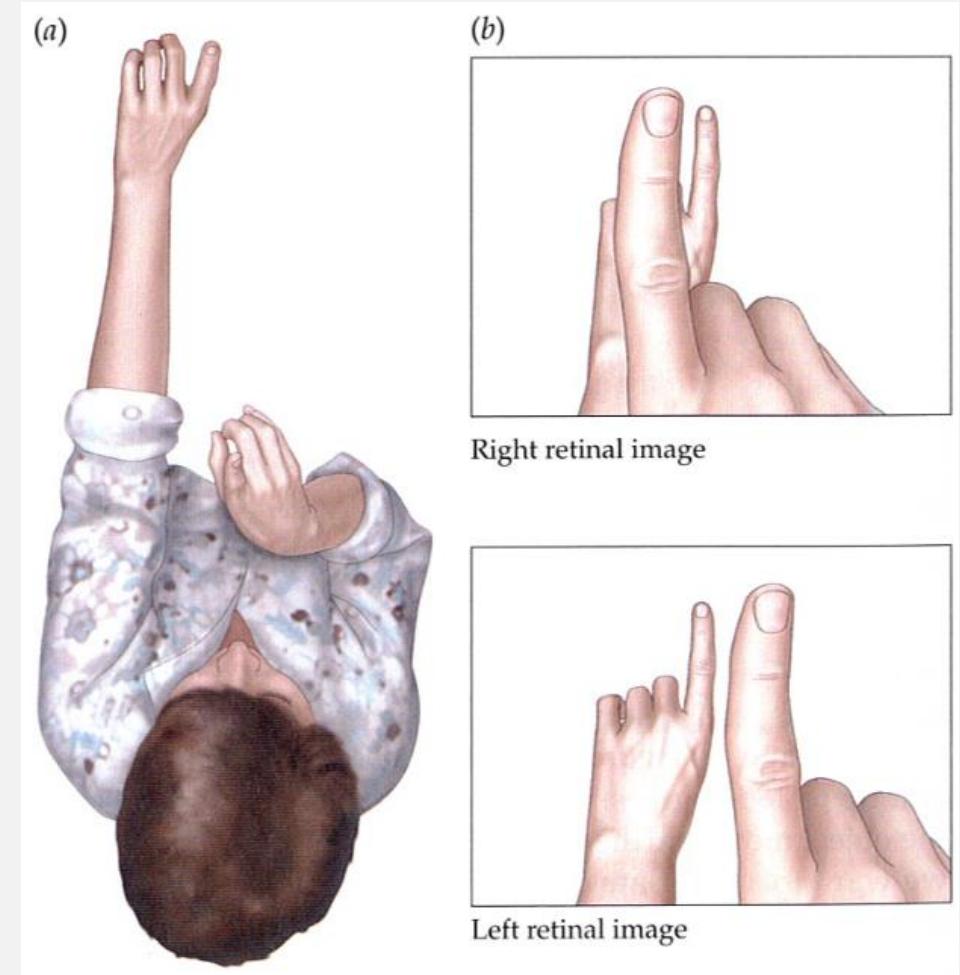
Space Perception

- The ability to perceive and interact with the structure of space is one of the fundamental goals of the visual system.
- Our visual system reconstructs the world from two non-Euclidean inputs, the two **distinct retinal images**.

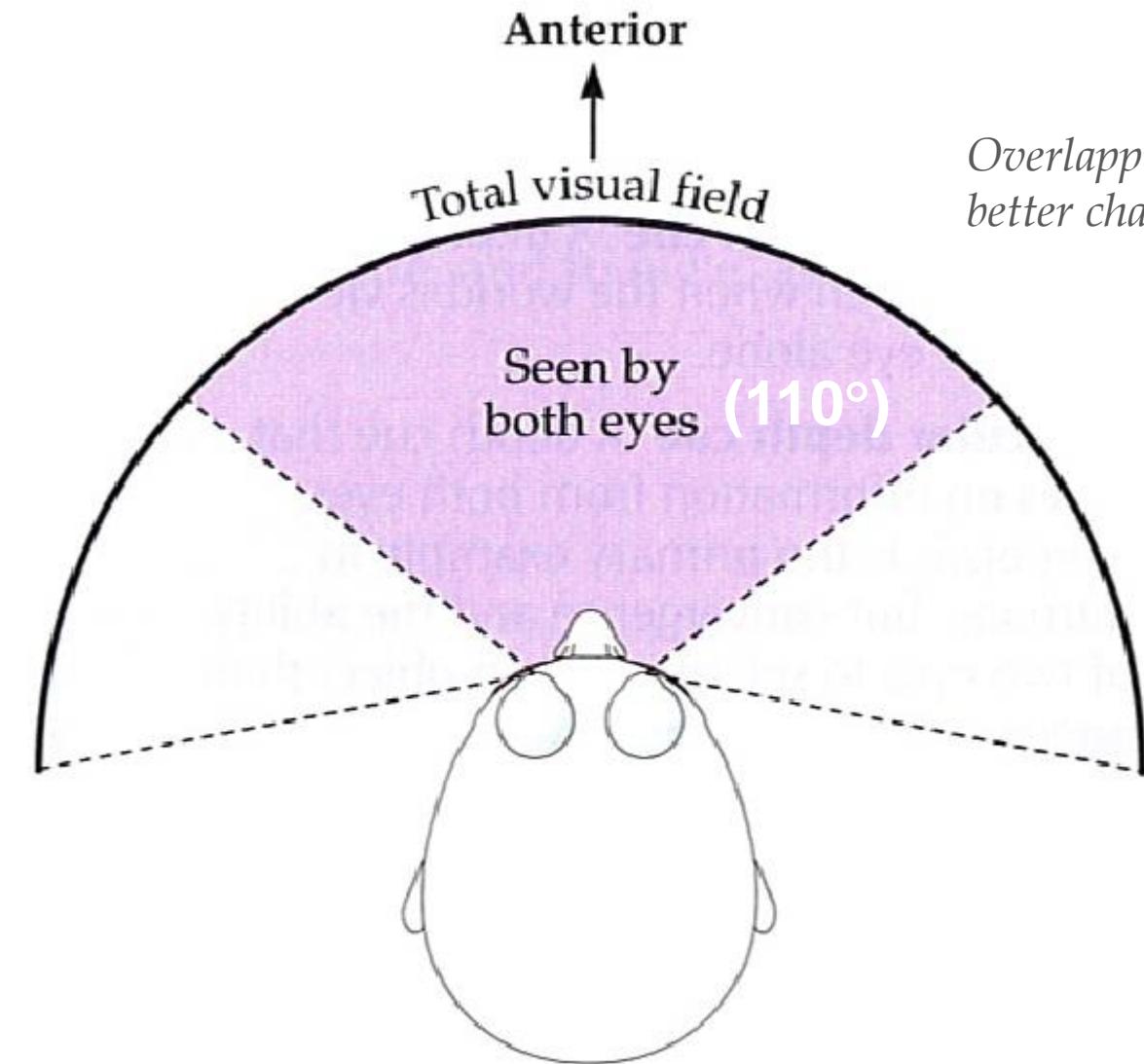


Binocular Vision

- Two retinal images are different because the retinas are in slightly different places.
- The combination of signals from each eye makes performance on many tasks better with both eyes than with either eye alone.



Binocular vision



Overlapping binocular visual fields give predator a better chance to spot small fast-moving objects

Vertical visual field: 60 ° up 80 ° down

Binocular Disparity

■ Binocular disparity:

- the differences between the two retinal images of the same scene

■ Monocular:

- with one eye

■ Stereopsis:

- the ability to use binocular disparity as a cue to depth.

■ Note that, although stereopsis adds richness to depth perception, it is not a necessary condition for depth perception.

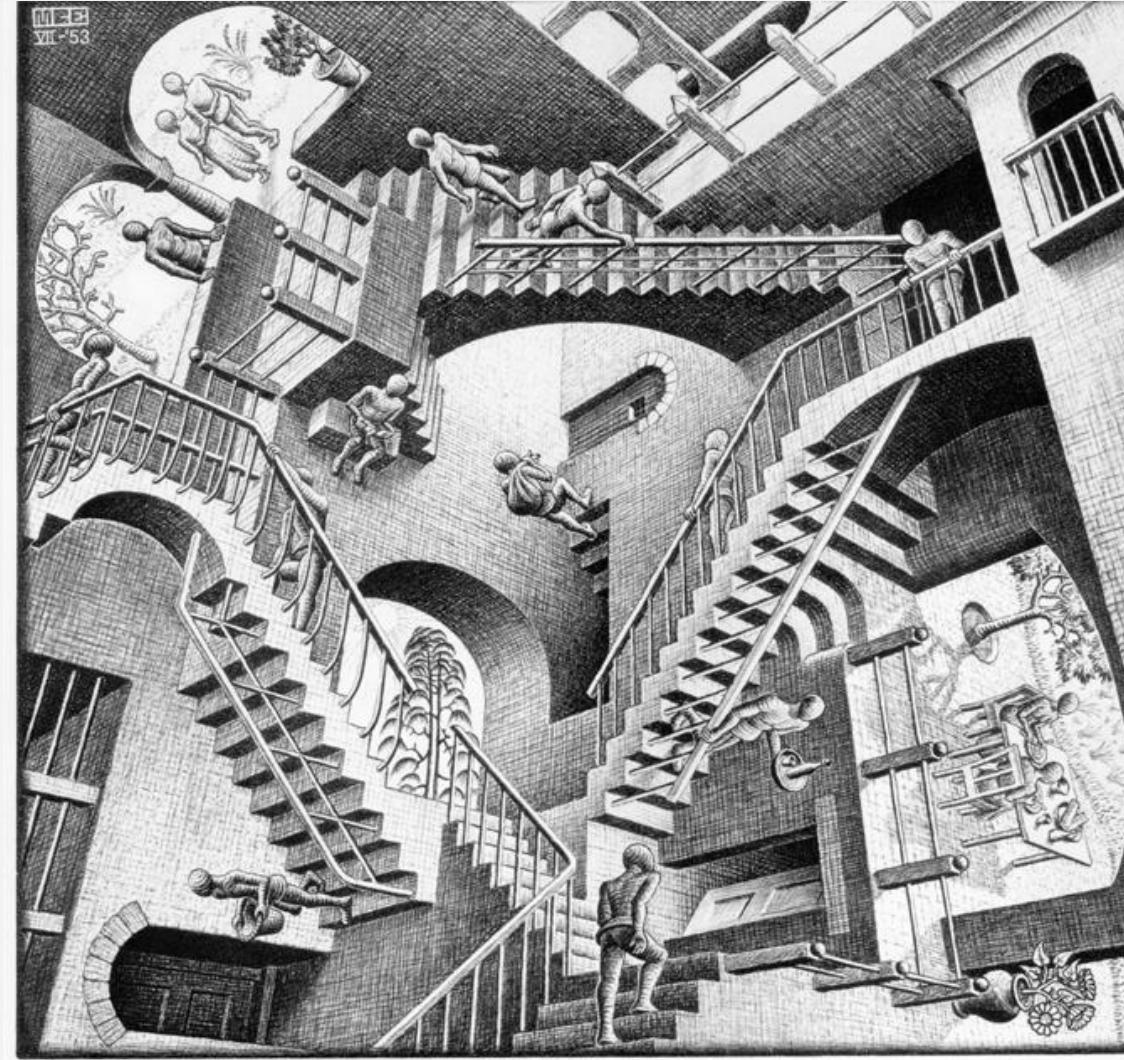
Monocular Cues to 3D Space

- Every view of the world provides multiple depth cues.
- Usually, the cues reinforce each other, and can be combined to produce a convincing and reliable representation of 3D world.
- Occasionally, however, the cues are contradictory.
 - Escher fools us by deliberately manipulating depth cues and other visual inferences. He arranges sensible local cues into a globally impossible story.

Monocular Cues to 3D space

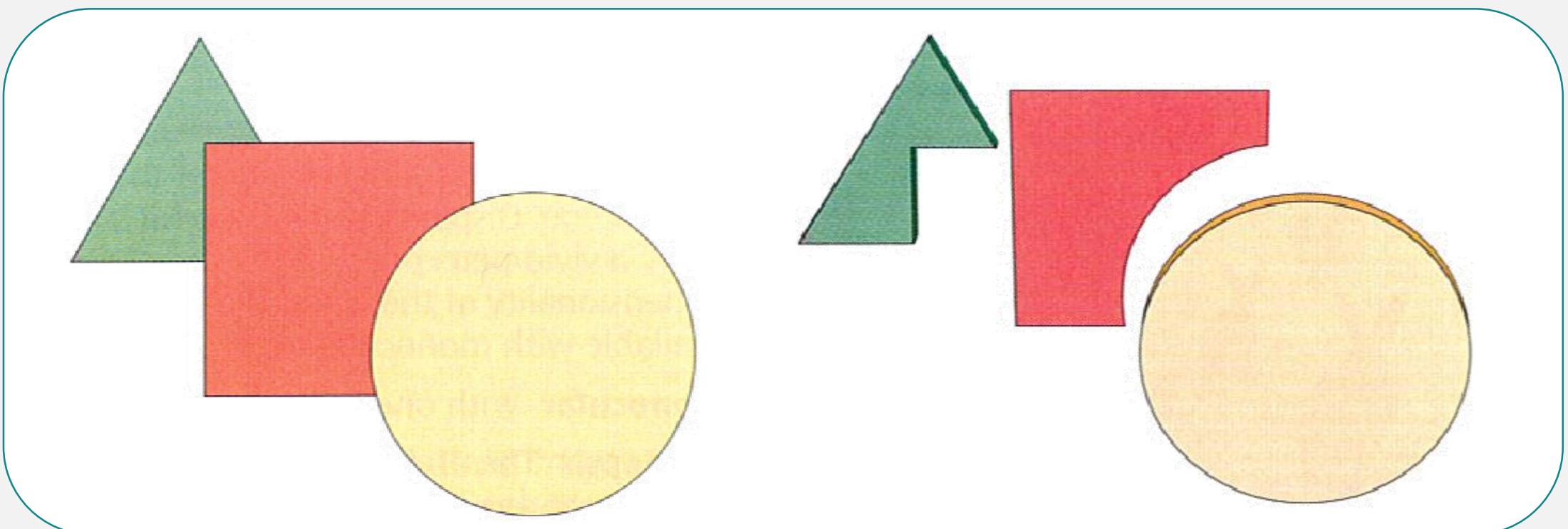


Monocular Cues to 3D space



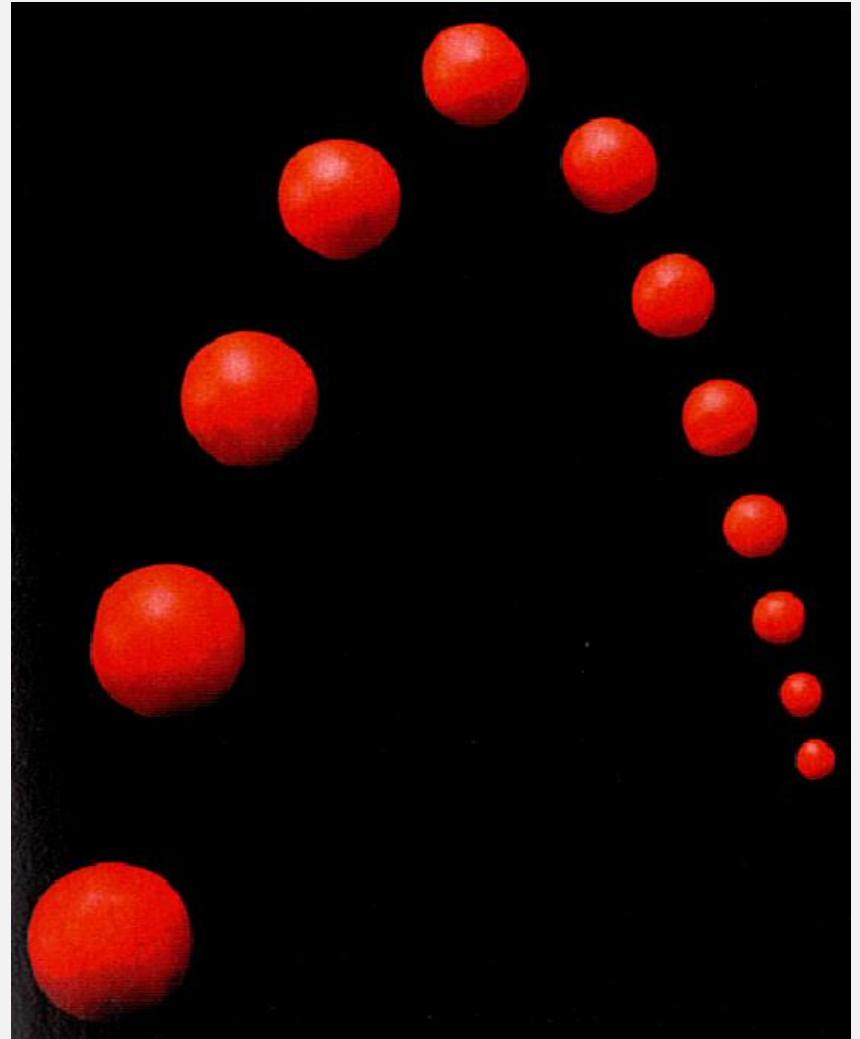
Occlusion

- Occlusion gives relative position of objects as a depth cue.
- It occurs in almost every scene and some argues that it is the most reliable depth cue.

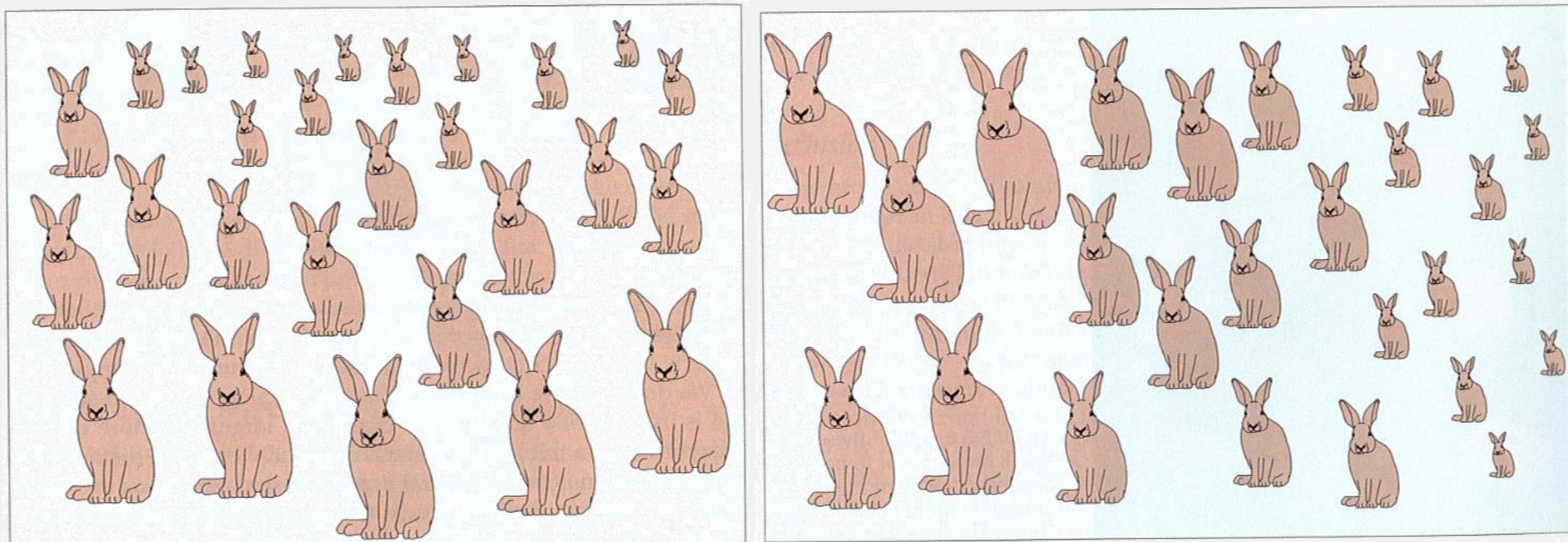


Size and position cues

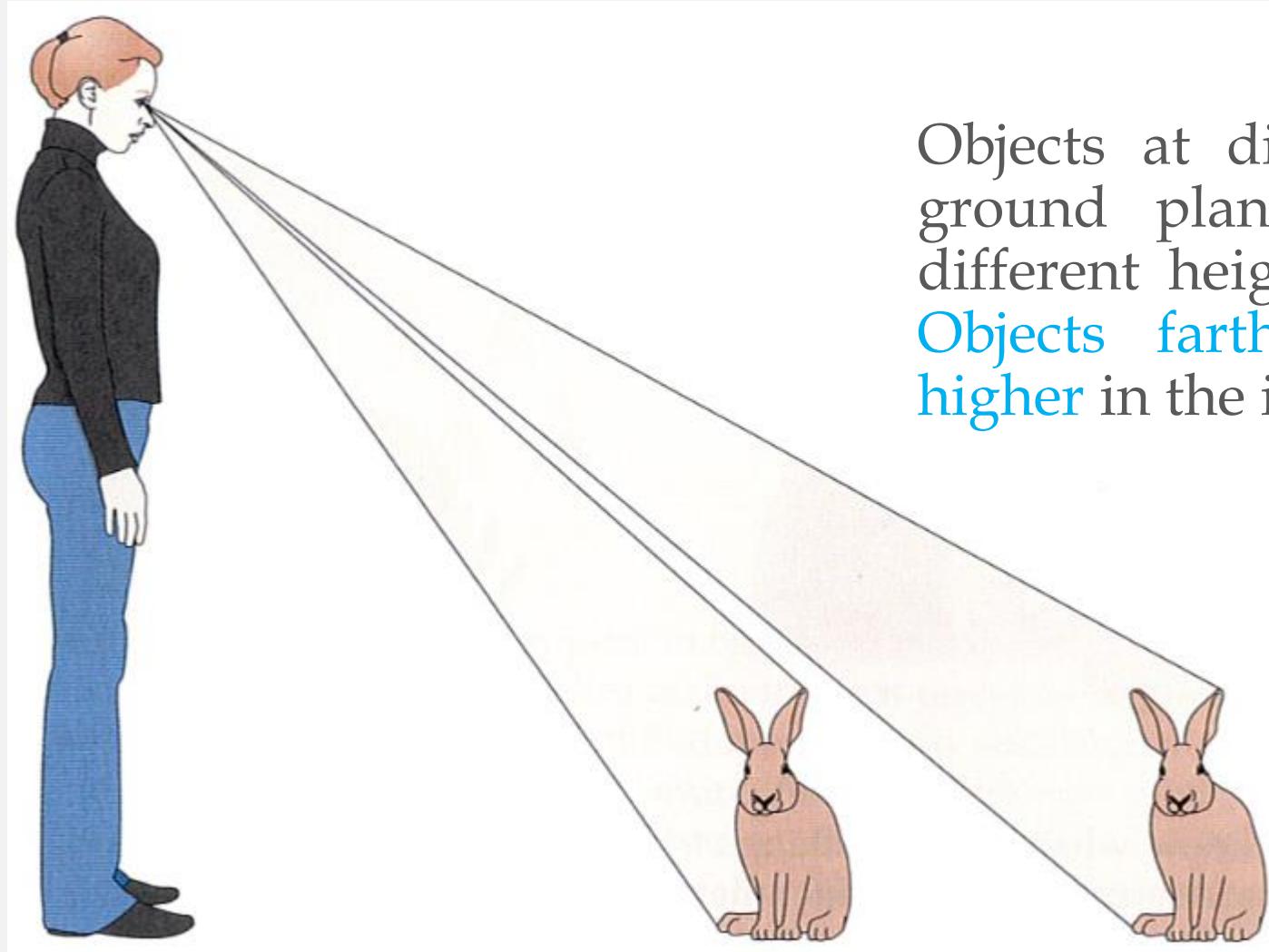
- We have projective geometry embedded.
 - Smaller things are farther away.



Texture Gradient



Relative Height



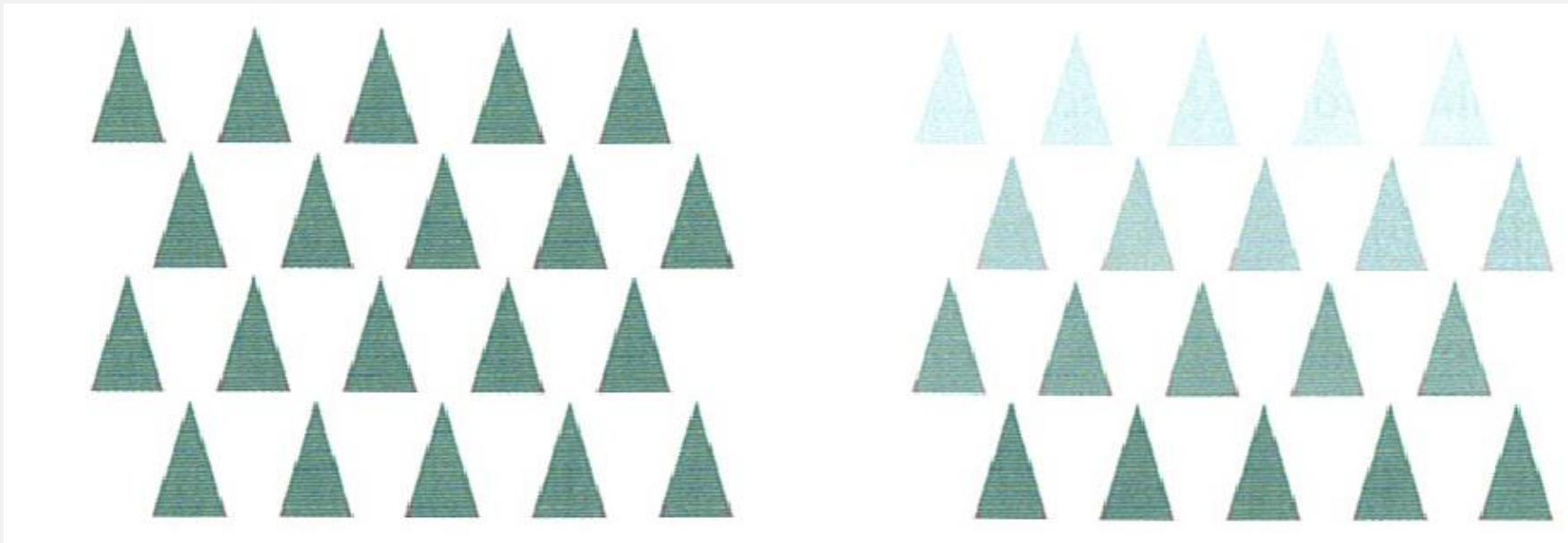
Objects at different distances on the ground plane will form images at different heights in the retinal image. Objects **further away** will be seen as **higher** in the image.

Familiar Size



Familiar size: a depth cue based on knowledge of the typical size of objects.

Aerial Perspective

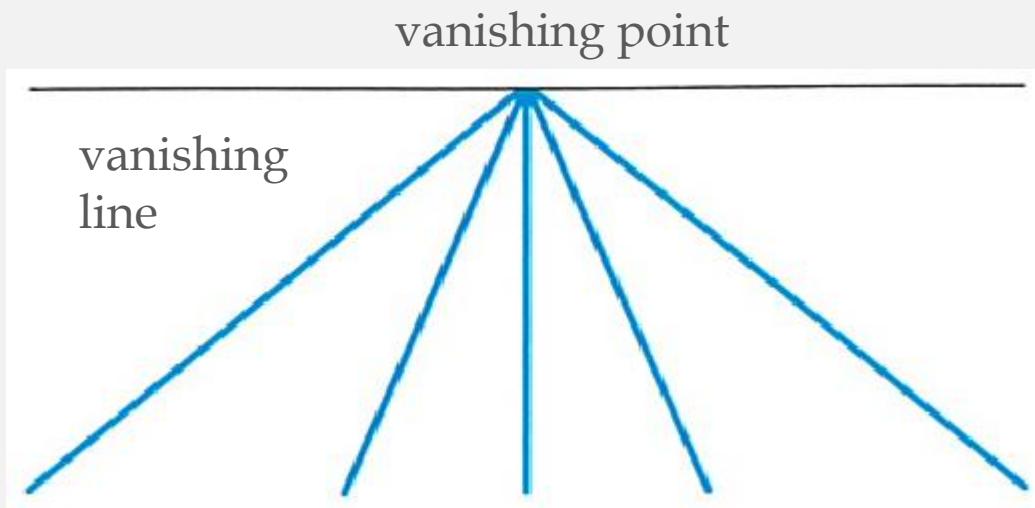


Aerial (haze) perspective: light is scattered by the atmosphere, and more light is scattered when we look through more atmosphere.

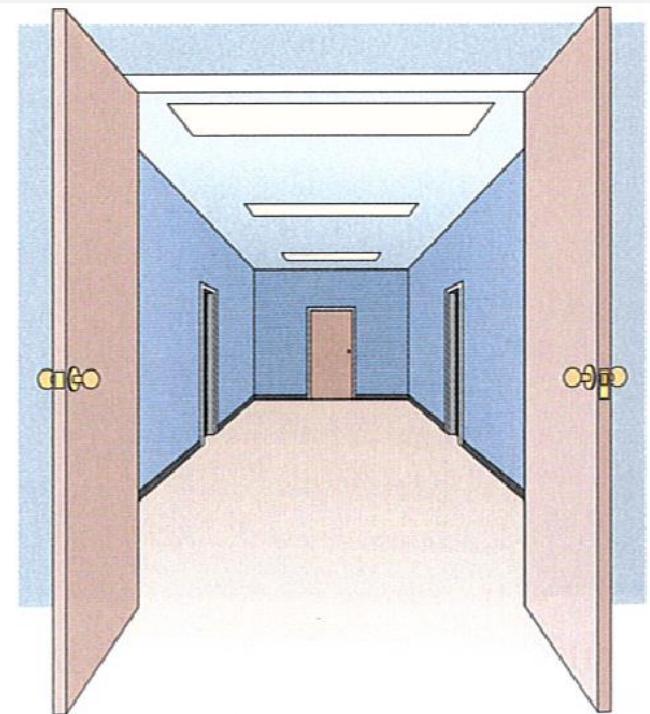
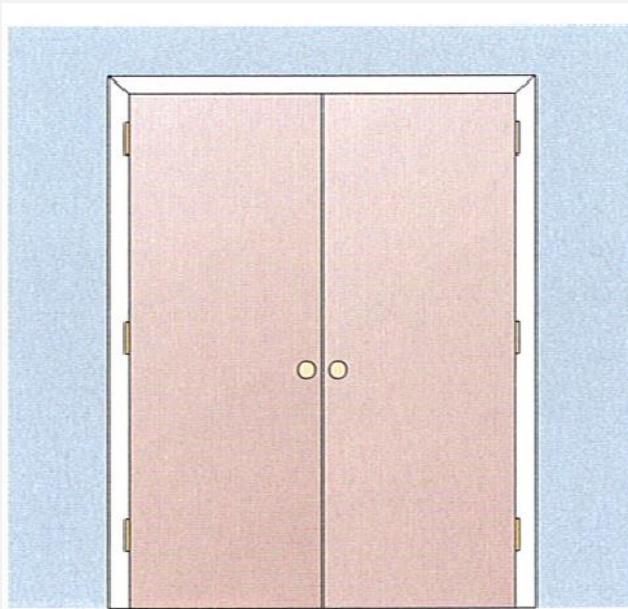
Aerial Perspective



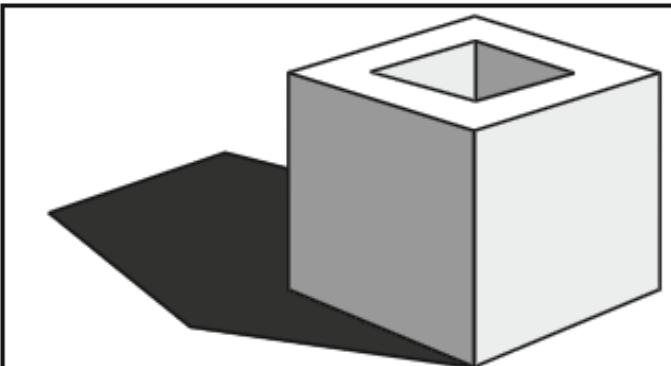
Linear Perspective



Parallel lines in 3D world appear to converge in the 2D image, except for the ones parallel to the image plane.



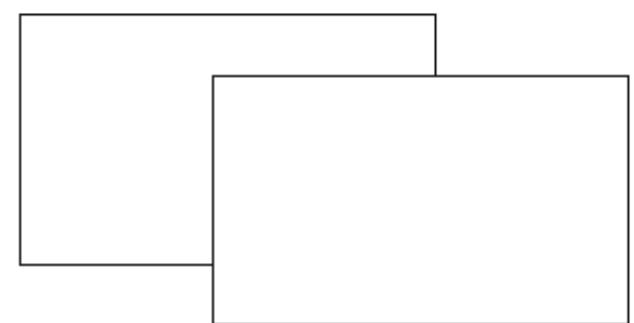
Monocular Cues



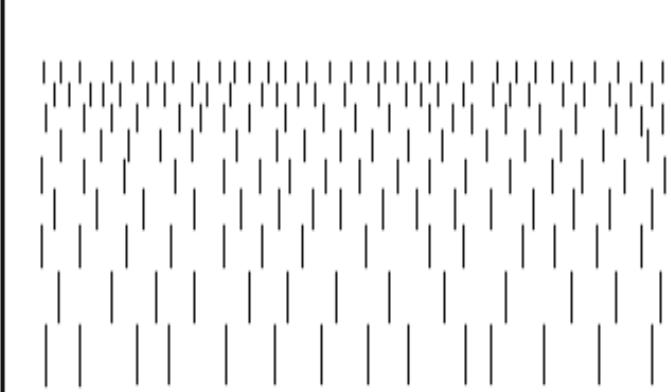
Light and shade



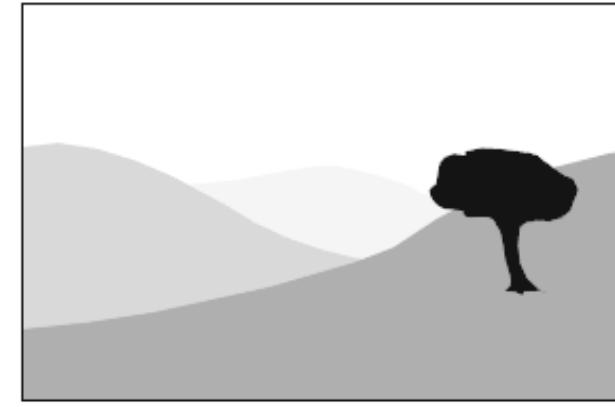
Relative size



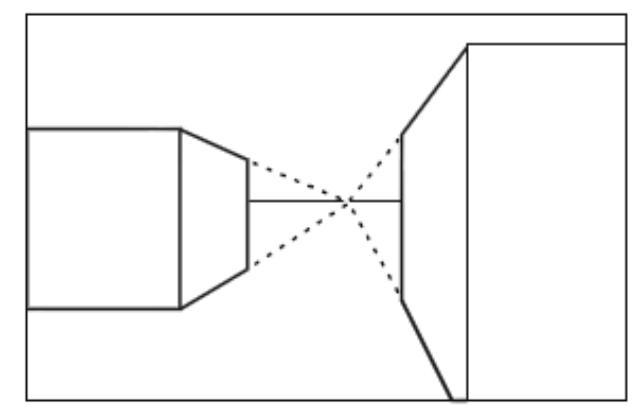
Interposition



Textural gradient



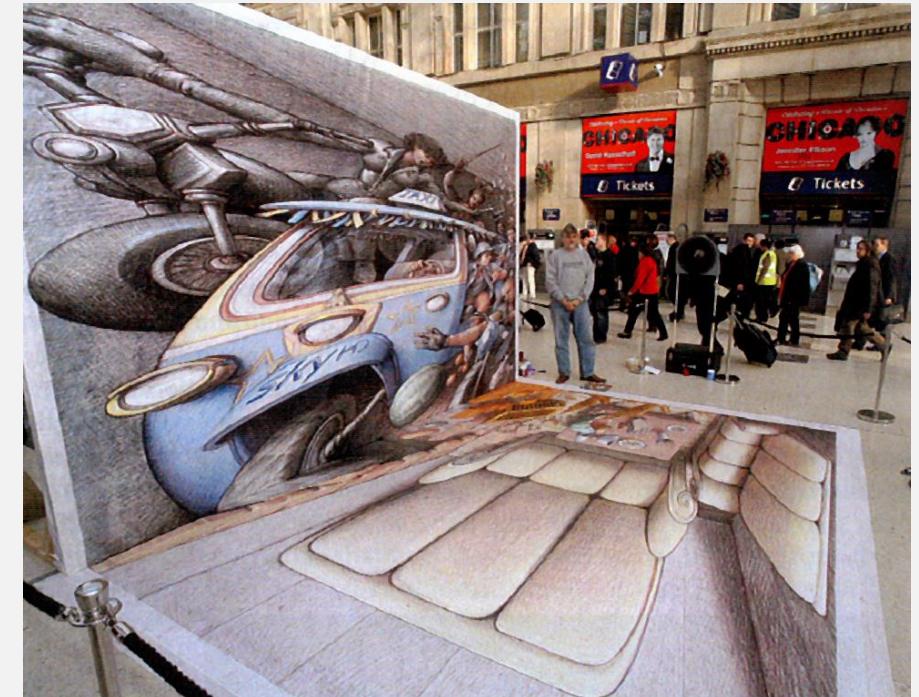
Aerial perspective



Perspective

Pictorial Depth Cues

- All these monocular cues are pictorial depth cues produced by the projection of the 3D world onto the 2D surface of the retina.
- Combined with proper shading, these cues could be effective in illustrating 3D.



