Stereoscopic 3D video for the human eyes

Frédéric Devernay

with
Sergi Pujades
Elise Mansilla
Loïc Lefort
Martin Guillon
Matthieu Volat
Sylvain Duchêne
Adrian Ramos-Peron

INRIA

July 2011

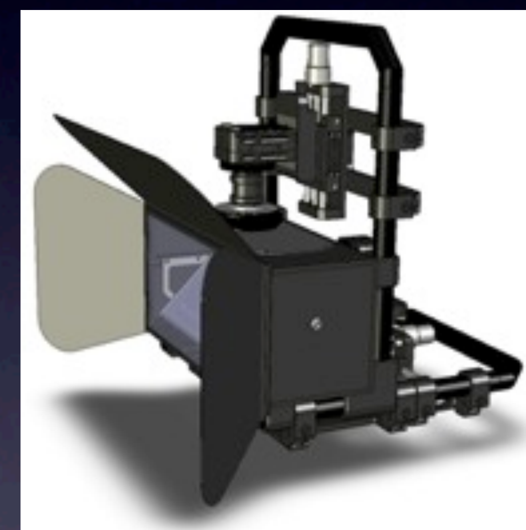
Stereoscopic cinema

- Movie made using two cameras in stereoscopic configuration
- Not the same as:
  - free-viewpoint video (hundreds of cameras in linear or array arrangement)
  - 3-D video from multiple views

History

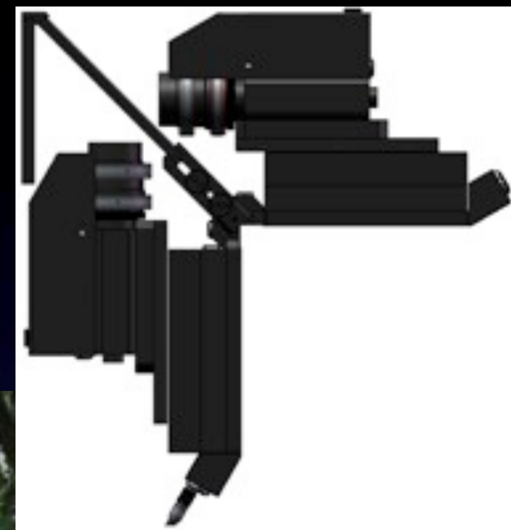
- 1922: first public projection (*The Power of Love*, anaglyph)
- 1952: first feature-length movie (*Bwana Devil*)
- 1954: Hitchcock's *Dial M for Murder*
- 1980's: Rebirth of 3-D, IMAX-3D
- 2003-: Digital 3-D (Spy Kids 3-D, U2 3D, animated 3-D movies by Disney et al.)
- 2009: Coraline, Avatar, live sports events...

3-D cameras: Fixed/manual interocural



US motion-control

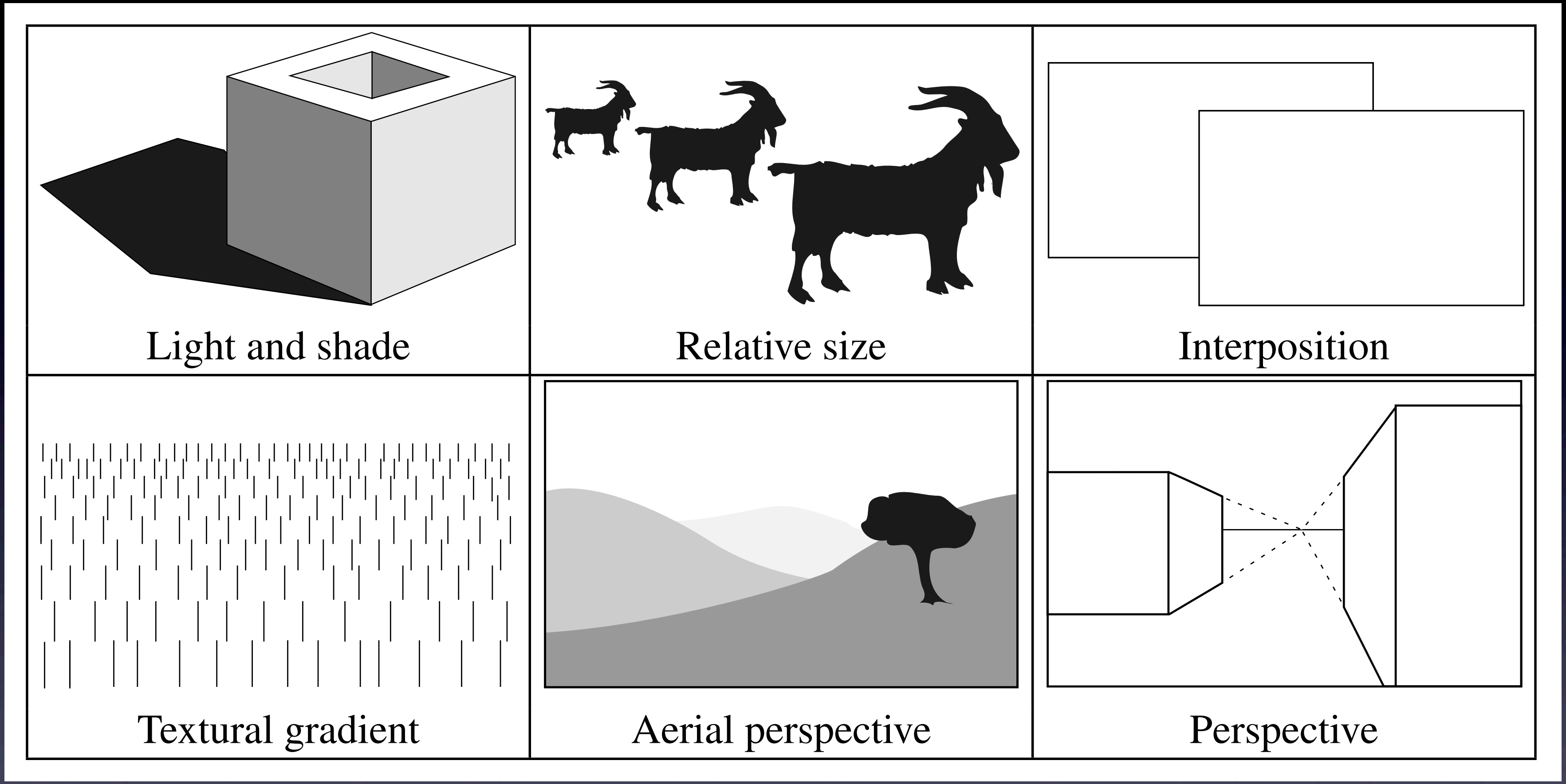
Binocle motion-control systems



Why do we see 3D?

- NOT because we have two eyes...

Three-Dimensional Depth Cues



And also **motion parallax**, **depth of field**, and... **stereoscopy**

Depth of field as a depth cue: focus matters!

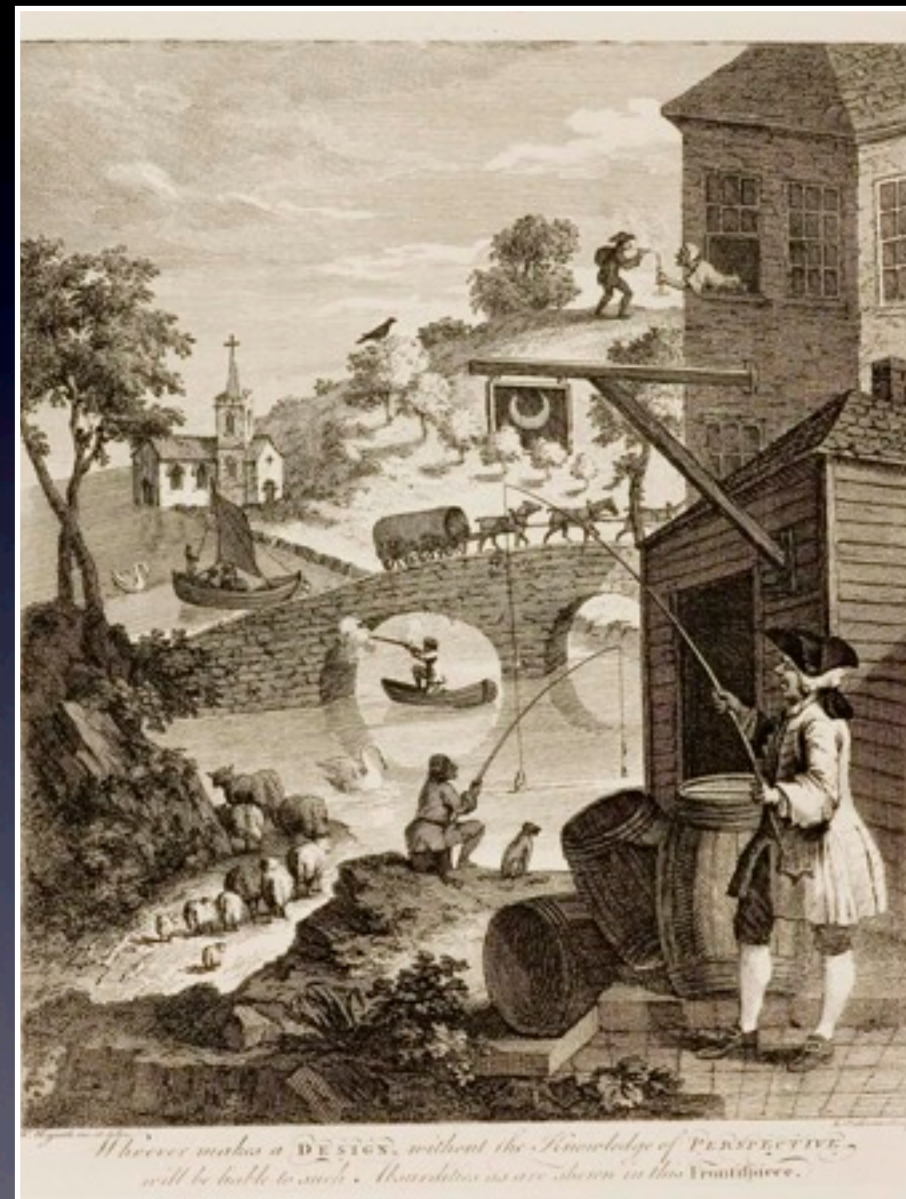
Conflicting depth cues

- The 9 cues may give opposite indications on the scene geometry

- The **pseudoscope** (Wheatstone) - reverse left and right eyes - causes *closer* objects to seem even *bigger*:

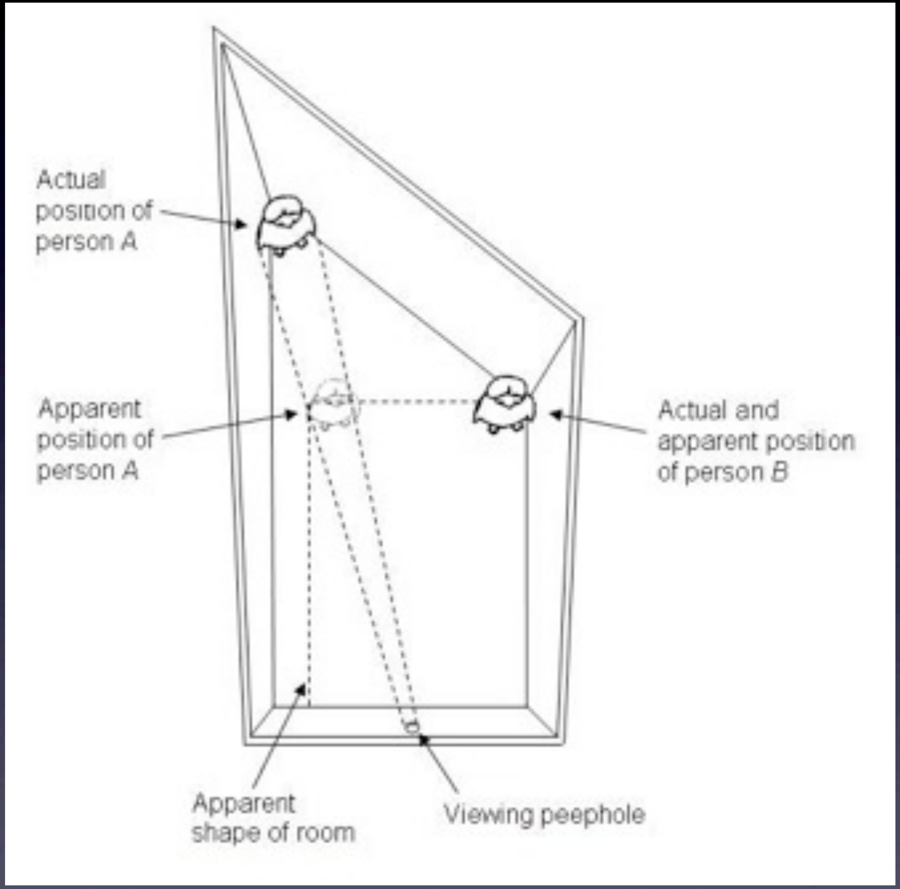
    - big in the image

    - binocular disparity indicates they are also far away



William Hogarth, 1754

Conflicting cues: Ames room



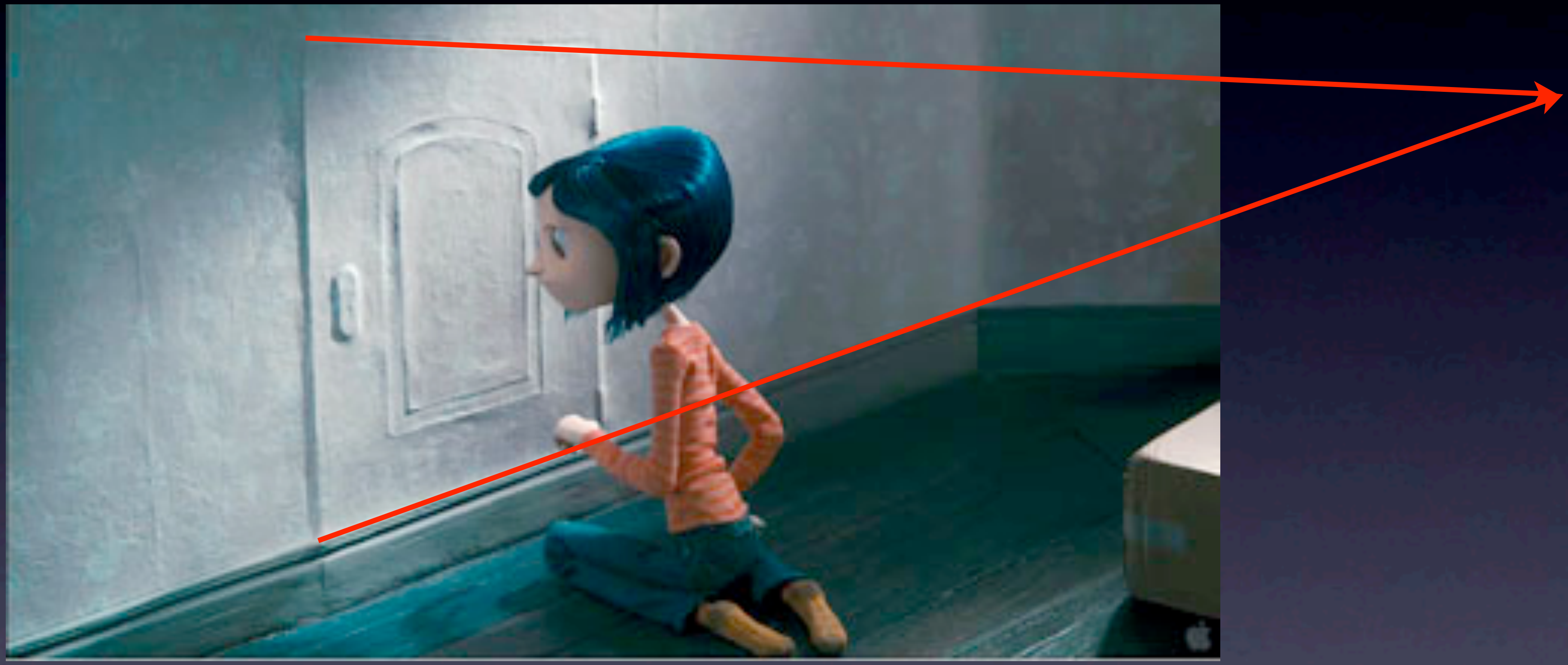
Used in *Lord of the Rings*,
Eternal Sunshine of the Spotless Mind...

