
Computer Vision
Thursday, August 30

Today

- Course overview
- Requirements, logistics
- Image formation

Introductions

- **Instructor**: Prof. Kristen Grauman
  grauman @ cs
  TAY 4.118, Thurs 2-4 pm
- **TA**: Sudheendra Vijayanarasimhan
  svnaras @ cs
  ENS 31 NQ, Mon/Wed 1-2 pm
- **Class page**: Check for updates to schedule, assignments, etc.

http://www.cs.utexas.edu/~grauman/courses/378/main.htm

Introductions

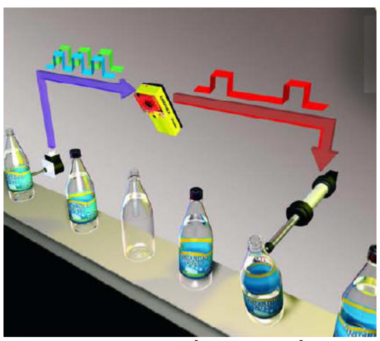
Computer vision

- Automatic understanding of images and video
- Computing properties of the 3D world from visual data
- Algorithms and representations to allow a machine to recognize objects, people, scenes, and activities.

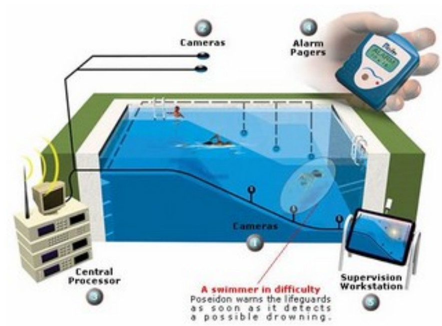
Why vision?

- As image sources multiply, so do applications
  - Relieve humans of boring, easy tasks
  - Enhance human abilities
  - Advance human-computer interaction, visualization
  - Perception for robotics / autonomous agents
- Possible insights into human vision

Some applications



Factory – inspection
(Cognex)



Monitoring for safety
(Poseidon)

Surveillance

Visualization
and tracking

License plate reading

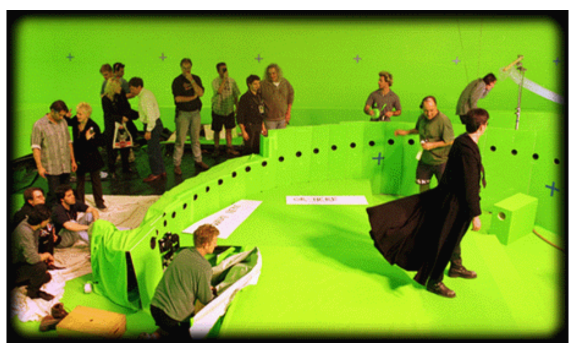
Visualization

Some applications

Autonomous robots

Navigation, driver safety

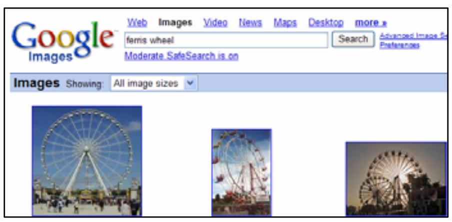
Assistive technology



Visual effects
(the Matrix)

Medical
imaging

Some applications



Multi-modal interfaces

Situated search

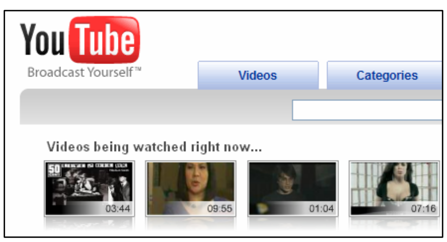
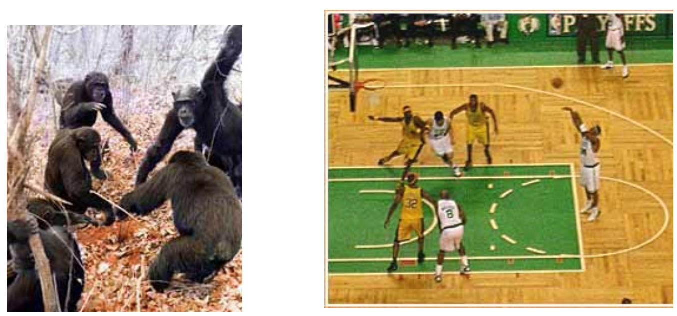


Image and video databases - CBIR



Tracking, activity recognition

Why is vision difficult?

- Ill-posed problem: real world much more complex than what we can measure in images
  - 3D → 2D
- Impossible to literally “invert” image formation process