Pictorial Depth Cues



© 2007 Kurt Wenner



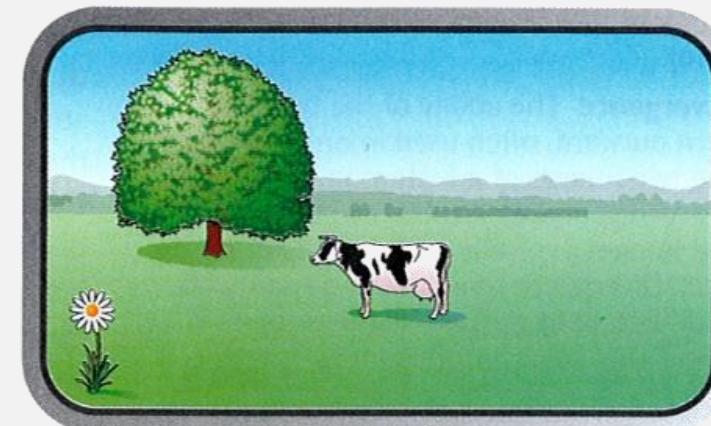
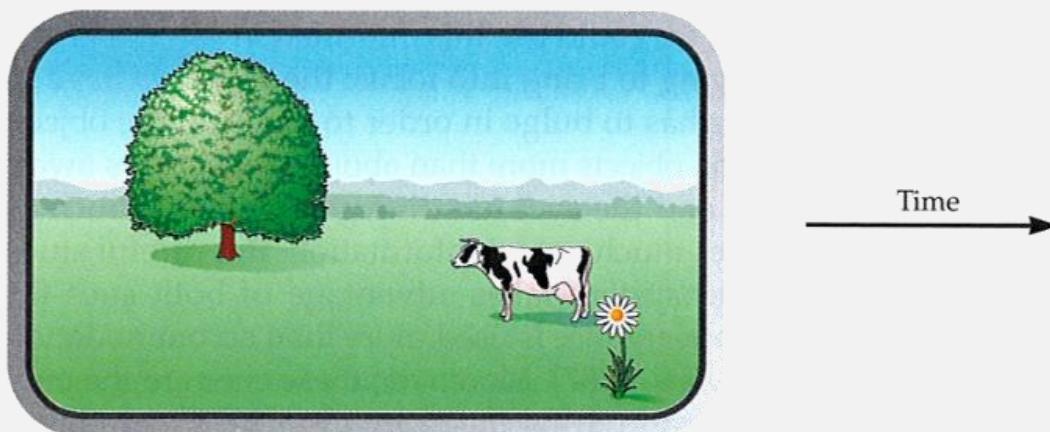
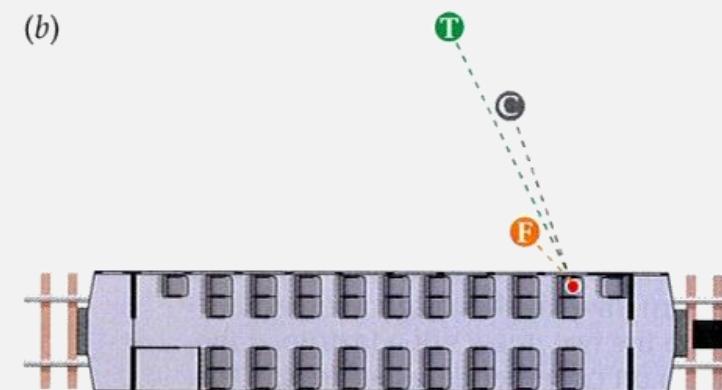
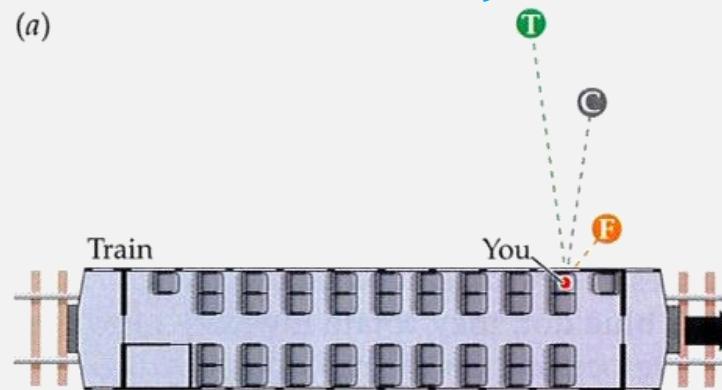
© 2007 Kurt Wenner

Monocular cues

- Because there are lots of monocular cues and they are not less important than binocular ones, some images could look more stereoscopic than others.
- Example, 2D-to-3D conversion is easier for some images but more difficult for others.

Motion Cues

- Motion parallax is a non-pictorial depth cue. When your head moves, **closer objects move faster** than more distant ones.

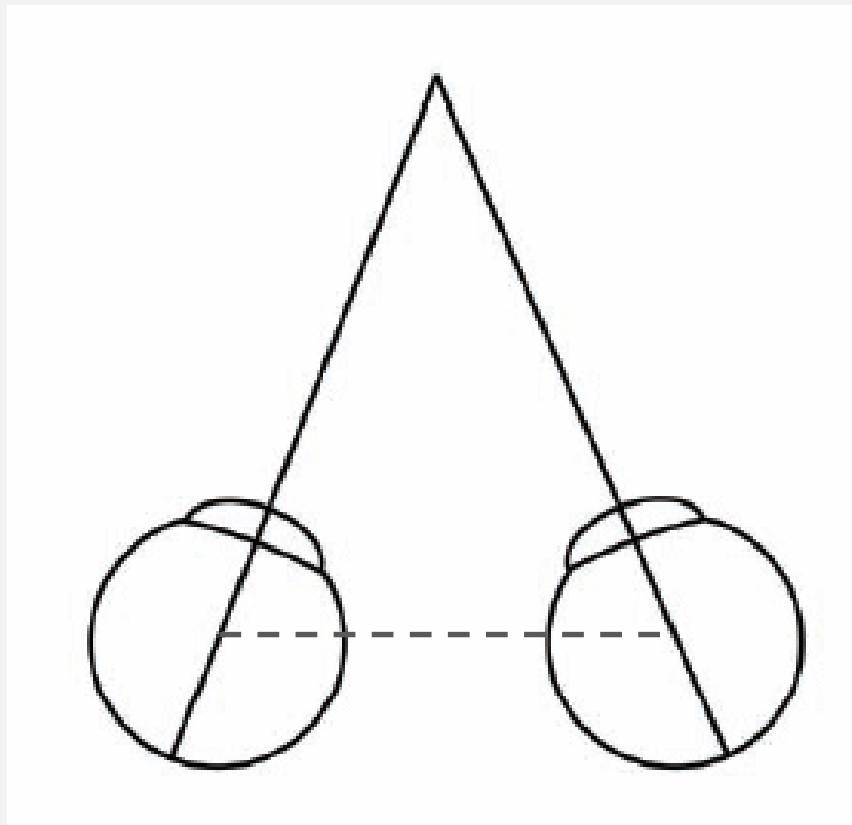


Motion Cues

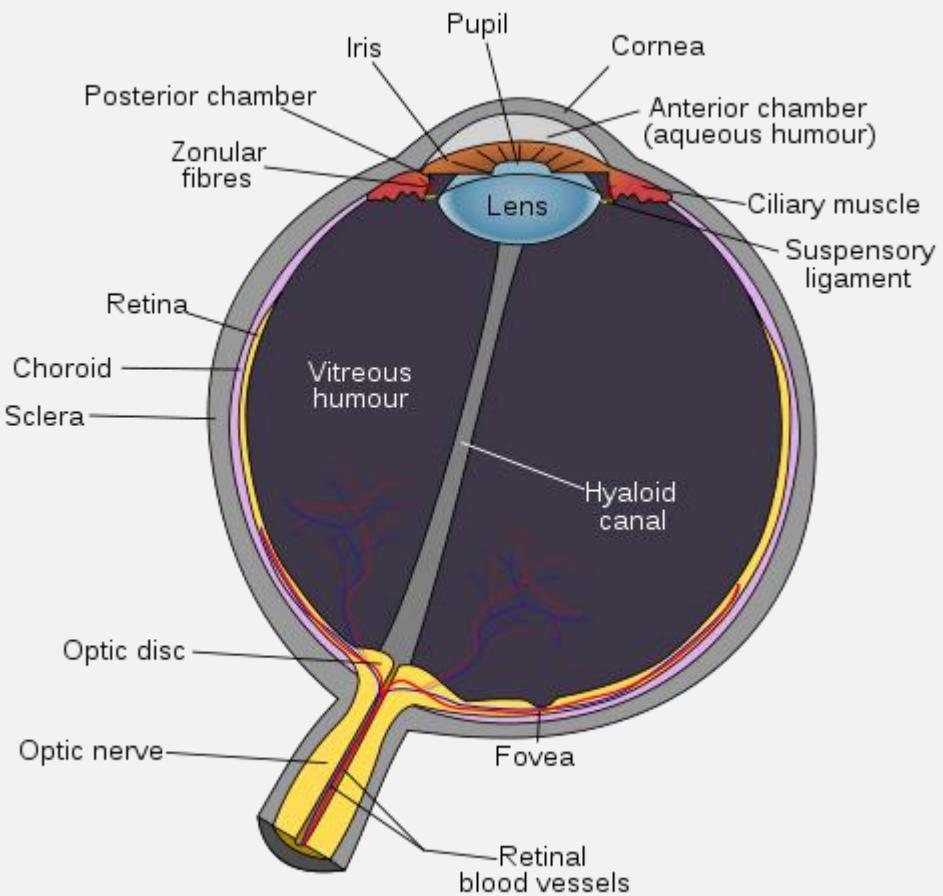
- A very effective depth cue but it relies on head movements.
- Some 3D games are designed this way.



Binocular Vision

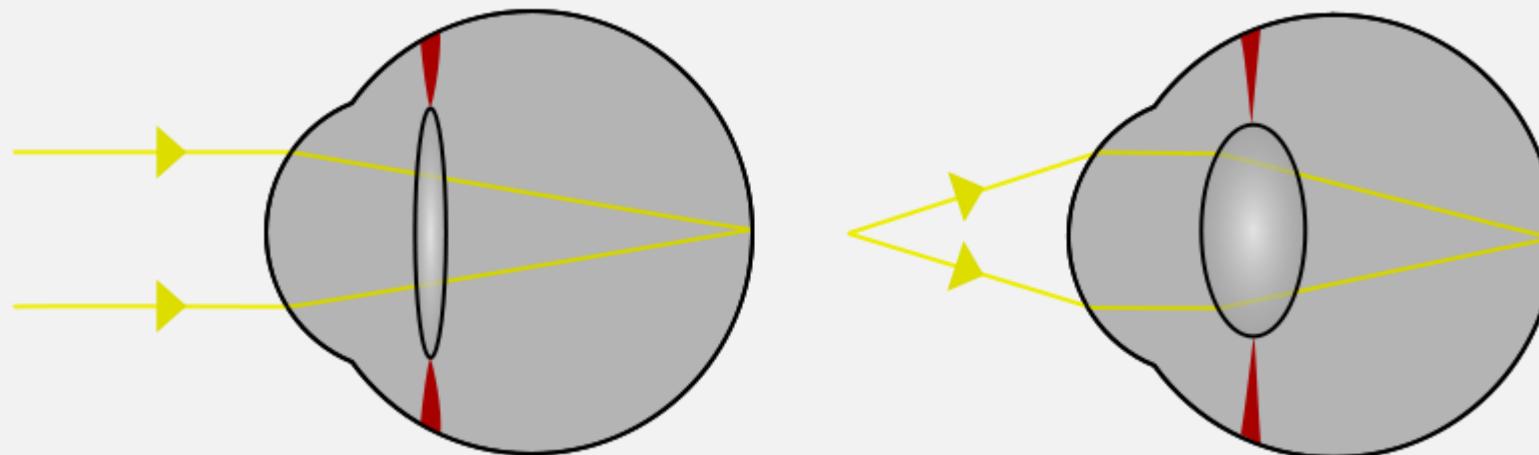


interocular distance ~6.5cm

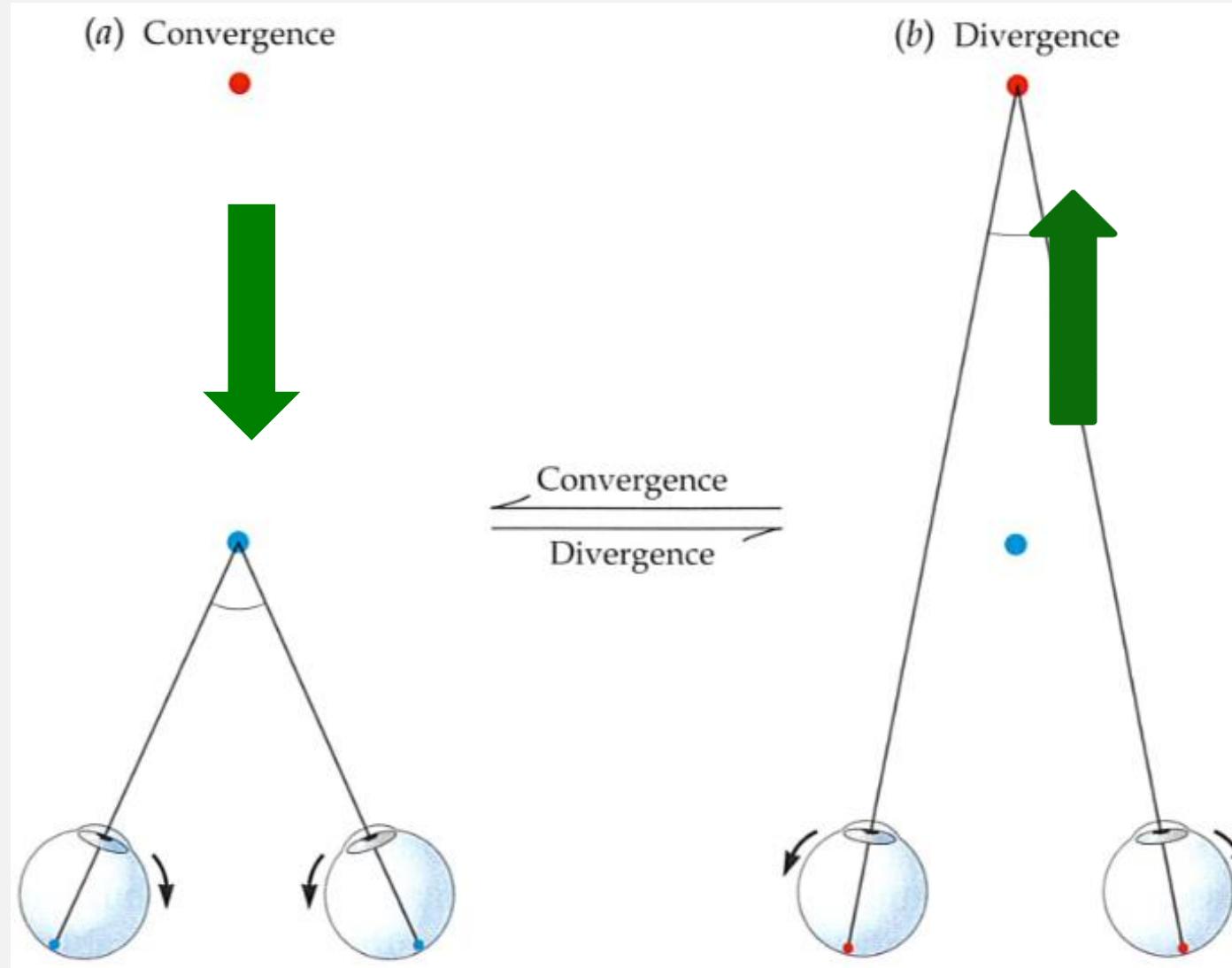


Accommodation and Convergence

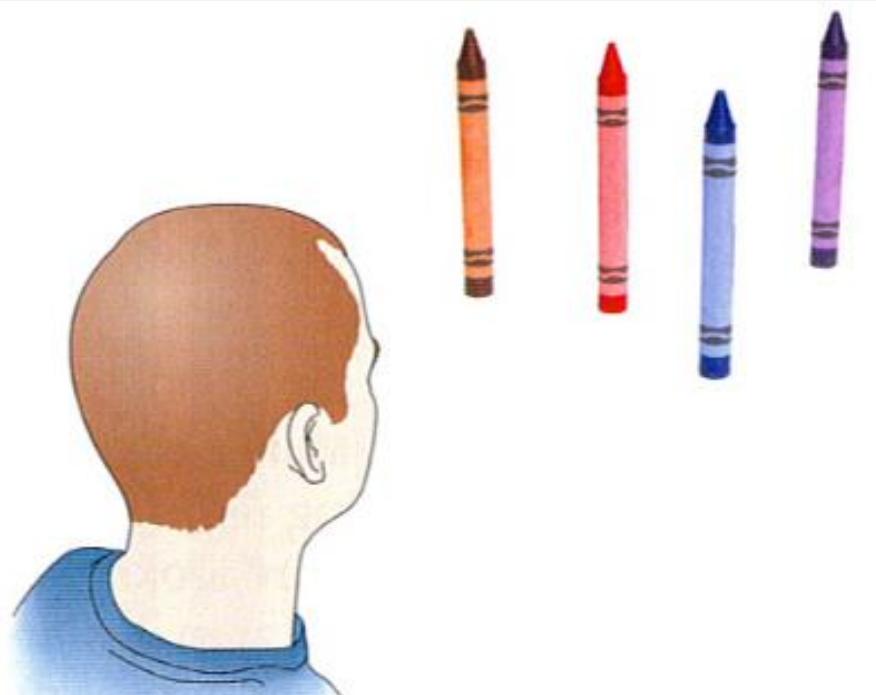
- Human can perceive depth by accommodation and convergence.
- Eyes need to be focused to see objects at different distances clearly.
- Human eye focuses via a process called accommodation, in which lens gets fatter as we direct our gaze toward nearer objects.



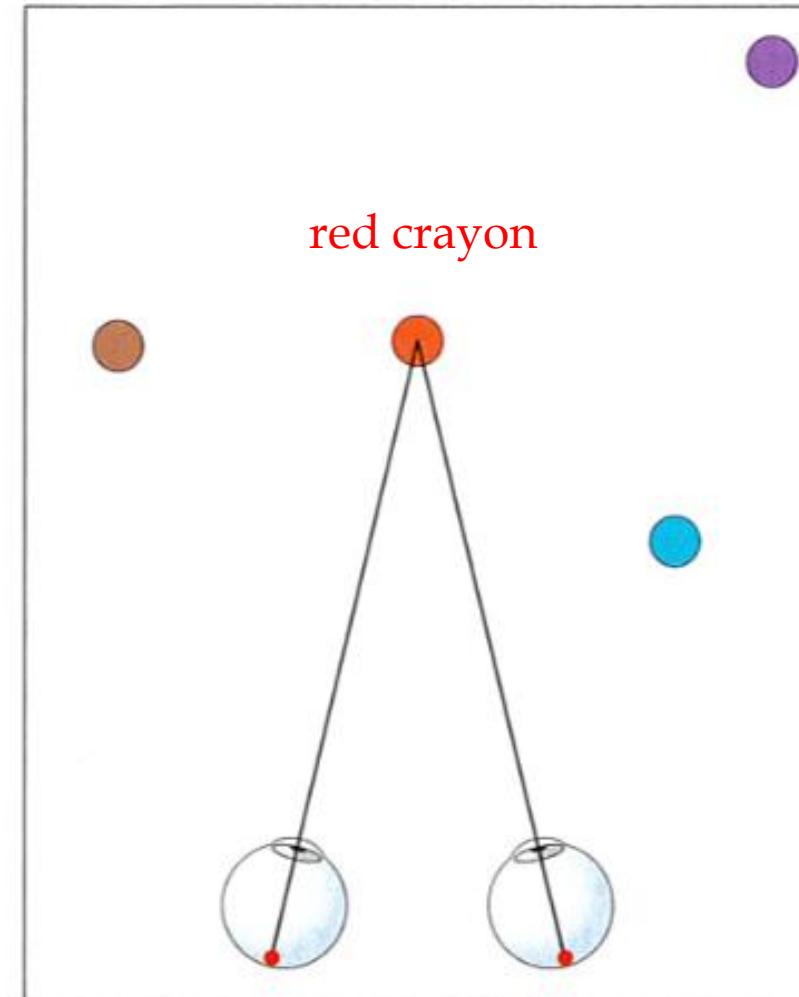
Accommodation and Convergence



Binocular Vision

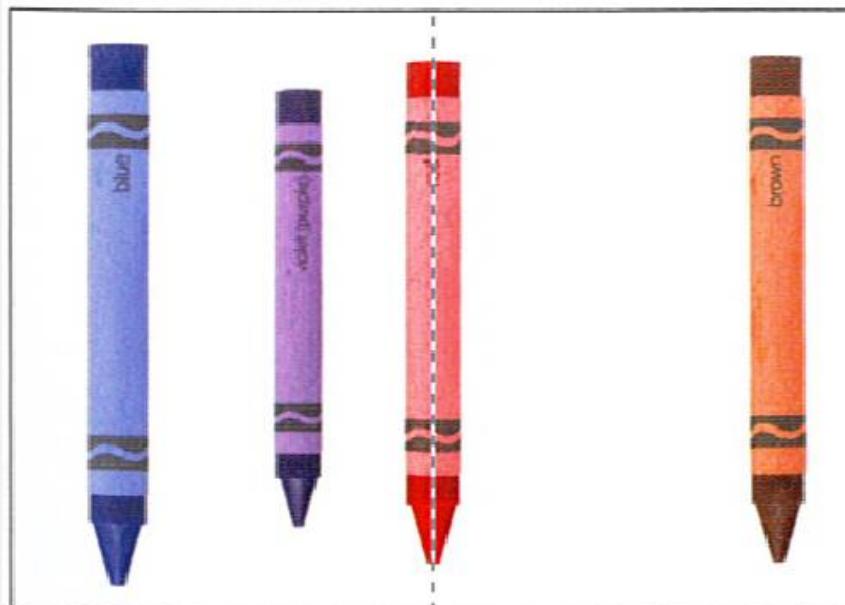


Assume that the viewer is focusing on the red crayon.



Binocular Vision

- Note that the retinal images are inverted. The object of our gaze falls on the fovea, center of the retina.
- The blue one happens to fall on **corresponding retinal points**.

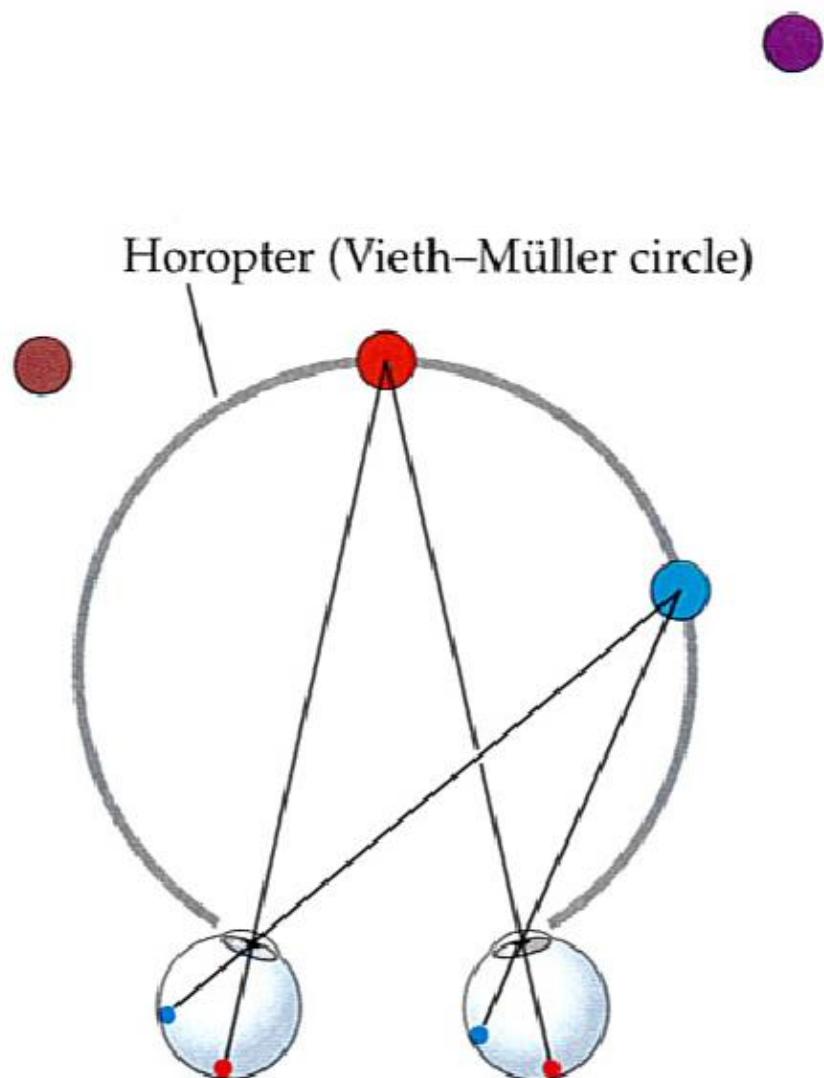


Left retinal image



Right retinal image

Binocular Vision

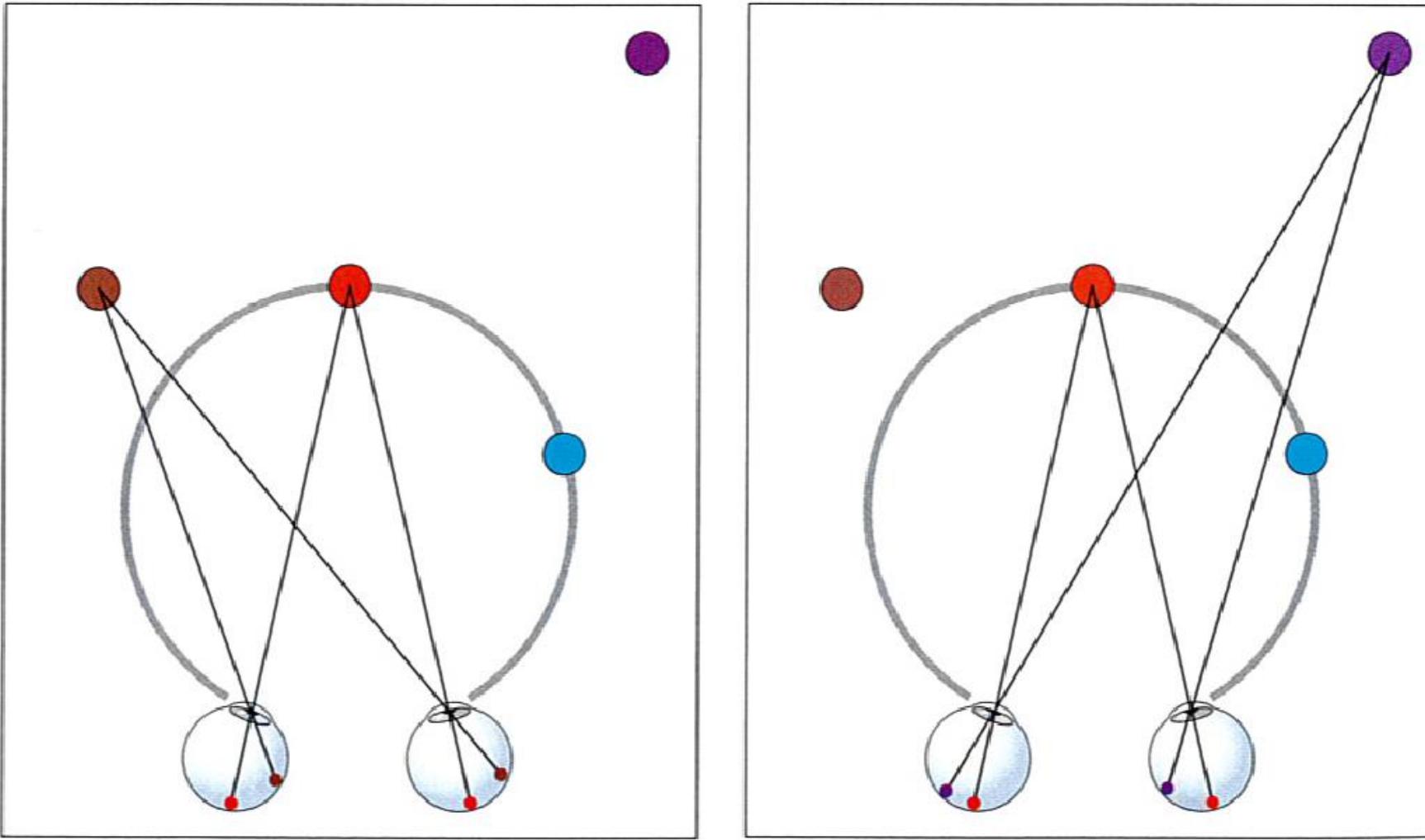


Horopter:
the surface with zero disparity.

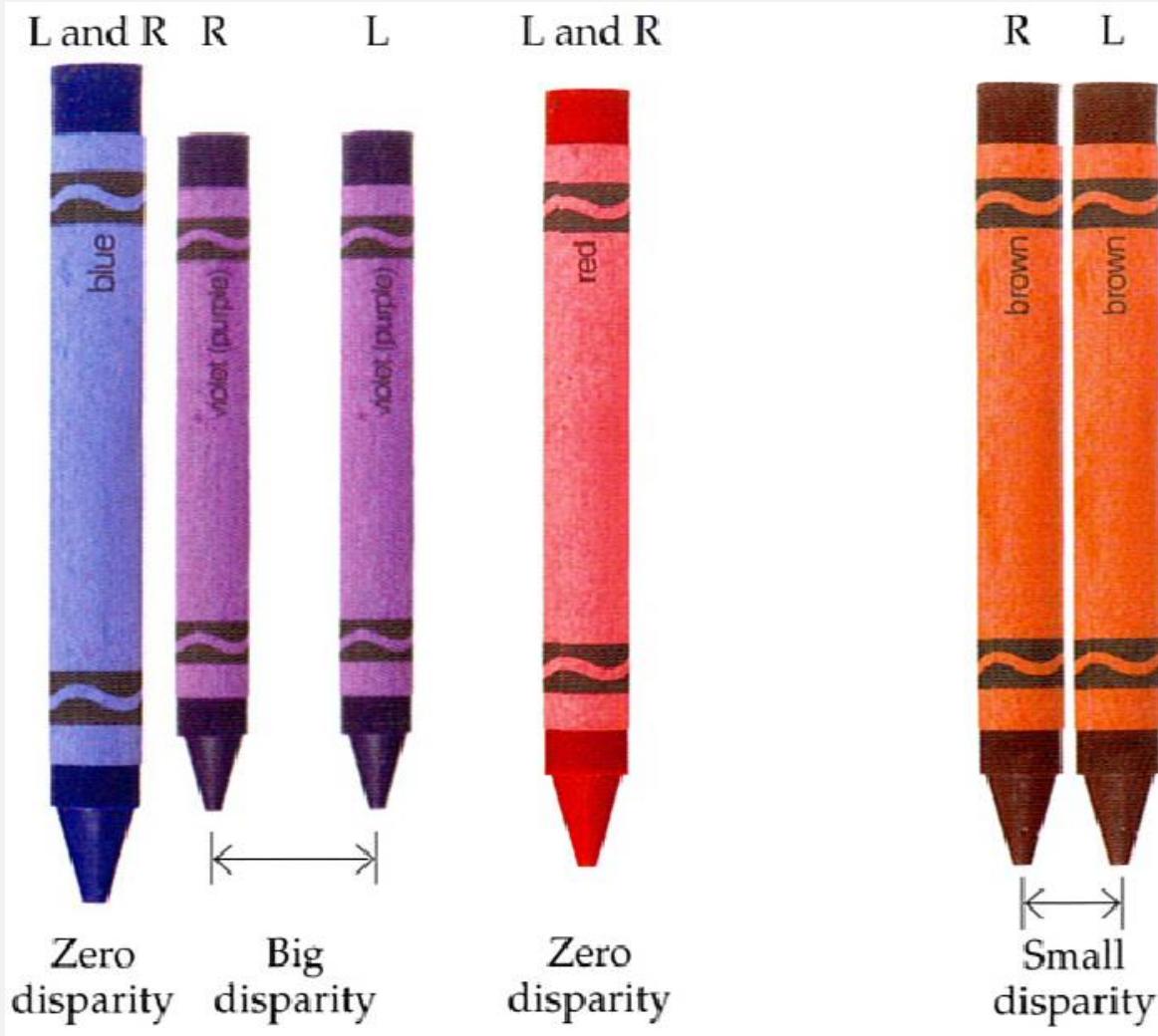
Panum's fusional area:
the region of space, in front of and
behind the horopter, within which
binocular vision is possible.

Diplopia:
double vision

Binocular Vision



Binocular Vision



Binocular Vision



left view



right view

Binocular Vision



Binocular Vision



Binocular Vision

- Our nervous system cannot measure the angle very accurately. Thus, we can only perceive relative depth.
- The role of eye movement is to bring the images within Panum's fusional area.