Stereoscopic conflicting cues: Coraline 3D

Coraline (H. Selick & P. Kozachik)

2 vanishing points in the same 3-D scene

Stereoscopic conflicting cues: Coraline 3D

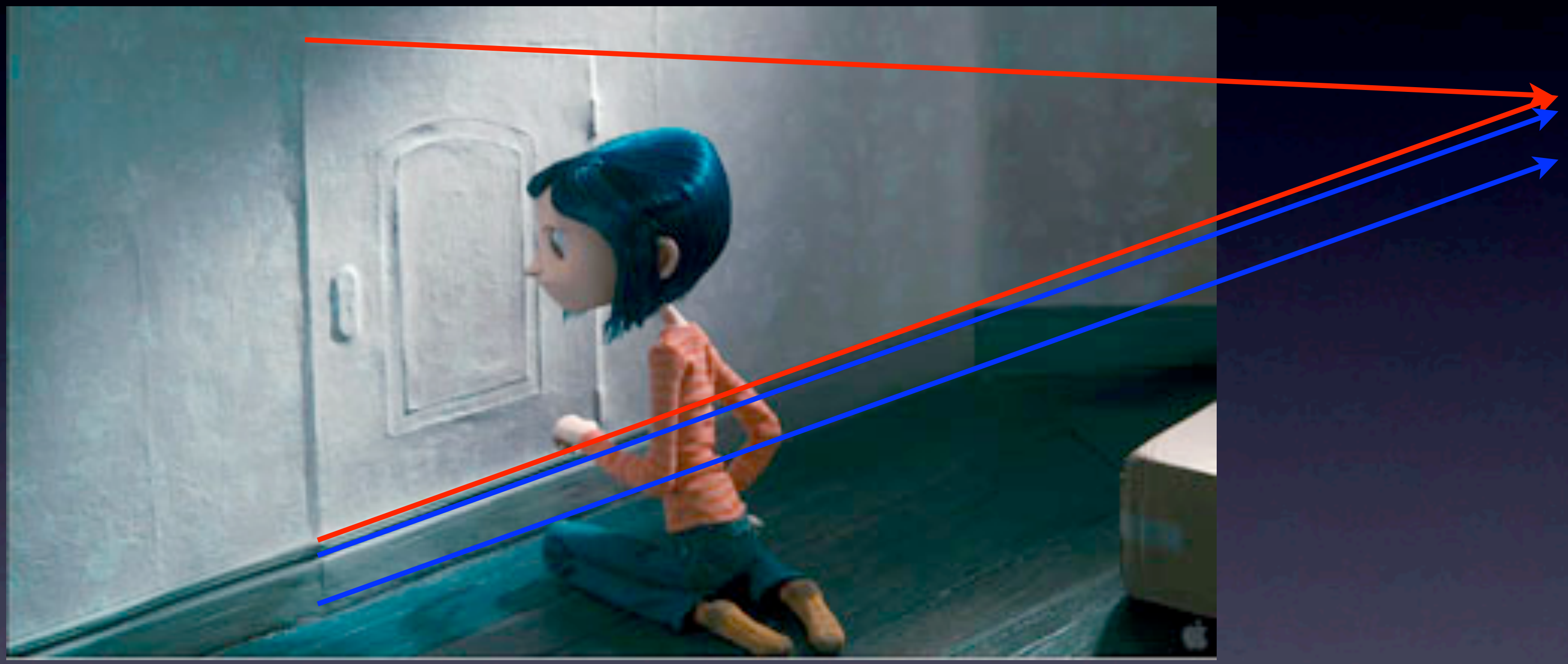


Coraline (H. Selick & P. Kozachik)

2 vanishing points in the same 3-D scene

INRIA

Stereoscopic conflicting cues: Coraline 3D

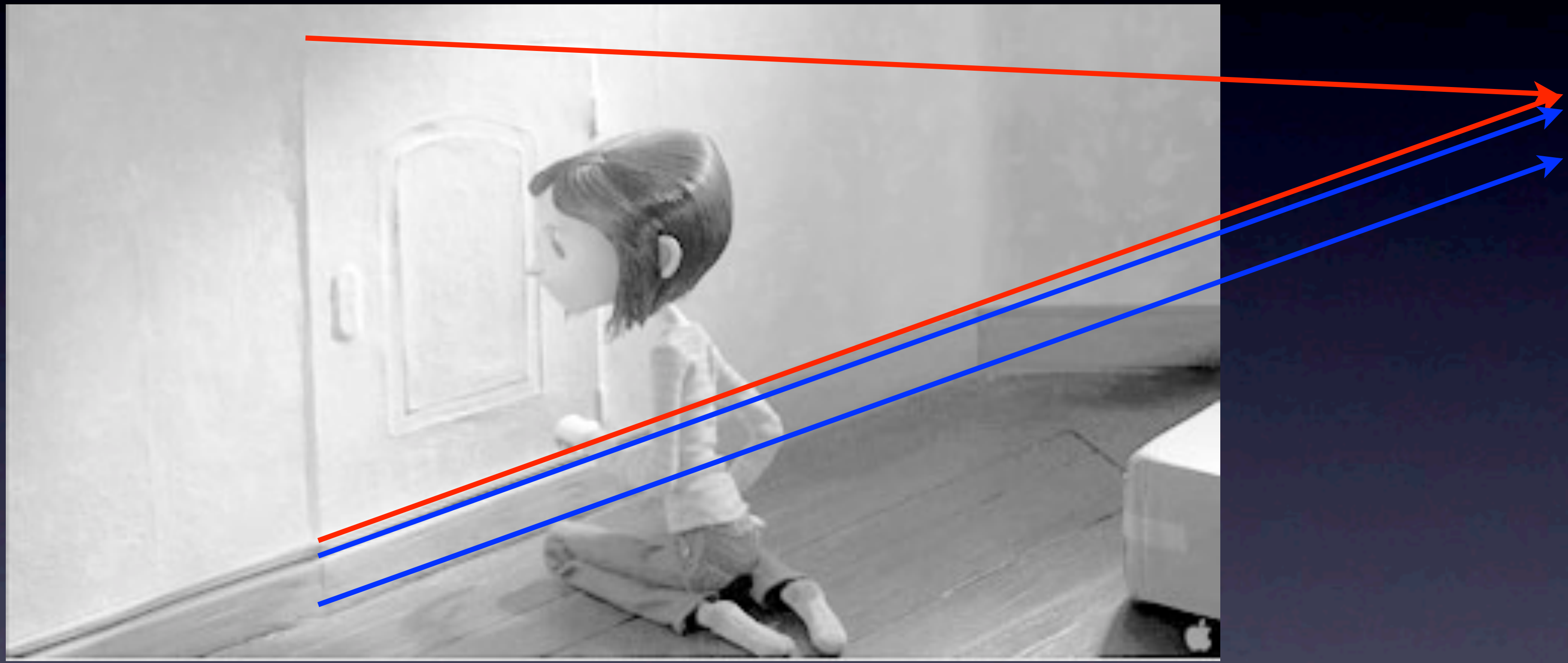


Coraline (H. Selick & P. Kozachik)

2 vanishing points in the same 3-D scene

INRIA

Stereoscopic conflicting cues: Coraline 3D



Coraline (H. Selick & P. Kozachik)

2 vanishing points in the same 3-D scene

Stereo-specific video processes

Stereo-specific video processes

- Correcting causes of visual fatigue

INRIA

Stereo-specific video processes

- Correcting causes of visual fatigue
- Color-balancing left and right cameras

Stereo-specific video processes

- Correcting causes of visual fatigue
- Color-balancing left and right cameras
- Adapt the movie to the screen size

Stereo-specific video processes

- Correcting causes of visual fatigue
- Color-balancing left and right cameras
- Adapt the movie to the screen size
- Global 3-D changes (interocular, infinity...)

Stereo-specific video processes

- Correcting causes of visual fatigue
- Color-balancing left and right cameras
- Adapt the movie to the screen size
- Global 3-D changes (interocular, infinity...)
- Local 3-D changes (3-D touchup)

Stereo-specific video processes

- Correcting causes of visual fatigue
- Color-balancing left and right cameras
- Adapt the movie to the screen size
- Global 3-D changes (interocular, infinity...)
- Local 3-D changes (3-D touchup)
- Playing with the depth of focus

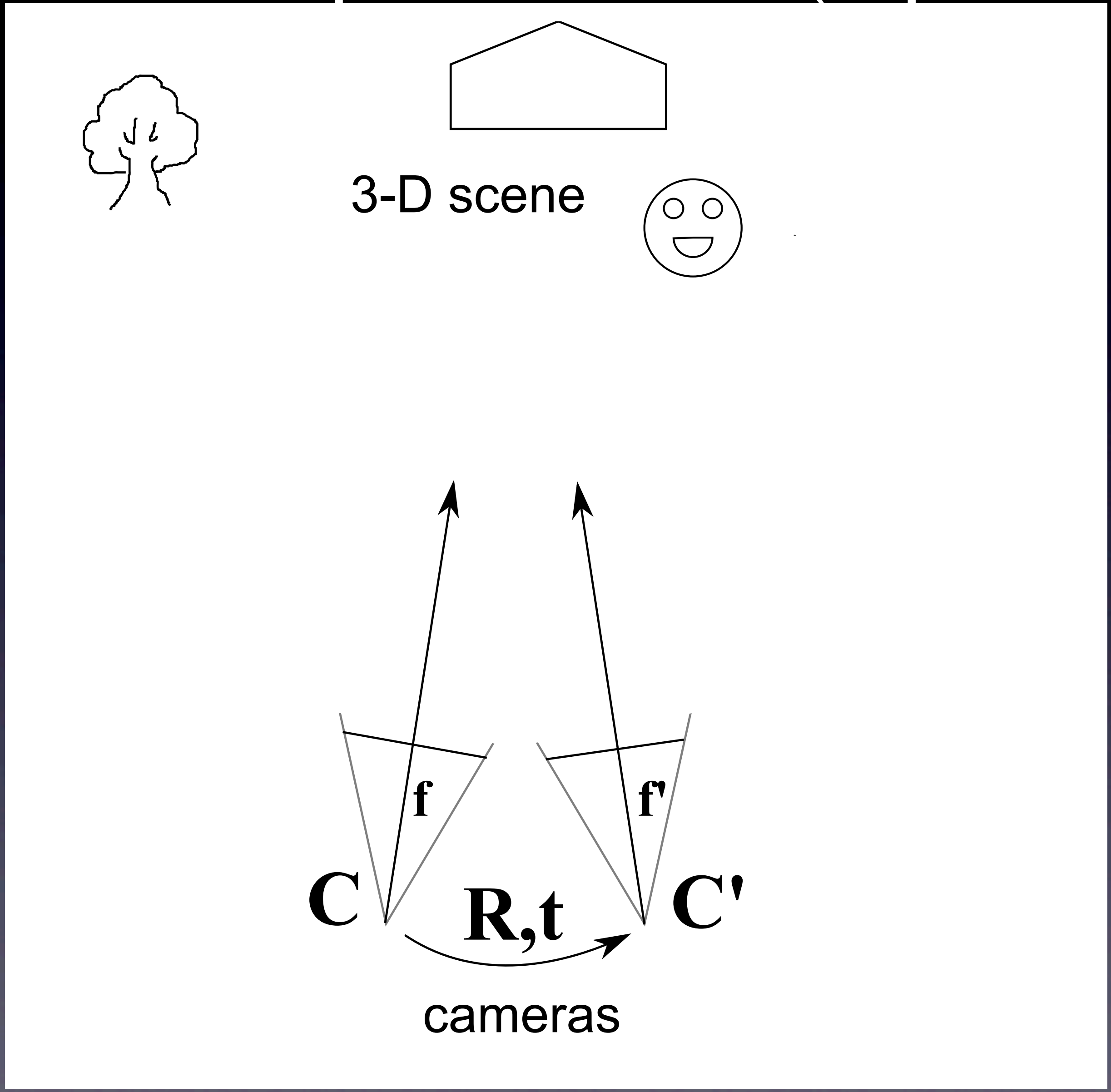
Stereo-specific video processes

- Correcting causes of visual fatigue
- Color-balancing left and right cameras
- Adapt the movie to the screen size
- Global 3-D changes (interocular, infinity...)
- Local 3-D changes (3-D touchup)
- Playing with the depth of focus
- Playing with the proscenium

Stereo-specific video processes

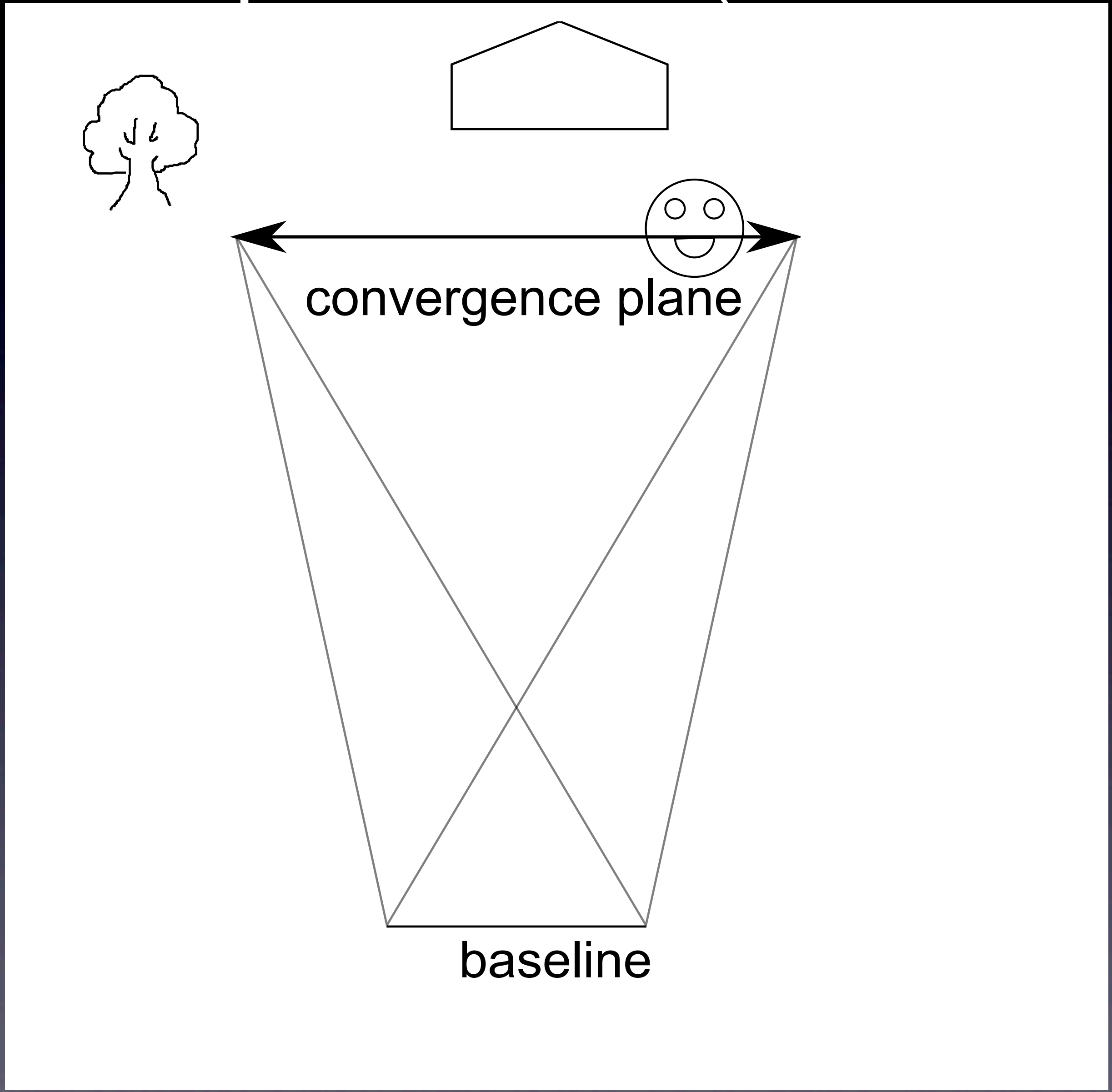
- Correcting causes of visual fatigue
- Color-balancing left and right cameras
- Adapt the movie to the screen size
- Global 3-D changes (interocular, infinity...)
- Local 3-D changes (3-D touchup)
- Playing with the depth of focus
- Playing with the proscenium
- 3-D compositing (real or CG scenes)