Challenges: robustness

Illumination

Object pose

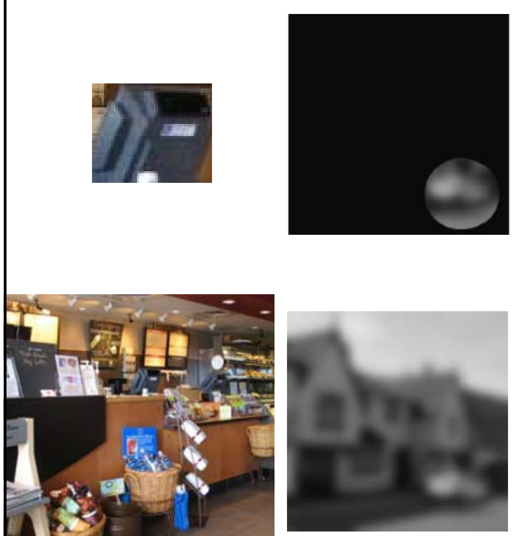
Clutter

Occlusions

Intra-class appearance

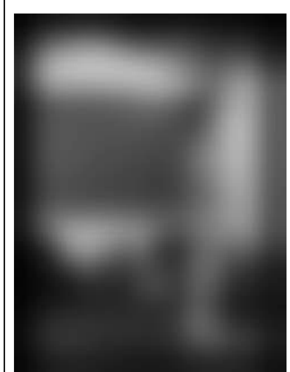
Viewpoint

Challenges: context and human experience



Context cues

Function



Dynamics

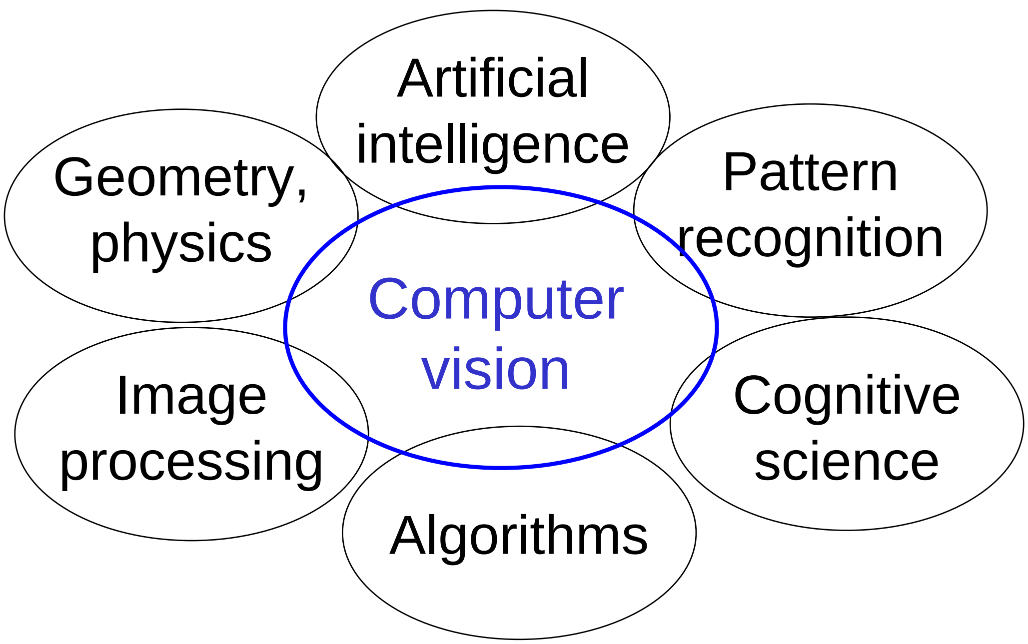
Challenges: complexity

- Thousands to millions of pixels in an image
- 3,000-30,000 human recognizable object categories
- 30+ degrees of freedom in the pose of articulated objects (humans)
- Billions of images indexed by Google Image Search
- 18 billion+ prints produced from digital camera images in 2004
- 295.5 million camera phones sold in 2005
- About half of the cerebral cortex in primates is devoted to processing visual information [Felleman and van Essen 1991]

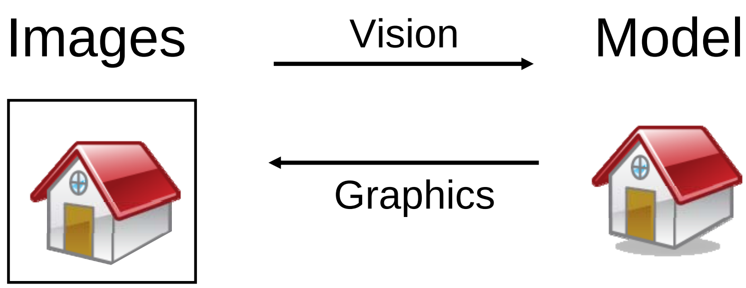
Why is vision difficult?

- Ill-posed problem: real world much more complex than what we can measure in images
  - 3D → 2D
- Not possible to “invert” image formation process
- ***Generally requires assumptions, constraints; exploitation of domain-specific knowledge***

Related disciplines



Vision and graphics



Inverse problems: analysis and synthesis.

Research problems vs. application areas

- Feature detection
- Contour representation
- Segmentation
- Stereo vision
- Shape modeling
- Color vision
- Motion analysis
- Invariants
- Uncalibrated, self-calibrating systems
- Object detection
- Object recognition

- Industrial inspection and quality control
- Reverse engineering
- Surveillance and security
- Face, gesture recognition
- Road monitoring
- Autonomous vehicles
- Military applications
- Medical image analysis
- Image databases
- Virtual reality

List from [Trucco & Verri 1998]

Goals of this course

- Introduction to primary topics
- Hands-on experience with algorithms
- Views of vision as a research area

Topics overview

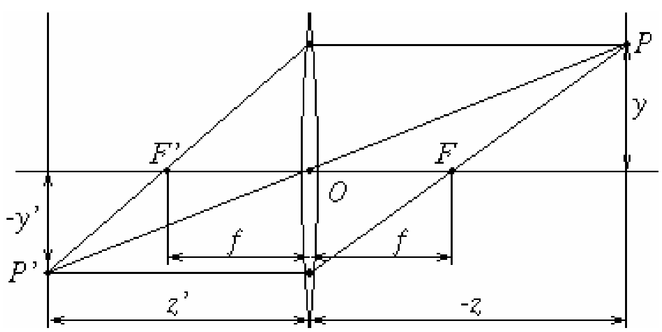
- Image formation, cameras
- Color
- Features
- Grouping
- Multiple views
- Recognition and learning
- Motion and tracking

We will not cover (extensively)

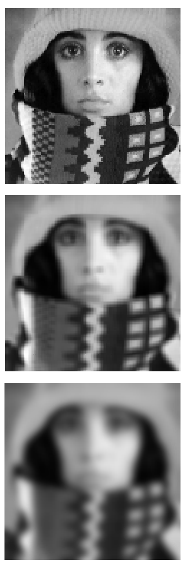
- Image processing
- Human visual system
- Particular machine vision systems or applications

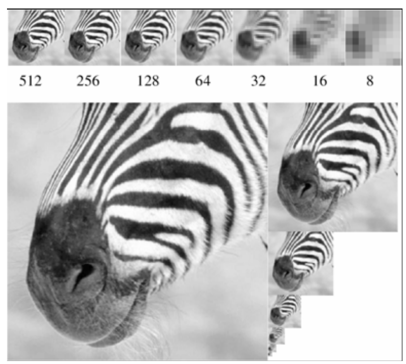
Image formation

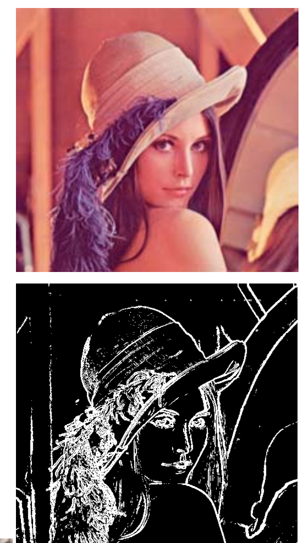
- Inverse process of vision: how does light in 3d world project to form 2d images?



Features and filters

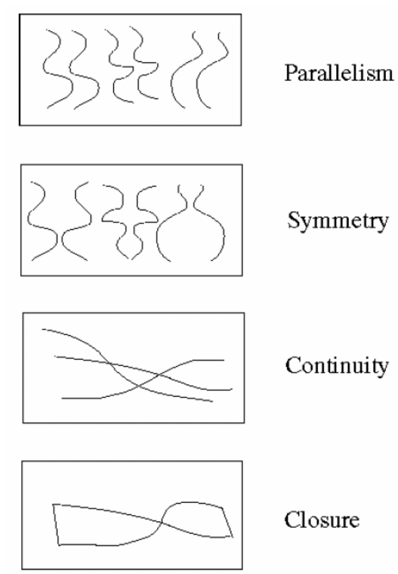


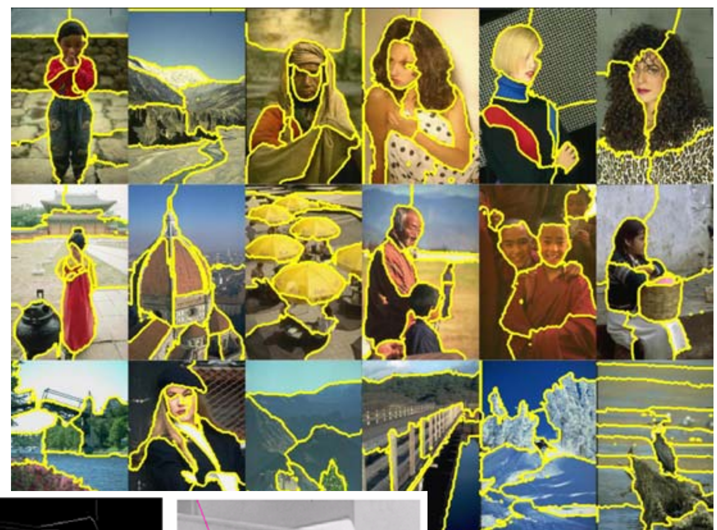




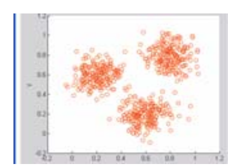
Transforming and describing images; textures and colors