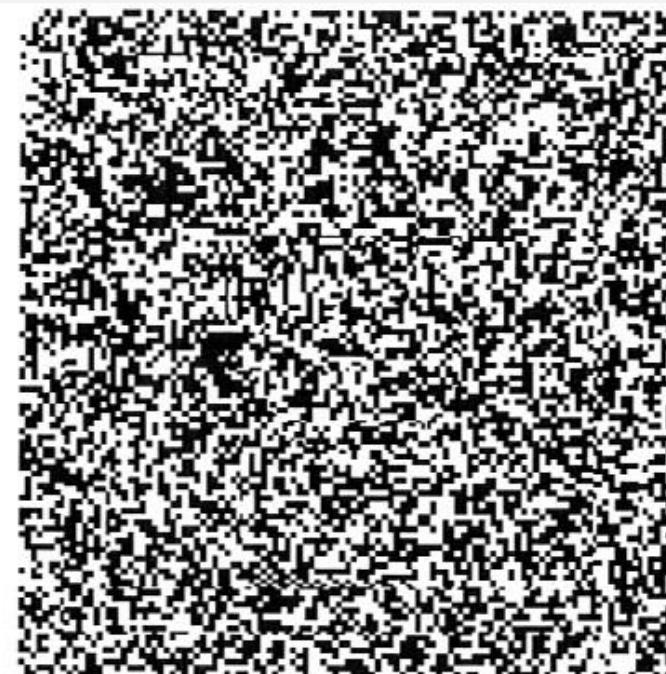
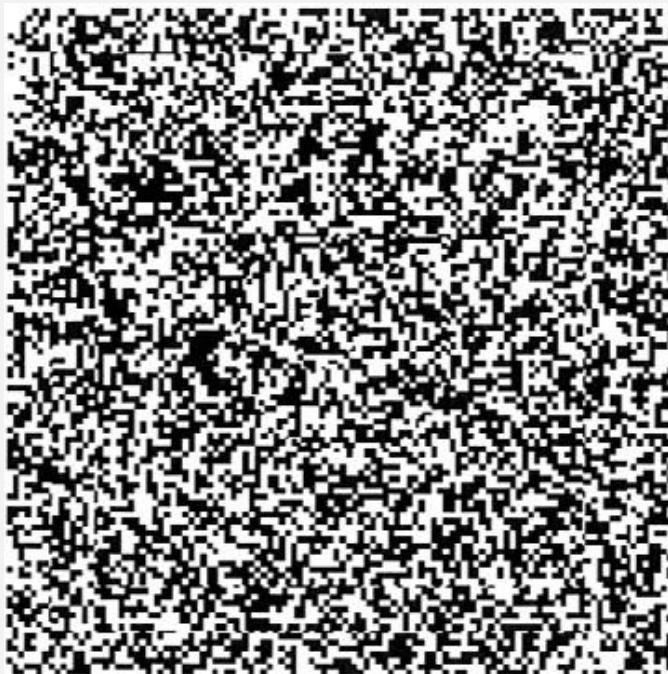
Random Dot Stereograms

- For 100 years after the invention of stereoscope, it is supposed that stereopsis occurred relatively late in the processing of visual stimuli.
- We recognize facial features and then use them to find depth.



Random Dot Stereograms

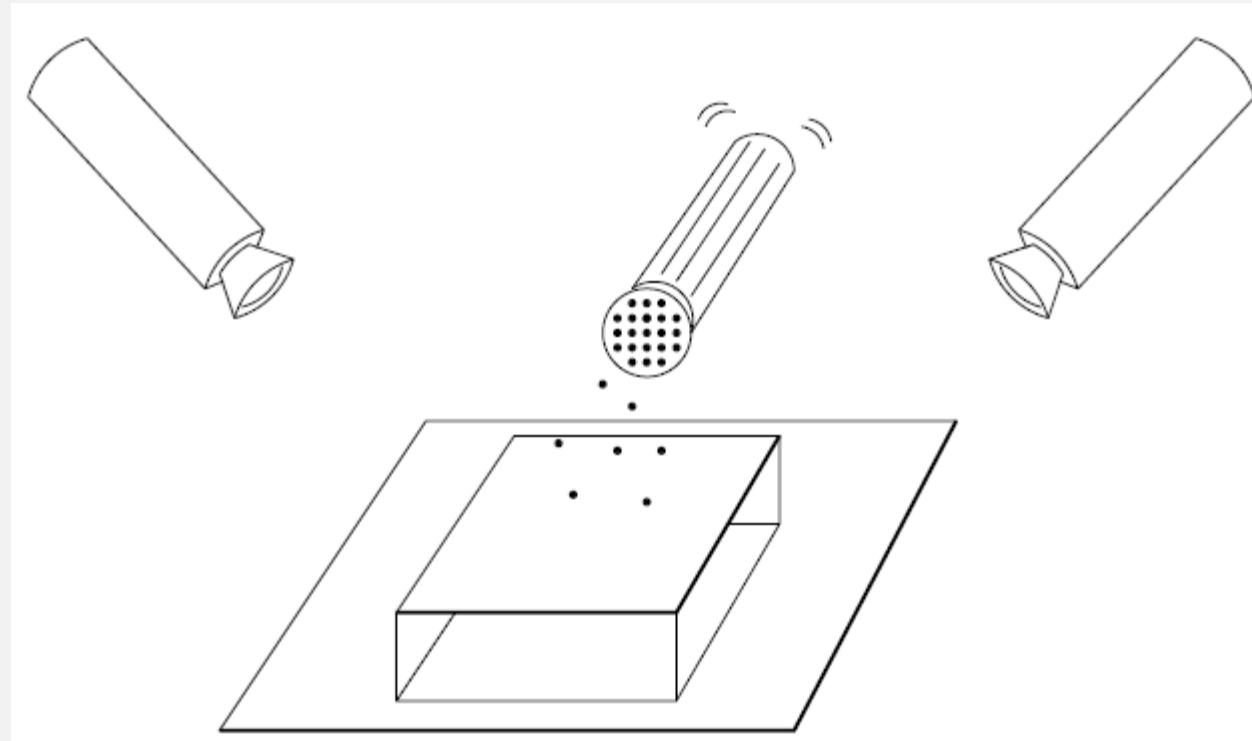
- Julesz thought that stereopsis might help reveal camouflaged objects (the example of cats and mice) and invented random dot stereograms.



There is no monocular cue, yet we can recognize the shape.
Disparity alone is enough for stereopsis!

Random Dot Stereograms

- Paint the scene in white and spray it with pepper.



Summary

- Monocular cues:

- occlusion, size and position cues, aerial perspective, linear perspective.

- Motion cues

- Accommodation and convergence cues

- Binocular cues

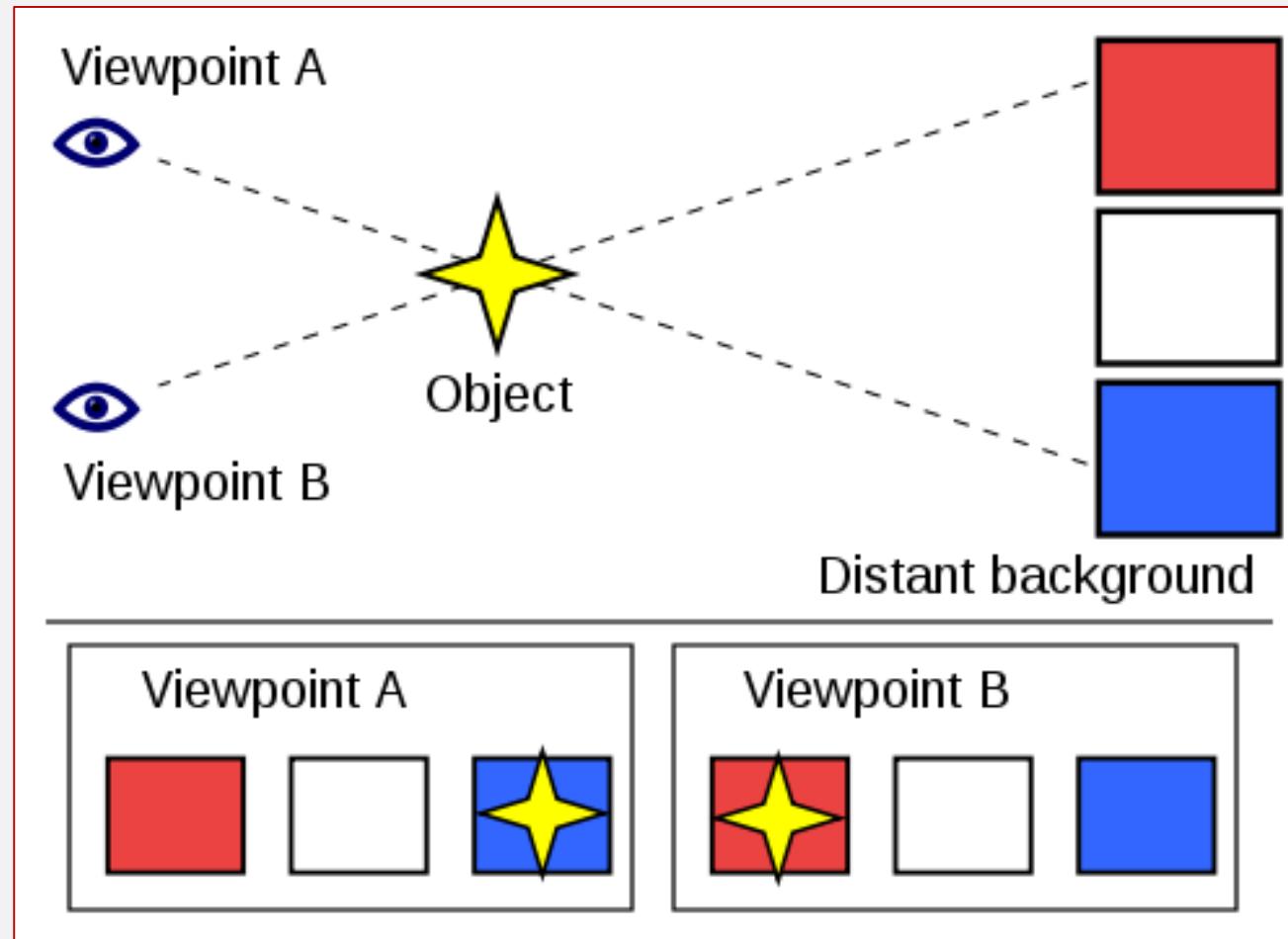
- resolve stereo correspondence problem, Panum's area.

Stereoblindness

- An inability to make use of binocular disparity as a depth cue.
- Approximately 3% to 5% of the population lacks stereoscopic depth perception.

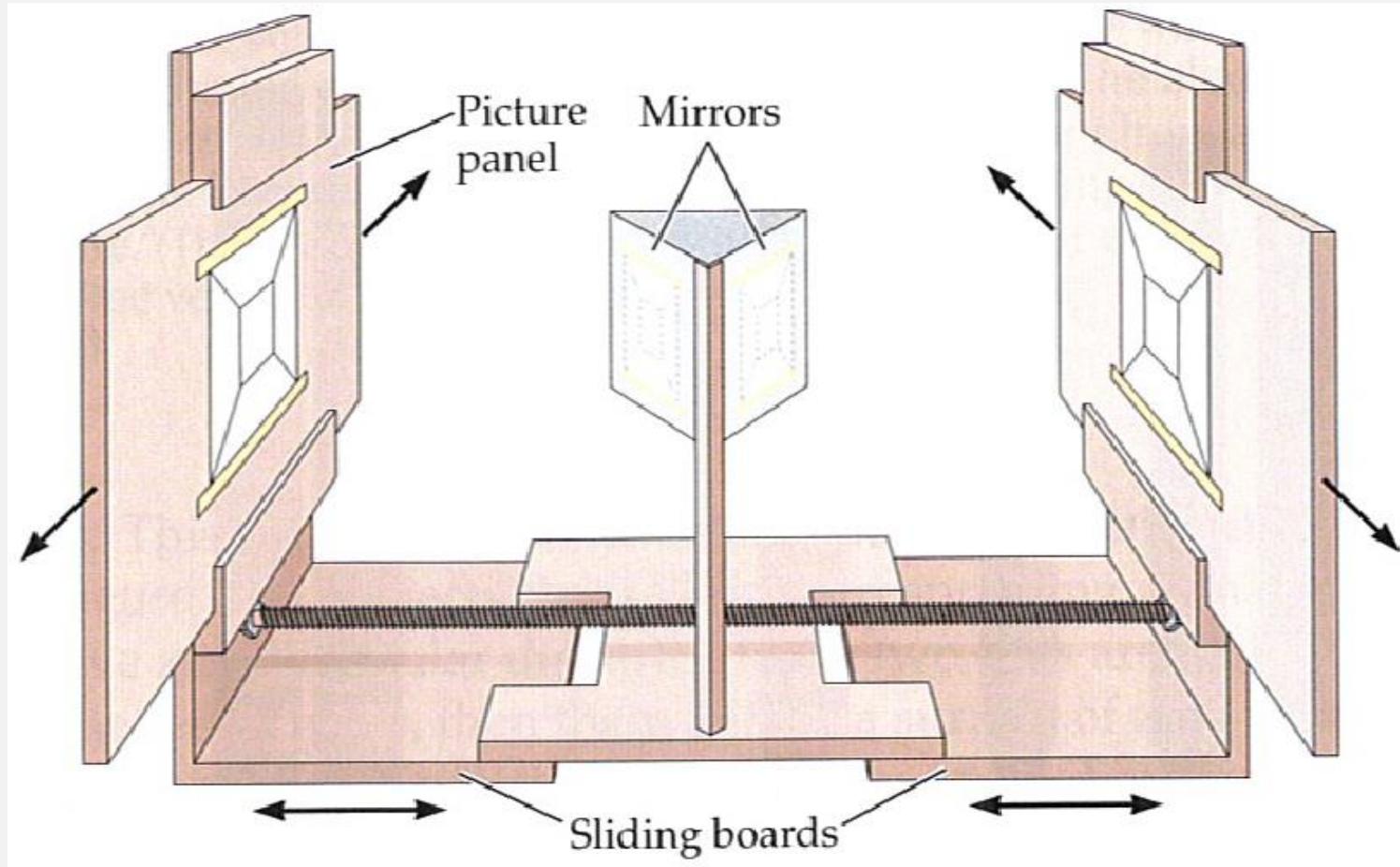
3D displays

Parallax



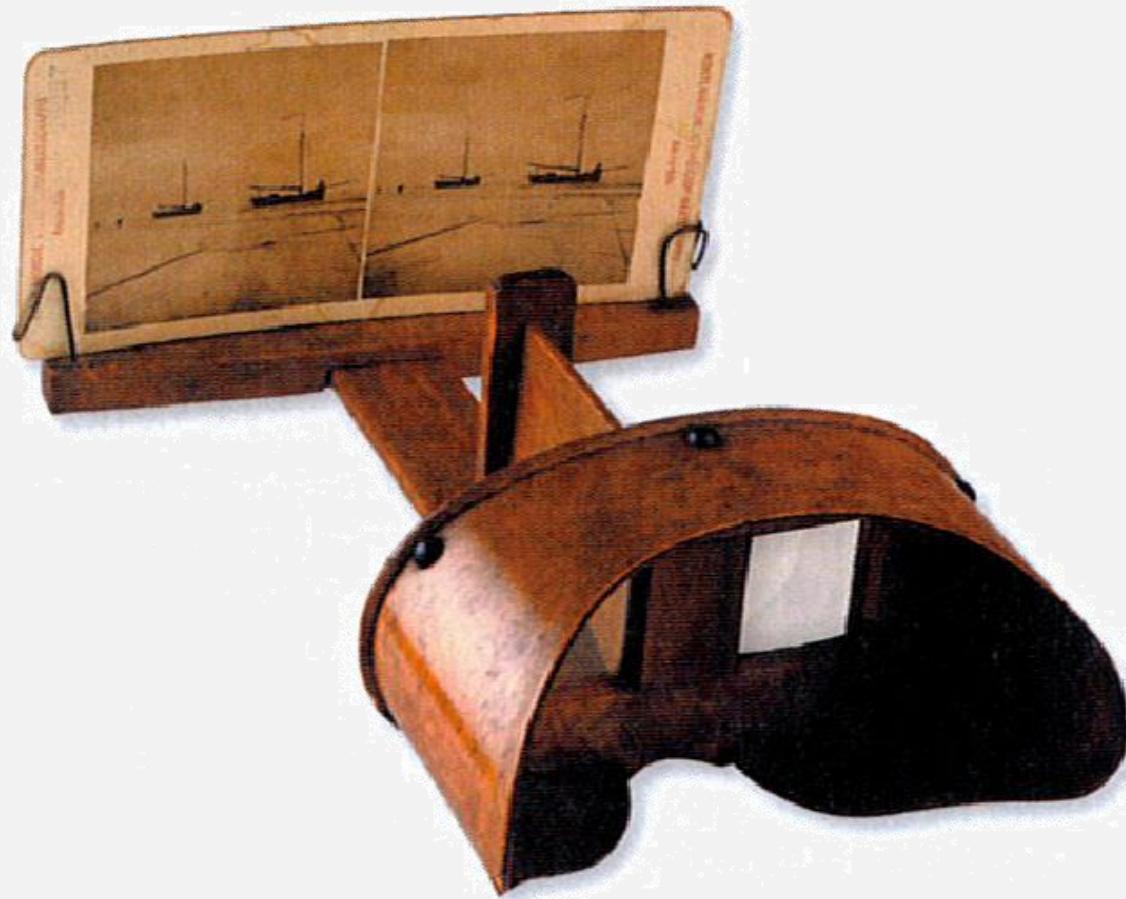
Stereoscope

- Invented by Sir Charles Wheatstone in 1830s.
- Our visual system treats binocular disparity as a depth cue, no matter it is produced by actual or simulated images.



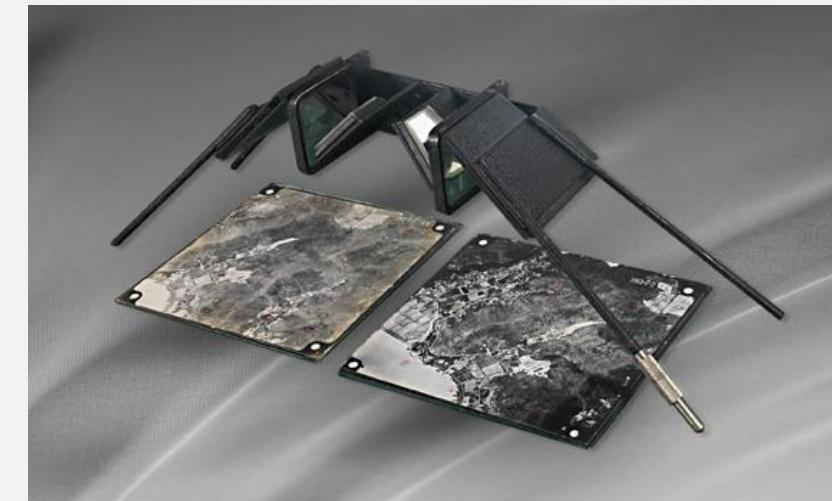
Stereoscope

- Invented in 1850s, it is stereopsis for the masses.

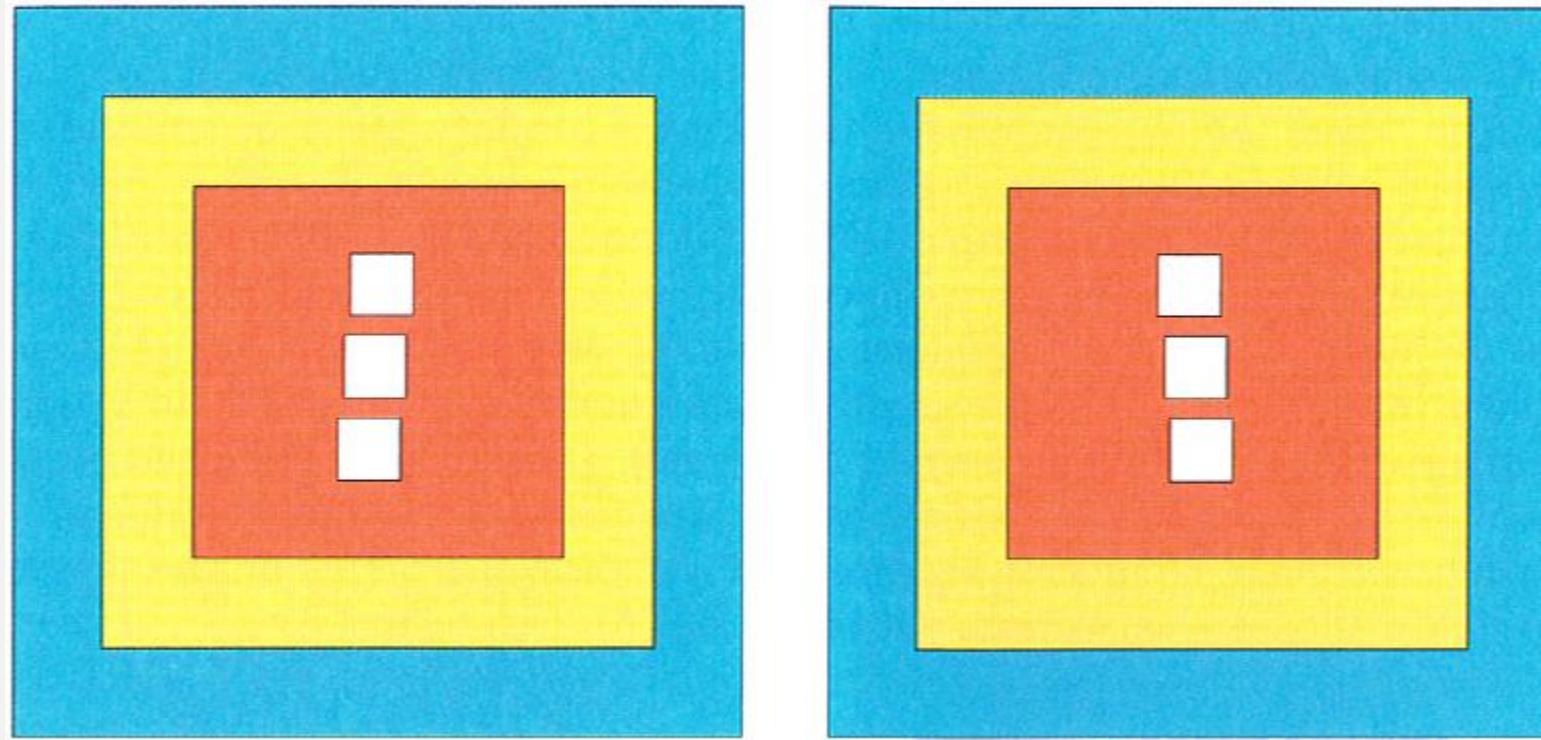


3D Displays

- Monocular cues can be produced by rendering/capturing the contents correctly.
- Most 3D displays enrich space perception by exploiting binocular vision. Thus, they have to present different contents to each of both eyes.



Free Vision



Free vision: the technique of converging (crossing) or diverging the eyes in order to view a stereogram without a stereoscope.

平行視線法



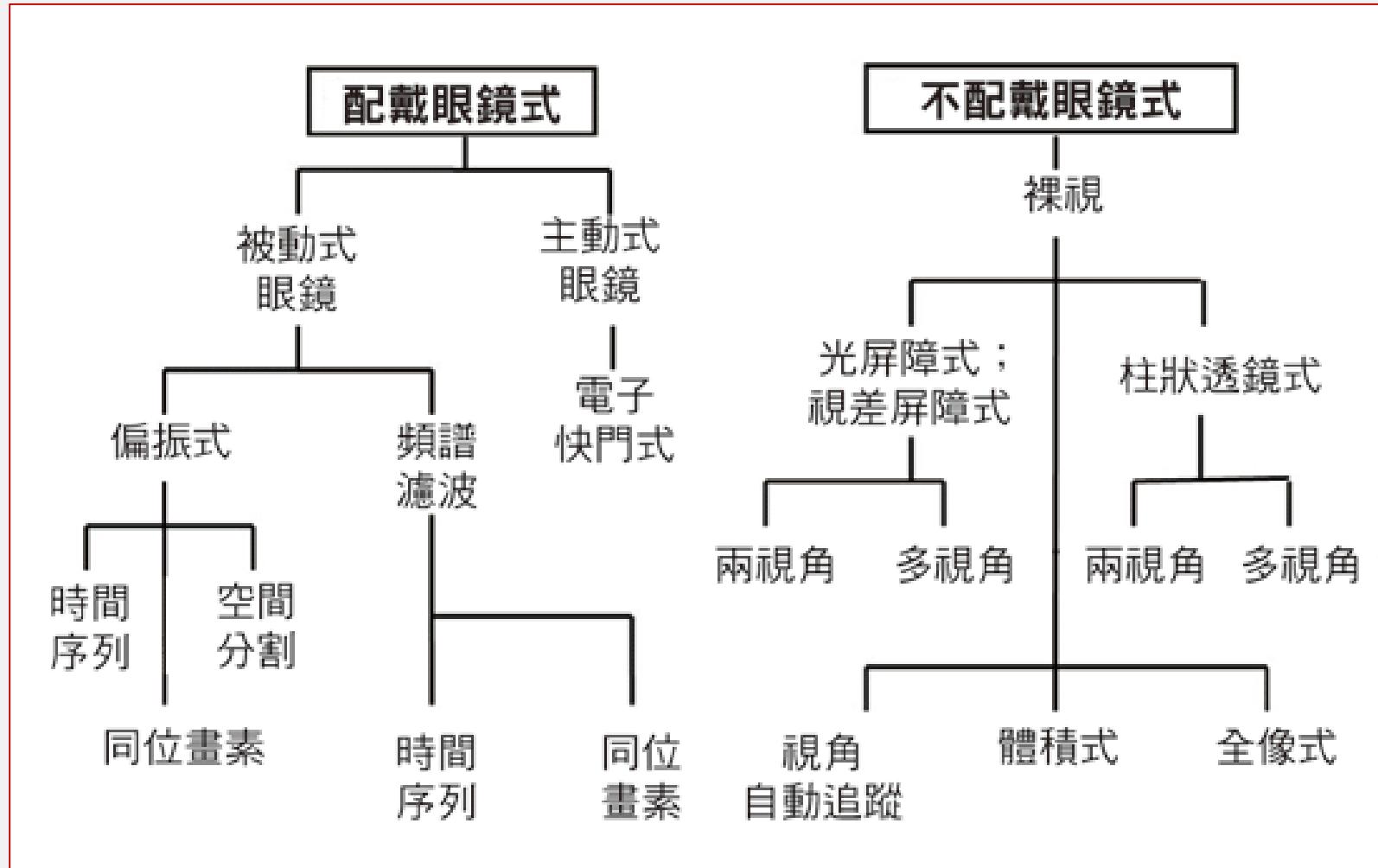
交叉視線法



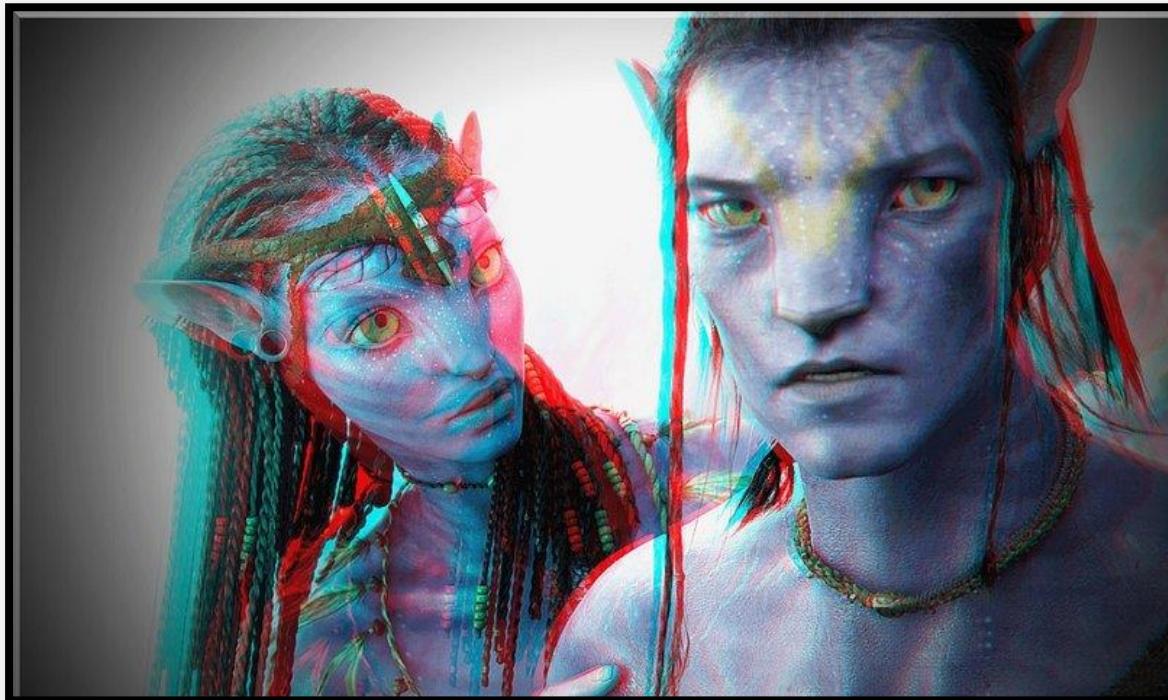
雙鏡筒式



3D Displays Nowadays



Anaglyph Glasses

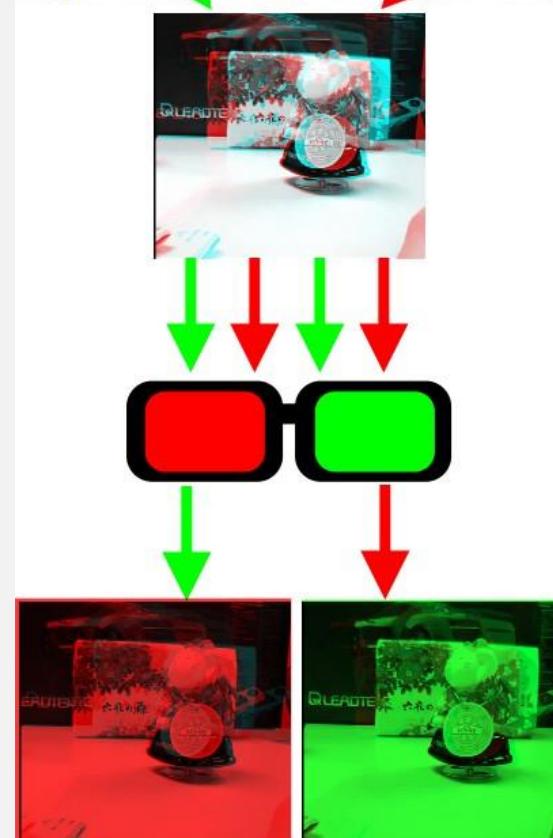


Pros: cheap (home-made)
Cons: without colors
bad 3D

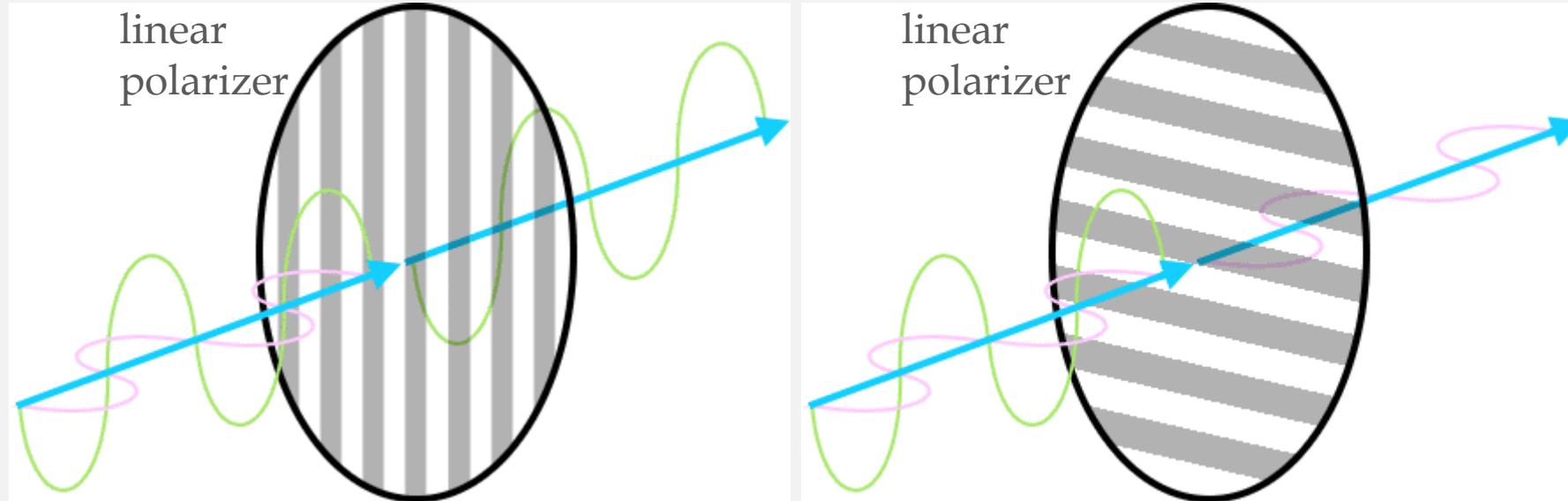
Anaglyph glasses



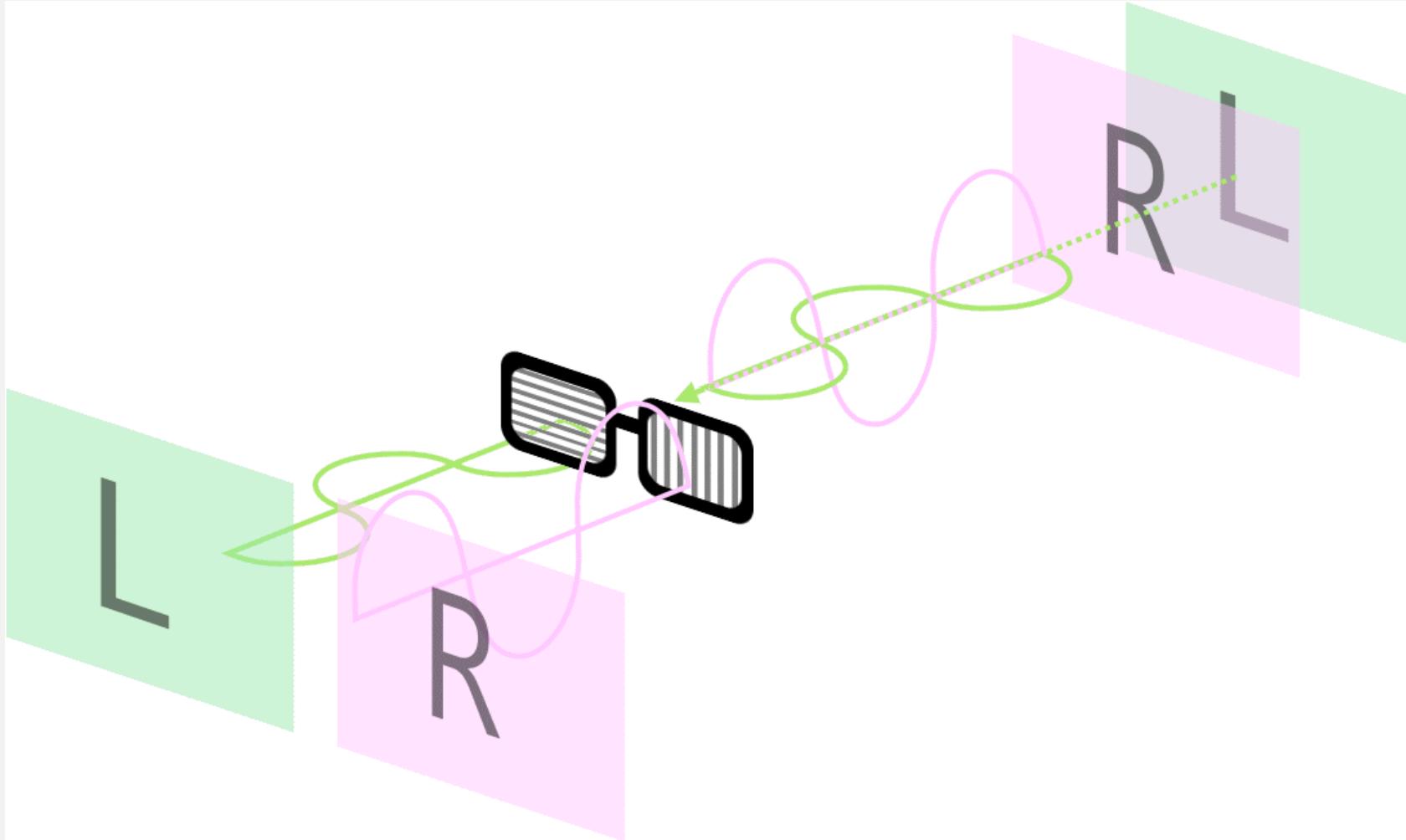
Many color formats
Supported by YouTube and
Google StreetView



Polarization Glasses



Polarization Glasses



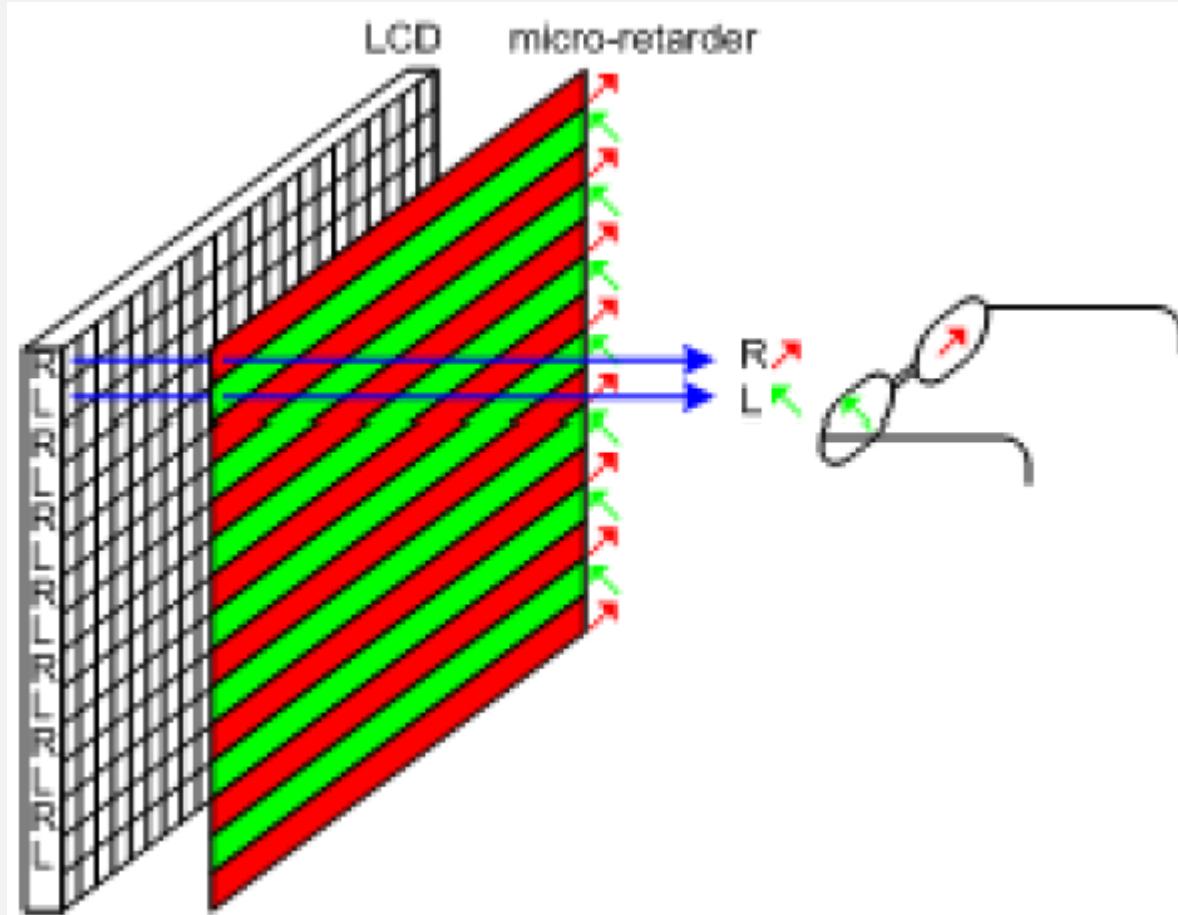
Polarization Glasses

- How to display two polarized frames?