The shooting geometry: classical representation (top view)



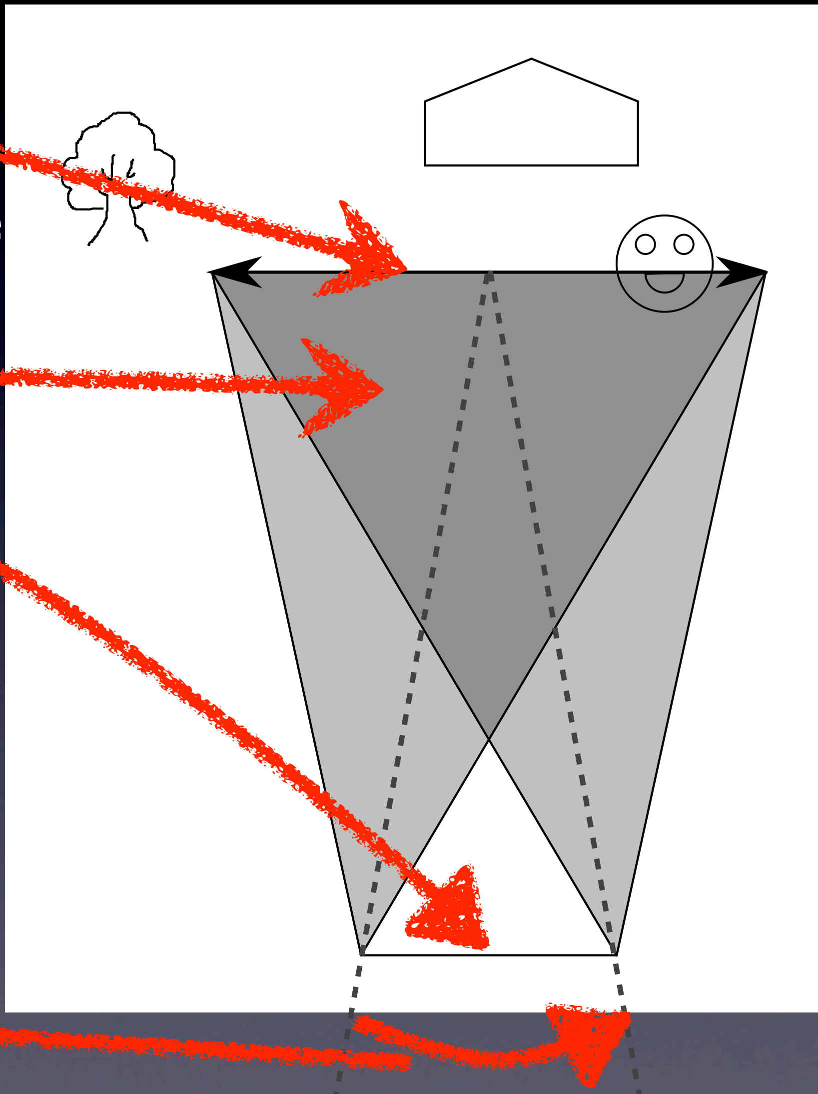
INRIA

The shooting geometry: simplified representation (rectified images)



A few definitions

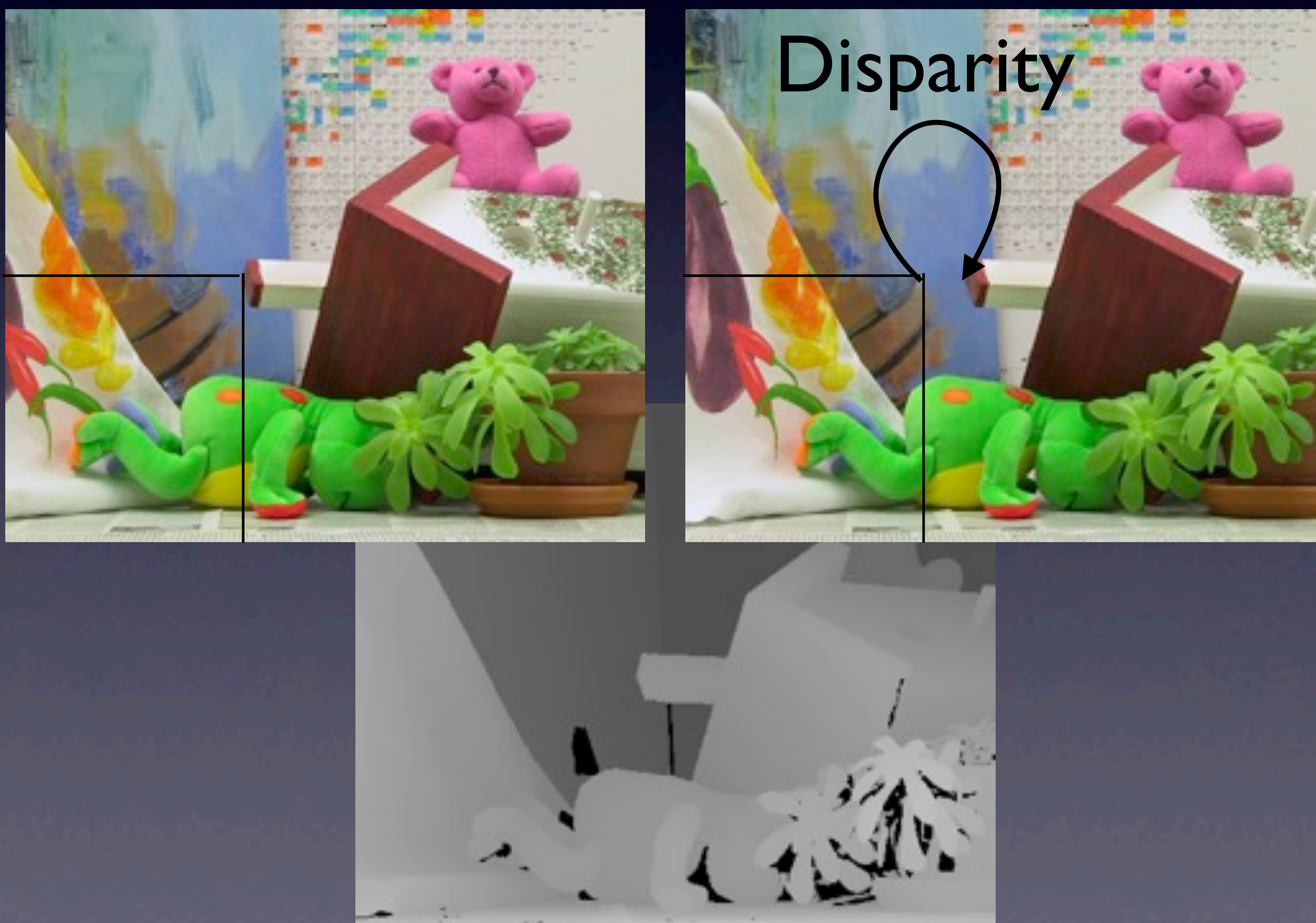
- Screen plane ... in the viewer space
- Plane of convergence .. in the scene space
- 3-D cone
- Interocular / Interaxial
  - bigger than 65mm (can be 30m)□ hyperstereo
  - smaller than 65mm (can be 0cm) □ hypostereo
- Convergence

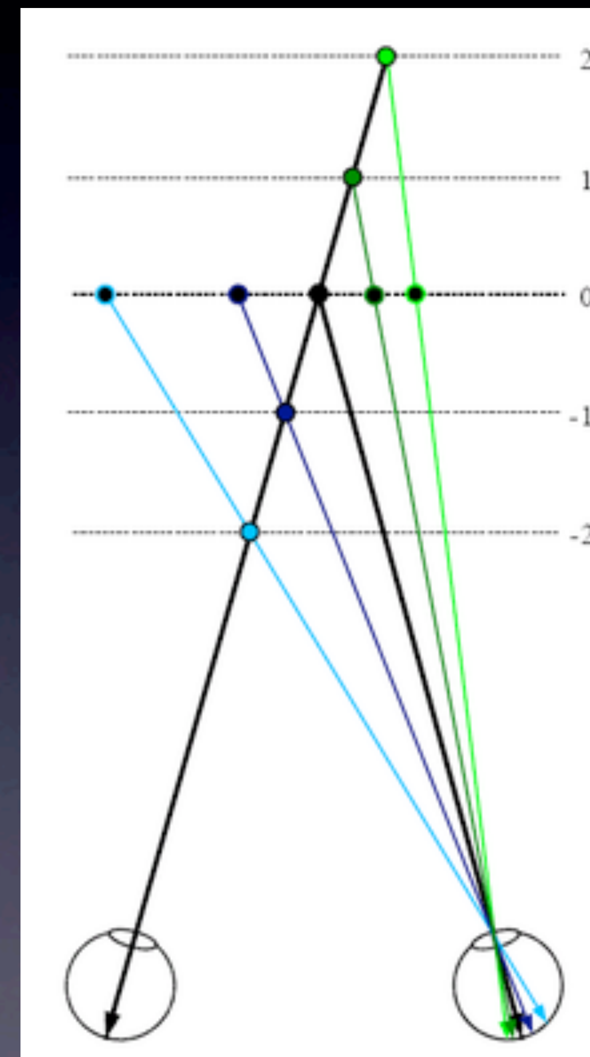


INRIA

Binocular disparity: how stereopsis works

- Objects at different depths cause different disparities





left view

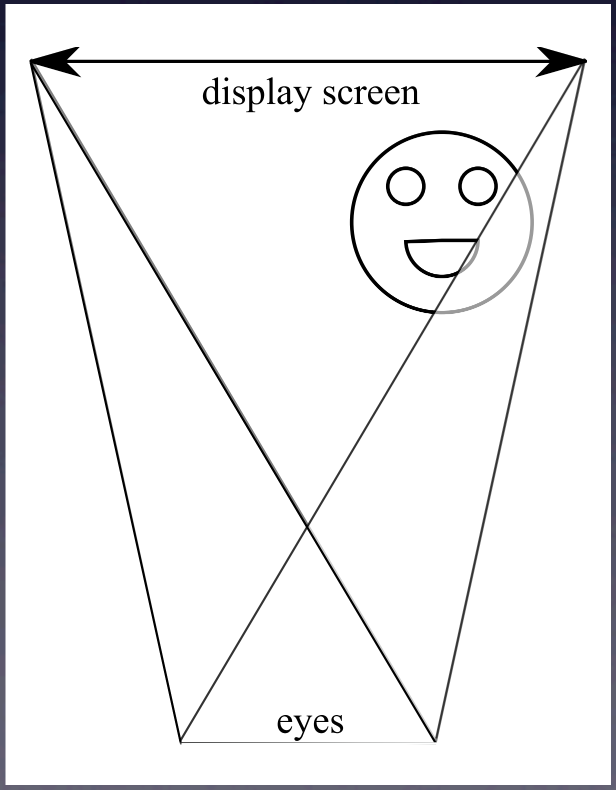
right view

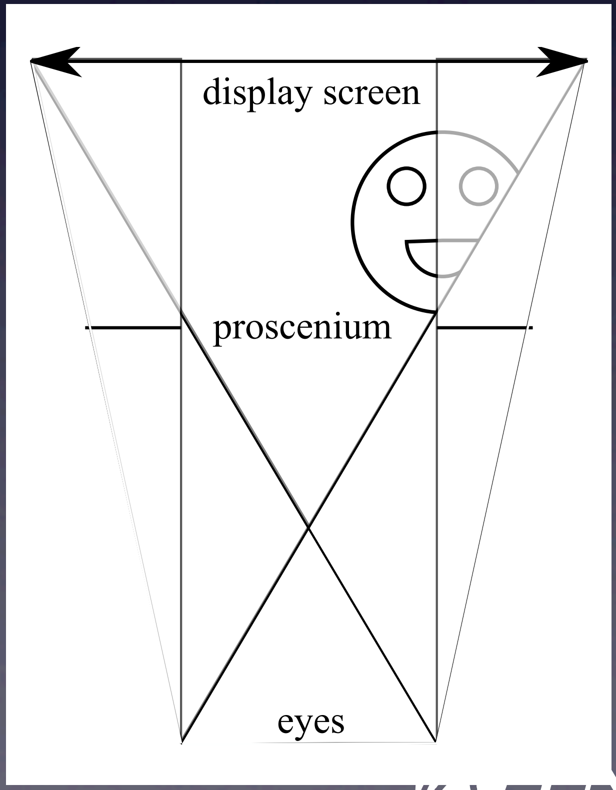
The proscenium arch (or stereoscopic window)

The stereoscopic display is a window on the world

If object closer than convergence plane touches the image borders...