Grouping



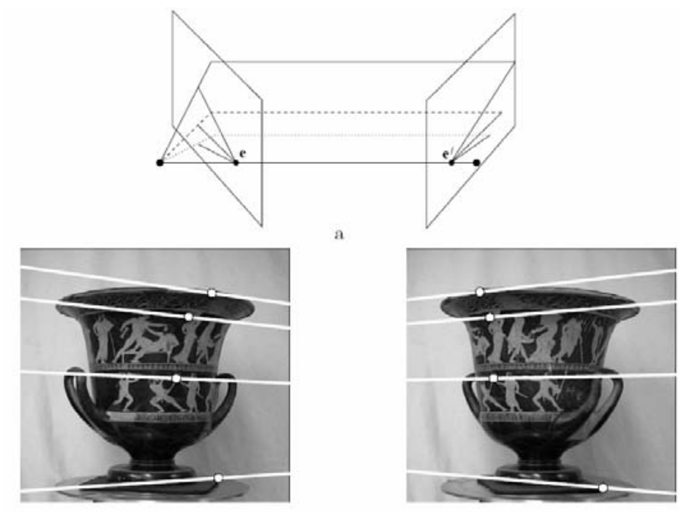


[fig from Shi et al]

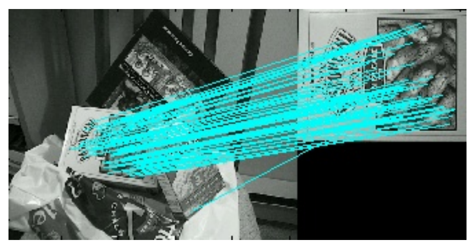


Clustering, segmentation, fitting; what parts belong together?

Multiple views

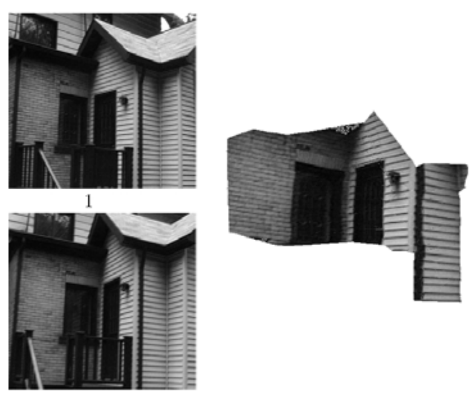


Hartley and Zisserman



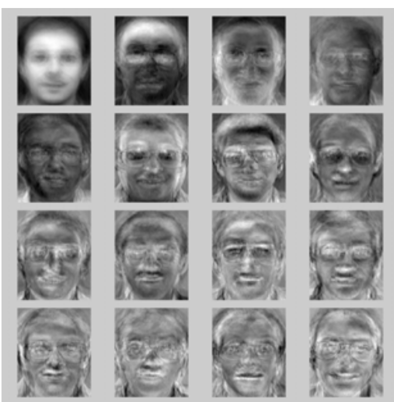
Lowe

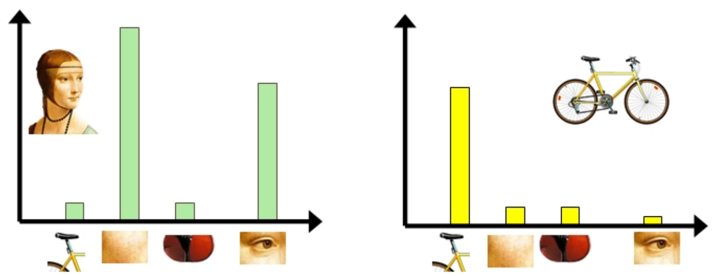
Multi-view geometry and matching, stereo



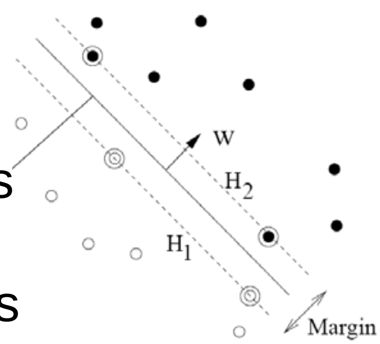
Tomasi and Kanade

Recognition and learning



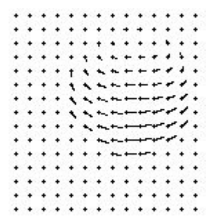


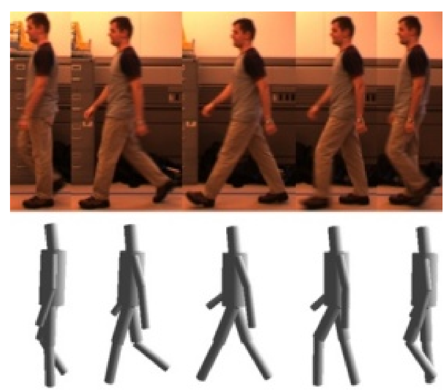
Shape matching, recognizing objects and categories, learning techniques



Motion and tracking

Tracking objects, video analysis, low level motion





Tomas Izo

Requirements

- Biweekly (approx) problem sets
  - Concept questions
  - Implementation problems
- Two exams, midterm and final
- Current events (optional)

In addition, for graduate students:

- Research paper summary and review
- Implementation extension

Grading policy

Final grade breakdown:

- Problem sets (50%)
- Midterm quiz (15%)
- Final exam (20%)
- Class participation (15%)

Due dates

- Assignments due before class starts on due date
- Lose half of possible remaining credit each day late
- Three free late days, total

Collaboration policy

You are welcome to discuss problem sets, but all responses and code must be written individually.

Students submitting solutions found to be identical or substantially similar (due to inappropriate collaboration) risk failing the course.