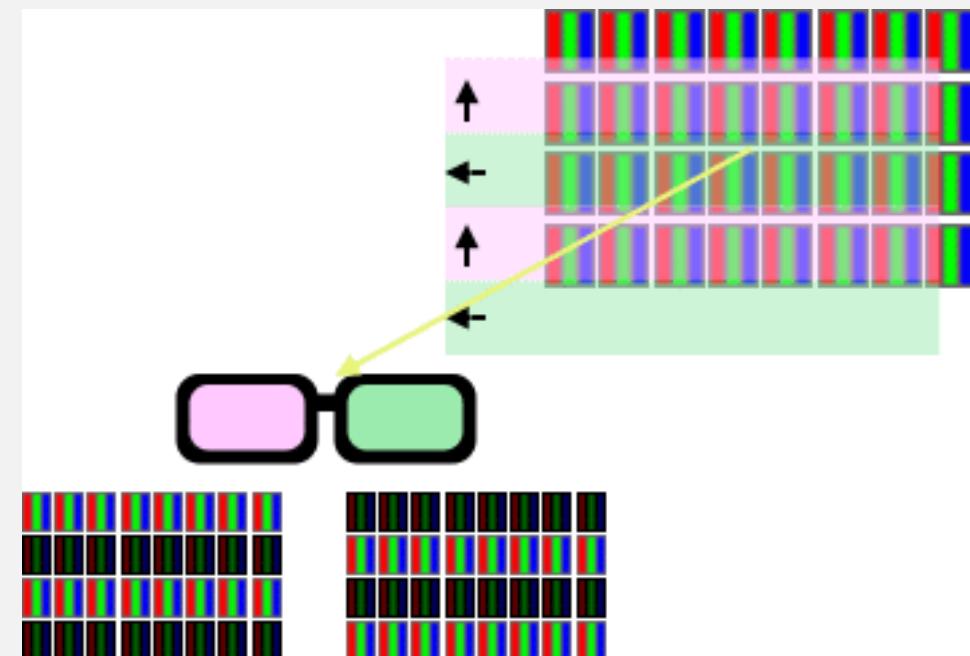


- Need accurate calibration so that the frames are aligned.
- Need non-depolarized screen.
- Cheaper glasses. Used in theater.

Polarization glasses



Pros: cheaper,
no need to sync
Cons: lower resolution

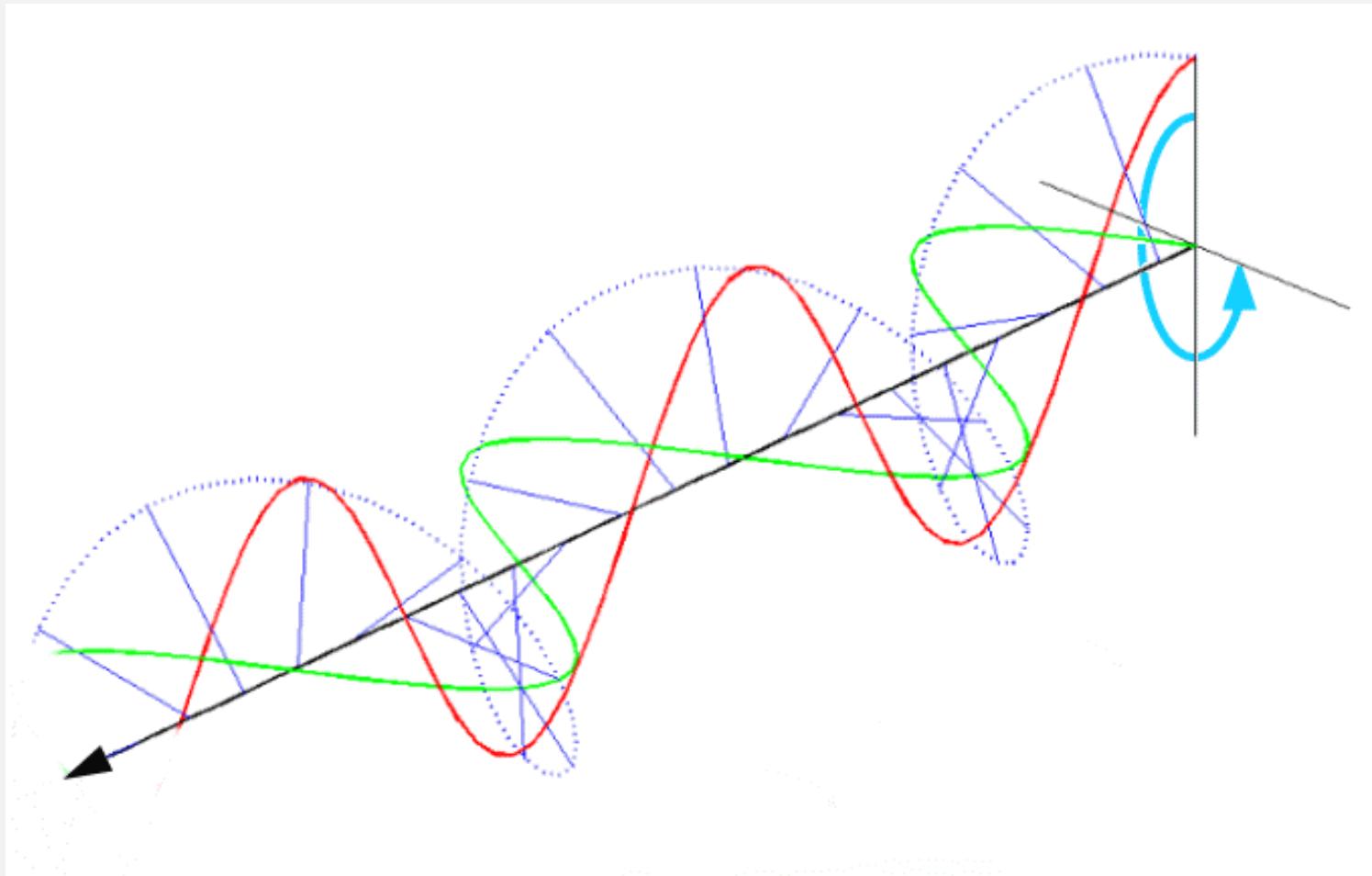


Problems with linear polarizer

- Non-aligned viewers might see cross-talk



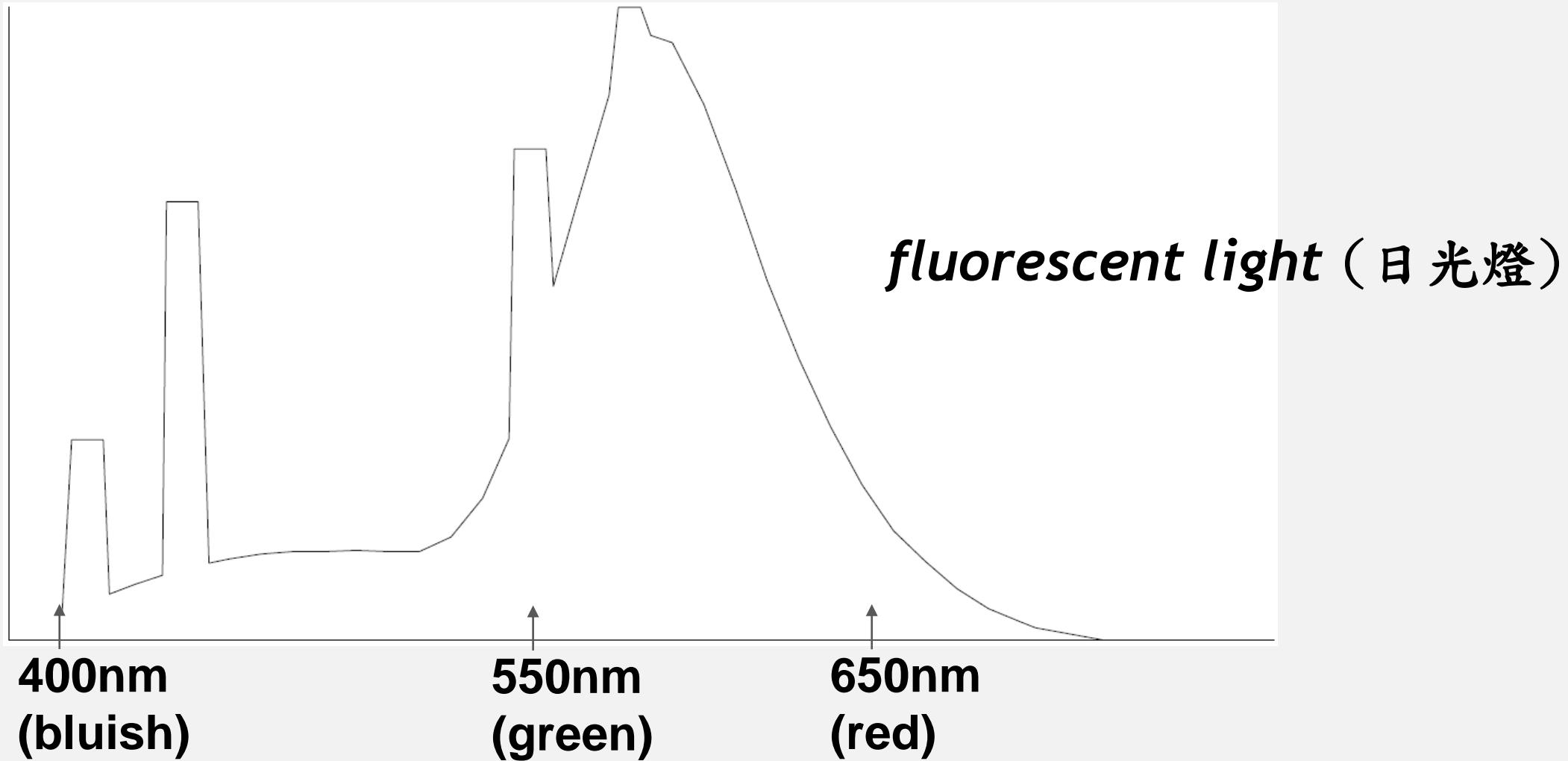
Circular Polarizer



Spectral Glasses

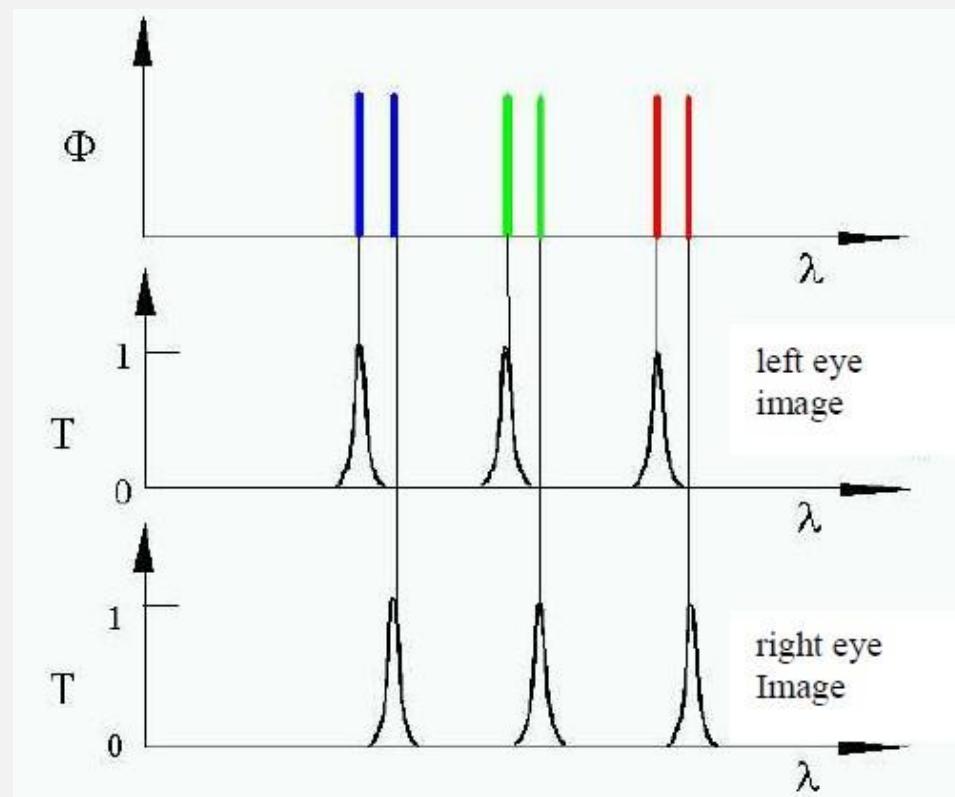
- Wavelength multiplexing.
 - [Infitec GmbH.](#)
 - Adapted by [Dolby 3D digital cinema.](#)
- Advanced anaglyph by dividing colors better.

Spectral Power Distribution



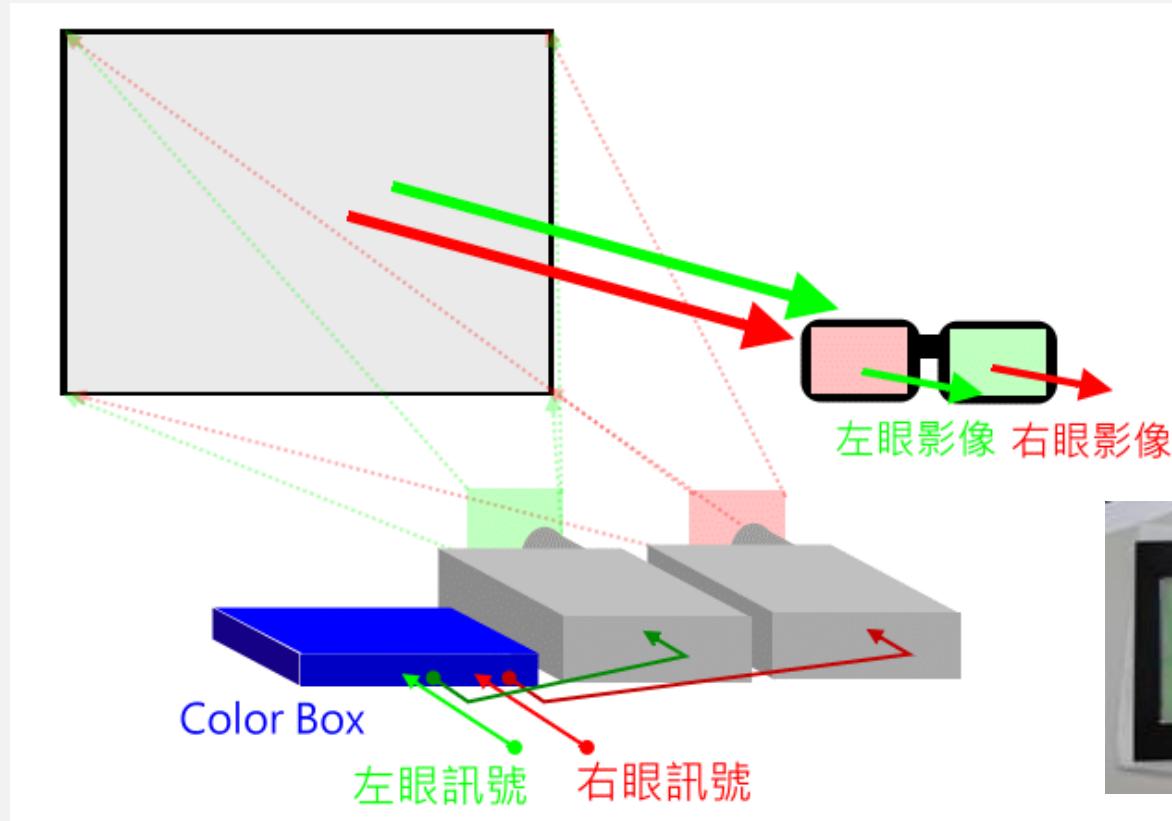
Spectral Glasses

- Although light is a complex combination of electromagnetic waves of different wavelengths, it can be represented by three primary colors due to human perception.

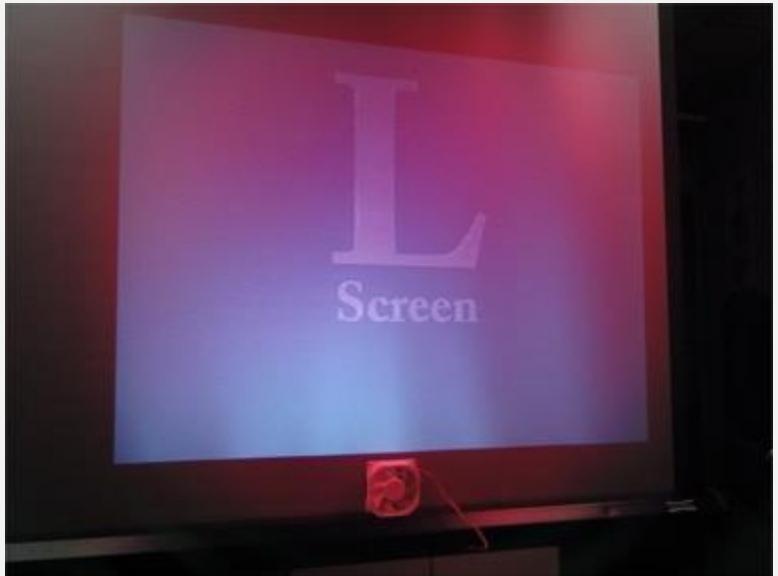


Spectral Glasses

- Filters are added into projectors and glasses so that only lights of specific wavelengths can pass by. The color box adjusts colors.



Spectral Glasses



Dolby 3D Digital Cinema

- Only requires one projector.



Pros: no need for special screen
Cons: expensive filters
darker
specular

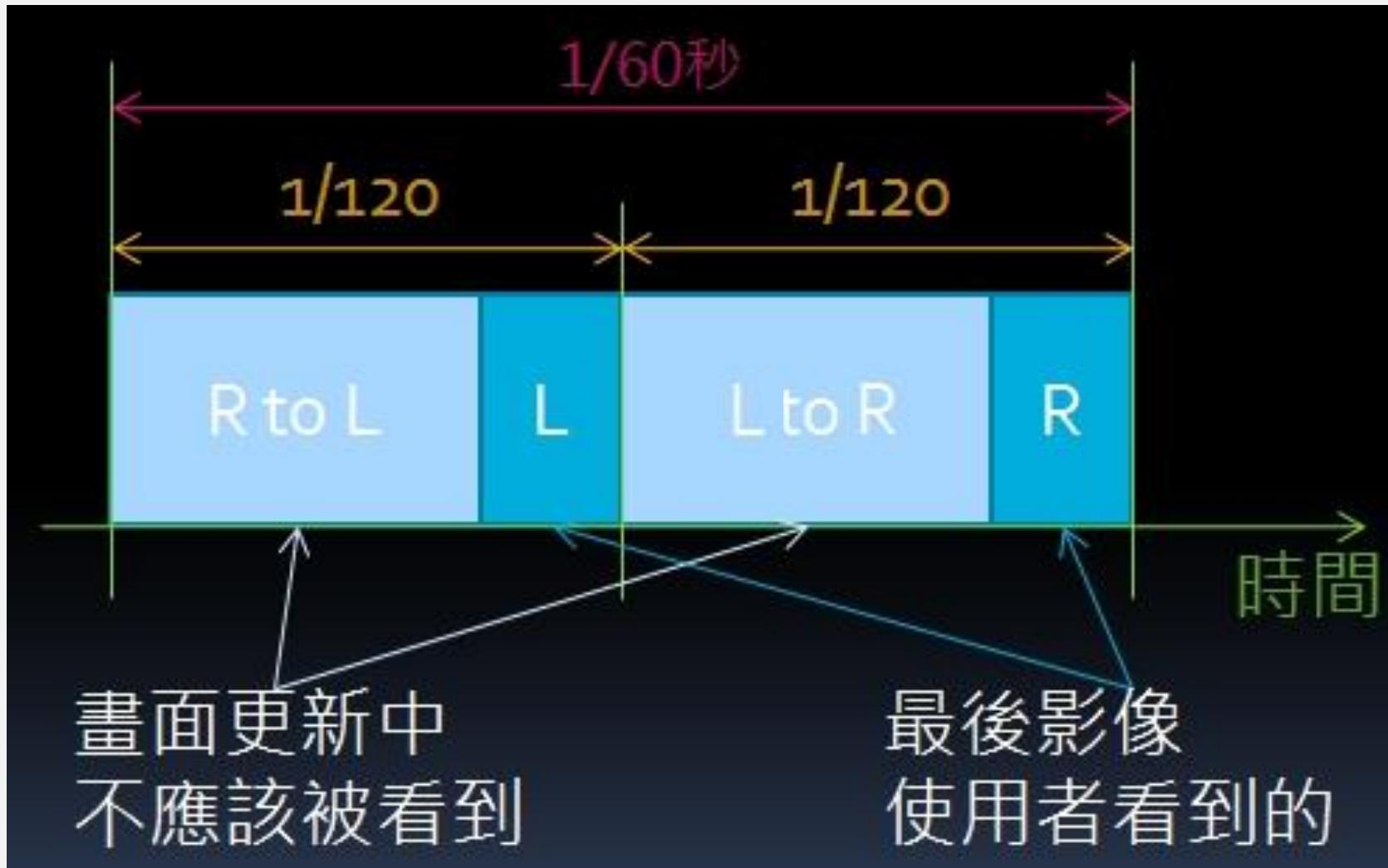
Shutter Glasses

- Twice frequency (usually 120Hz).
- Liquid crystal. Needs to sync.
- Persistence of vision (視覺暫留)



Pros: good 3D
Cons: need to sync
darker
expensive
not good for multi-user

Shutter Glasses



Head Mounted Displays (HMD)



Pros: very good 3D
could be used with head trackers

Cons: expensive
heavy
closed
single-user

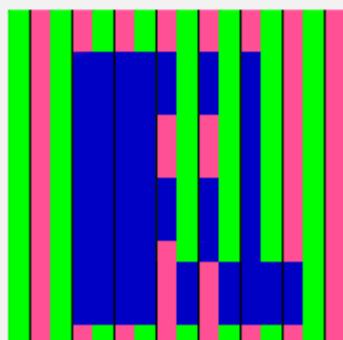
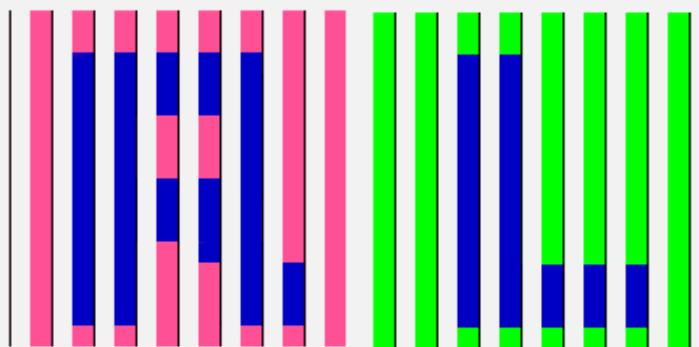
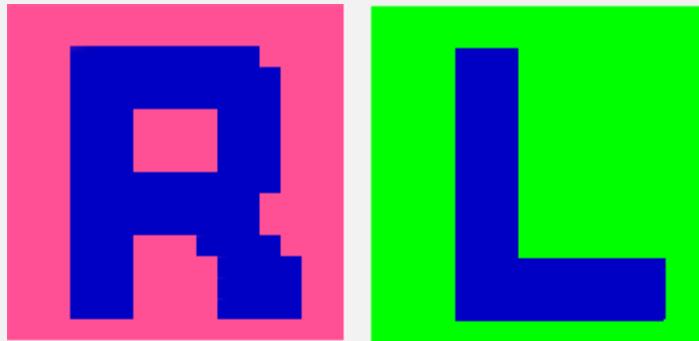
Head Mounted Displays (HMD)



Autostereoscopic

- Control lights to radiate to specific directions by accurate optics calculation.
 - Spatial-multiplexed
 - Time-multiplexed

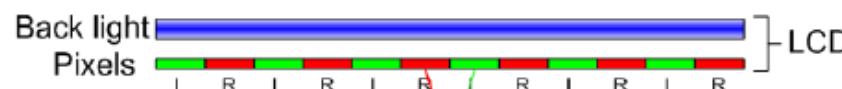
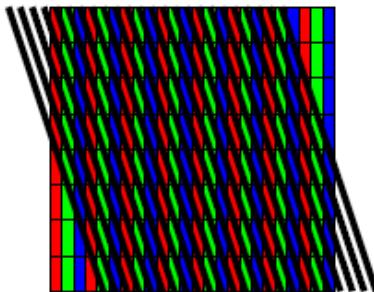
Spatial-multiplexed



- How to let left eye see only the left image and right eye see only the right one?

Autostereoscopic

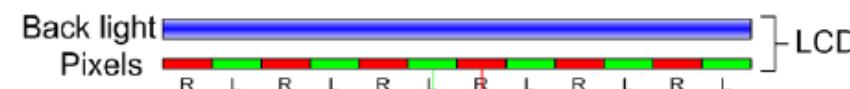
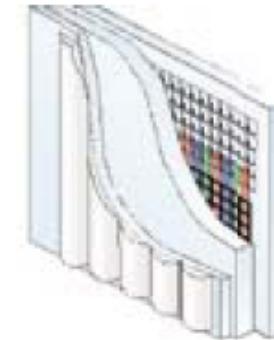
● Barriers(Slanted)



Barriers

Left eye viewing zone Right eye viewing zone

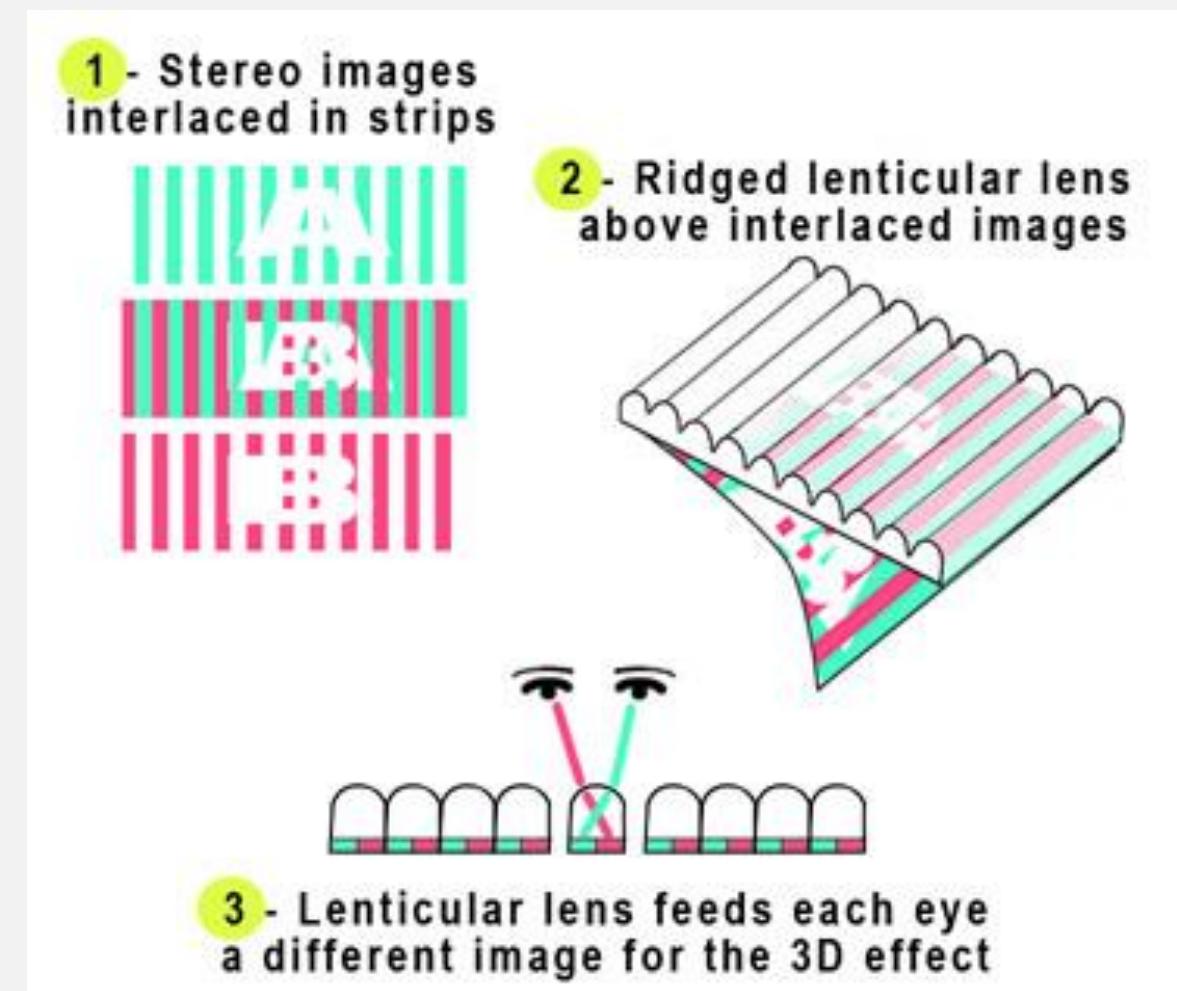
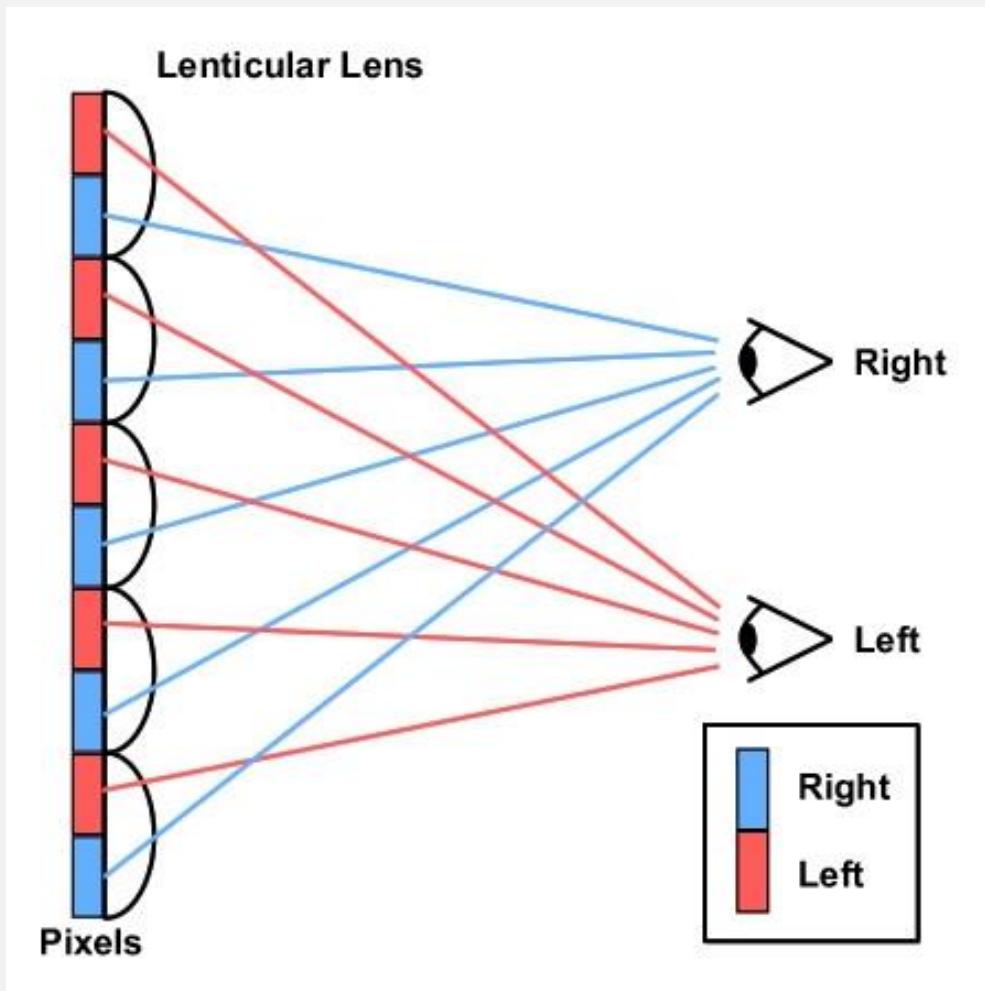
● Lenticulars