☐ Add black borders to move proscenium arch closer



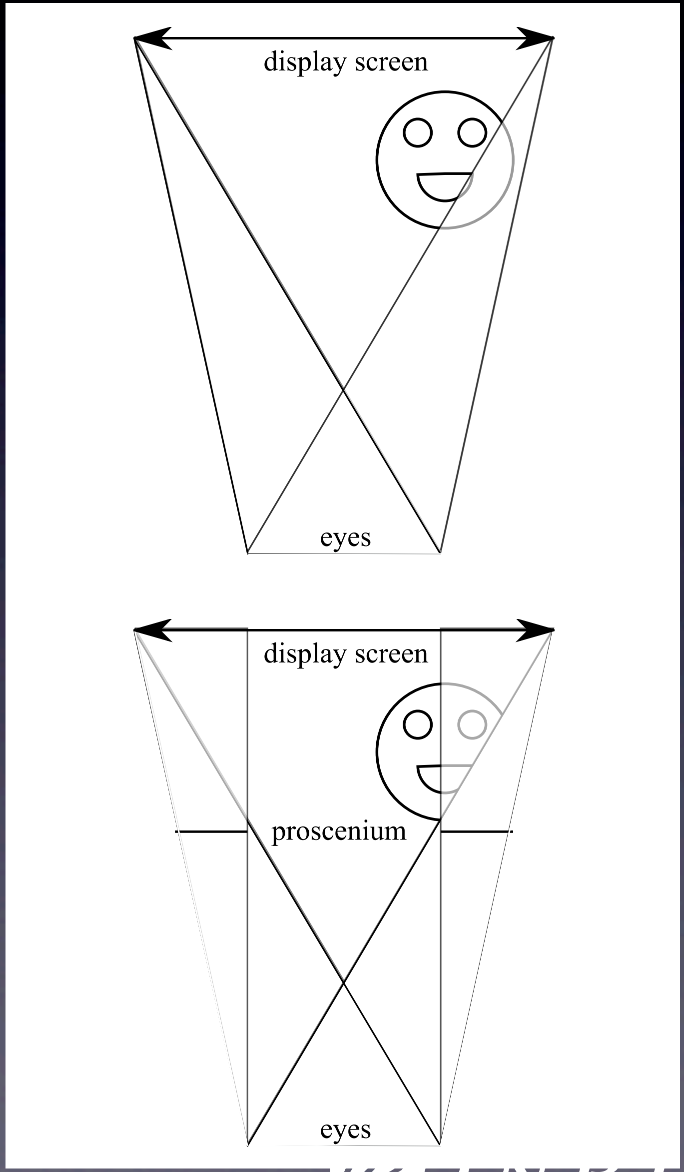


Visual fatigue: a critical point

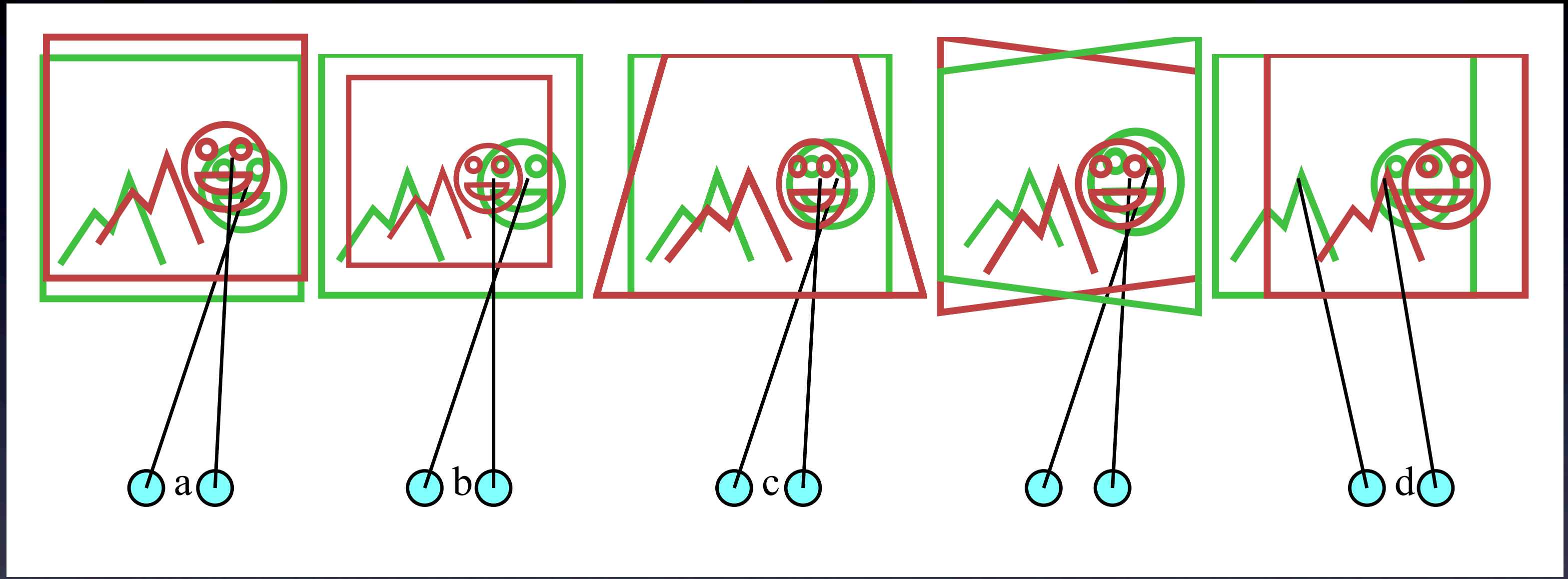
- Can lead to:
  - a simple headache
  - temporary or permanent damage to the oculo-motor system (especially on children)
- Probably a public health problem (just as the critical fusion frequency on CRT screens...)

Some sources of visual fatigue

- Crosstalk
- Breaking the proscenium rule (stereoscopic window violation)
- Horizontal disparity limits
- Vertical disparity
- Vergence-accomodation conflicts



Visual fatigue: geometric differences



a. vertical shift
b. size difference
c. distortion difference
d. keystone (toed-in cameras)
e. horizontal shift (divergence...)

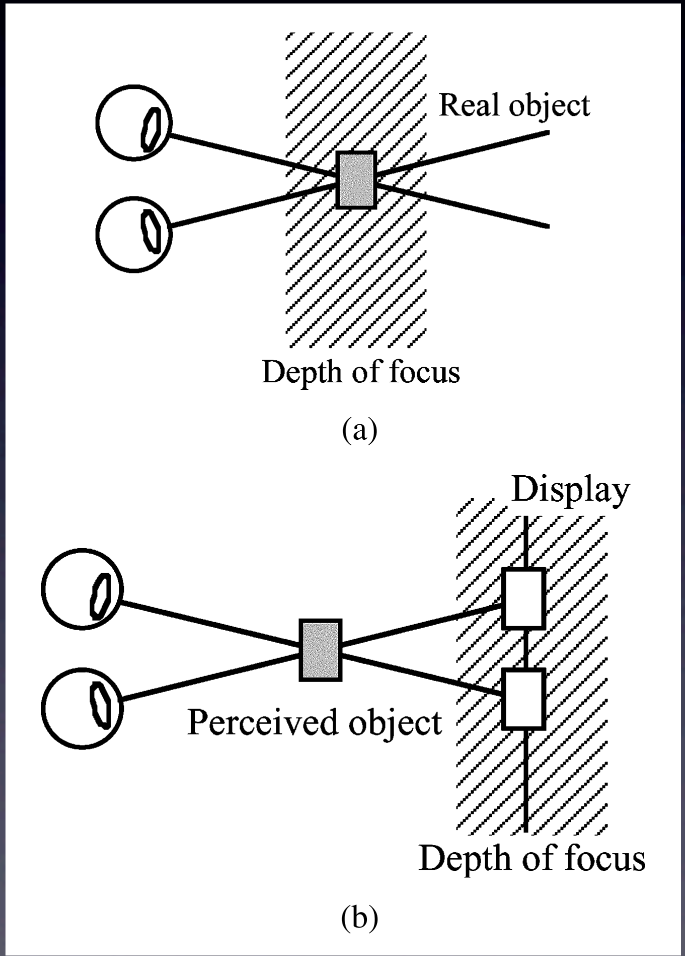
INRIA

Visual fatigue:
accommodation and convergence discrepancy

distance of accommodation
= distance to screen
≠ distance of convergence

Different display
⇒ Different depth of field:

- Human DOF=0.2-0.3D (diopter=1/m)
- 3DTV (3.5m): 2m → 12m
- Movie theater (16m): 4m → infinity



Emoto *et al.* 2005

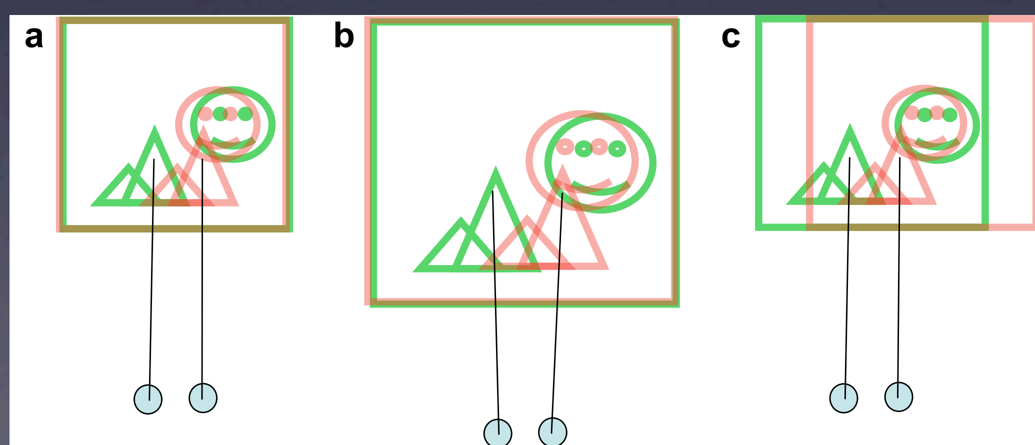
INRIA

Visual fatigue: screen size effects

One 3-D movie, different screens □ risk of divergence

Shifting the images solves divergence issue, but creates other problems:

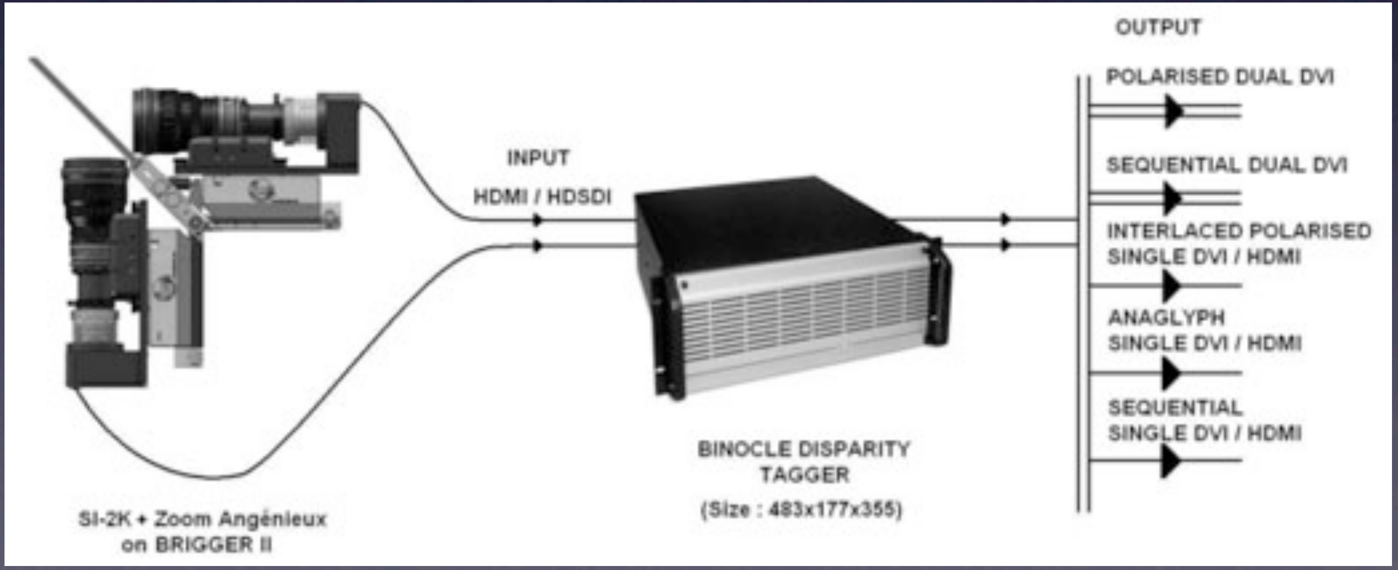
- Breaks the stereoscopic window
- Causes depth distortions



Ukai & Howarth 2008

Correcting geometric differences: the problem

- Mechanics and optics are intrinsically imprecise
- Check that the 3D movie can be comfortably viewed on a given screen (movie theater or 3DTV)
- On output, disparity must be purely horizontal
- **Transform the images to remove geometric differences**



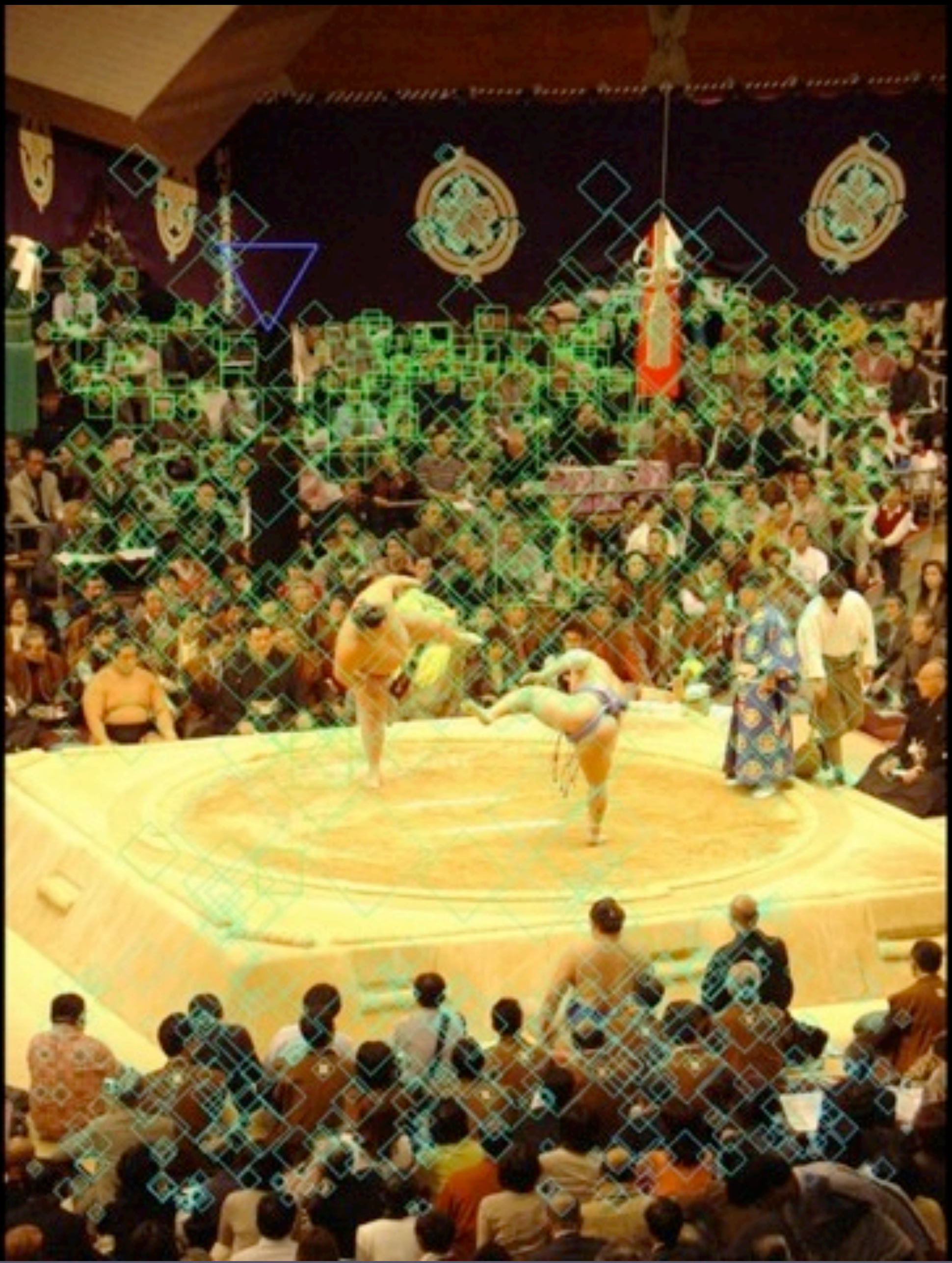
Disparity Tagger: The Binocle / INRIA solution