Current events (optional)

- Any vision-related piece of news; may revolve around policy, editorial, technology, new product, …
- Brief overview to the class
- Must be current
- No ads

- Email relevant links or information to TA

Paper review guidelines

- Thorough summary in your own words
- Main contribution
- Strengths? Weaknesses?
- How convincing are the experiments? Suggestions to improve them?
- Extensions?
- 4 pages max

- **May require reading additional references**

Miscellaneous

- Check class website
- Make sure you get on class mailing list
- No laptops in class please

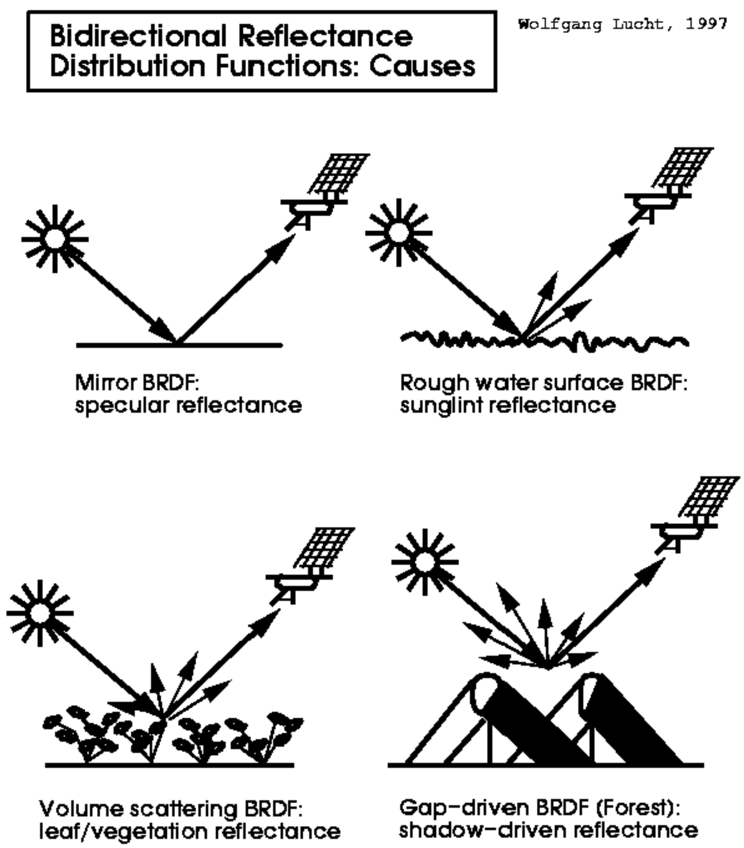
- Feedback welcome and useful

Image formation

- How are objects in the world captured in an image?

Physical parameters of image formation

- Photometric
  - Type, direction, intensity of light reaching sensor
  - Surfaces' reflectance properties
- Optical
  - Sensor's lens type
  - focal length, field of view, aperture
- Geometric
  - Type of projection
  - Camera pose
  - Perspective distortions



Radiometry

- Images formed depend on amount of light from light sources and surface reflectance properties (See F&P Ch 4)

Light source direction

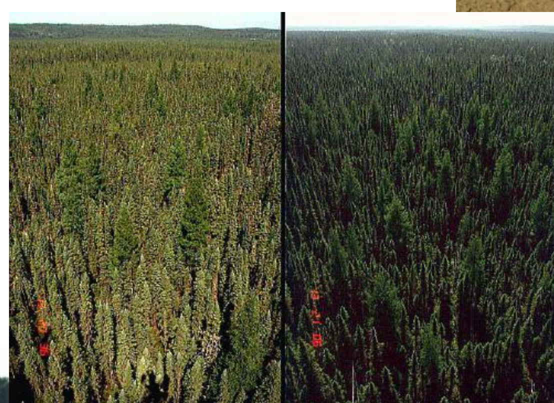
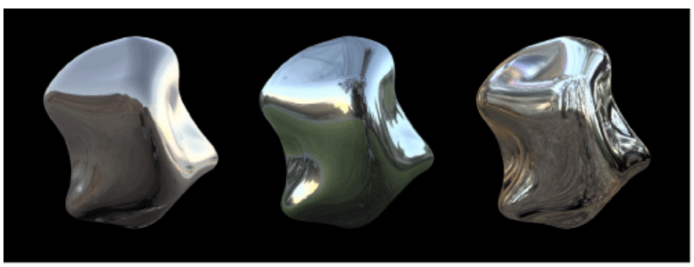


Image credit: Don Deering

Surface reflectance properties



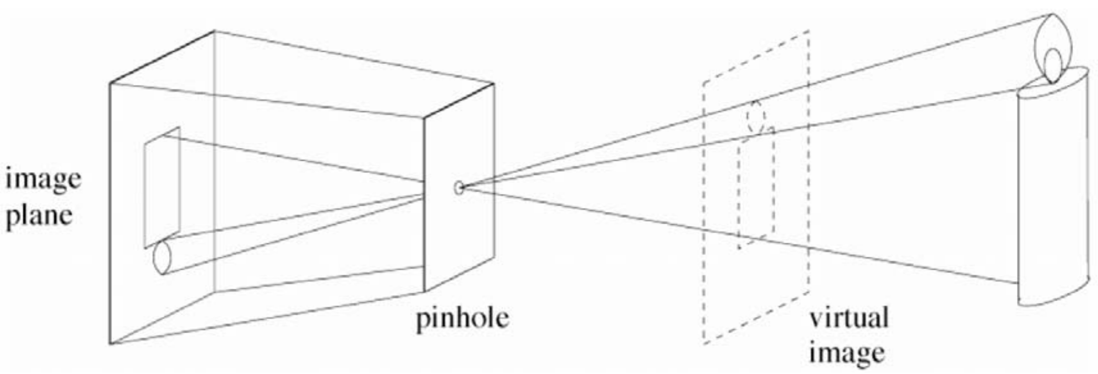
Specular

[fig from Fleming, Torralba, & Adelson, 2004]

Lambertian

Perspective projection

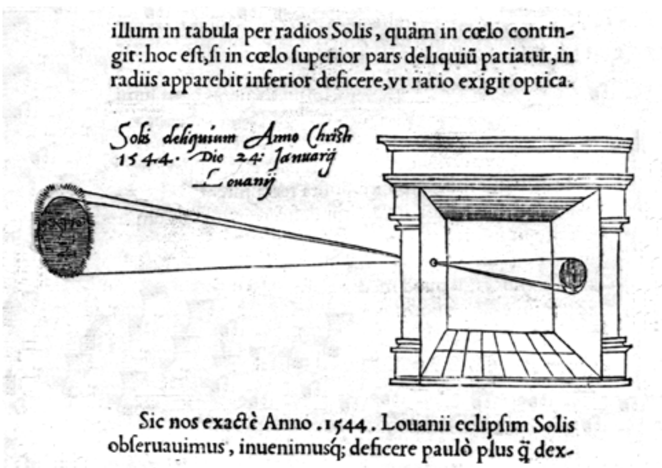
- Pinhole camera: simple model to approximate imaging process



If we treat pinhole as a point, only one ray from any given point can enter the camera

Forsyth and Ponce

Camera obscura

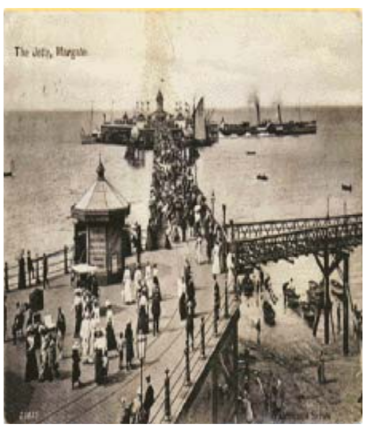


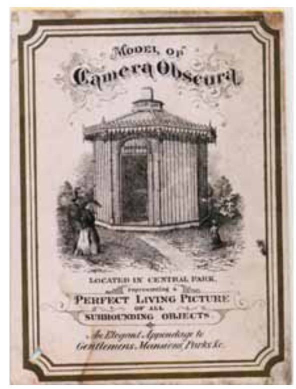
In Latin, means
'dark room'

"**Reinerus Gemma-Frisius**, observed an eclipse of the sun at Louvain on January 24, 1544, and later he used this illustration of the event in his book <u>De Radio Astronomica et Geometrica</u>, 1545. It is thought to be the first published illustration of a camera obscura..."
Hammond, John H., <u>The Camera Obscura, A Chronicle</u>

http://www.acmi.net.au/AIC/CAMERA_OBSCURA.html

Camera obscura





Jetty at Margate England, 1898.