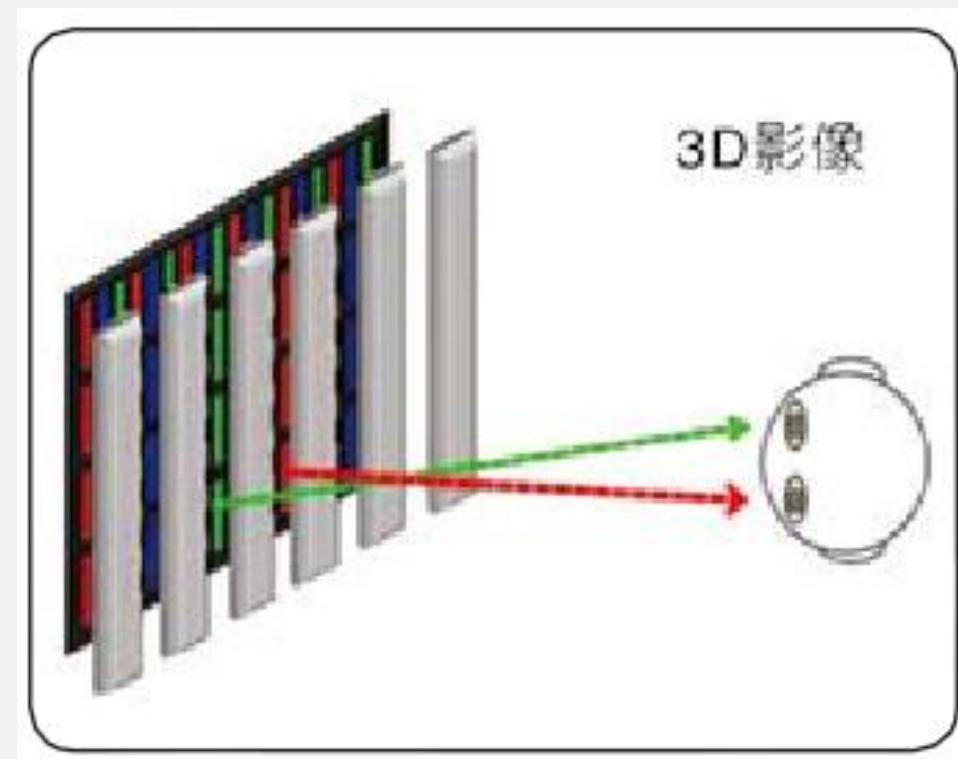
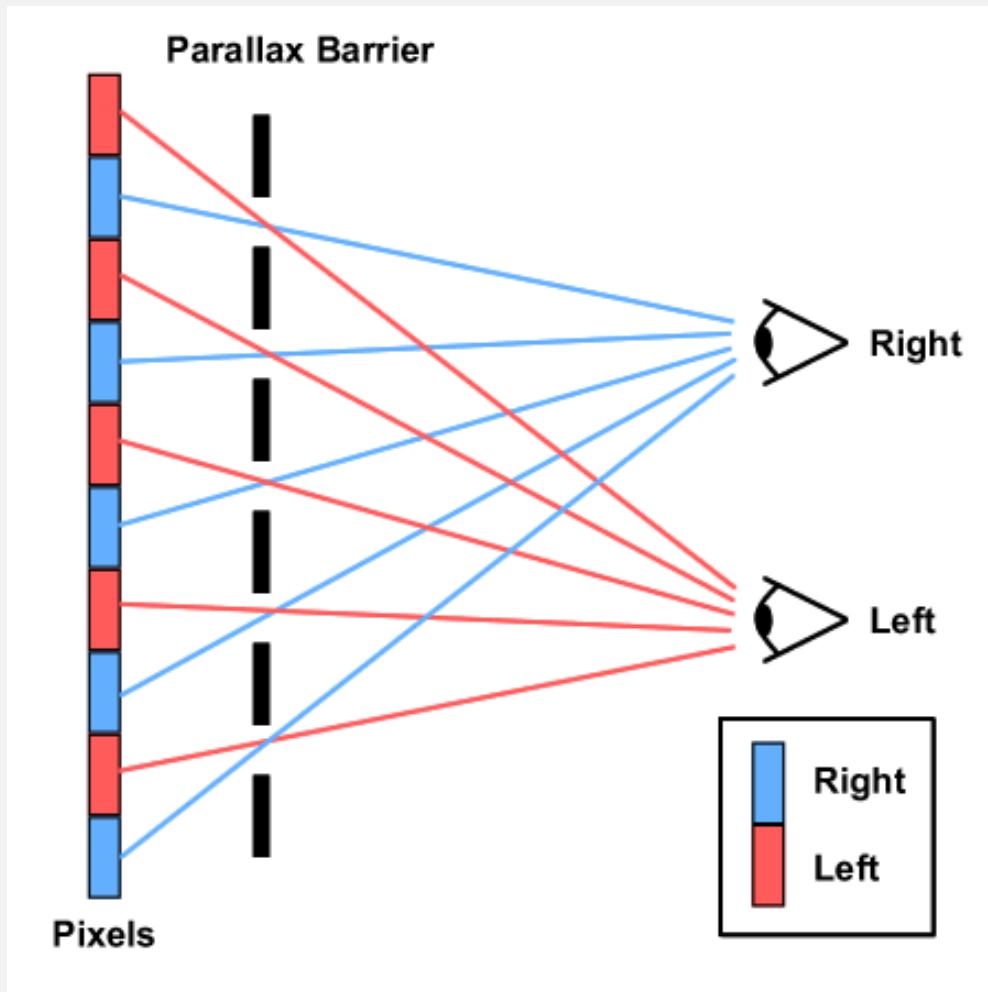
Lenticulars

Left eye viewing zone Right eye viewing zone

Lenticular Lenses

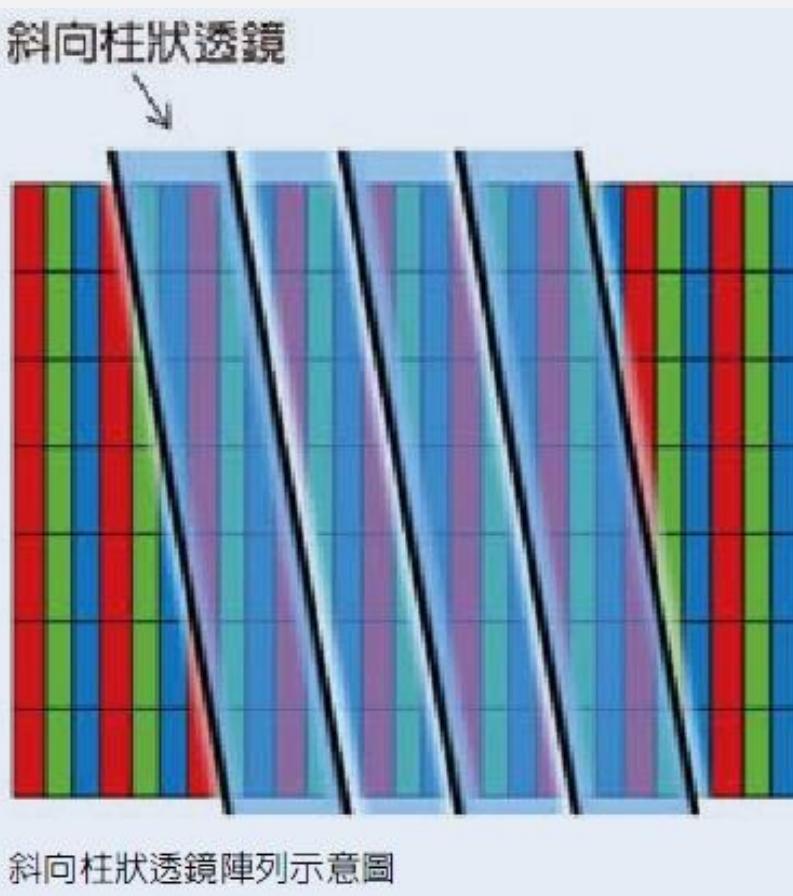


Parallax Barrier

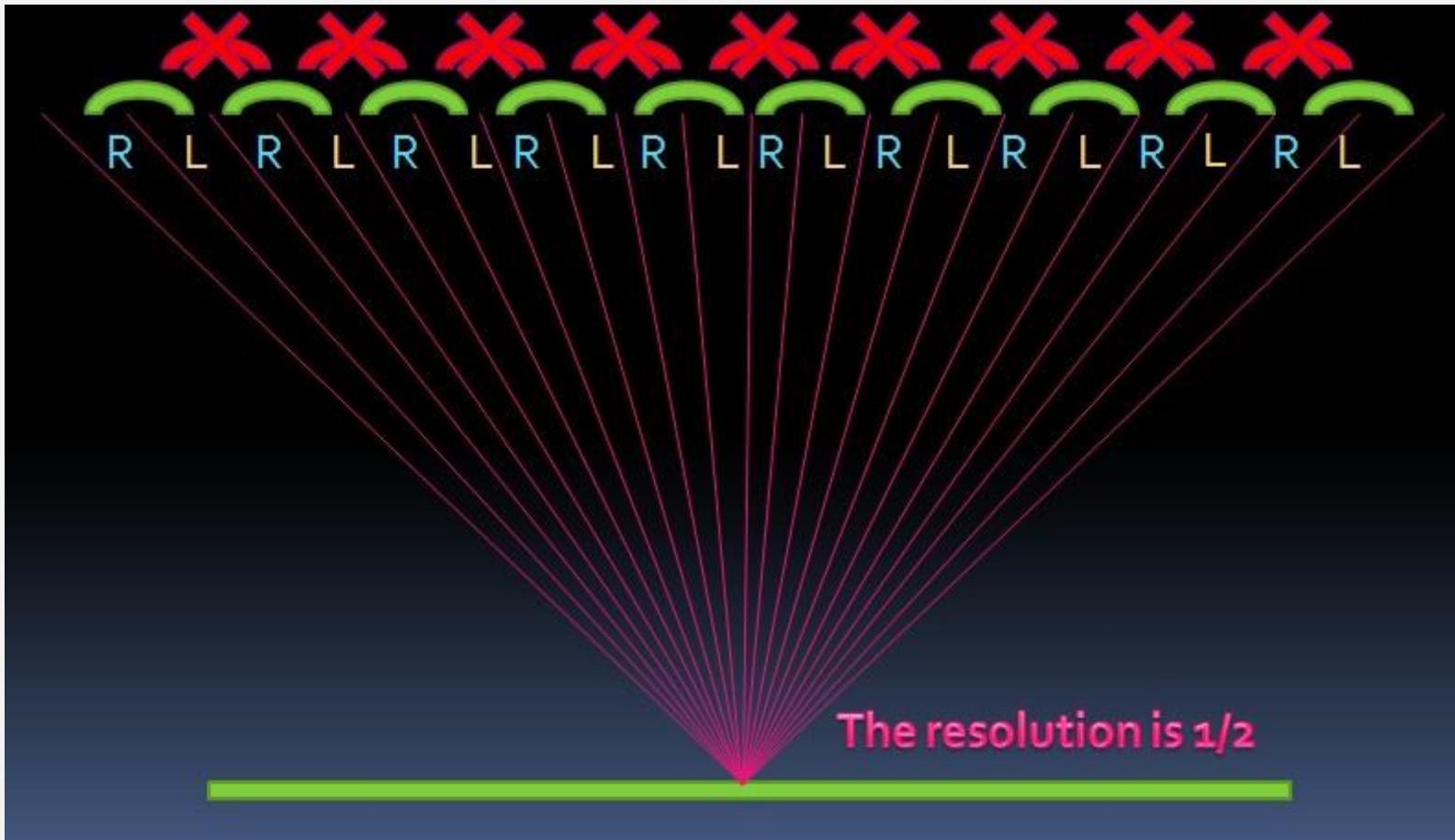


Slanted

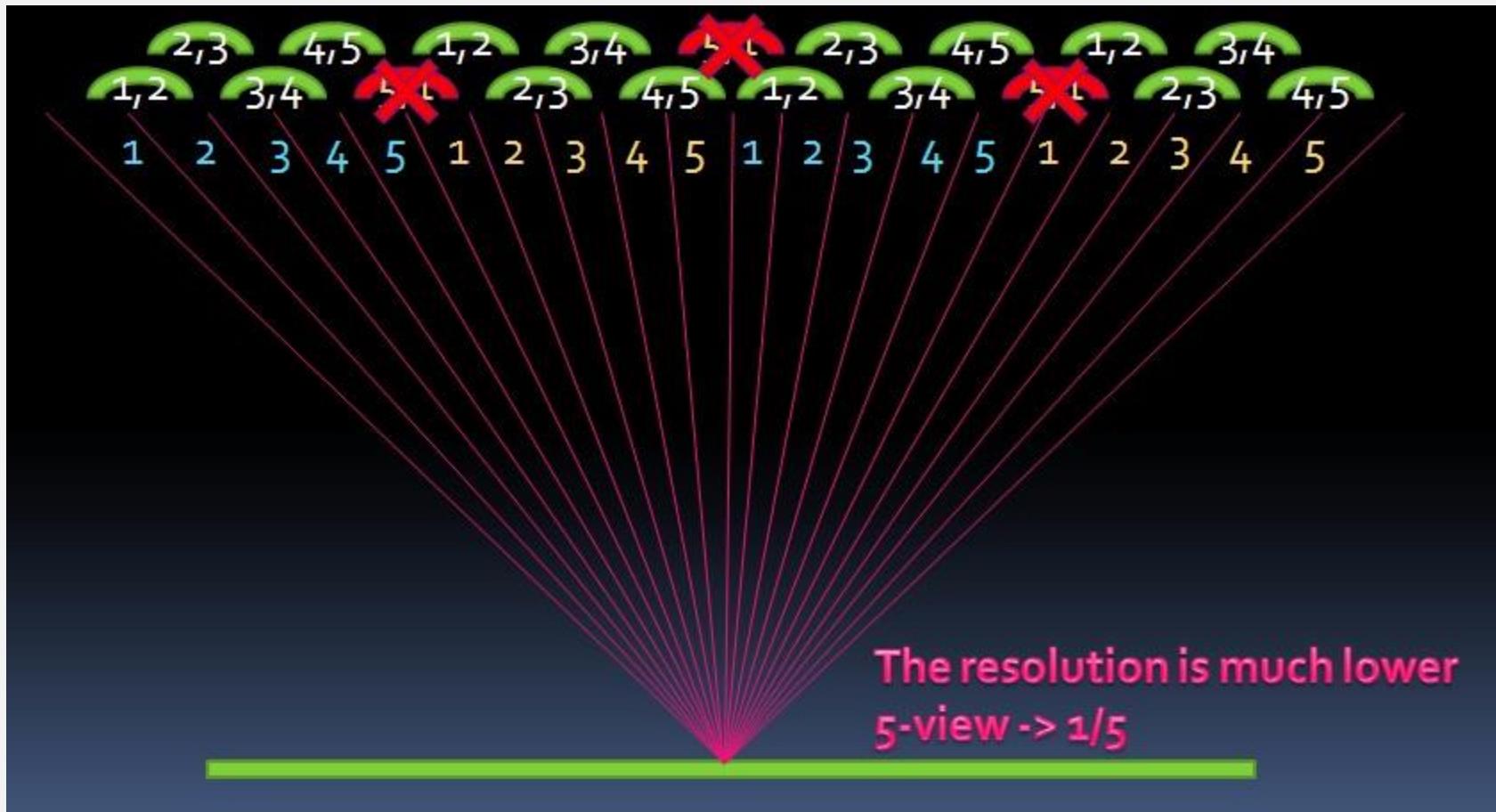
- To make more consistent horizontal and vertical aspect ratio.



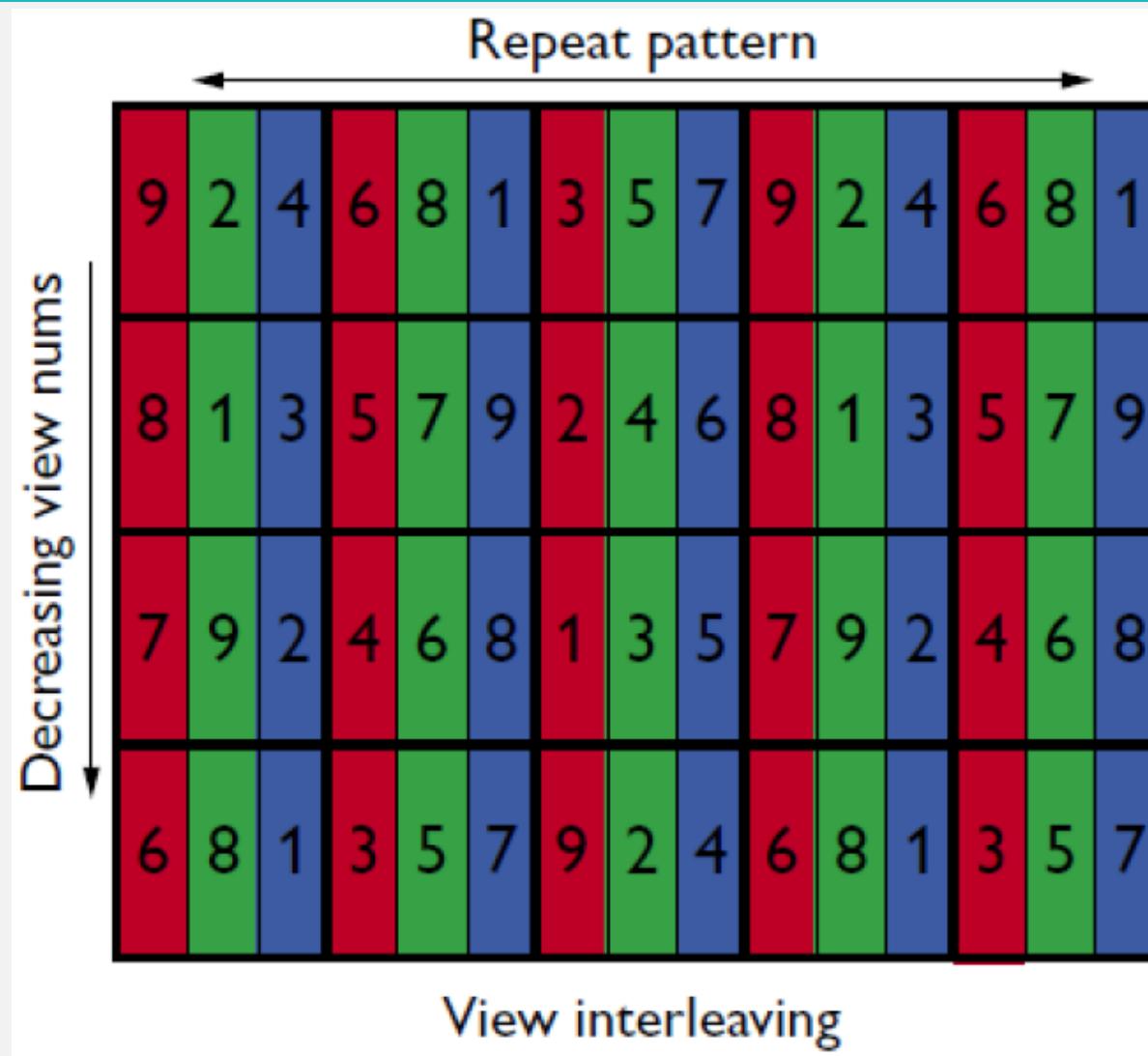
2-view



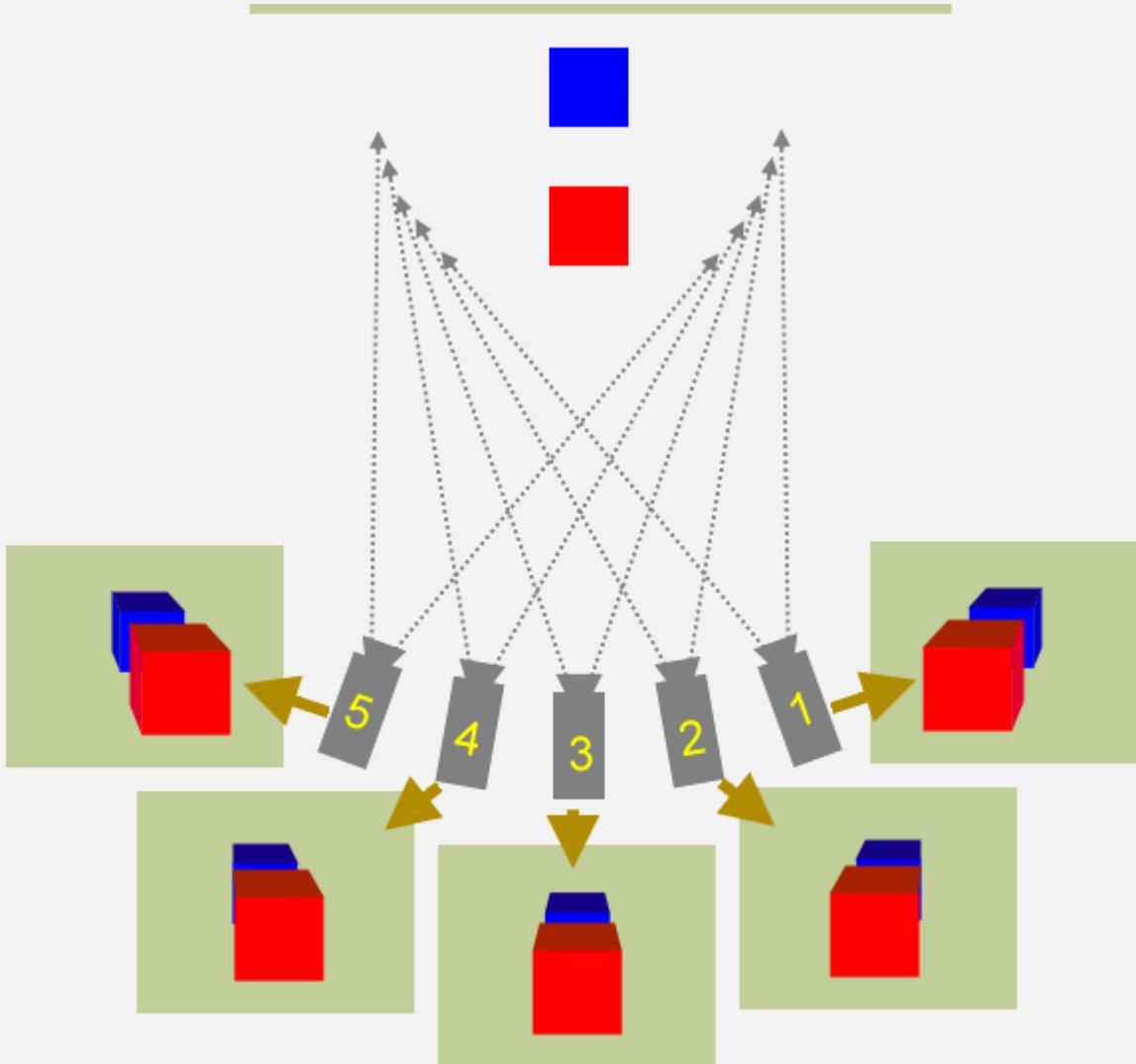
Multiple-view



Philips 9-view Format



Multiple-view

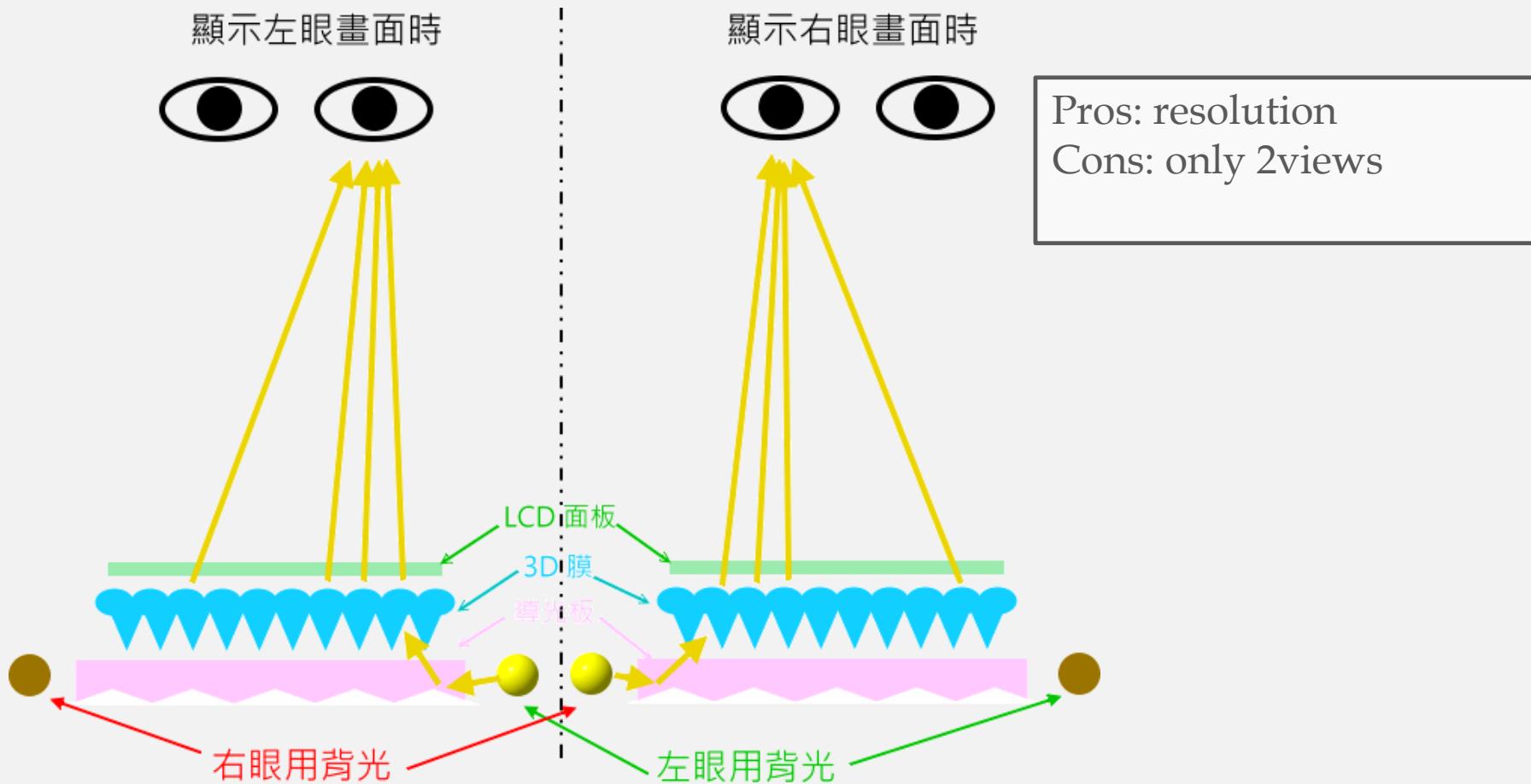


- Need more inputs.
- Reduced resolution.

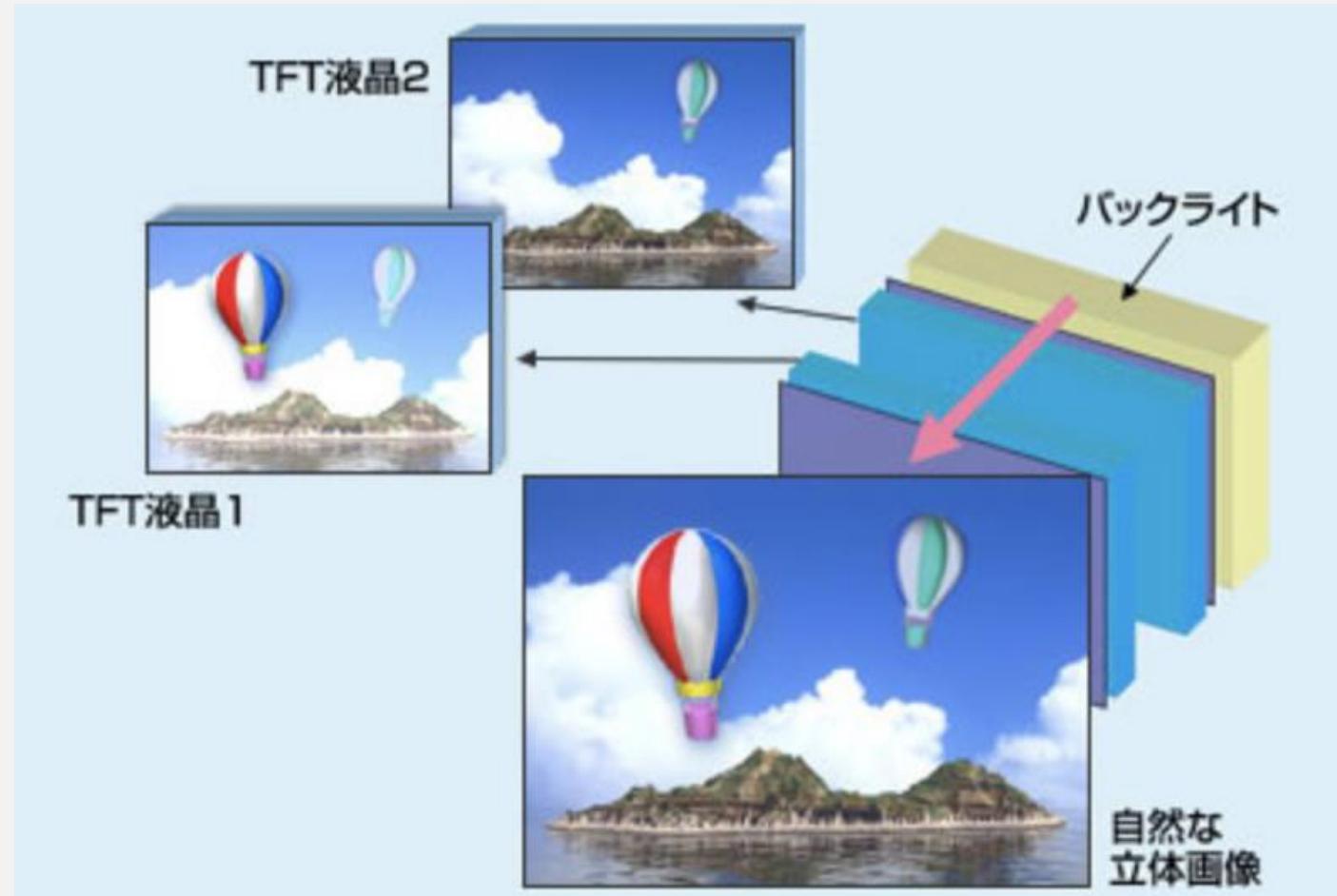
Pros: no glasses
multi-user
Cons: location
bad 3D