- Detect remarkable points or regions in both images
- Match these points and regions
- Compute image transformations to remove vertical disparities
- Real-time correction of HD-SDI stereoscopic streams (2 x 1080i60)

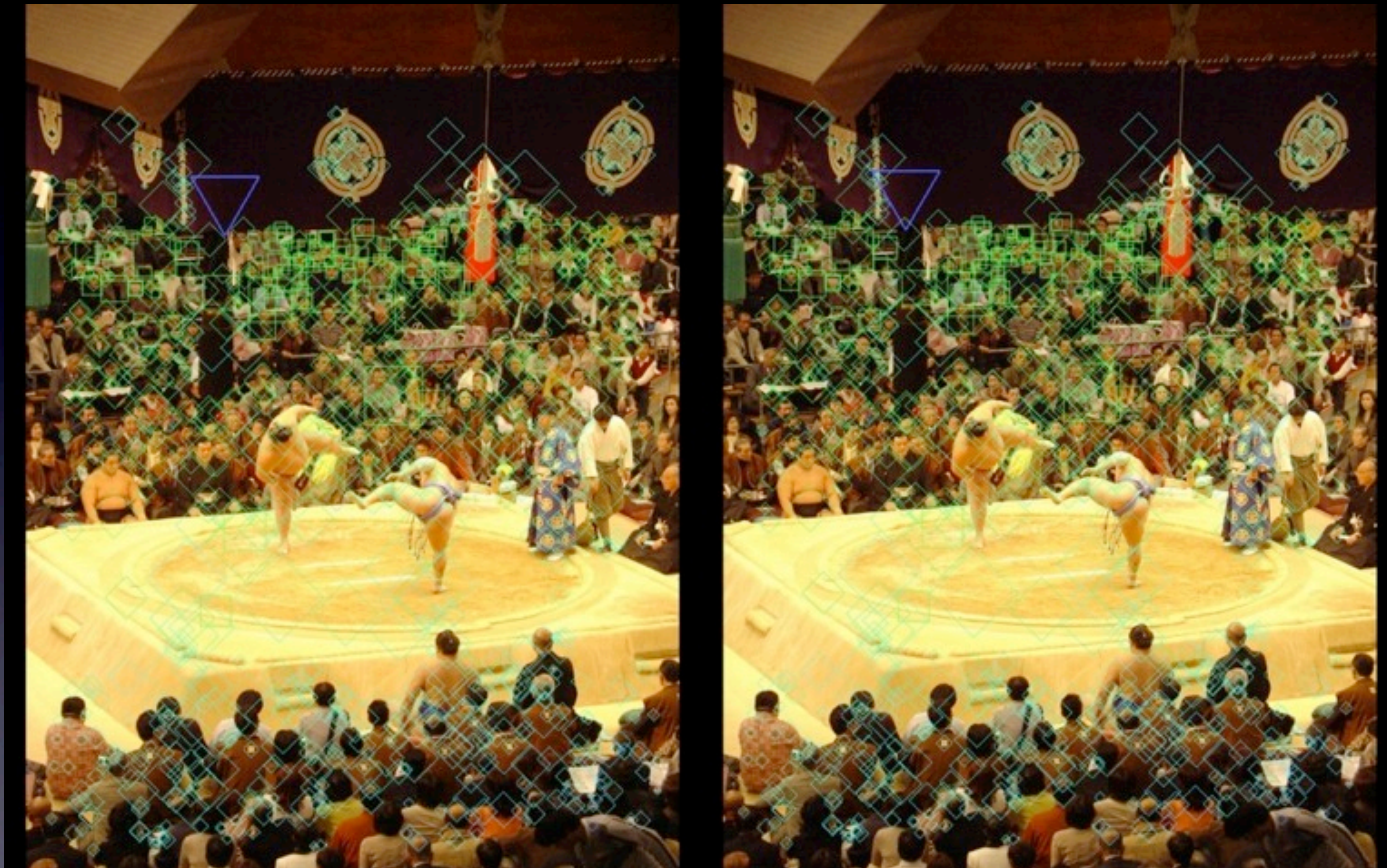
INRIA

Research or Engineering?

- Based on state-of-the-art Computer Vision techniques:
  - SIFT/SURF detector/descriptor + matching
  - F-matrix by RANSAC/PROSAC
  - Stereo pair rectification
- But still hard to implement in practice
  - Must be robust to any kind of images
  - Rectification for cinema imposes constraints (aspect ratio, no black borders)

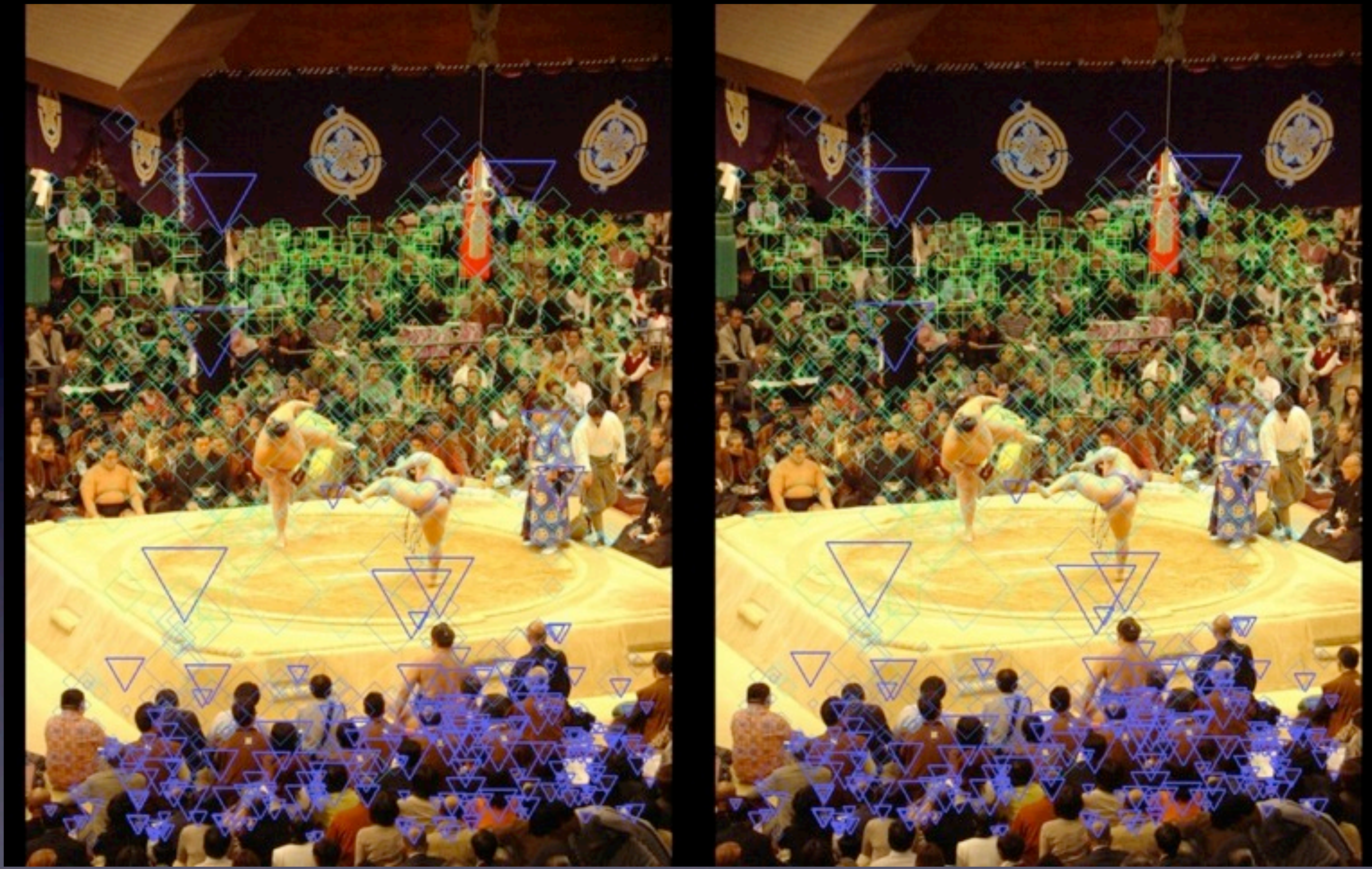


INRIA

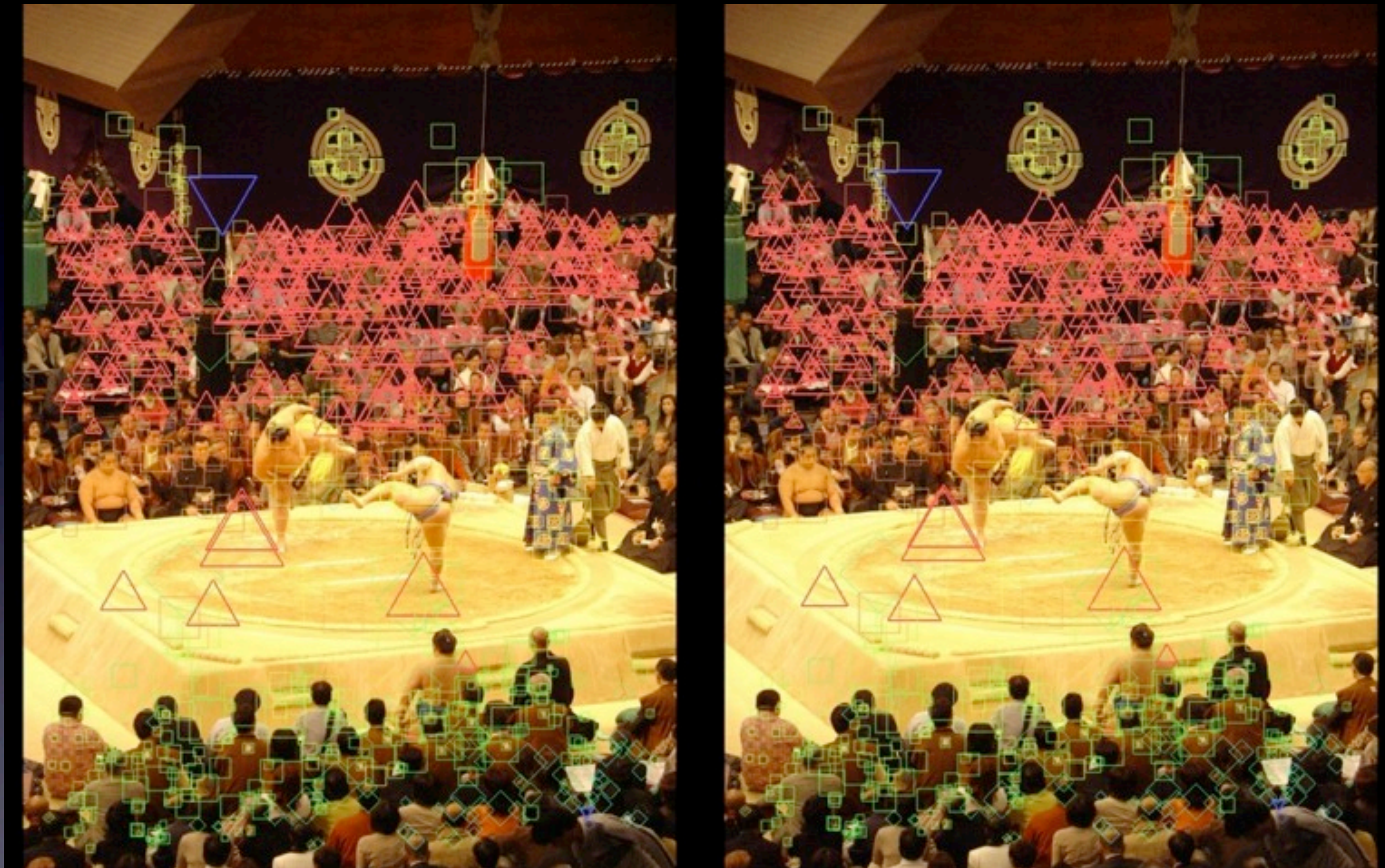


Alerts for a 4m wide screen

INRIA



Alerts for a 10m wide screen: crowd too close!

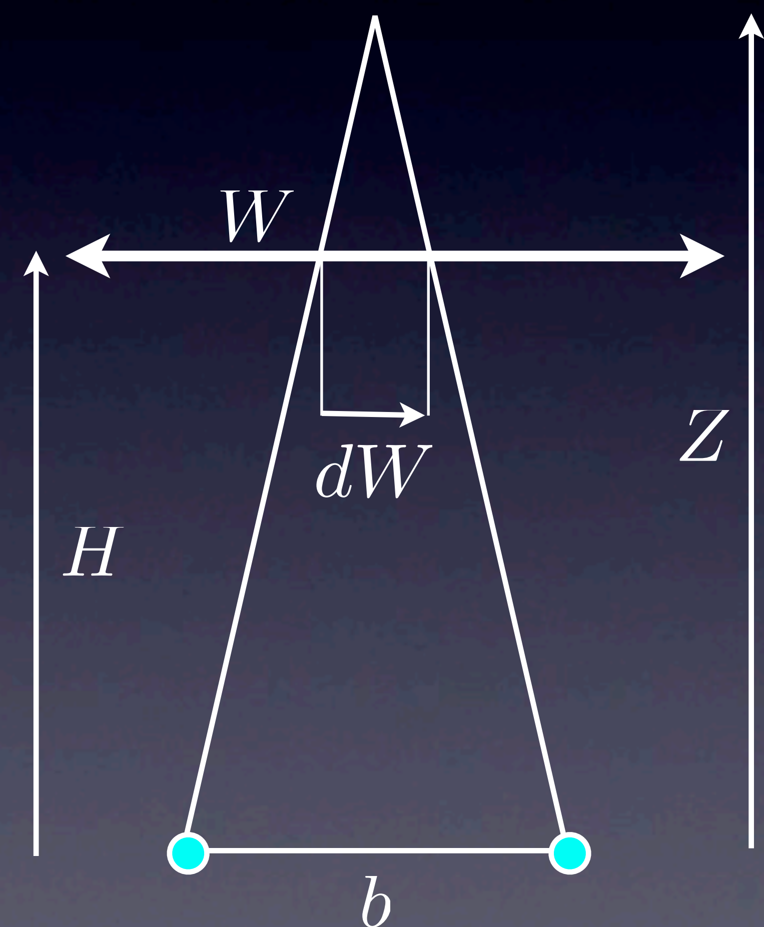


Alerts for a 10m wide screen + shift: divergence!