Around 1870s

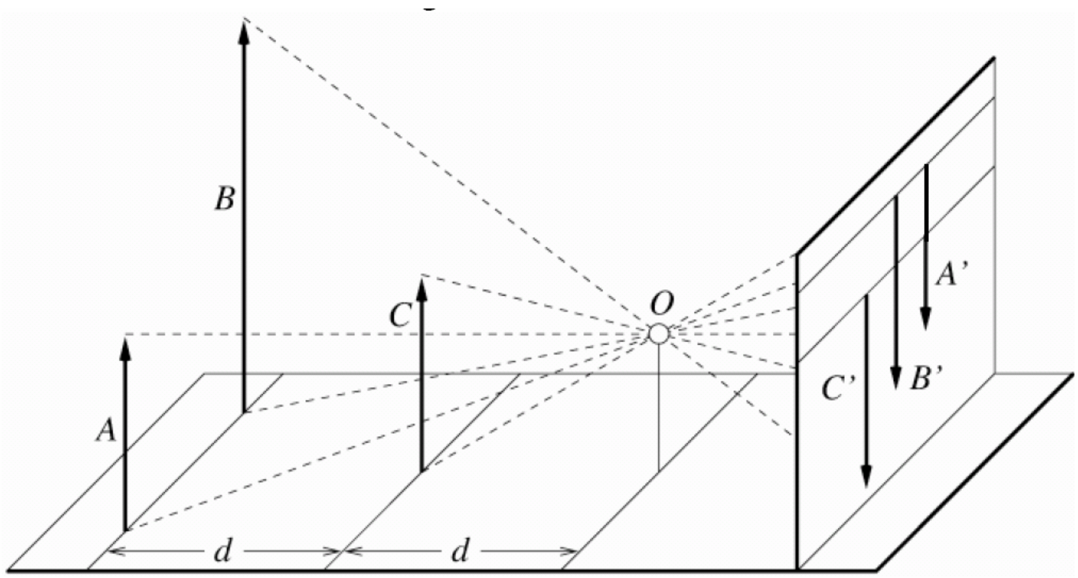
An attraction in the late 19th century

http://brightbytes.com/cosite/collection2.html

Adapted from R. Duraiswami

Perspective effects

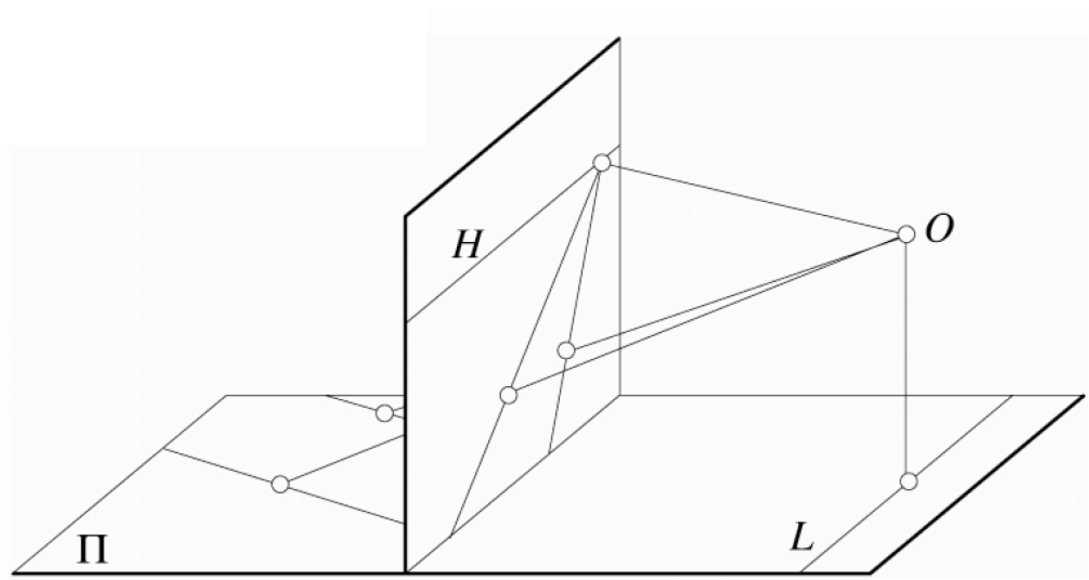
- Far away objects appear smaller



Forsyth and Ponce

Perspective effects

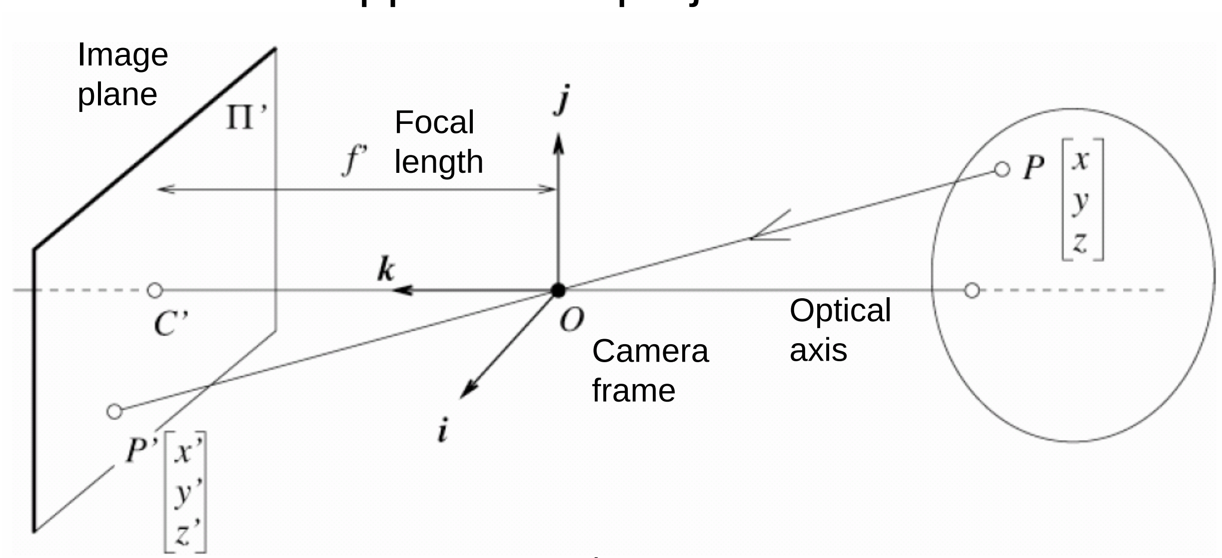
- Parallel lines in the scene intersect in the image