Time-multiplexed

■ Directional back light unit 3D film



Depth Fused 3D



Comparisons



Common 3D formats

- Side-by-side
- Multi-view
- 2D+Z

2-view

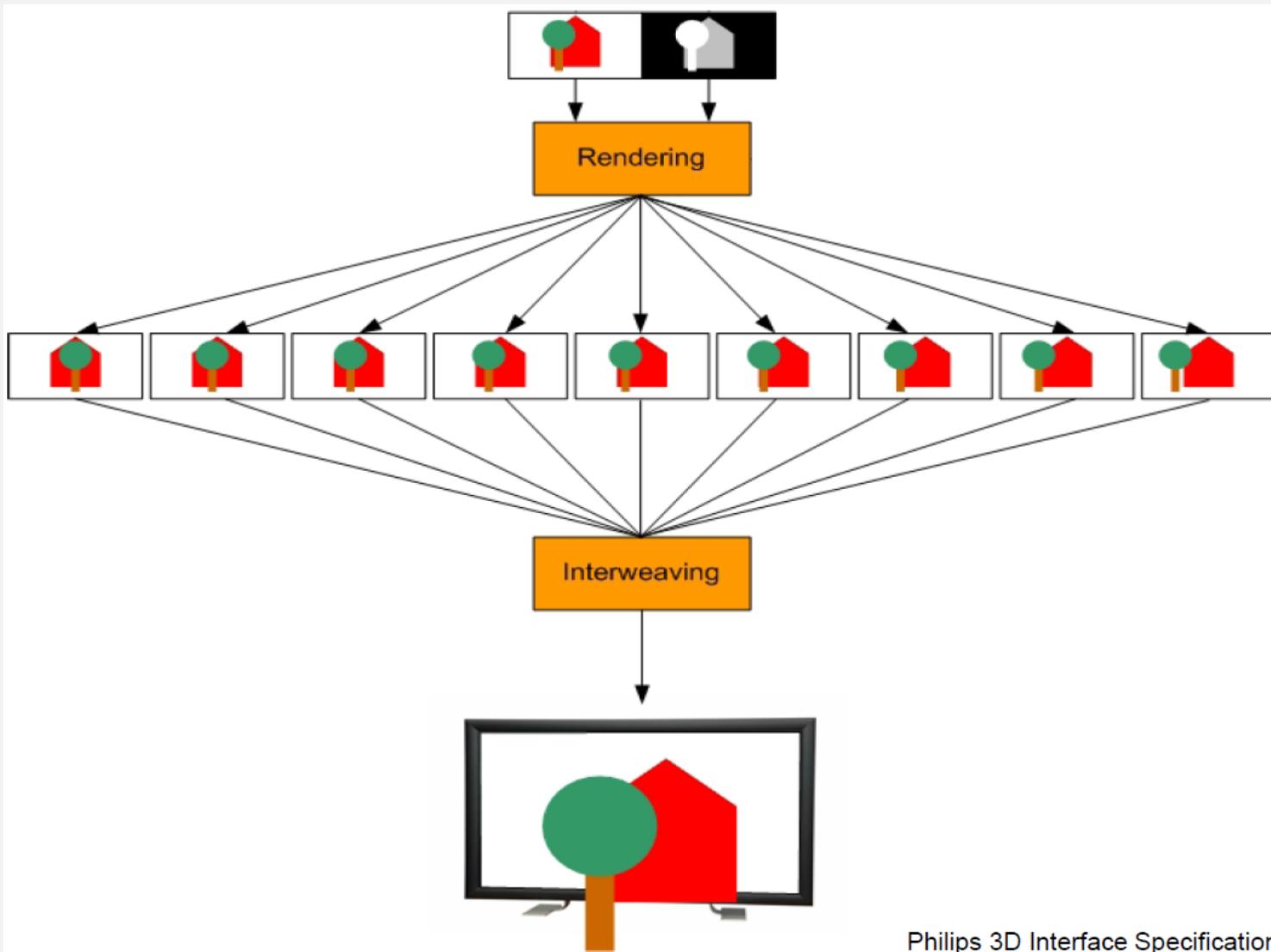


Multi-view



Newsgight 8-View

2D+Z

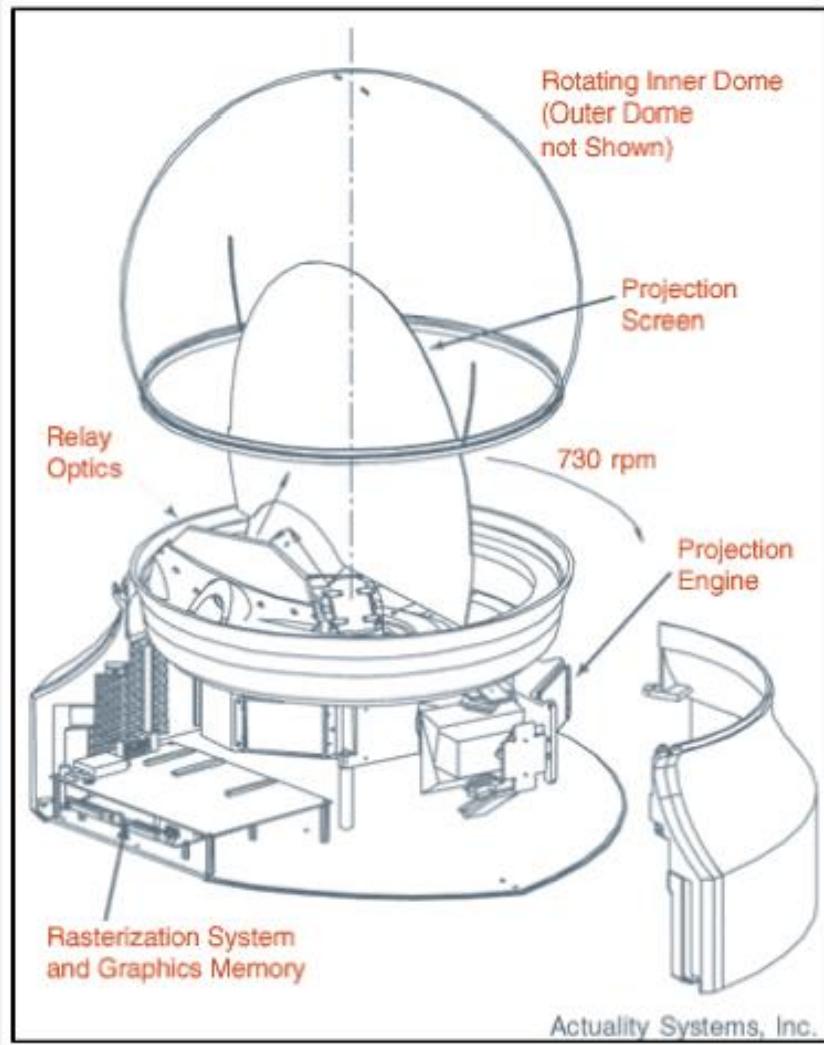


Volumetric Displays

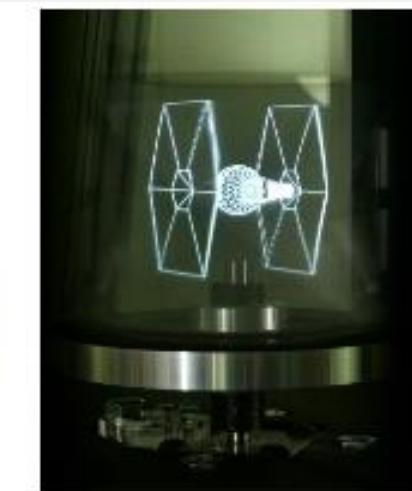
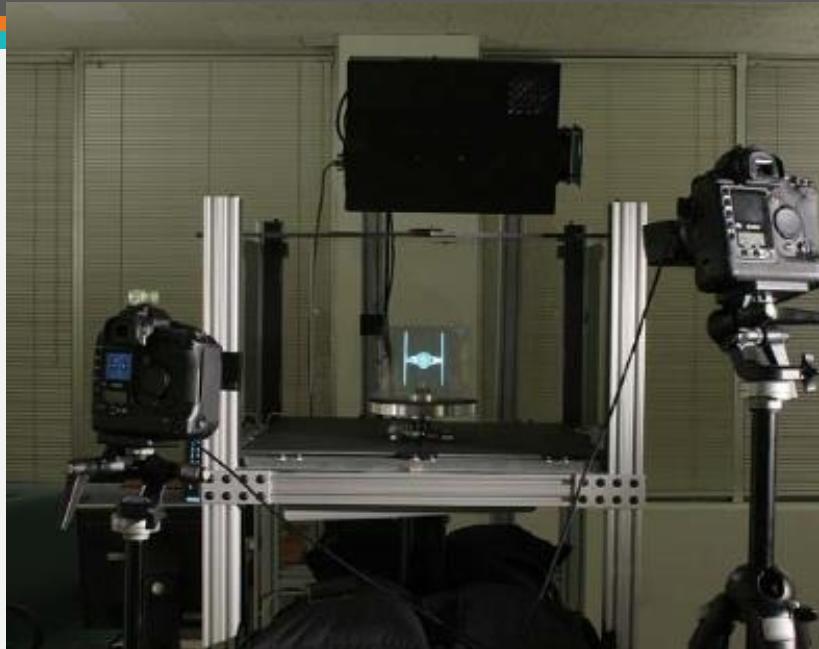
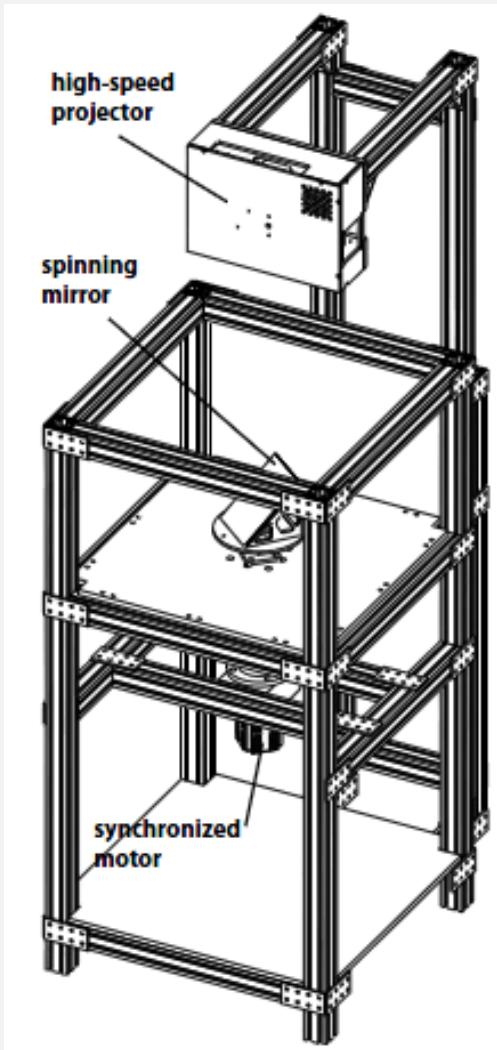
- Non-volumetric displays will make viewers fatigue after long viewing time because of inconsistency between focus and convergence.
- Volumetric displays will be better in this aspect but it is much more expensive and requires more data consumption (more views are required).
- Pros: good 3D, no glasses, multi-user
- Cons: often with limited size, suitable only for objects, not scenes



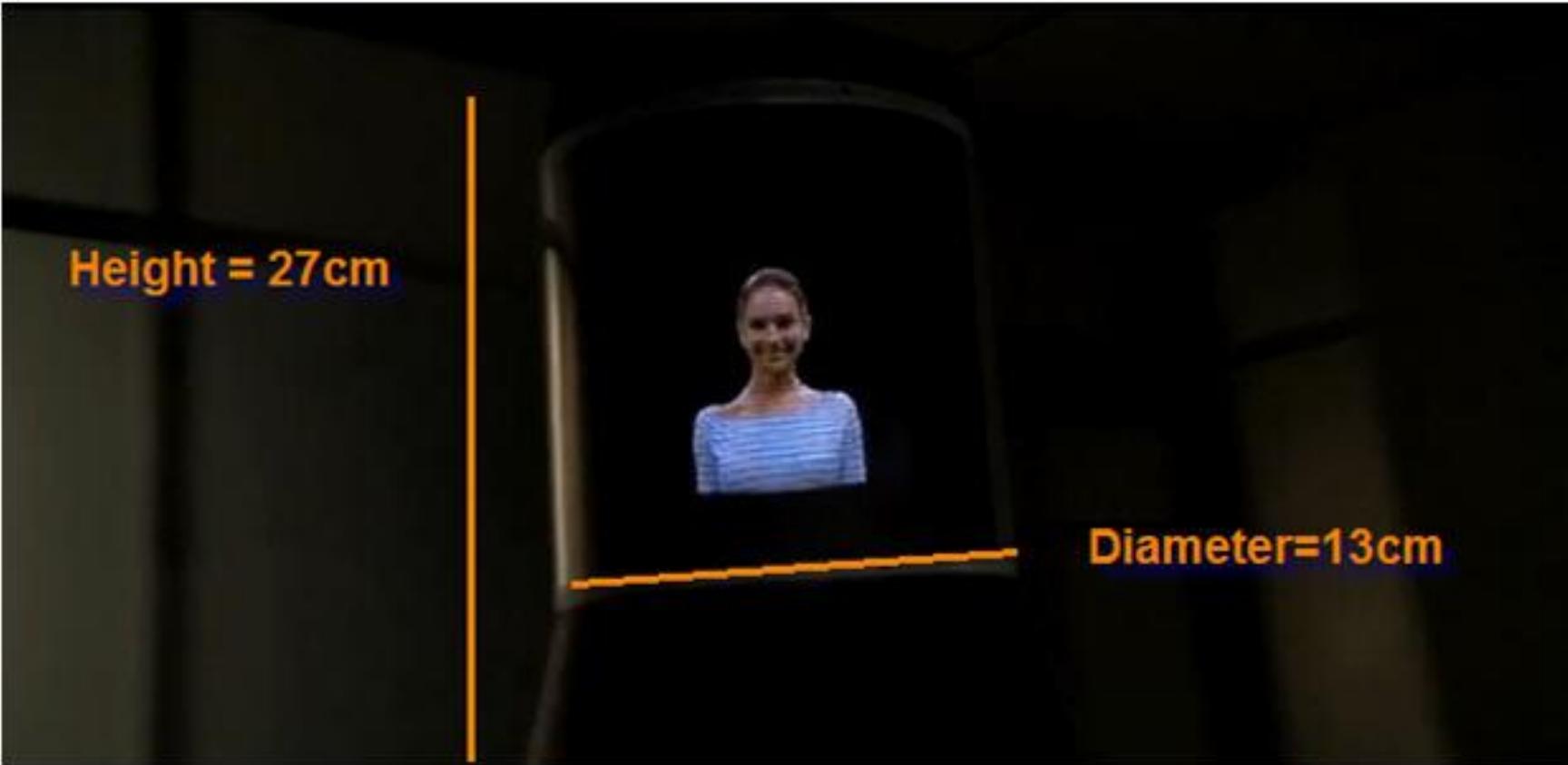
Actuality System



USC ICT



SONY RayModeler



The RayModeler is able to generate a true 3-D image with depth where a different proportionally correct view is visible from each side of the display.

Summary

- Many 3D displays will be produced in the coming years.
- Glass-equipped 3D display technology is very matured.
- Autostereoscopic displays need more time and will be used for advertisement first.
- 3D contents are the major bottleneck.
- 3D cameras are on the corner.

3D Cinematography



Capture Stereoscopic Photos

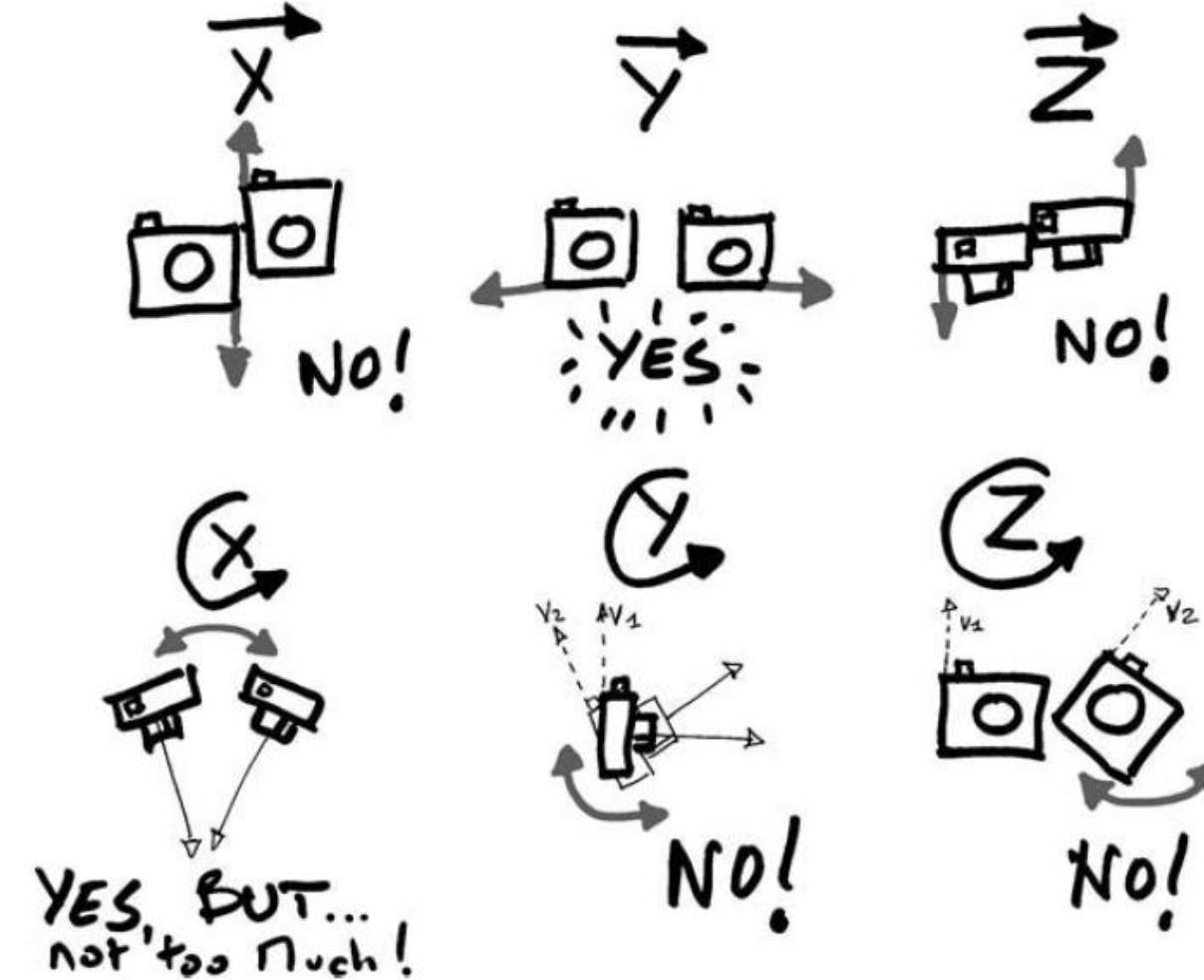
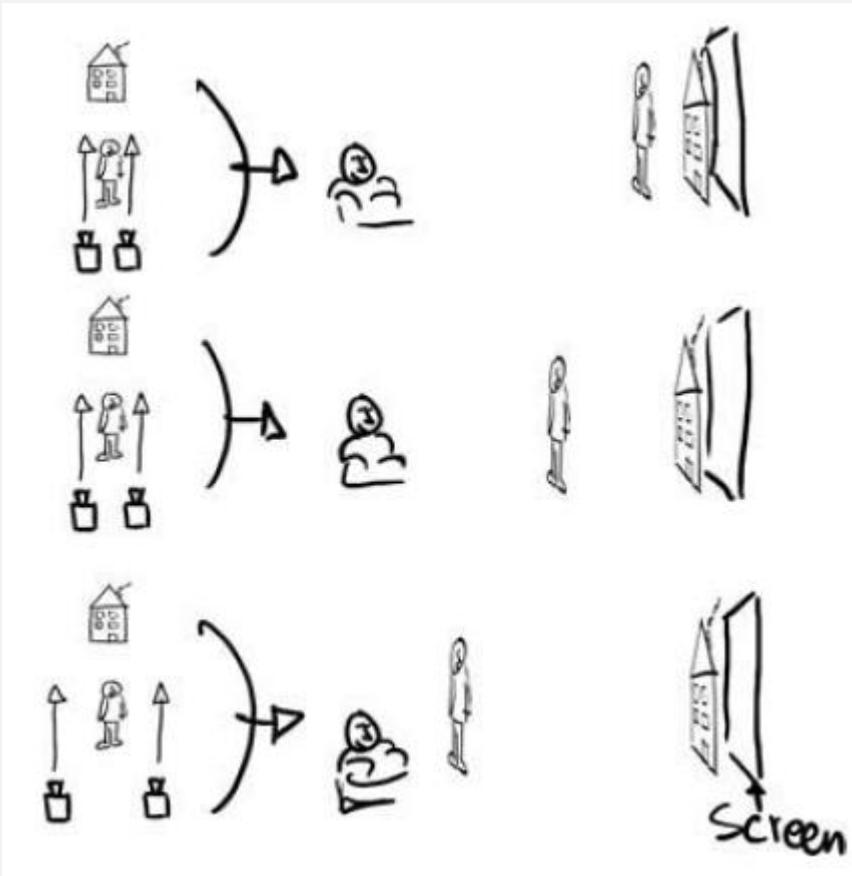


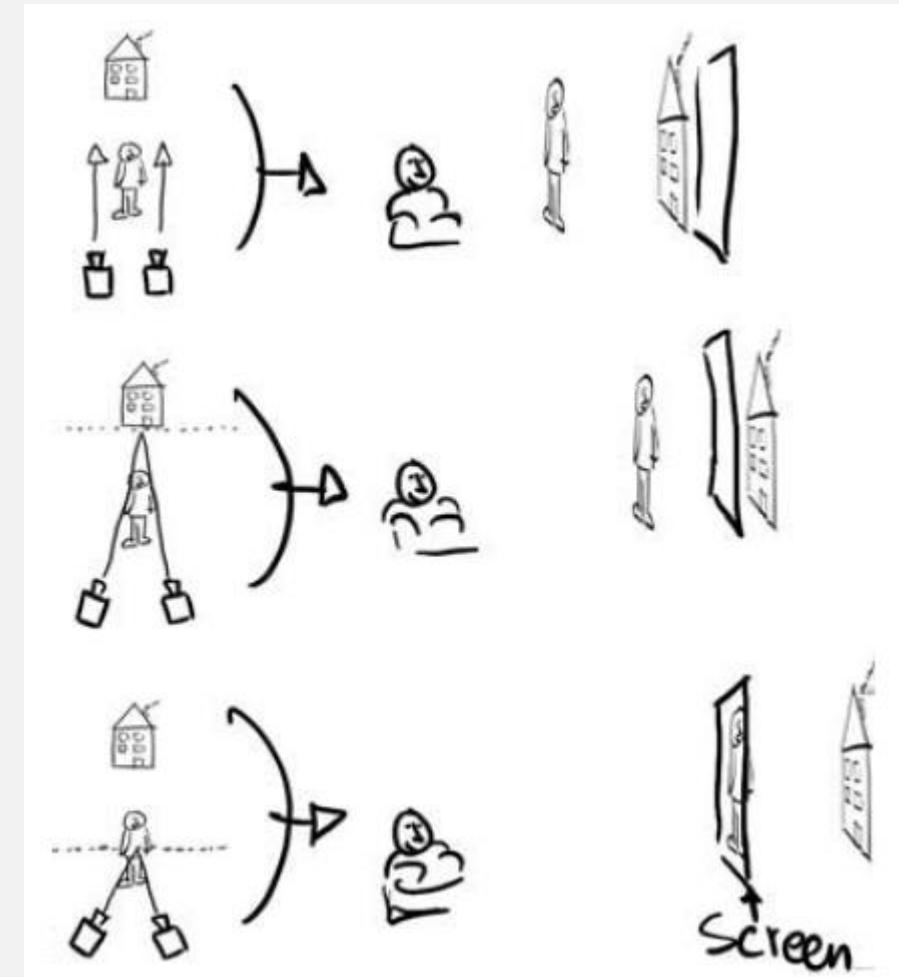
FIGURE 4.4

Of all six possible rotations and translations, only one translation and one rotation are appropriate in 3D.

Capture Stereoscopic Photos



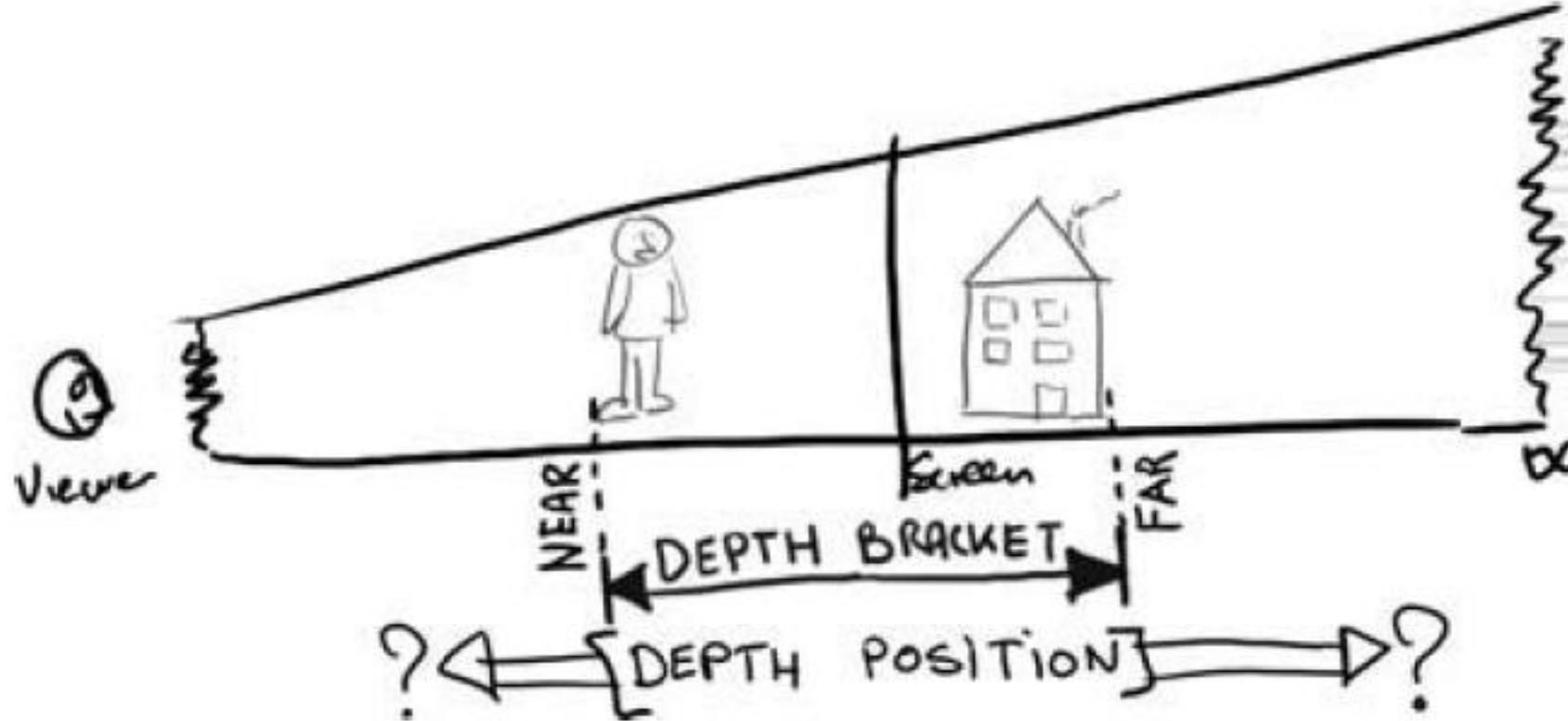
adjust baseline



adjust convergence

A Few Terms

■ Depth bracket



Sources of Visual Fatigue

- The most important point to be considered in stereoscopic cinema.
- Symptoms:
 - conscious (headache, tiredness, soreness of the eyes) or
 - unconscious (perturbation of oculomotor system)
- Some even reported the case of an infant whose oculomotor system was permanently disturbed by viewing a stereoscopic movie.

Sources of Visual Fatigue

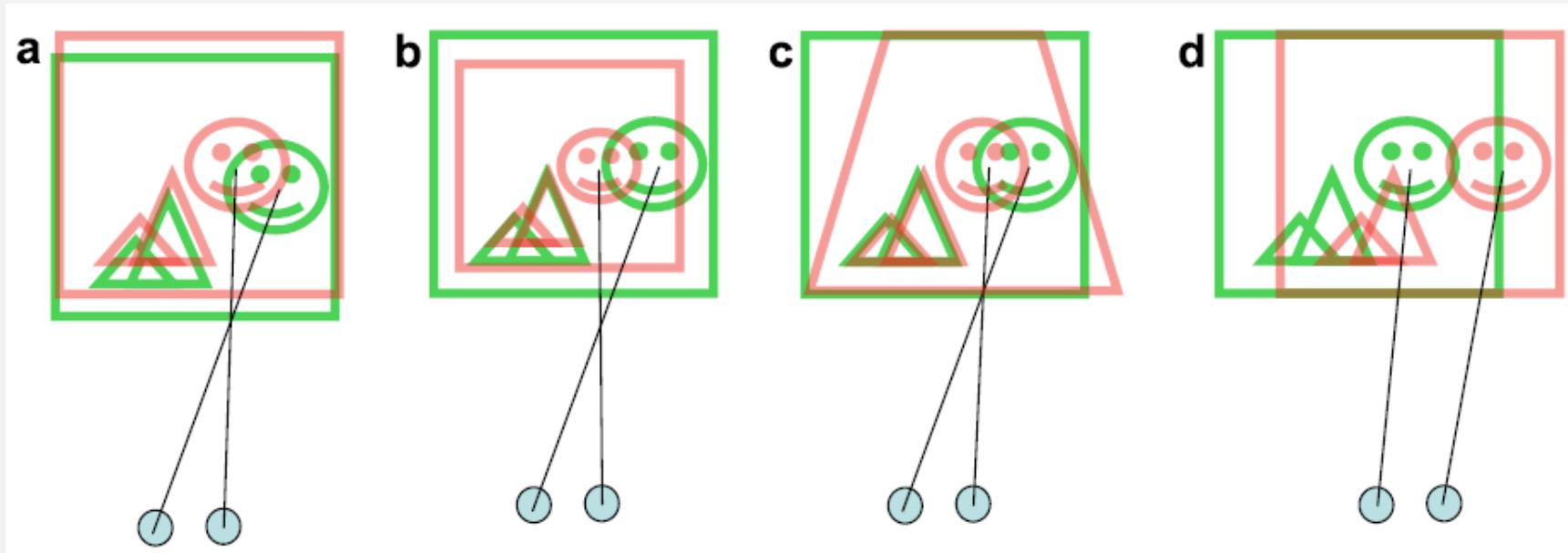
- **Binocular asymmetry:**

- photometric or geometric differences between the left and right retinal images.

- Kooi and Toet experimentally measured thresholds on various asymmetries that will lead to visual inconfort.

- Human visual system is most sensitive to vertical binocular disparities.
 - 35 arcmin horizontal disparity range is quite acceptable and 70 arcmin disparity is too much to be viewed.

Sources of Visual Fatigue



vertical
shift

size
difference

distortion
difference

horizontal
shift
(diverged)

Table 4

Threshold values for each of the binocular image manipulations

Image manipulation	Threshold value
Rotations (1°)	$>1^\circ$
<i>Other distortions</i>	
Overall magnification	2.5%
Trapezoid	$>1\text{PD}$
Meridional horizontal	$>3\%$
Meridional vertical	$<3\%$
<i>Shifts</i>	
Horizontal	Between 2 and 3PD
Vertical	$<1\text{PD}$
Stereo images	Between $2\times$ and $4\times$ hyper-stereo
<i>Crosstalk</i>	
Low disparity	$>5\%$
High disparity	5%
<i>Filter</i>	
Contrast difference	Between 25 and 50%
Luminance difference (-25%)	$>25\%$
Black and white representation	OK
Gaussian blur	<1 pixel
Local luminance	Between 4 and 8 bit digitization
Color asymmetry	Less than red/green separation

Sources of Visual Fatigue

■ Crosstalk (crossover or ghosting):

- typical values for crosstalk are 0.1-0.3% with polarization-based systems, and 4-10% with LCD shutter glasses.
- It could be reduced by a preprocess called ghost-busting.

■ Breaking the proscenium rule (breaking the stereoscopic window):

- a simple solution is to float the window

Sources of Visual Fatigue

■ Horizontal disparity limits:

- the eyes should not diverge and there is a certain limit below which human visual system can fuse

■ Vertical disparity:

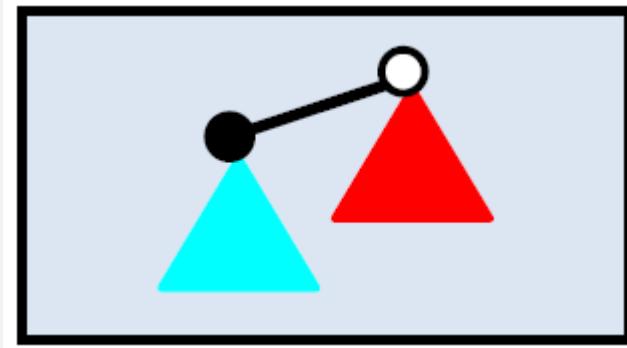
- causes torsion motion of the ocular globes and is only tolerable for short time intervals.

■ Vergence-accommodation conflicts:

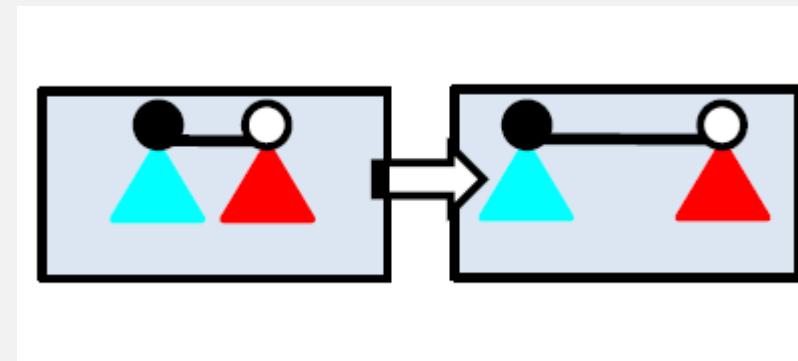
- focus distance of the eyes is not consistent with their convergence angle.
- happen very often for stereoscopic displays and could be relaxed by using the depth of field of the visual system.

Stereopsis

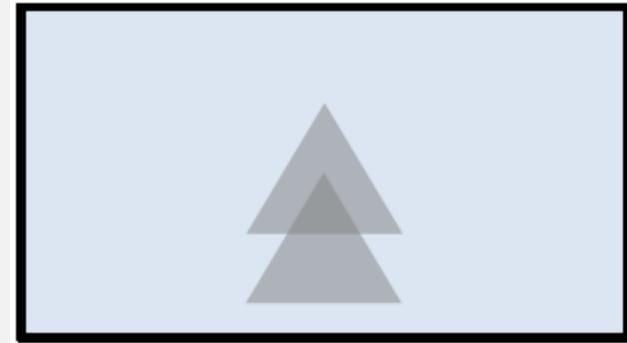
- Stereopsis could be broken for the following reasons



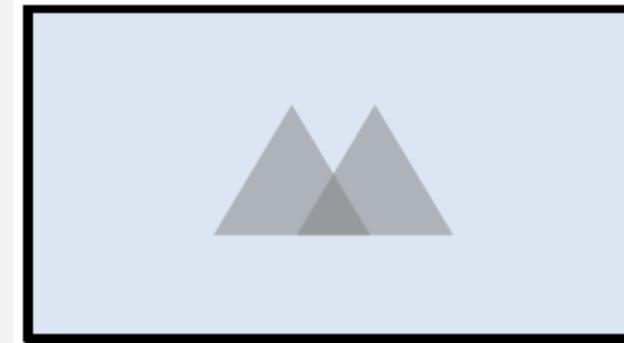
not aligned



large disparity



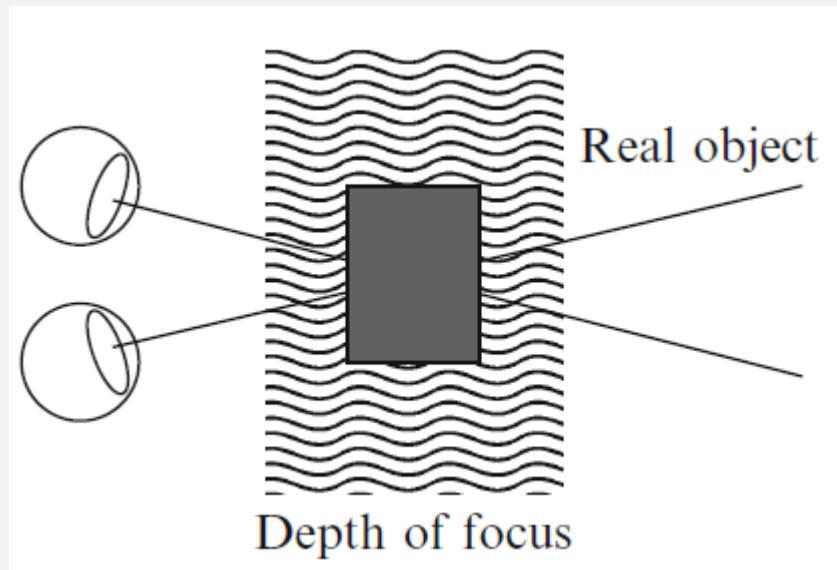
ghosting



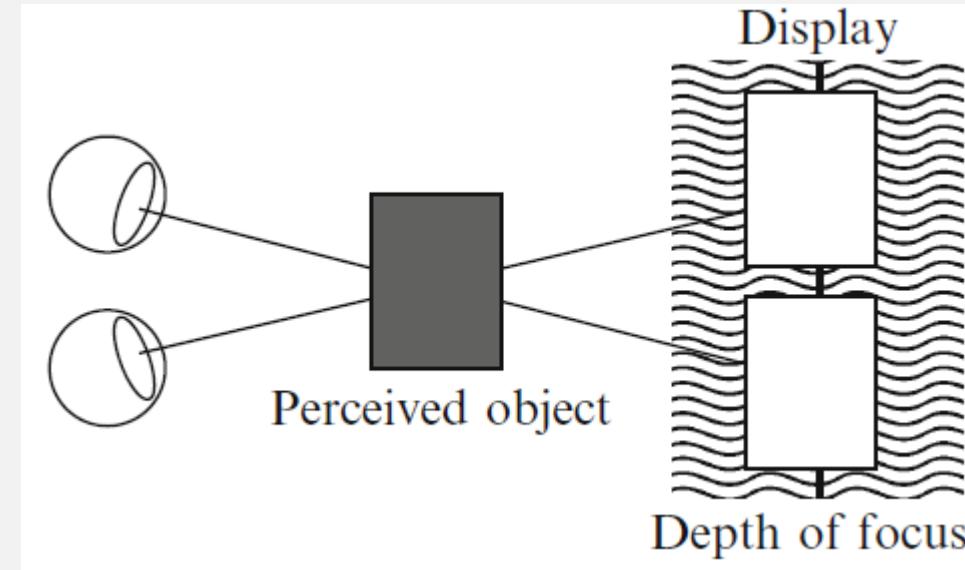
failure to fuse

Vergence-accommodation

- There is an area around it where vergence and accommodation agree, which is called **zone of comfort**.
- This discrepancy could damage the visual acuity before the age of 8.



real world

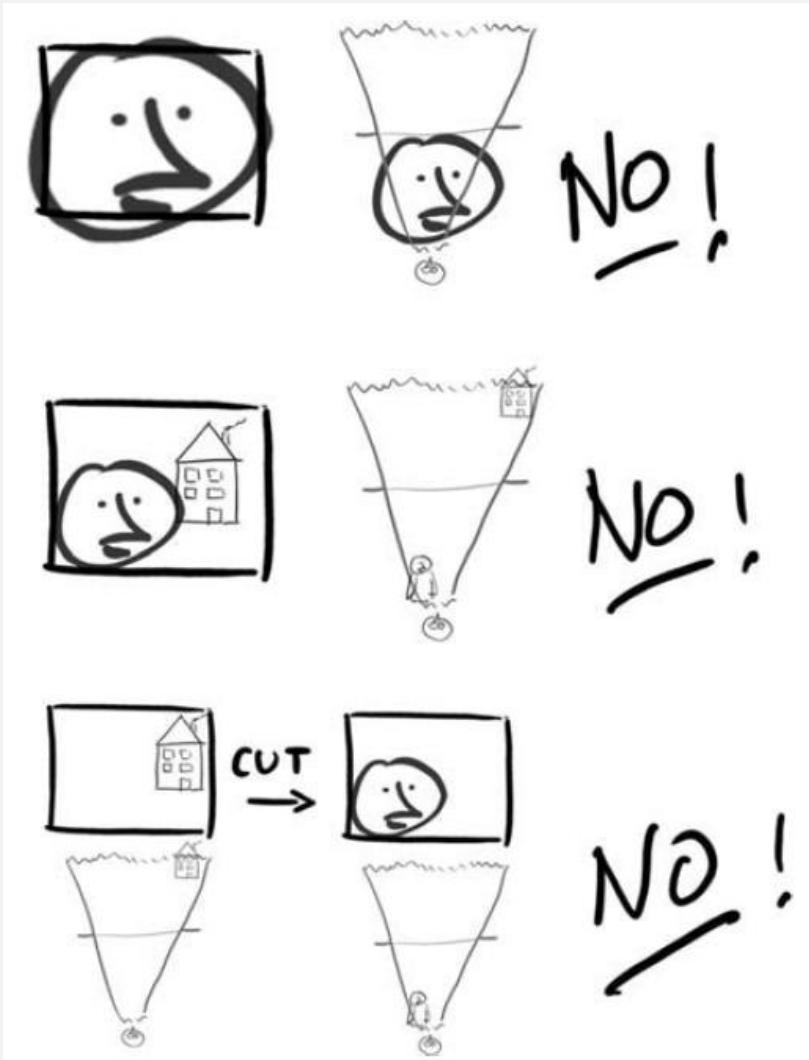


stereoscopic displays

Principles to avoid fatigue

- Maintain coordination among views
- Have a continuous depth chart
- Place rest area between strong 3D shots
- Use shallow depth of field for excessive depth range
- Respect to stereoscopic window

Some other rules



1. Objects can't cross the edges of the screen. As an example, an extreme close-up with the talent's face reaching all four sides of the frame must be set behind the screen.
2. You can't look at something reaching far inside the theater, in front of a background that is far behind the screen.
3. You can't jump cut, such as from a shot centered inside the theater, to a shot far behind the screen.