INRIA

Shooting/viewing geometries

	camera (without primes)	display (with primes)
b	camera interocular	eye interocular
H	convergence distance	screen distance
W	width of convergence plane	screen size
Z	real depth	perceived depth
d	disparity (as a fraction of W)	



INRIA

Depth and disparity

Triangles ABC and ADE are homothetic:

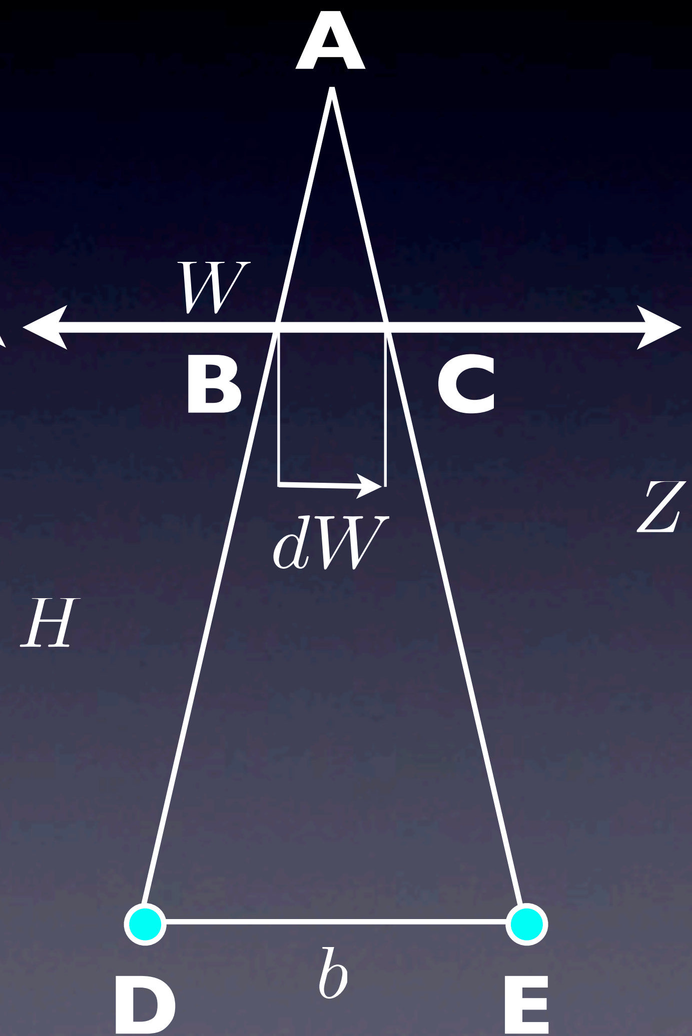
$$\frac{Z-H}{Z} = \frac{dW}{b}$$

which is easily rewritten as

$$d = \boxed{\frac{b}{W}}\frac{Z-H}{Z}$$

or

$$Z = \frac{H}{1-\boxed{\frac{W}{b}}d}$$



Perceived depth

b, W, H, Z : Camera

b', W', H', Z' : Display

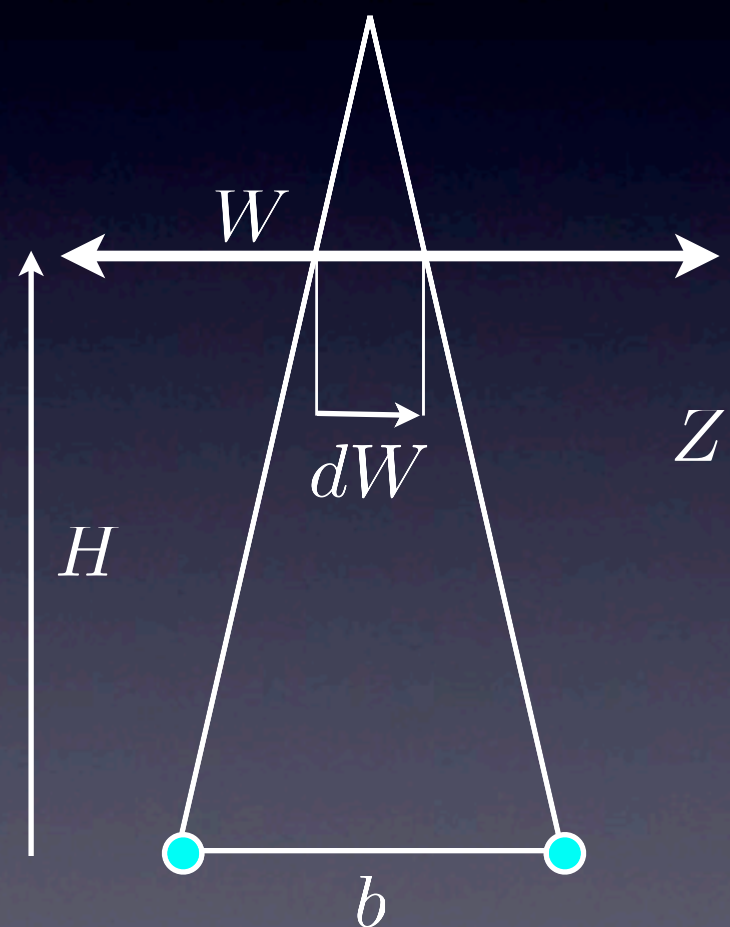
$d' = d$: Disparity (no shift)

a) compute disparity from real depth:

$$d = \frac{b}{W}\frac{Z-H}{Z}$$

b) compute perceived depth from disparity:

$$Z' = \frac{H'}{1 - \frac{W'}{b'}d}$$



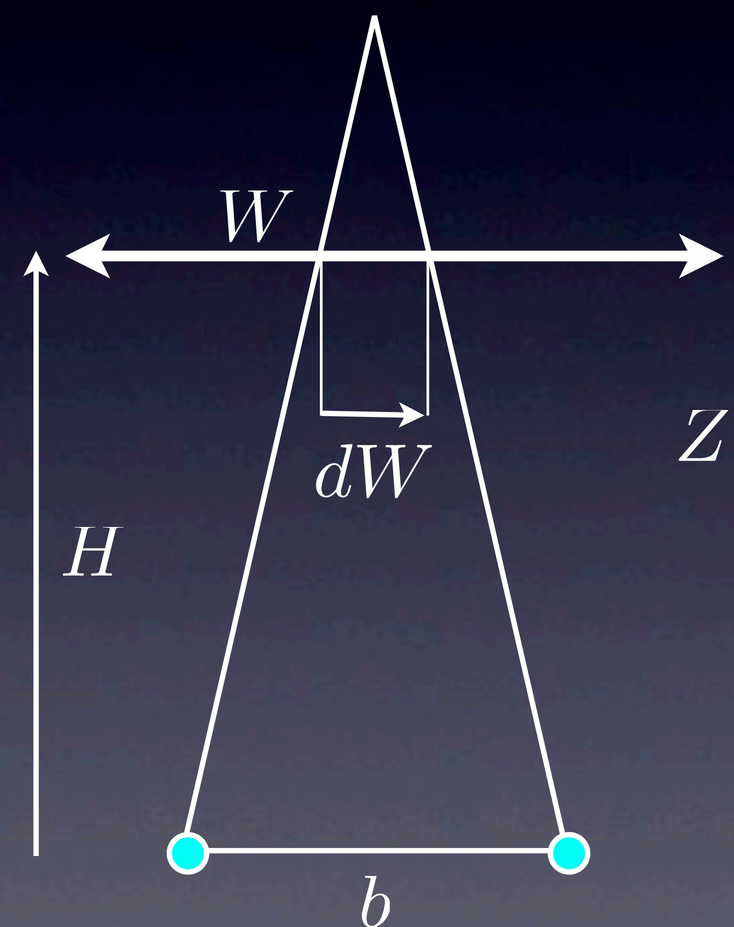
INRIA

Perceived depth (2)

b, W, H, Z : Camera
b', W', H', Z' : Display

c) Finally, eliminate disparity:

$$Z' = \frac{H'}{1 - \frac{W'}{b'} \frac{b}{W} \frac{Z-H}{Z}}$$



INRIA

Perceived vs. real depth

$$Z' = \frac{H'}{1 - \frac{W'}{b'} \frac{b}{W} \frac{Z-H}{Z}}$$

- The relation between Z and Z' is not linear, except if $\frac{W}{b} = \frac{W'}{b'}$, in which case:

$$Z' = Z \frac{H'}{H}$$

- Infinity is perceived at $Z' = \frac{H'}{1 - \frac{W'}{b'} \frac{b}{W}}$

- Divergence happens when Z' becomes negative (divergence at Z=infinity iff $\frac{b'}{W'} < \frac{b}{W}$)

INRIA

Image scale ratio

- enlargement/reduction in image plane (X-Y) of an object at depth Z (disparity d) wrt an object at H (0):

$$\sigma' = \frac{s'}{s} = \frac{H'}{Z'}\frac{Z}{H} = \frac{1 - dW'/b'}{1 - dW/b}.$$

The roundness factor

- Idea: we film a sphere, does it look like an rugby ball or a flat disc?

$$\rho = \frac{\partial Z'}{\partial Z} / \frac{\partial X'}{\partial X} = \frac{\partial Z'}{\partial Z} / \frac{W'/s'}{W/s}$$

- In the screen plane:

$$\rho_{\text{screen}} = \frac{W}{W'} \frac{\partial Z'}{\partial Z}_{(Z=H)} = \frac{b}{H} \frac{H'}{b'}$$

INRIA

The canonical (linear) setup

$$\frac{W}{b} = \frac{W'}{b'}, \qquad Z' = Z\frac{H'}{H}$$

- Keeping proportions, $\rho = 1$:

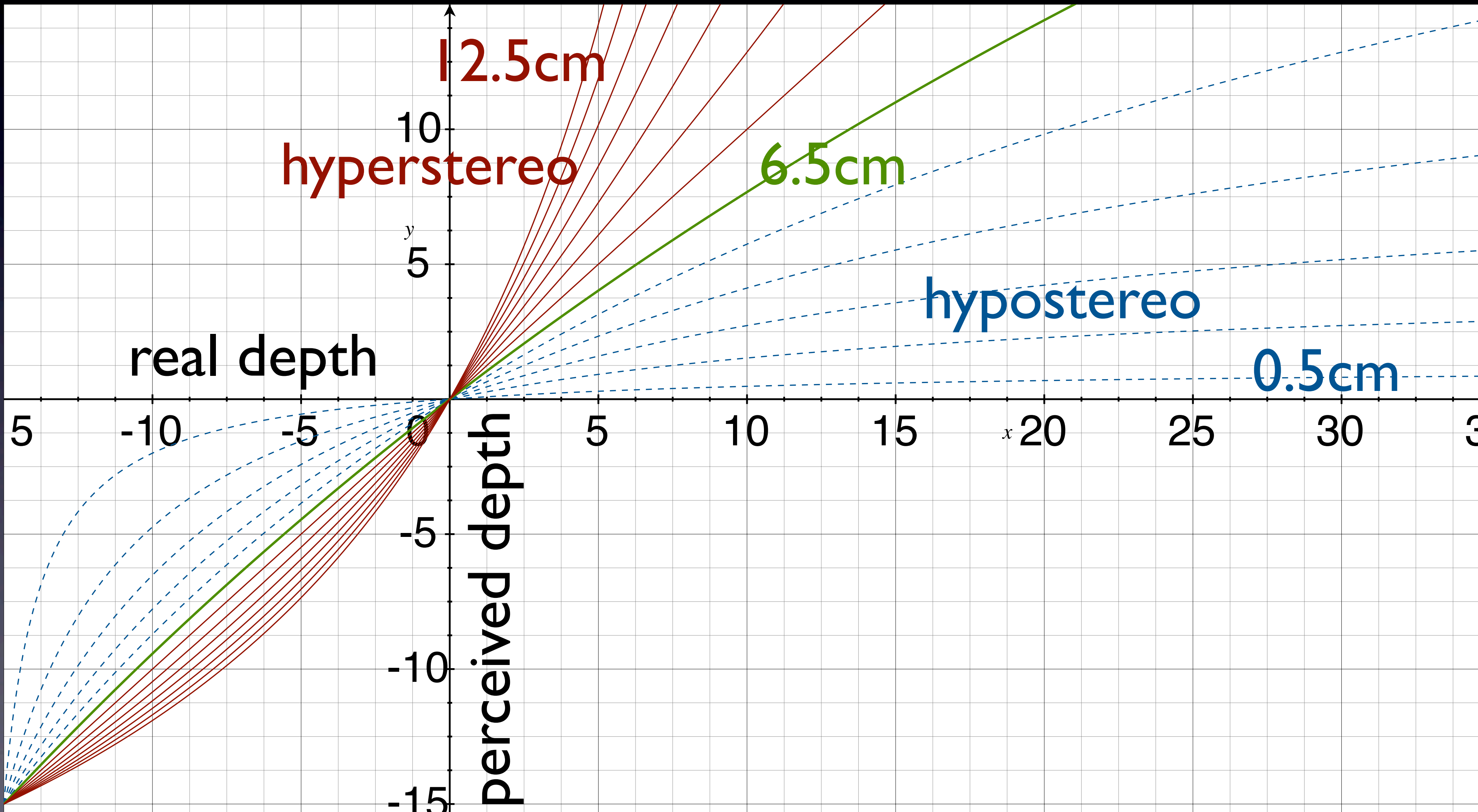
$$\frac{W'}{W} = \frac{H'}{H} \left(= \frac{b'}{b} \right)$$

- The *only* depth-preserving and roundness-preserving setup: Scale factor between filming setup and display setup (fixed FOV, baseline depends on the width of the convergence plane)

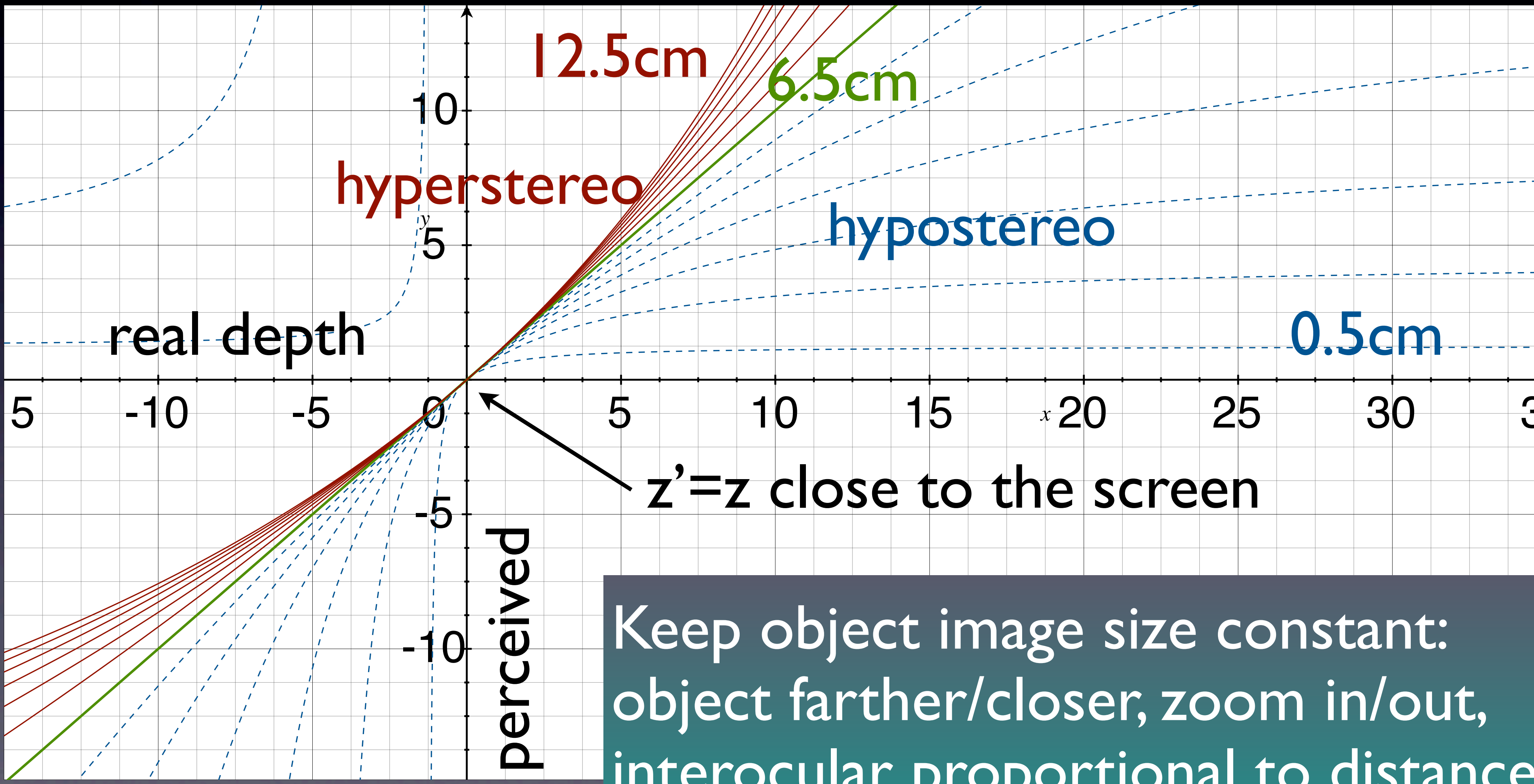
Case study

- b = b’ = 6.5cm
- W = W’ = 10m
- H = H’ = 15m
- no image shift
- depth is measured from plane of convergence / screen, for comparison purposes

Global depth modifications: changing b (camera interocular)



Global depth modifications:
H = αb



Keep object image size constant: object farther/closer, zoom in/out, interocular proportional to distance

Global depth modifications:

- Depth transformations are nonlinear:
  the perceived space is a homographic transform of the real space
- Shooting from farther away while zooming in with a bigger interocular doesn't distort (much) depth:
  That's probably the right way to zoom in - the baseline should be proportional to the convergence distance, but be careful with divergence at Z=infinity!