Forsyth and Ponce

Perspective projection equations

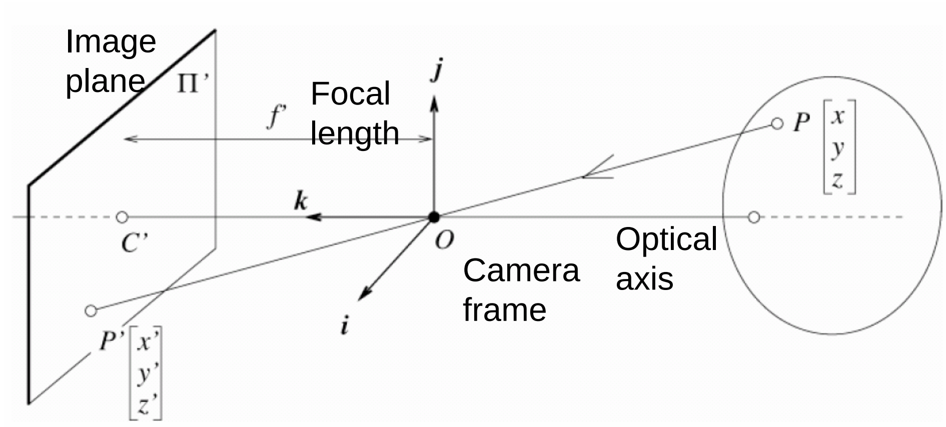
- 3d world mapped to 2d projection



Board

Forsyth and Ponce

Perspective projection equations



$$(x, y, z) \rightarrow (f'\frac{x}{z}, f'\frac{y}{z})$$

Scene point → Image coordinates

Non-linear

Forsyth and Ponce

Projection properties

- Many-to-one: any points along same ray map to same point in image
- Points → points
- Lines → lines (collinearity preserved)
- Distances and angles are **not** preserved
- Degenerate cases:
  – Line through focal point projects to a point.
  – Plane through focal point projects to line
  – Plane perpendicular to image plane projects to part of the image.

Perspective and art

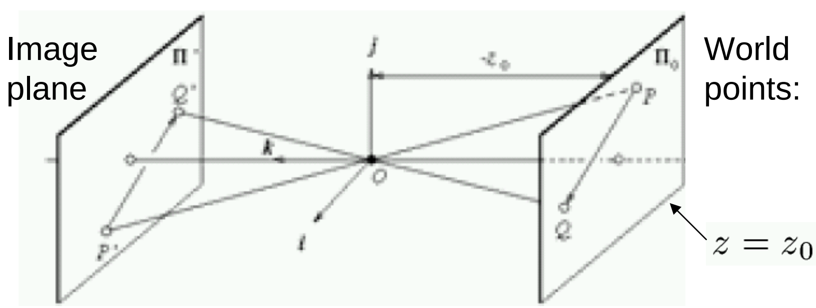
- Use of correct perspective projection indicated in 1st century B.C. frescoes
- Skill resurfaces in Renaissance: artists develop systematic methods to determine perspective projection (around 1480-1515)

Raphael

Durer, 1525

Weak perspective

- Approximation: treat magnification as constant
- Assumes scene depth << average distance to camera
- Makes perspective equations linear

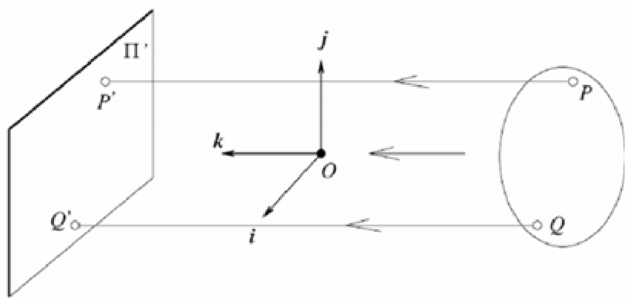


$$x' = f\frac{x}{z} \approx \frac{f}{z_0}x$$

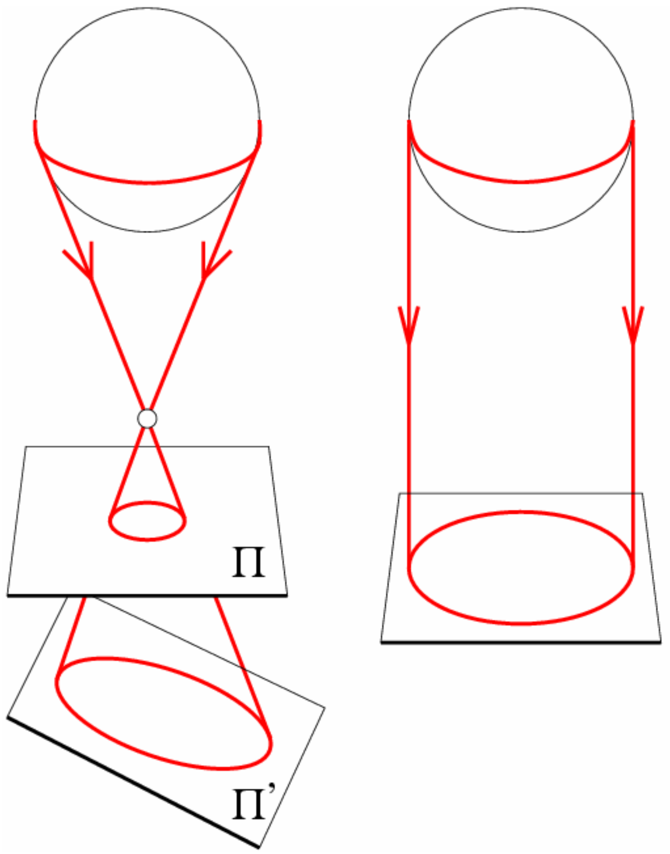
$$y' = f\frac{y}{z} \approx \frac{f}{z_0}y$$

Orthographic projection

- Given camera at constant distance from scene
- World points projected along rays parallel to optical access
- Limit of perspective projection as $f \rightarrow \infty$



$$x' = x$$
$$y' = y$$



Planar pinhole perspective

Orthographic projection

From M. Pollefeys

Which projection model?

- Weak perspective:
  - Accurate for small, distant objects; recognition
  - Linear projection equations - simplifies math
- Pinhole perspective:
  - More accurate but more complex
  - Structure from motion

Pinhole size / aperture

Fig. 5.96 The pinhole camera. Note the variation in image clarity as the hole diameter decreases. [Photos courtesy Dr. N. Joel, UNESCO.]