Excessive depth range

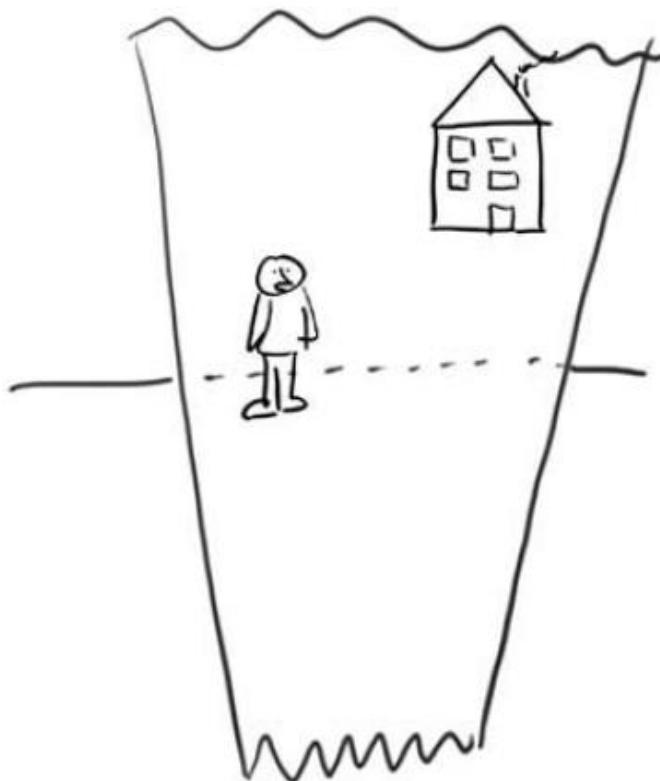


FIGURE 2.17

Background and foreground are both in the comfort zone, and close enough to each other to be fused using the same eye convergence. This is adequate for a character reading a billboard.

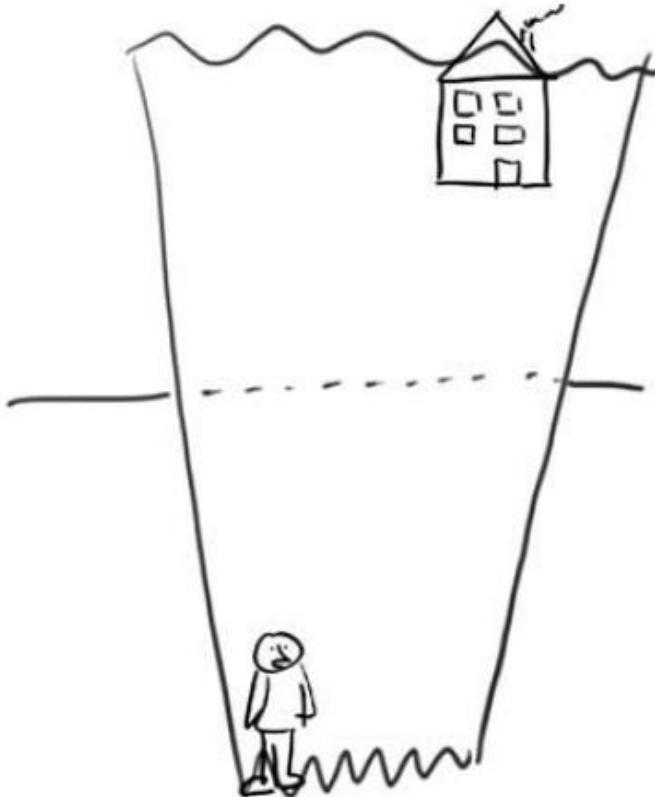
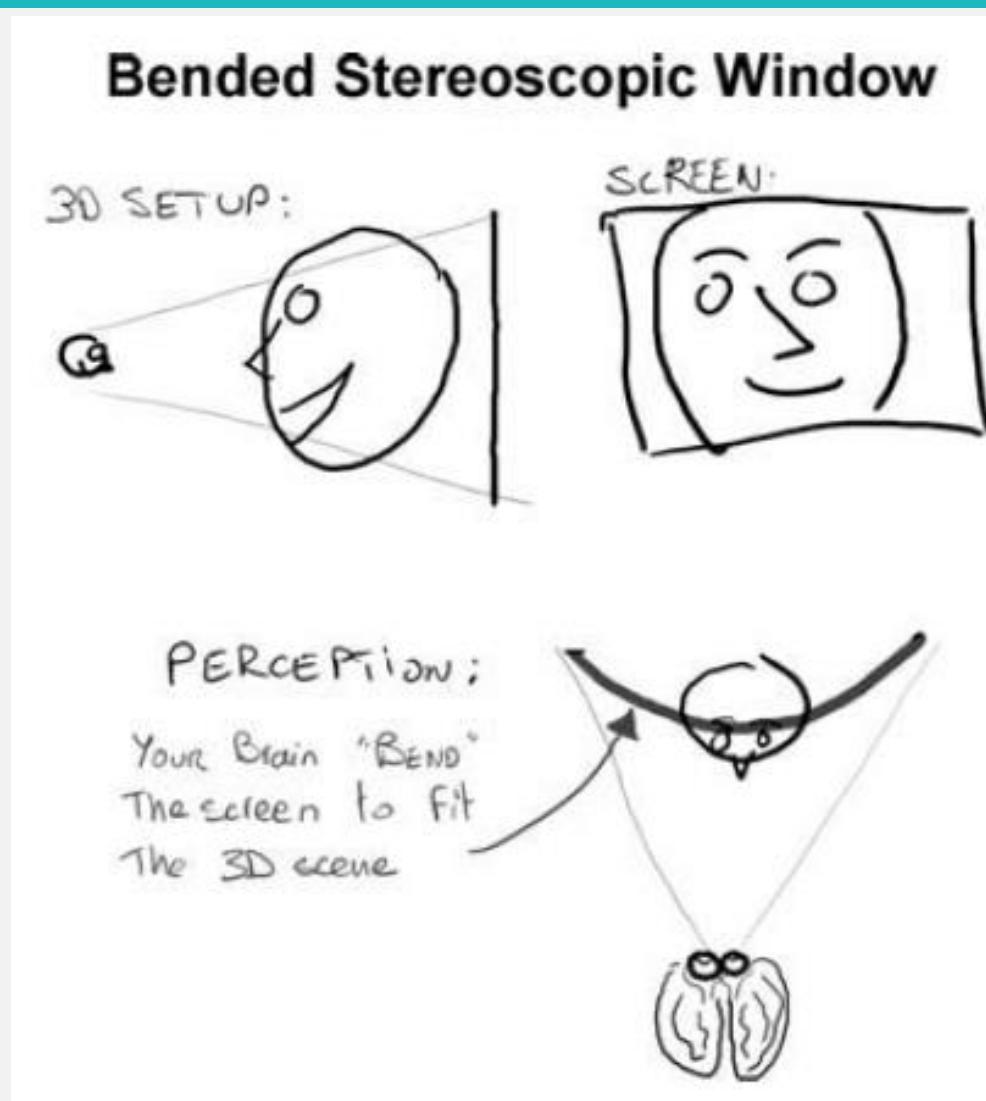


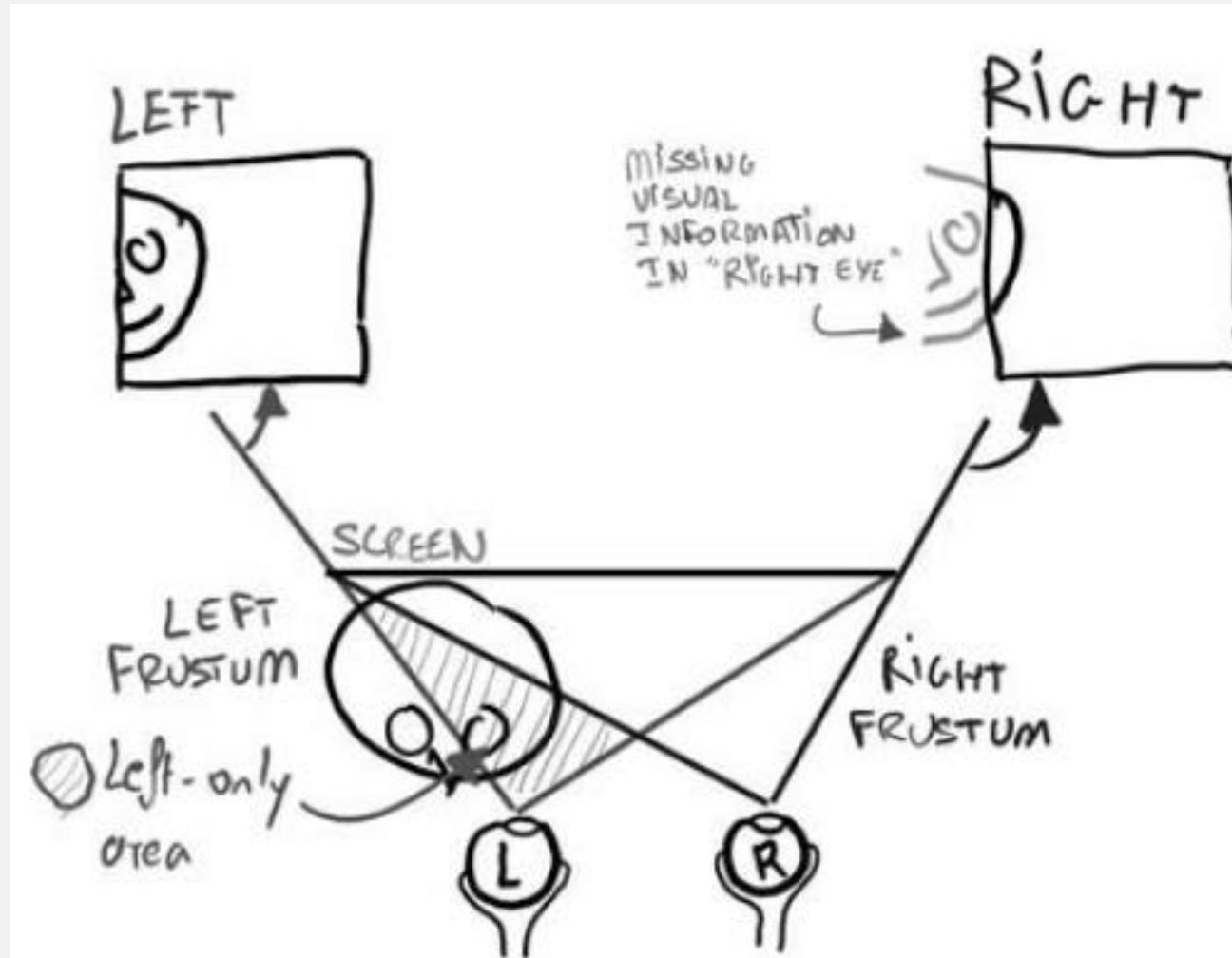
FIGURE 2.18

Background and foreground are both in the comfort zone, but too far away from each other to be fused using the same eye convergence. This is adequate for a character standing in front of a building.

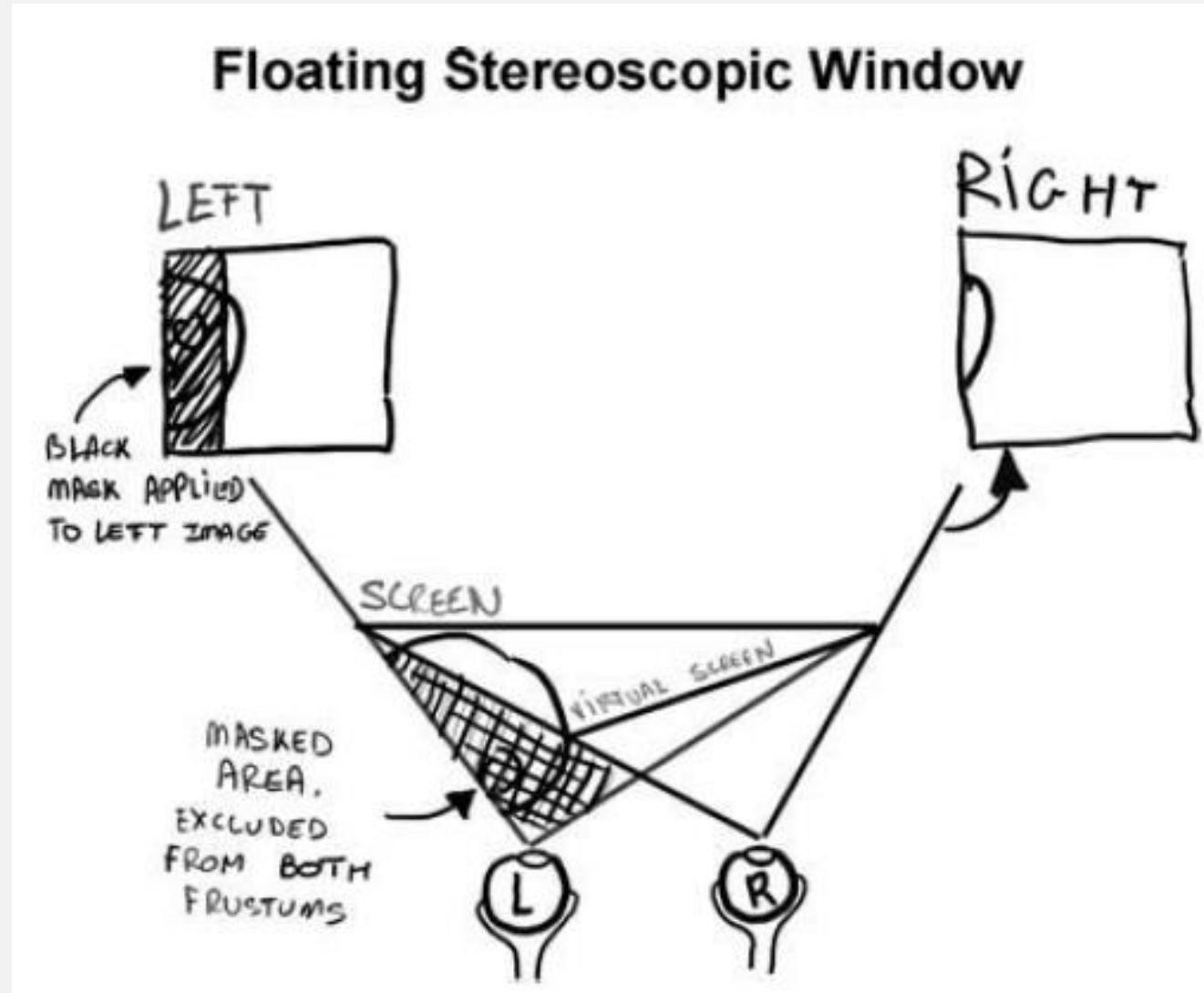
Bending the stereoscopic window



Breaking the stereoscopic window



Floating the stereoscopic window



Without floating

LEFT IMAGE



LEFT IMAGE OF FINGERS IS RIGHT MOST. RIGHT IS LEFT MOST
THIS 3D IMAGE IS IN FRONT OF THE SCREEN.

RIGHT IMAGE



3D IMAGE



FINGERS IMAGES ARE ASYMMETRICAL

With floating

LEFT IMAGE



↑ Masking on left side

RIGHT IMAGE



Masking on right side ↑

3D IMAGE



NOTE THE SYMMETRICAL
MASKING OF THE FINGERS

Stereoscopic Media Postprocessing



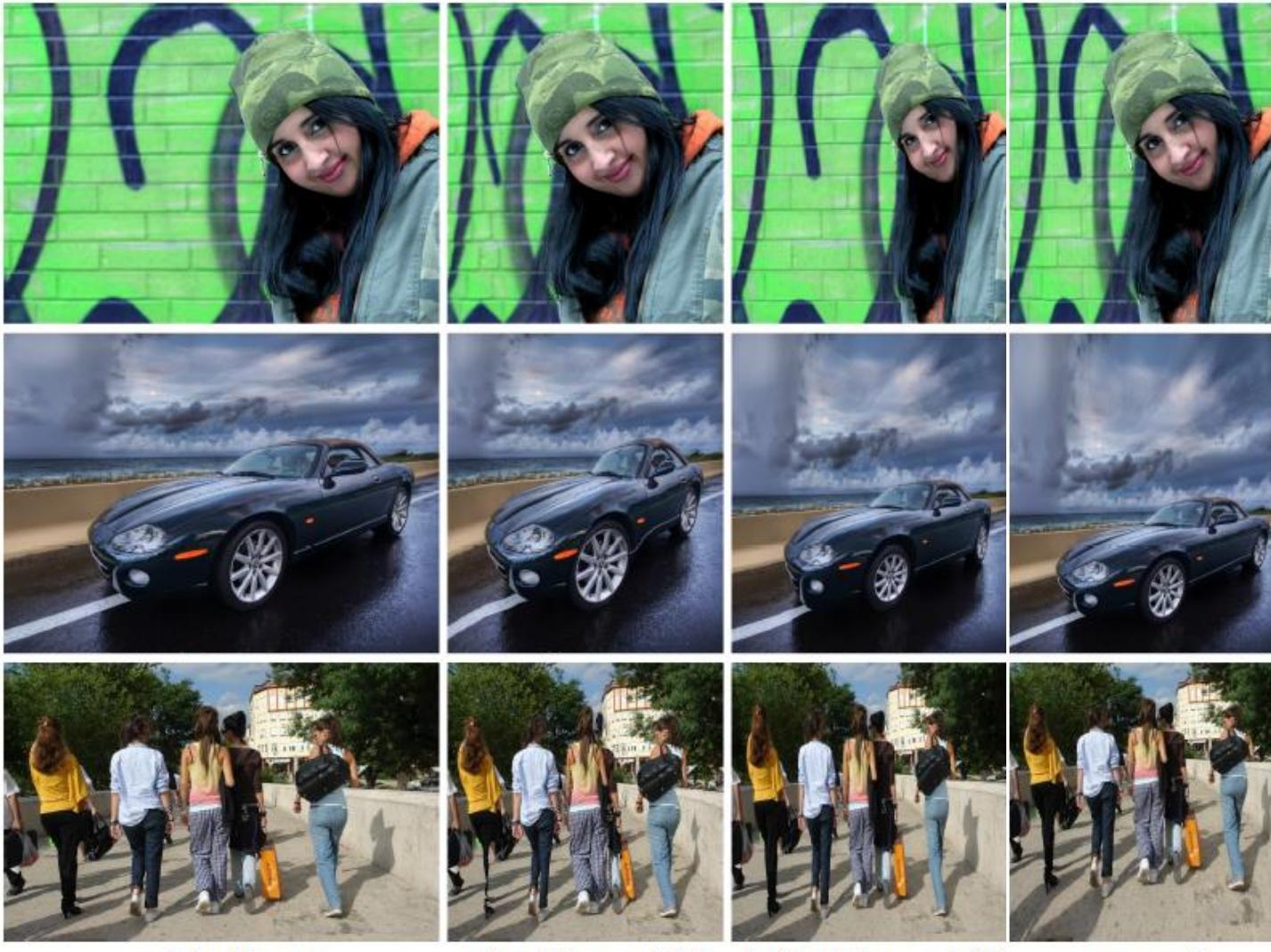
Display Adaptation



Content-aware display adaptation

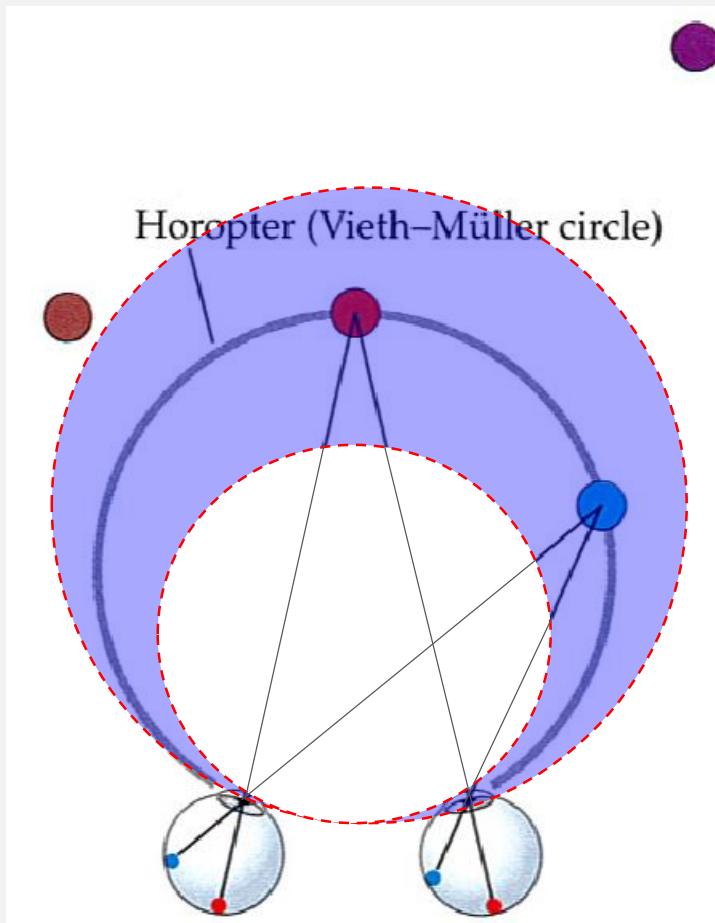
- Stereoscopic displays have different resolutions, aspect ratios and comfort zones.
- To display stereoscopic media properly on different displays, we need content-aware stereoscopic image display adaptation methods to perform **image retargeting** and **depth adaptation** simultaneously.

2D Media Retargeting



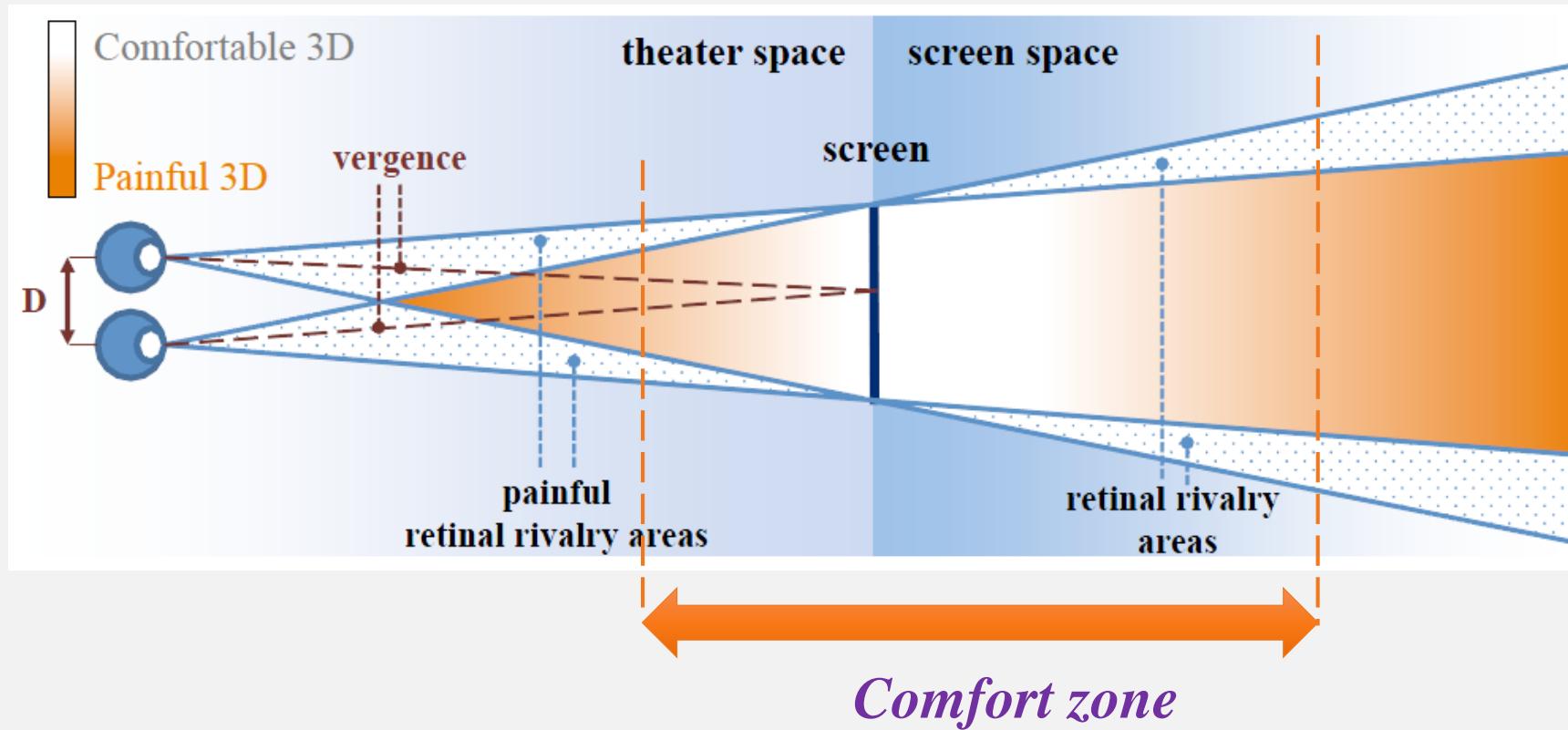
Depth Adaptation

- Adapt depths to the **comfort zone** to avoid visual discomfort such as blur and double vision.



Panum's fusional area

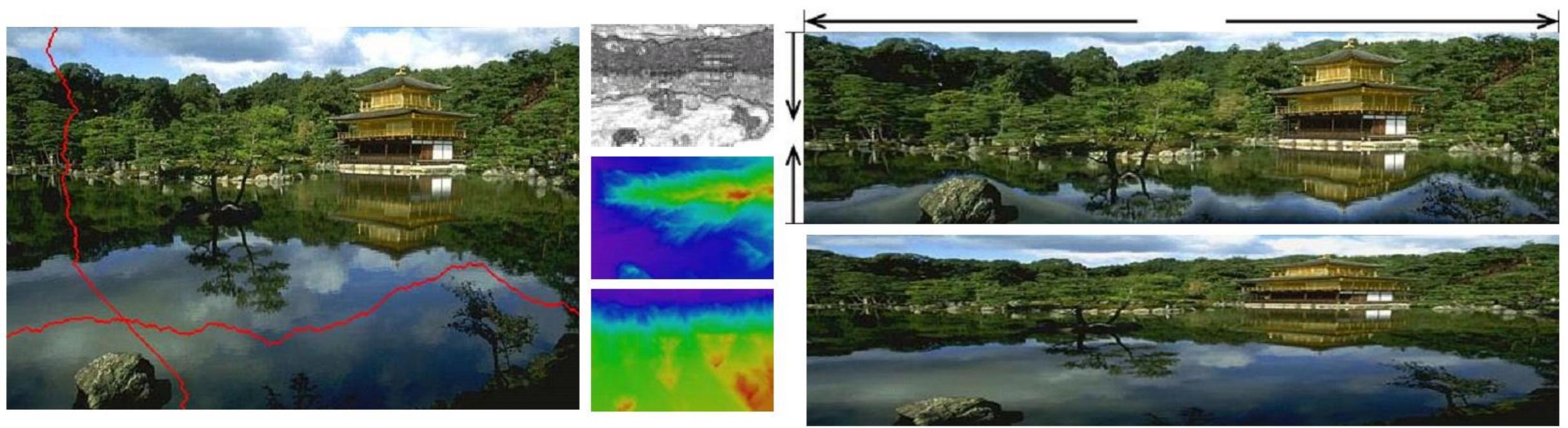
Stereoscopic Comfort Zone



Previous Work (2D image resizing)

■ Seam Carving for Content-Aware Image Resizing

[Avidan *et al.*, SIGGRAPH07]



Seam Carving for Content-Aware Image Resizing

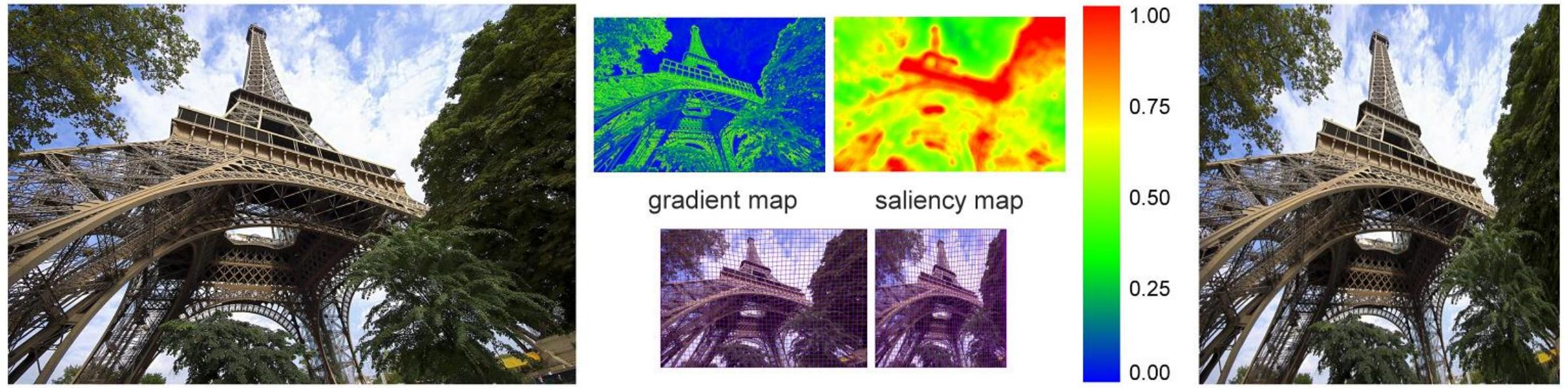


Shai Avidan
Mitsubishi Electric Research Lab
Ariel Shamir
The interdisciplinary Center & MERL

Previous Work (2D image resizing)

■ Optimized scale-and-stretch (OSS)

[Wang *et al.*, SIGGRAPH ASIA08]



Optimized scale-and-stretch for Image Resizing

Optimized Scale-and-Stretch for Image Resizing



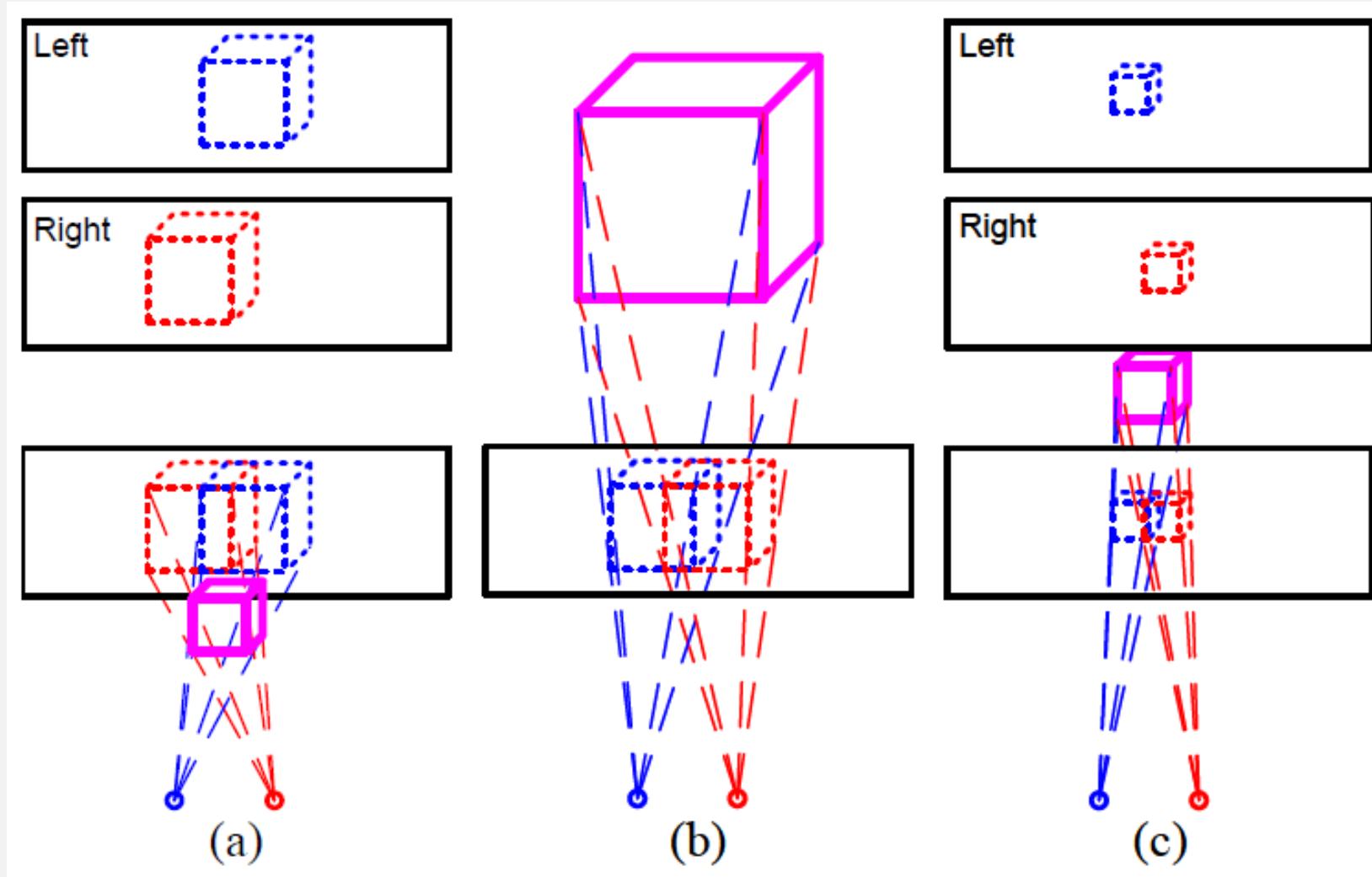
Yu-Shuen Wang¹ Chiew-Lan Tai² Olga Sorkine³ Tong-Yee Lee¹

¹National Cheng Kung
University

²Hong Kong University of
Science & Technology

³New York University

Previous Work (Stereo Resizing by Linear Shift)