Global depth modifications: the depth consistency rule

- Roundness of on-screen objects:

$$\rho_{\text{screen}} = \frac{b}{H}\frac{H'}{b'}$$

- $\rho_{\text{screen}} = 1,$ Depth consistency rule:

$$\boxed{\frac{b}{H} = \frac{b'}{H'}}$$

INRIA

Global depth modifications: the depth consistency rule

$$\frac{b}{H} = \frac{b'}{H'}$$

- Screen size does not matter! (except at Z=infinity, where divergence may occur on bigger screens)
- The screen distance dramatically influences the perceived depth, but it's usually constrained by the viewing conditions (movie theater vs. home cinema vs. TV)
- Since b' is fixed, what can we do to enforce the depth consistency rule, i.e. to produce the same 3-D experience in different environments?

INRIA

Perceived depth distortions: summary

- 3D geometry is not distorted if and only if shooting and viewing geometry are the same
  - used for IMAX-3D
  - impossible in real situations (sports...)
  - may break the stereoscopic window
- Objects don’t look «more 3D» on a bigger screen
- Distance is important: «more 3D» if screen farther
- **Novel view synthesis** is the only solution (requires depth map)

Fixing the roundness factor issue using novel view synthesis

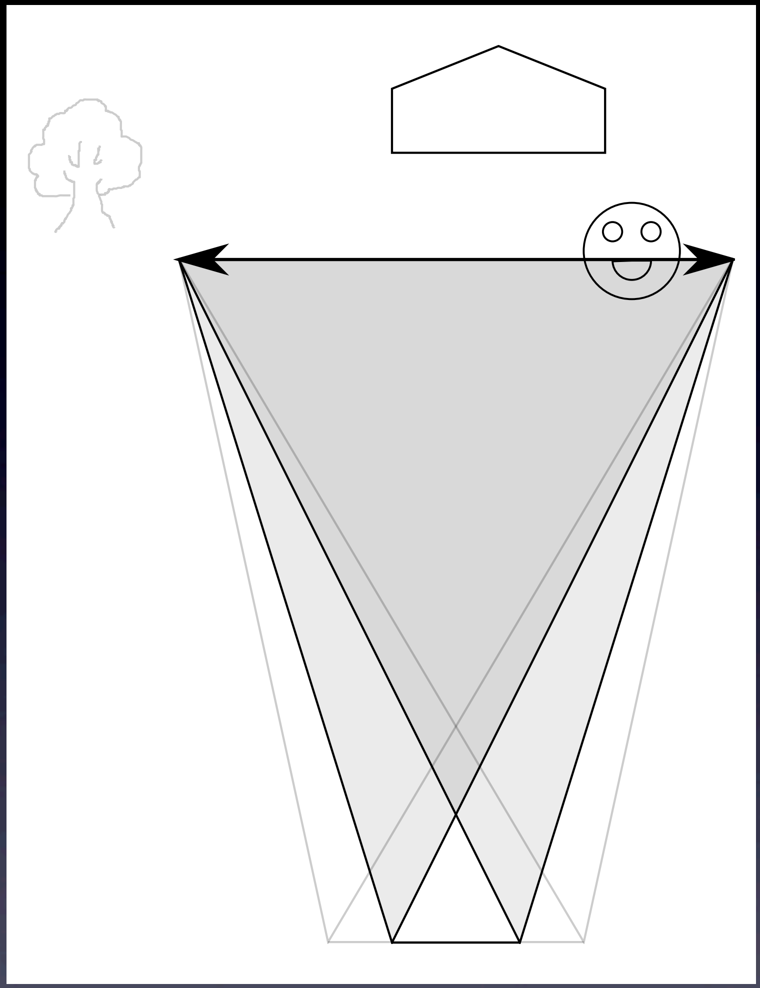
Changing the shooting parameters by post-processing the images to fix the roundness factor:

- **Baseline modification** (or view interpolation) corrects on-screen roundness factor, but distorts off-screen depth and image size, and may cause eye divergence
- **Viewpoint modification** gives perfect depth and roundness factor, but difficult in practice because of large disoccluded areas
- We propose **depth-preserving disparity remapping** - fixes the on-screen roundness factor, no depth distortion, no eye divergence

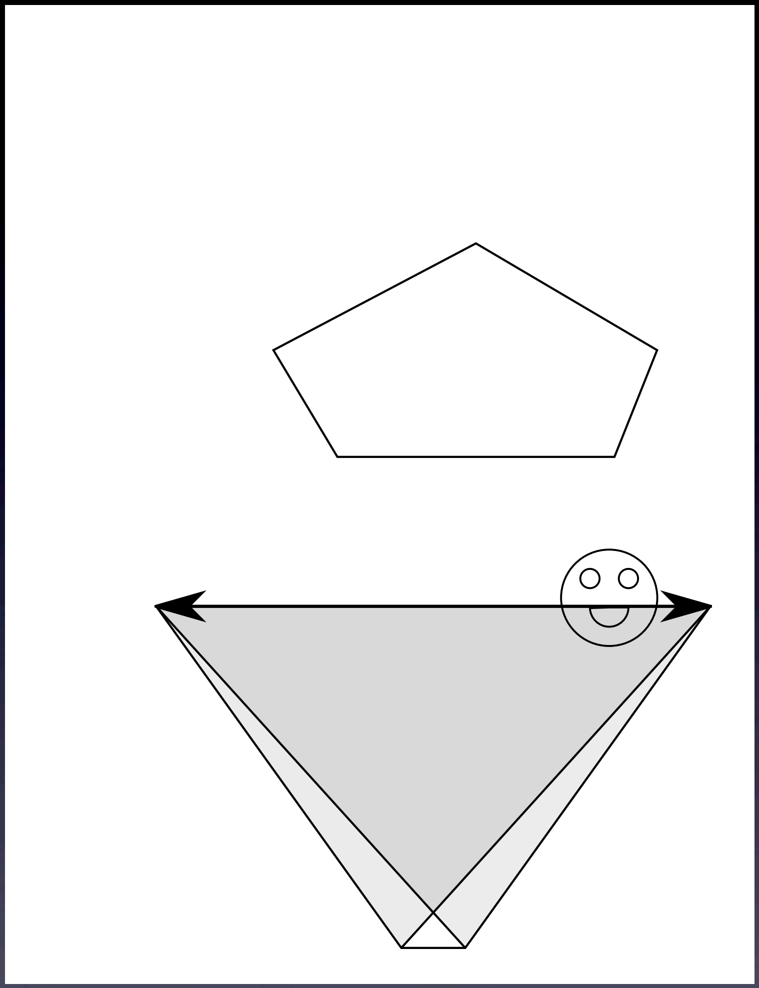
Baseline modification

- synthesized baseline b” computed by setting $\rho_{screen}=1$
- view interpolation (b” < b) or extrapolation (b” > b)
- symmetric or asymmetric (one view can be left untouched)

New view synthesis: baseline modification



Scene geometry



Viewing geometry

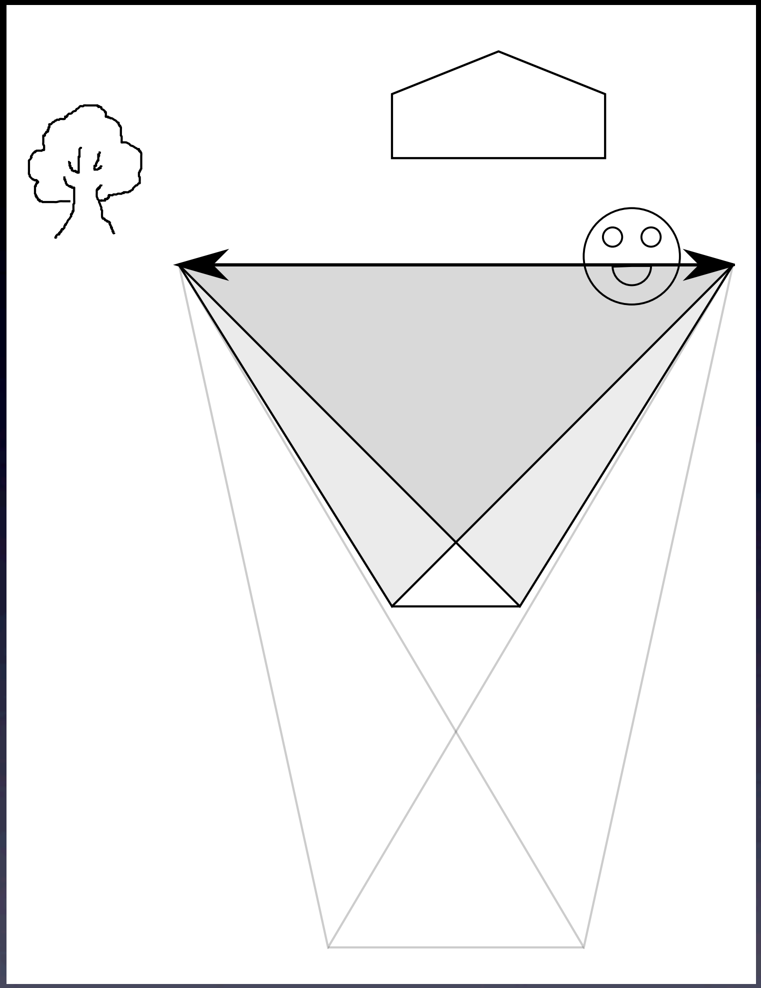
Objects on screen are not distorted, but everything else is **very** distorted! **Divergence** may happen!

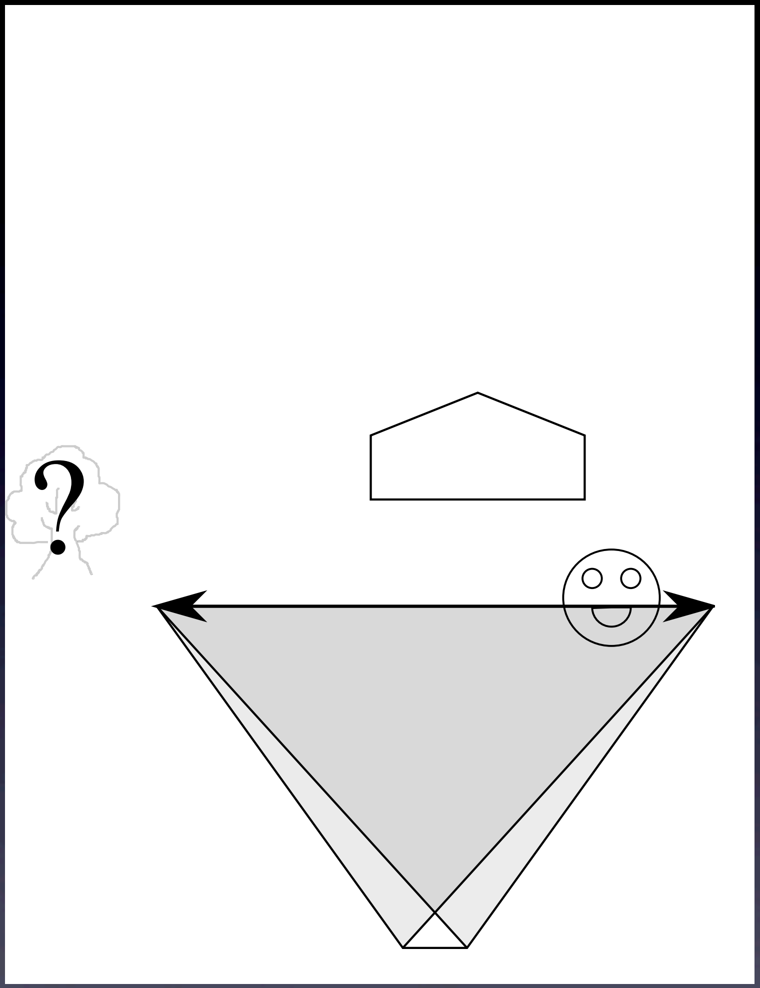
Viewpoint modification

- Synthesized geometry is homothetic to the viewing geometry.
- Both views must be synthesized (**symmetric**)
- Large scene parts that are not visible in the original views may become **disoccluded**

➡ Produces many holes and image artifacts...

New view synthesis: viewpoint modification


Scene geometry



Viewing geometry

No distortion at all, but many objects cannot be seen in the original images... bad solution!