Larger

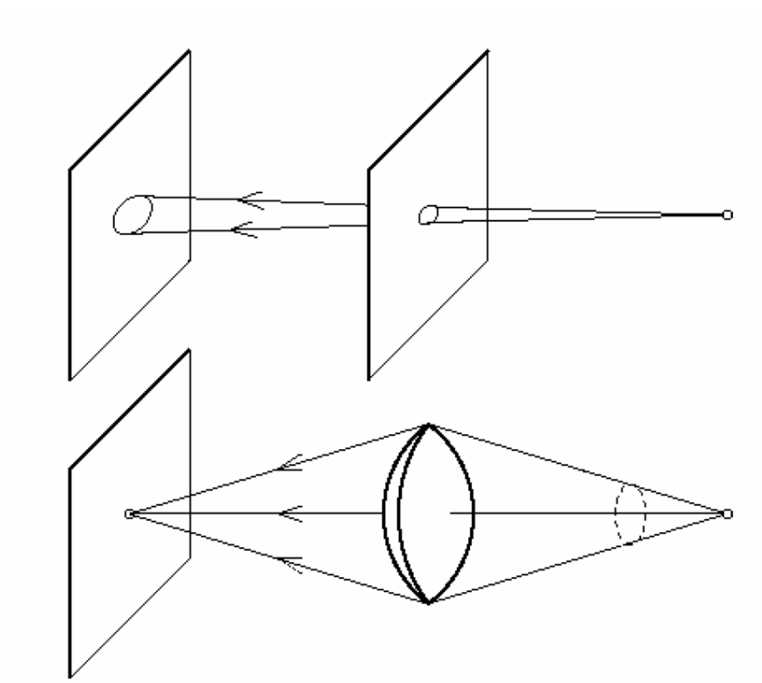
Brighter, blurrier

Dimmer, more focus

Dimmer, blur from defraction

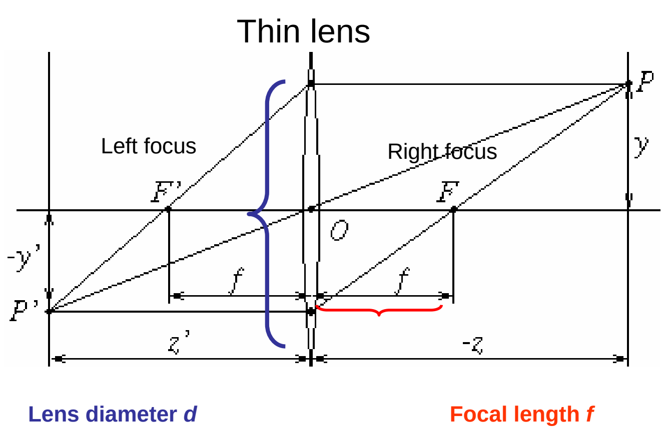
Smaller

Pinhole vs. lens



Cameras with lenses

- Gather more light, while keeping focus; make pinhole perspective projection practical



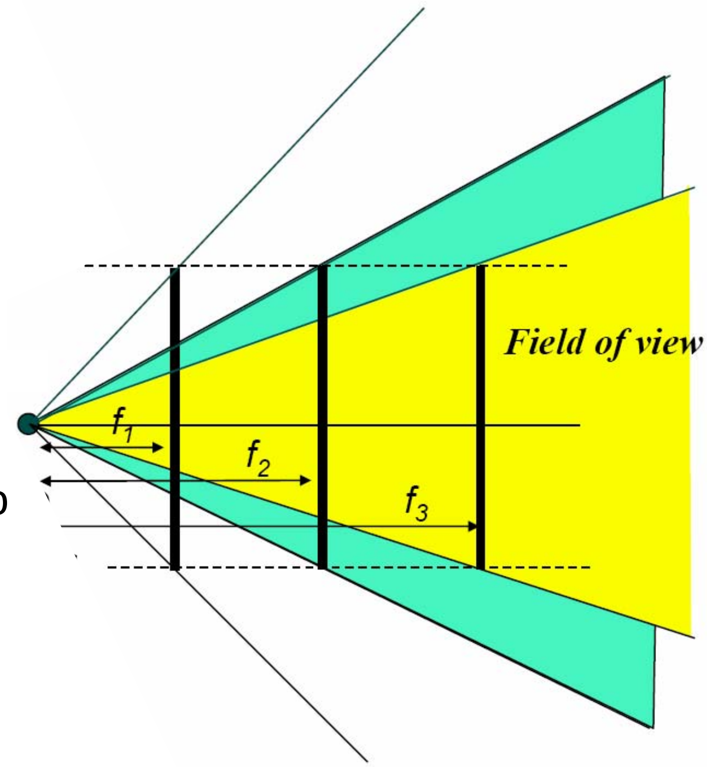
Rays entering parallel on one side go through focus on other, and vice versa.

In ideal case – all rays from P imaged at P'.

Field of view (portion of 3d space seen by camera) depends on *d* and *f.*

Field of view

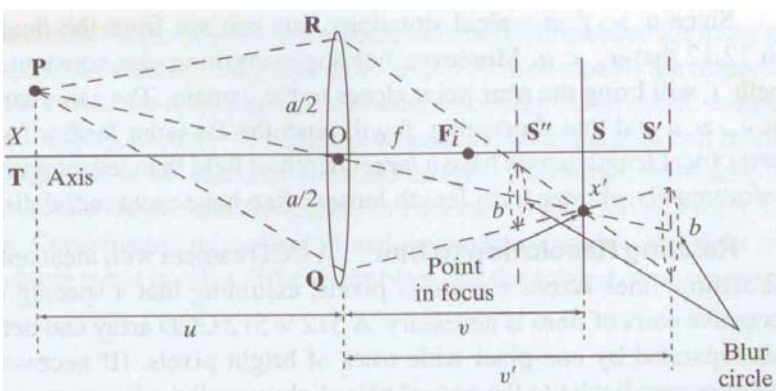
- As ***f*** gets smaller, image becomes more *wide angle* (more world points project onto the finite image plane)
- As ***f*** gets larger, image becomes more *telescopic* (smaller part of the world projects onto the finite image plane)



from R. Duraiswami

Focus and depth of field

- Depth of field: distance between image planes where blur is tolerable



← "circles of confusion" →

Thin lens: scene points at distinct depths come in focus at different image planes.

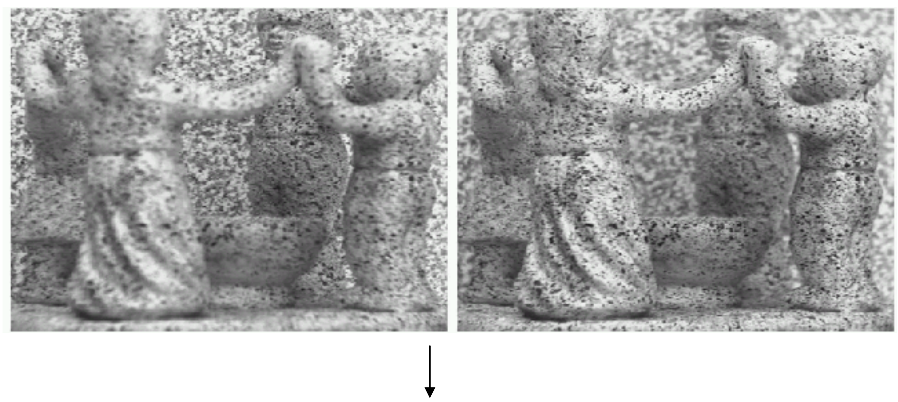
(Real camera lens systems have greater depth of field.)