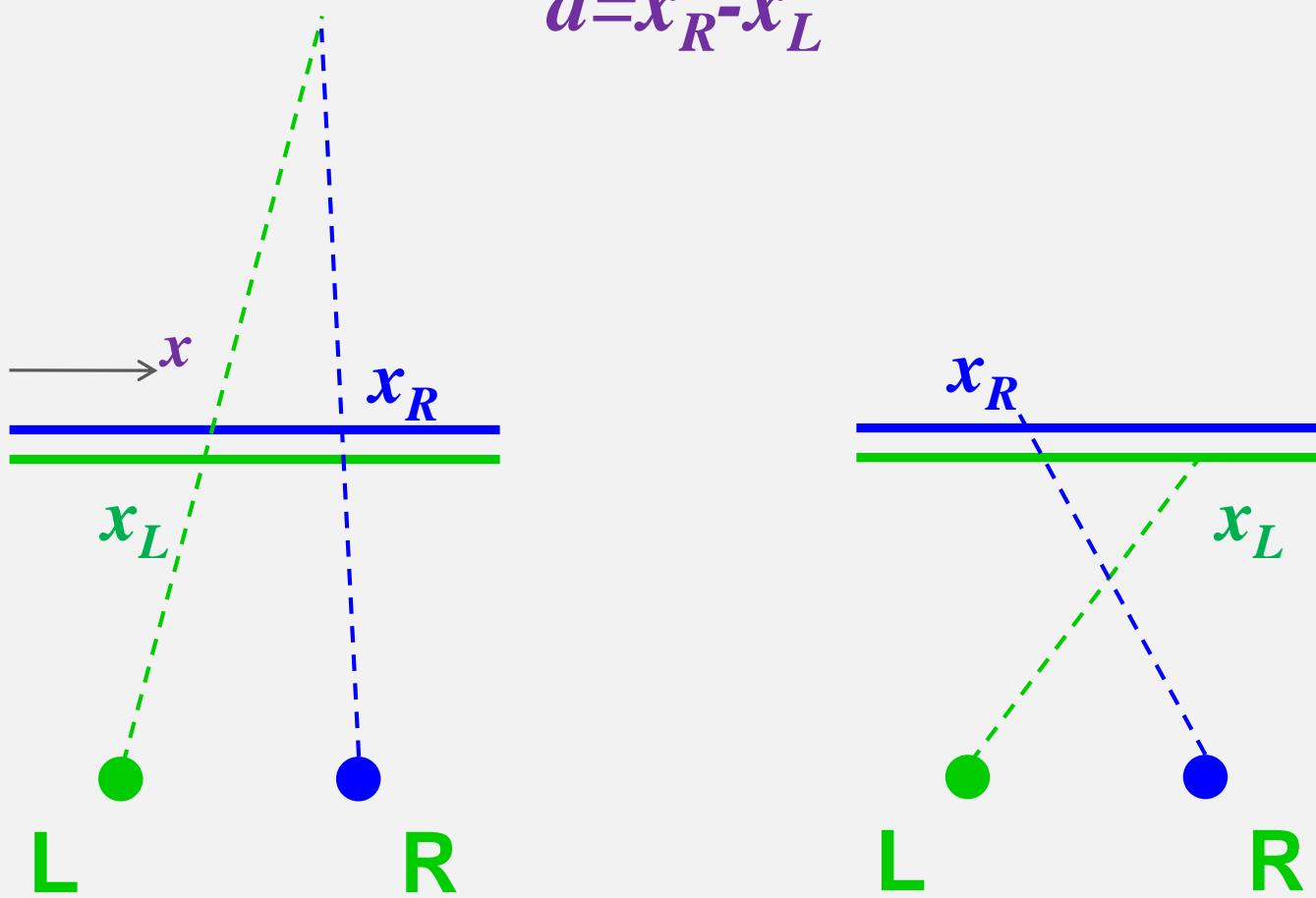


Content-aware Display Adaptation and Interactive Editing for Stereoscopic Images

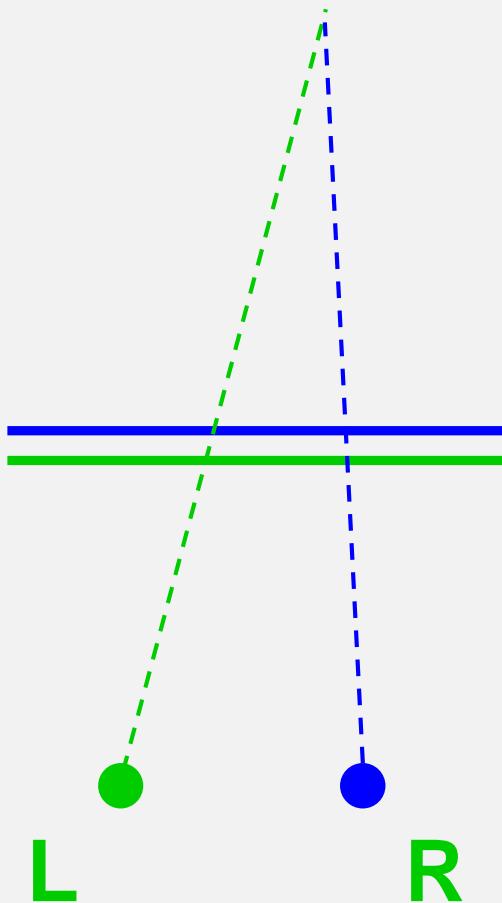


Mapping between disparities and depths

$$d = x_R - x_L$$



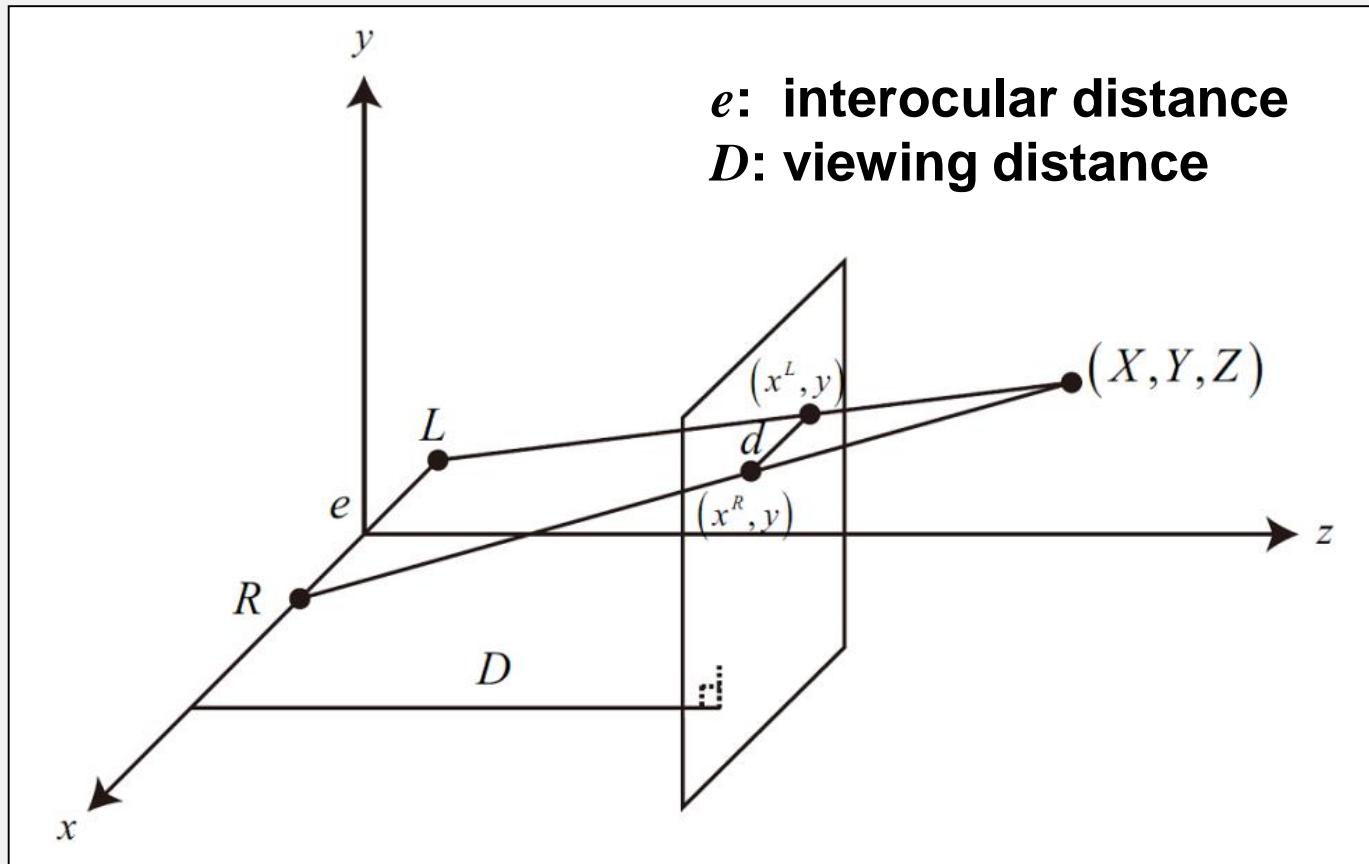
Mapping between disparities and depths



$$Z = \frac{eD}{e - d}$$

Z: perceived depth
d : screen disparity
e: interocular distance
D: viewing distance

Mapping between disparities and depths

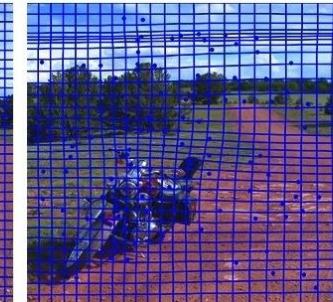
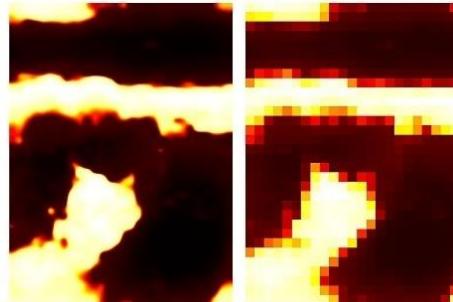


$$\begin{aligned}x^L &= -\frac{e}{2} + \left(X + \frac{e}{2}\right) \frac{D}{Z} \\x^R &= \frac{e}{2} + \left(X - \frac{e}{2}\right) \frac{D}{Z} \\y &= Y \frac{D}{Z}.\end{aligned}$$

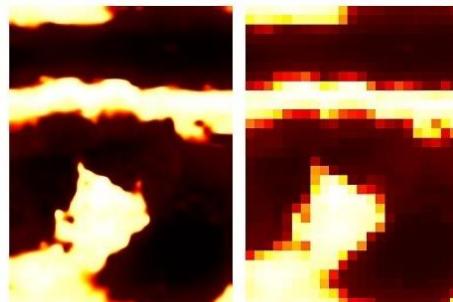
$$(X, Y, Z)^T = \frac{e}{e - d} \left(\frac{x^L + x^R}{2}, y, D \right)^T$$

Overview

Left



Right



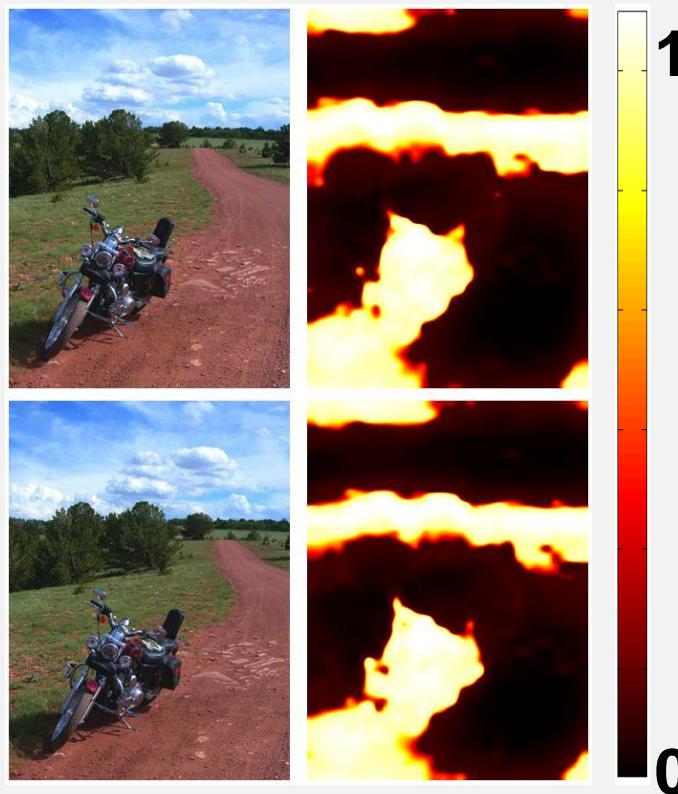
$$\{\mathbf{V}^L, \mathbf{V}^R\}$$

$$\{\tilde{\mathbf{V}}^L, \tilde{\mathbf{V}}^R\}$$

- Minimize $\Psi(\tilde{\mathbf{V}}^L, \tilde{\mathbf{V}}^R)$

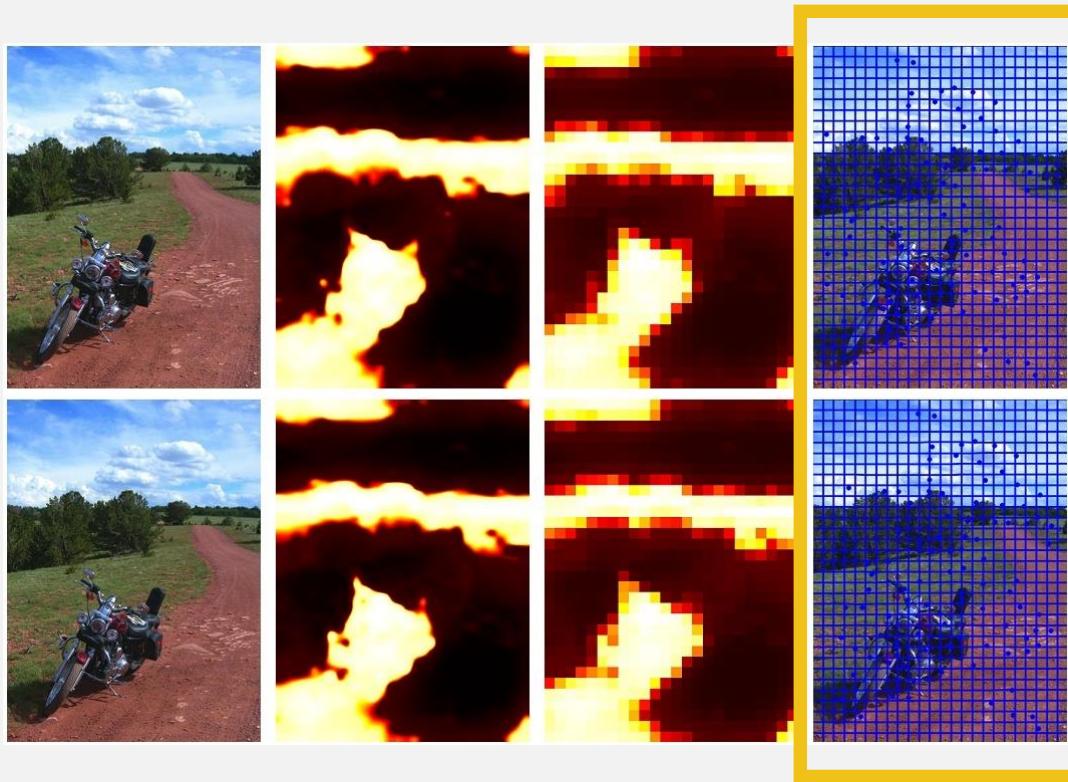
Saliency detection

- Graph-based visual saliency algorithm [Jonathan *et al.*, NIPS06]



Mesh representation

- An image is represented as a rectangular mesh



Quad importance

- Average of saliency values of all pixels in q

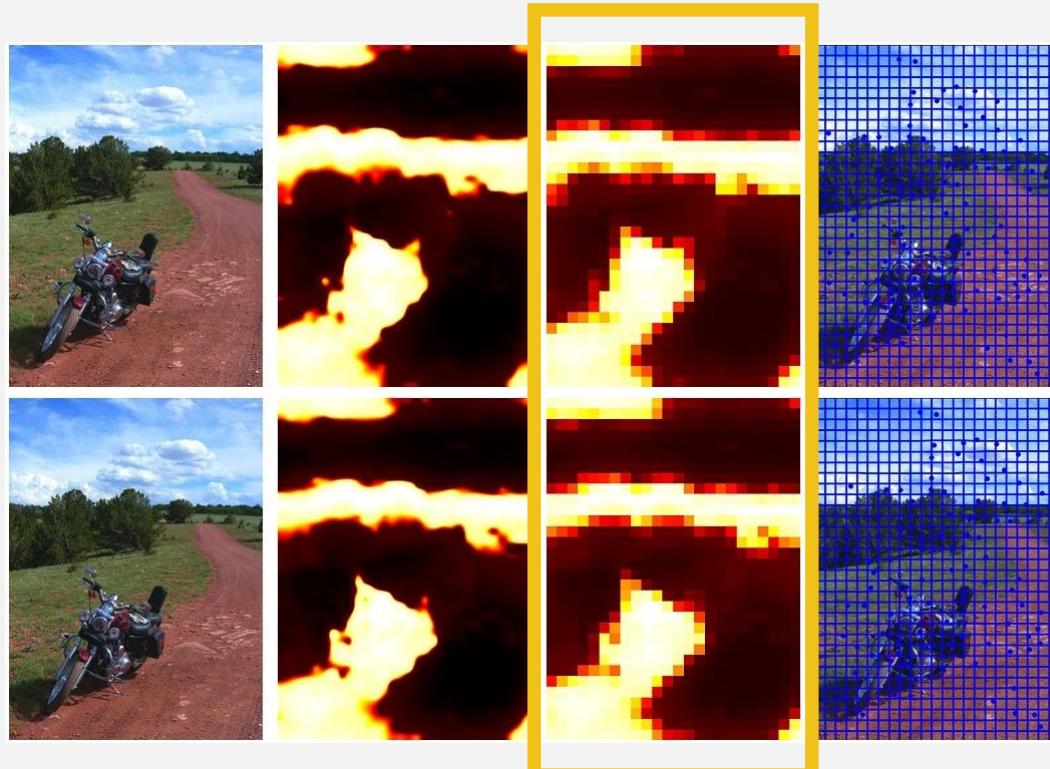


Image correspondence

- Build the stereoscopic constraints
- The state-of-the-art stereo methods are still far from perfection
- We only use **sparse features pairs**
 - Reliable
 - “Sparse correspondences + image warping” are often enough to fool human brain

Feature correspondence

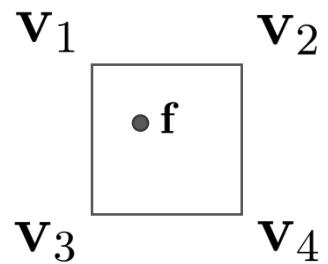
- Feature extraction
 - SIFT feature
- Feature matching
- Verification
 - Estimate fundamental matrix using RANSAC
- Non-maximal suppression



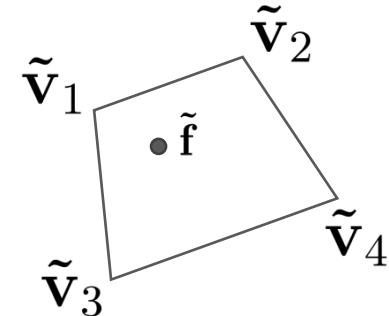
$$\mathbf{F} = \{(\mathbf{f}_i^L, \mathbf{f}_i^R)\}$$

Feature correspondence

$\tilde{\mathbf{f}}$ is expressed as a linear combination of $\tilde{\mathbf{v}}_i$



$$\mathbf{f} = 0.42\mathbf{v}_1 + 0.18\mathbf{v}_2 + 0.12\mathbf{v}_3 + 0.28\mathbf{v}_4$$



$$\tilde{\mathbf{f}} = 0.42\tilde{\mathbf{v}}_1 + 0.18\tilde{\mathbf{v}}_2 + 0.12\tilde{\mathbf{v}}_3 + 0.28\tilde{\mathbf{v}}_4$$

Energy minimization

■ Energy function

$$\Psi = \underline{\Psi_d + \lambda_b \Psi_b + \lambda_a \Psi_a + \lambda_c \Psi_c}$$

Content preservation

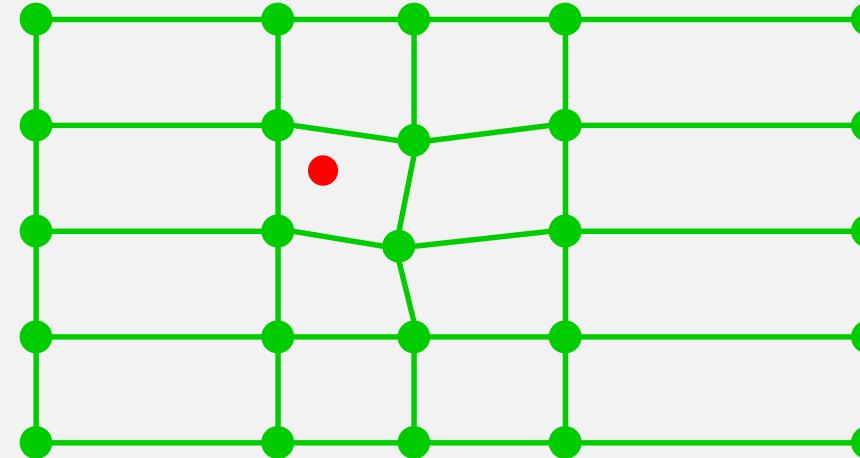
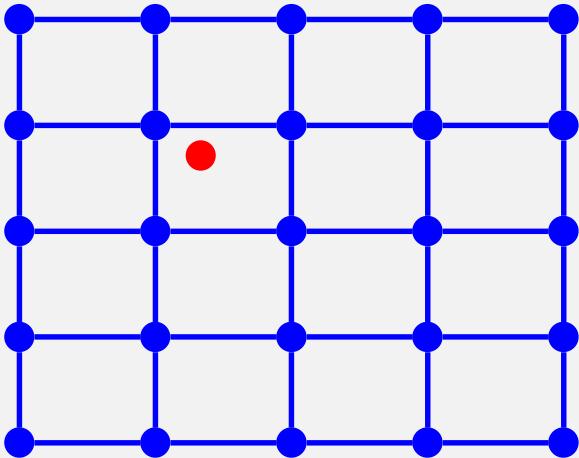
Ψ_d : Distortion energy

Ψ_b : Line bending energy

Ψ_a : Alignment energy

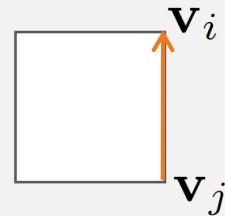
Ψ_c : Consistent disparity energy

Energy minimization



Distortion energy

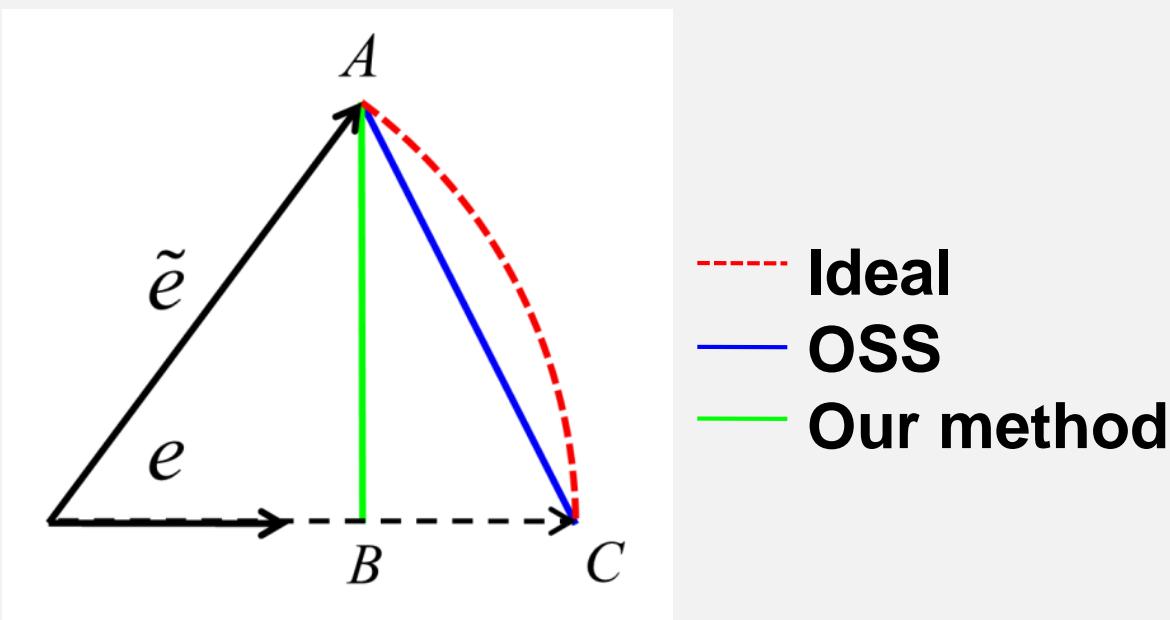
$$\Psi_q(q) = \sum_{(i,j) \in \mathbf{E}(q)} \| (\tilde{\mathbf{v}}_i - \tilde{\mathbf{v}}_j) - s_q (\mathbf{v}_i - \mathbf{v}_j) \|^2$$



- The optimal s_q is completely defined by \mathbf{v}_i and $\tilde{\mathbf{v}}_i$

Line bending energy

$$\Delta(\tilde{\mathbf{e}}) = \|s\mathbf{e} - \tilde{\mathbf{e}}\|^2$$



Energy minimization

- Energy function

$$\Psi = \Psi_d + \lambda_b \Psi_b + \underline{\lambda_a \Psi_a + \lambda_c \Psi_c}$$

Stereoscopic constraints

Ψ_d : Distortion energy

Ψ_b : Line bending energy

Ψ_a : Alignment energy

Ψ_c : Consistent disparity energy

Alignment energy

- Maintain purely horizontal parallax
(avoid vertical parallax)

$$\Psi_a = \sum_{i=1}^n \left(\tilde{\mathbf{f}}_i^L[y] - \tilde{\mathbf{f}}_i^R[y] \right)^2$$

Consistent disparity energy

- Maintain the relative shapes and depths

$$\Psi_c = \sum_{i=1}^n \left((sd_i + t) - \tilde{d}_i \right)^2 \quad \text{where } d_i = \mathbf{f}_i^R[x] - \mathbf{f}_i^L[x] \text{ and } \tilde{d}_i = \tilde{\mathbf{f}}_i^R[x] - \tilde{\mathbf{f}}_i^L[x].$$

- Keep the absolute shapes and depths

$$\Psi_c = \sum_{i=1}^n \left(d_i - \tilde{d}_i \right)^2$$

Energy minimization

- Energy function

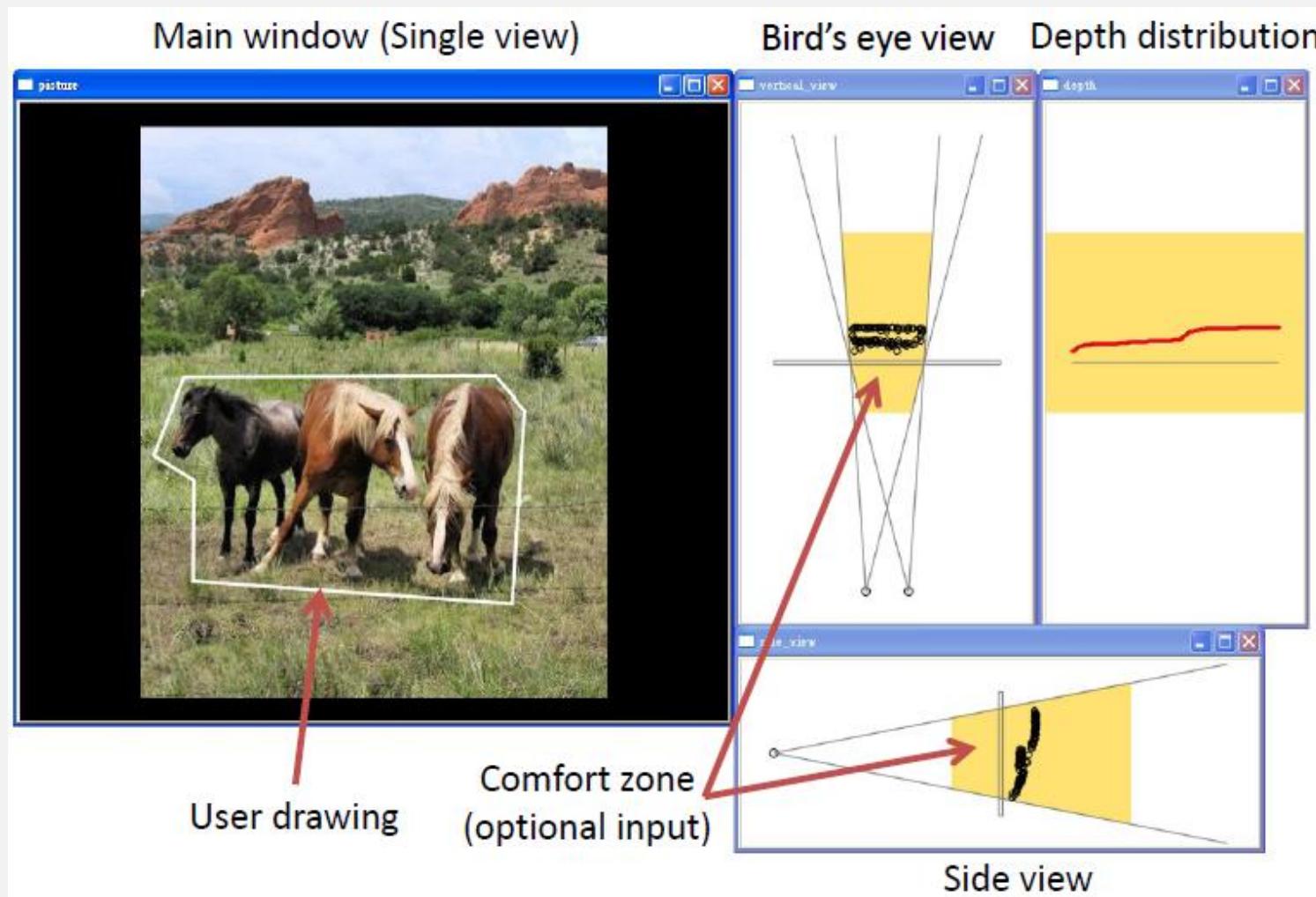
$$\Psi = \Psi_d + \lambda_b \Psi_b + \lambda_a \Psi_a + \lambda_c \Psi_c$$

- Linear least squares problem → closed form solution

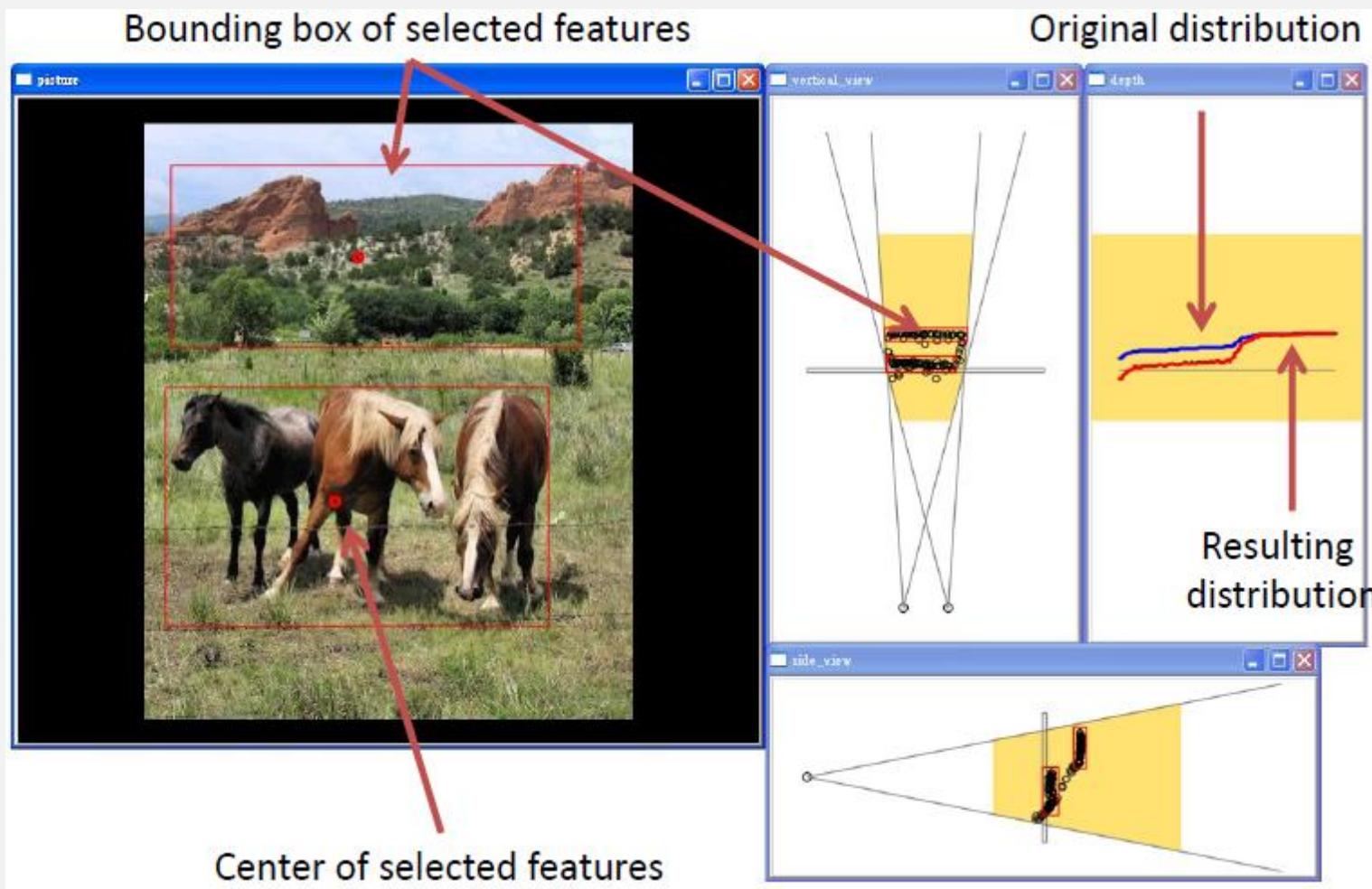
User editing

- User specifies depths (Z-axis)
- User specifies 3D position (XYZ-axis)

User editing



User editing

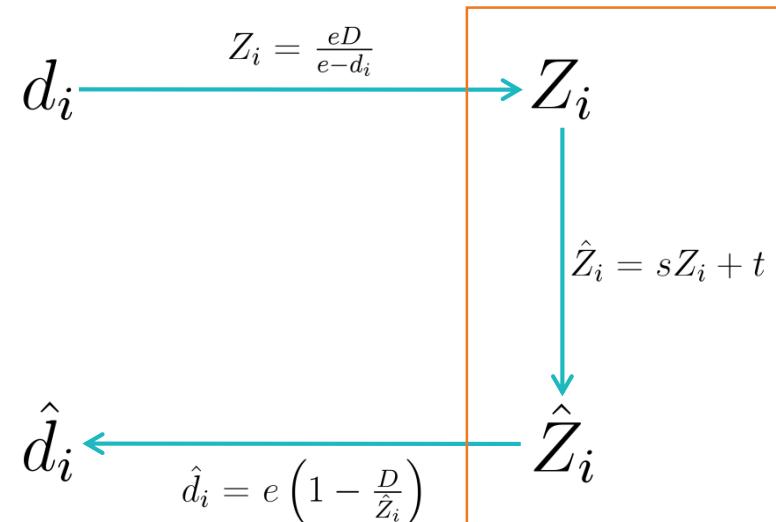


User editing

- User specifies depths (Z-axis)

$$\Psi_c = \sum_{i=1}^n \left(\tilde{d}_i - \hat{d}_i \right)^2$$

Target disparity



User editing

- User specifies 3D position (XYZ-axis)

$$\begin{aligned}\Psi_c = \sum_{i \in \hat{\mathbf{F}}} & \left(\|\tilde{\mathbf{f}}_i^L - \underline{\hat{\mathbf{f}}_i^L}\|^2 + \|\tilde{\mathbf{f}}_i^R - \underline{\hat{\mathbf{f}}_i^R}\|^2 \right) \\ & + \lambda \sum_{i \in \mathbf{F} - \hat{\mathbf{F}}} \left((sd_i + t) - \tilde{d}_i \right)^2.\end{aligned}$$

Target 2D position

Results



original



Scaling



Saliency



Our method