INRIA

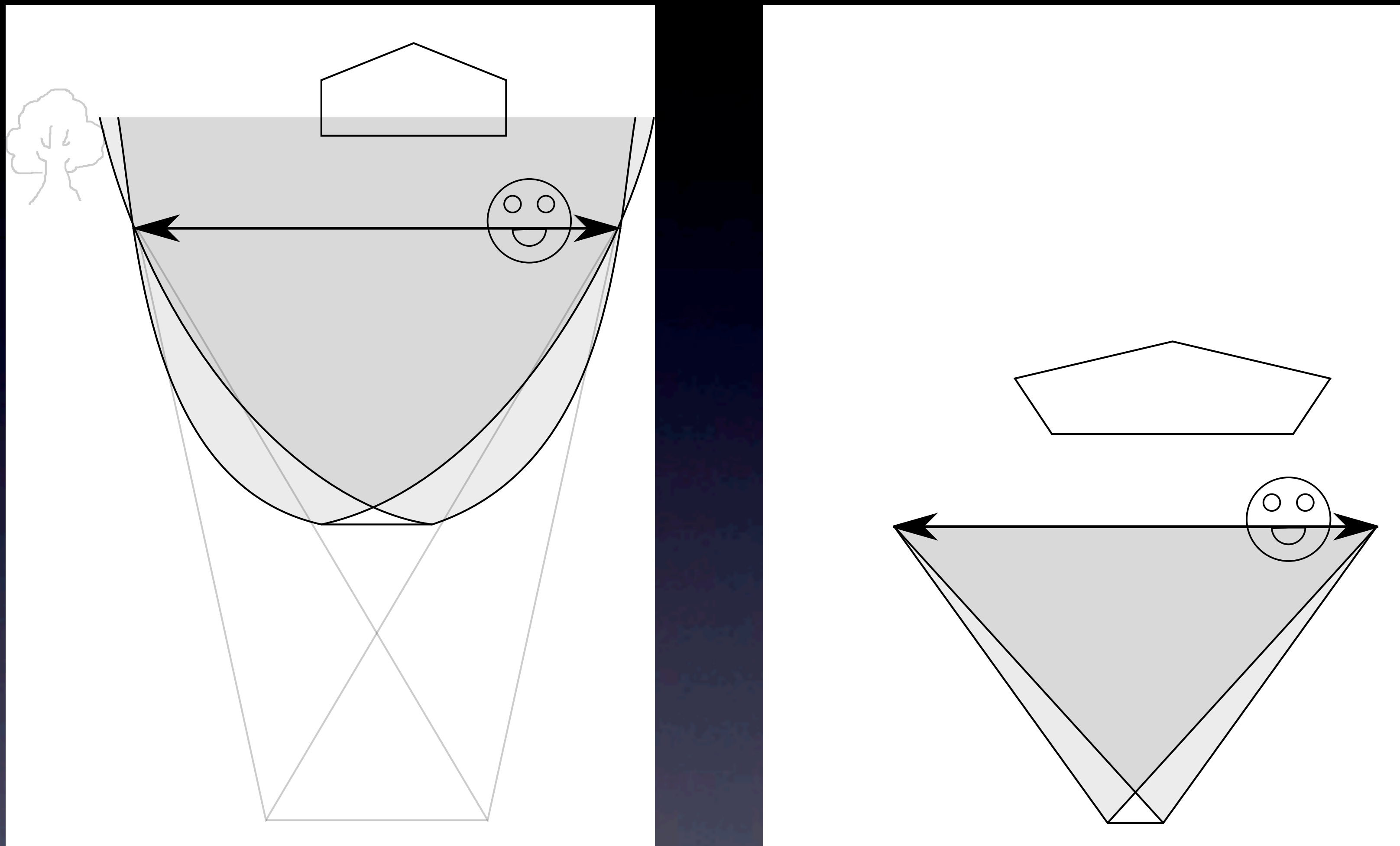
Depth-preserving

- Compute a disparity remapping function d''(d) so that
  $\rho_{screen} = 1$ and $Z' = \alpha Z$

➡ *same disparity as viewpoint modification, but no depth-dependent image scaling.*

- Depth is preserved, but image scale is not respected for off-screen objects
  \- *Just like when zooming with a 2-D camera.*

New view synthesis: disparity remapping

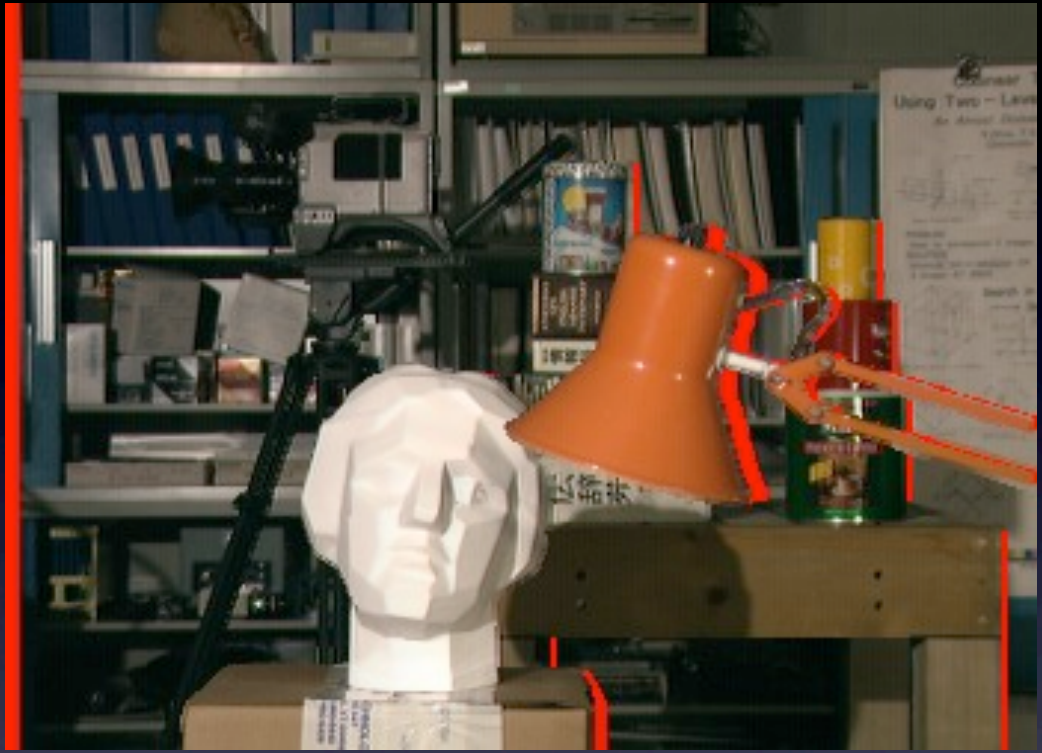


Scene geometry Viewing geometry

Best tradeoff: depth is not distorted, no divergence happens, only apparent width is distorted... like on any 2D image

INRIA

Example showing disoccluded areas



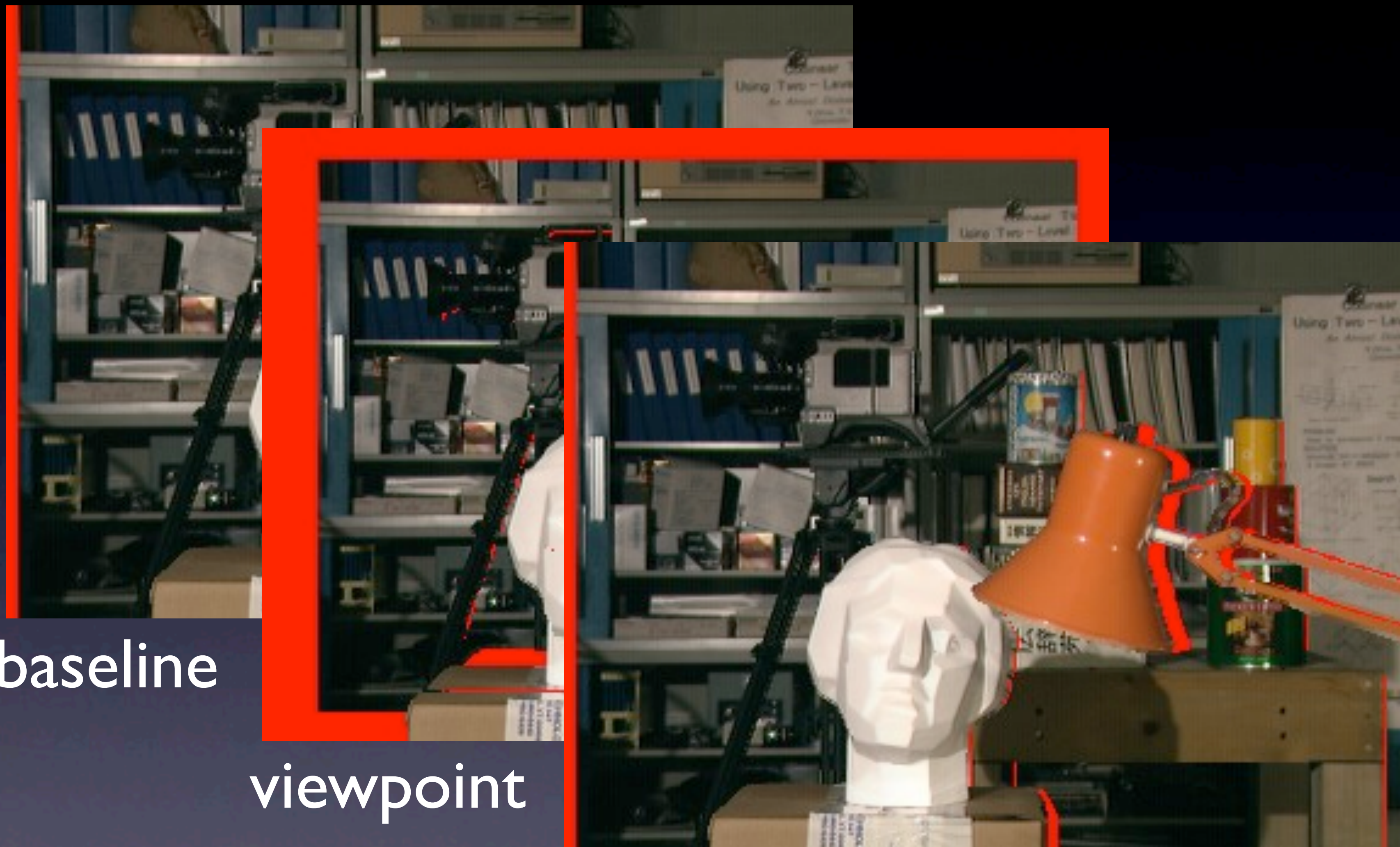
baseline

Example showing disoccluded areas

baseline

viewpoint

Example showing disoccluded areas



baseline

viewpoint

hybrid disparity remapping

Demo:

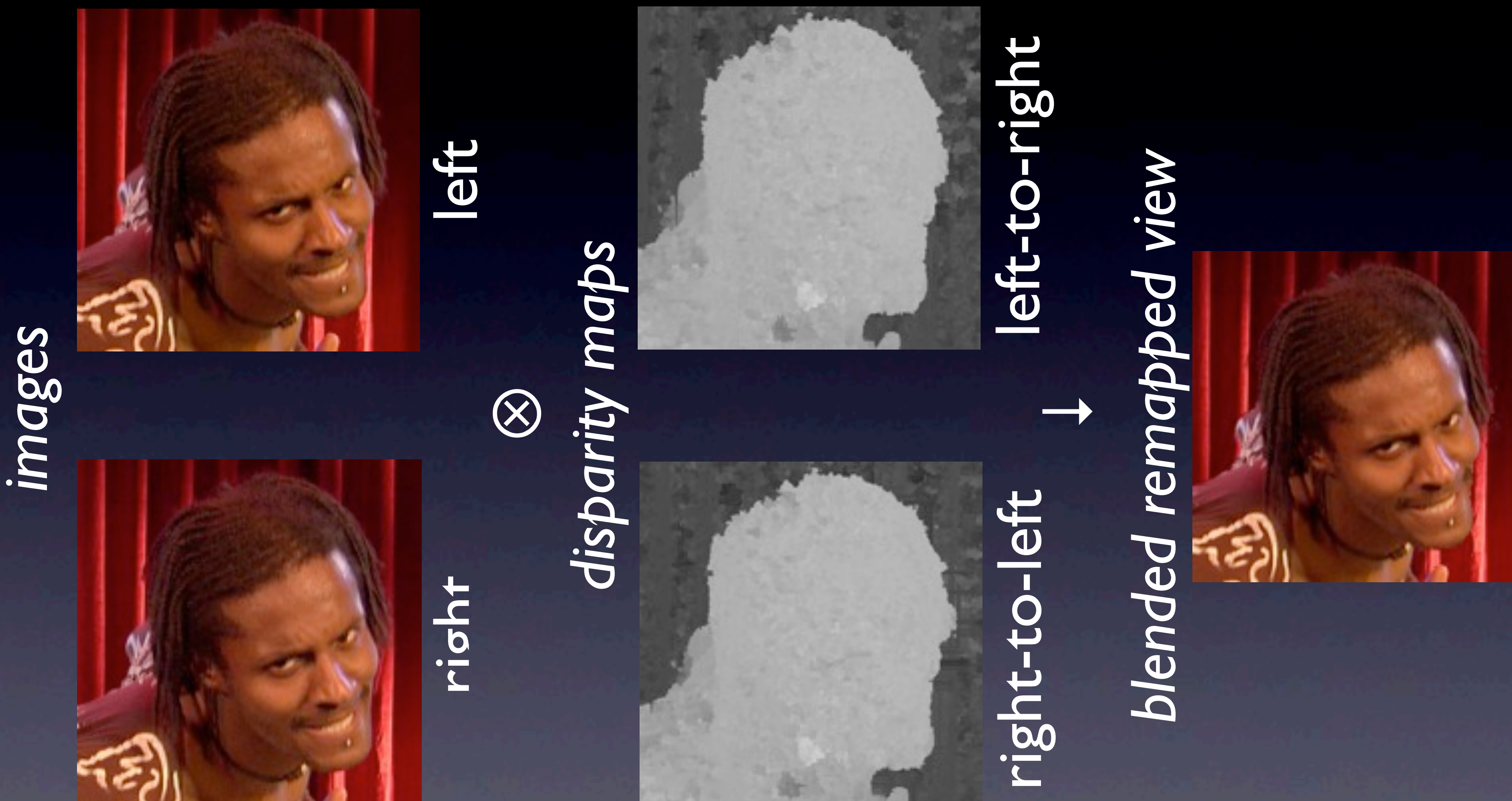
Perceived depth from stereopsis and depth-preserving disparity remapping

Dealing with the vergence-accomodation conflict

- Human depth of field for a screen at 3m is from 1.9m to 7.5m.
- Corresponds to disparities from -3.8cm to 2.6cm.
- In-focus objects should not be displayed out of this range!
- Hybrid disparity remapping can be used to adapt movies so that:
  - The on-screen roundness factor is 1
  - The disparity at infinity is no more than 2.6cm
- Just synthesize views for a screen at the same distance, but 2.5 times wider! (6.5/2.6=2.5)

INRIA

New View Synthesis from Stereo




Artifacts!

Artifacts detection and removal

Our approach:

- Use **asymmetric** synthesis, so that one view keeps the highest possible image quality
- **Detect** artifacts in the synthesized view
- **blur out** the artifacts by anisotropic filtering

Why it should work:

- This locally reduces the high frequency content on artifacts
- The visual system will use *other 3-D cues* from the other (original) view to perceive 3-D in these areas [Stelmach 2000,Seuntiens 2006]
- Temporal consistency *should not be critical* because of low spatial frequency (to be validated)

Detecting and removing artifacts

Comparison of interpolated image with the original images:

- **colors** should be **similar**
- **Laplacian** should be **similar** too: *an edge can not appear!*

We compute a **confidence map** combining both, and use it as the conduction in the Perona-Malik anisotropic diffusion/ blur equation:

$$\frac{\partial I}{\partial t} = \nabla \cdot (c(x,y,t)\nabla I) = c(x,y,t)\Delta I + \nabla c \cdot \nabla I$$

conduction

$c \in [0,1]$

Laplacian

gradients

Interpolated frame

Interpolated frame, artifacts removed

INRIA

Interpolated frame

INRIA

Interpolated frame, artifacts removed

Novel view synthesis: summary

- **Depth map** accuracy is **not crucial**, but the **rendered quality is**
- **Hybrid disparity remapping** of stereoscopic content solves most issues caused by classical novel view synthesis methods.
- **Asymmetric synthesis** helps preserving perceived quality.
- **Artifact removal** is performed by detecting and blurring out artifacts in the synthesized view

Work In Progress:

- Video-rate depth map computation on the GPU with accurate depth boundaries (currently 80ms in OpenCL on Quadro5000)
- Video-rate view synthesis integrated in a stereoscopic player (Bino) from left & right views and left & right disparity maps coded as H.264 videos

INRIA

More work in progress...

- Real-time monitoring:
  - focus and color differences between the cameras
- *Beyond the stereo rig*, novel camera setups:
  - for sports / wildlife (long focal length)
  - for production of glasses-free 3DTV content
- Post-production (with the artist in the loop):
  - stereo compositing, video cut-and-paste using stereo
  - relighting

INRIA

Thank you

Credits:

Yves Pupulin (Binocle) and Bernard Mendiburu

the *Stereocam SuperHD* RIAM project (2005-2008)

the *3DLive* FUI project (2009-2012)
www.3dlive-project.com

INRIA