Shapiro and Stockman

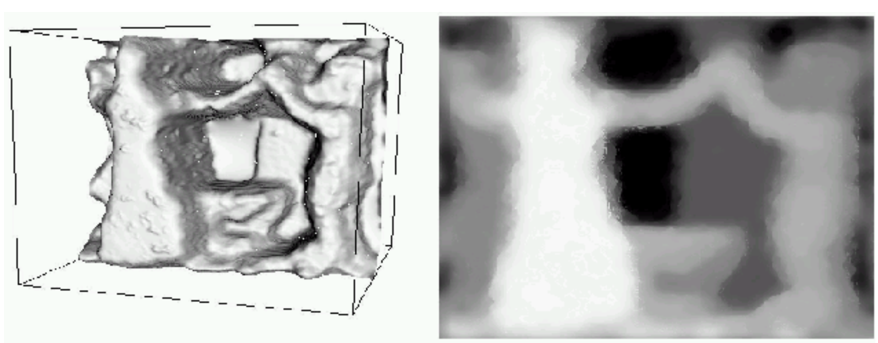
Focus and depth of field

Image credit: cambridgeincolour.com

Depth from focus



Images from same point of view, different camera parameters



3d shape / depth estimates

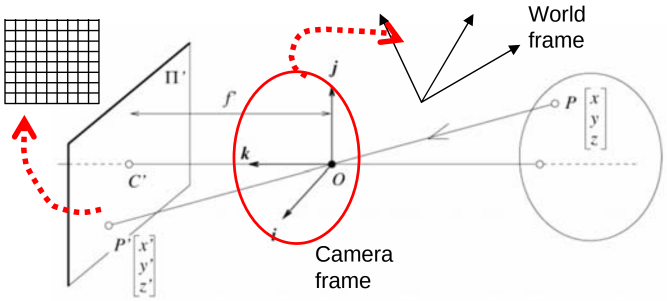
[figs from H. Jin and P. Favaro, 2002]

Camera parameters

- How do points in real world relate to positions in the image?
- Perspective equations so far in terms of *camera's* reference frame…

Camera parameters

Need to estimate camera's *intrinsic* and *extrinsic* parameters to calibrate geometry.



Extrinsic:
Camera frame ←→World frame

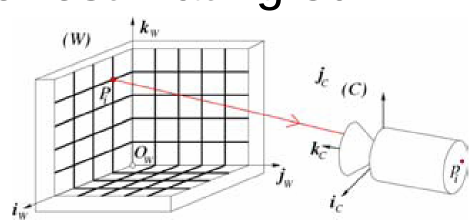
Intrinsic:
Image coordinates relative to camera ←→ Pixel coordinates

Camera calibration

- *Extrinsic* params: rotation matrix and translation vector
- *Intrinsic* params: focal length, pixel sizes (mm), image center point, radial distortion parameters

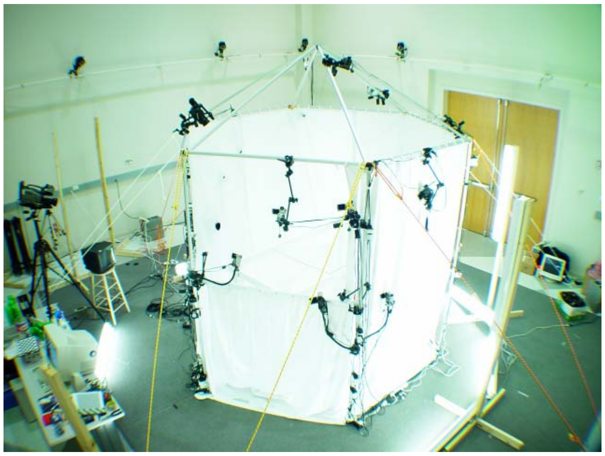
- Knowing the relationship between real world and image coordinates useful for estimating 3d shape

More on this later

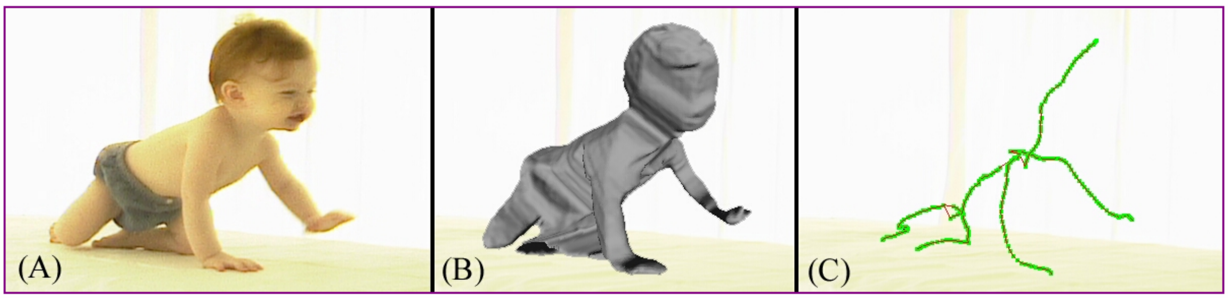


Articulated tracking

Demirdjian et al.

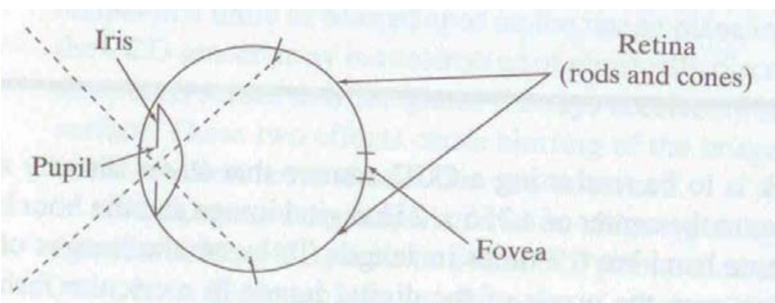


3d skeleton extraction



Brostow et al, 2004

Human eye



Pupil/Iris – control amount of light passing through lens

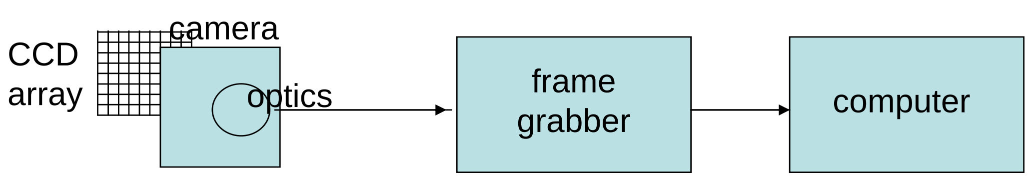
Retina - contains sensor cells, where image is formed

Fovea – highest concentration of cones

Shapiro and Stockman

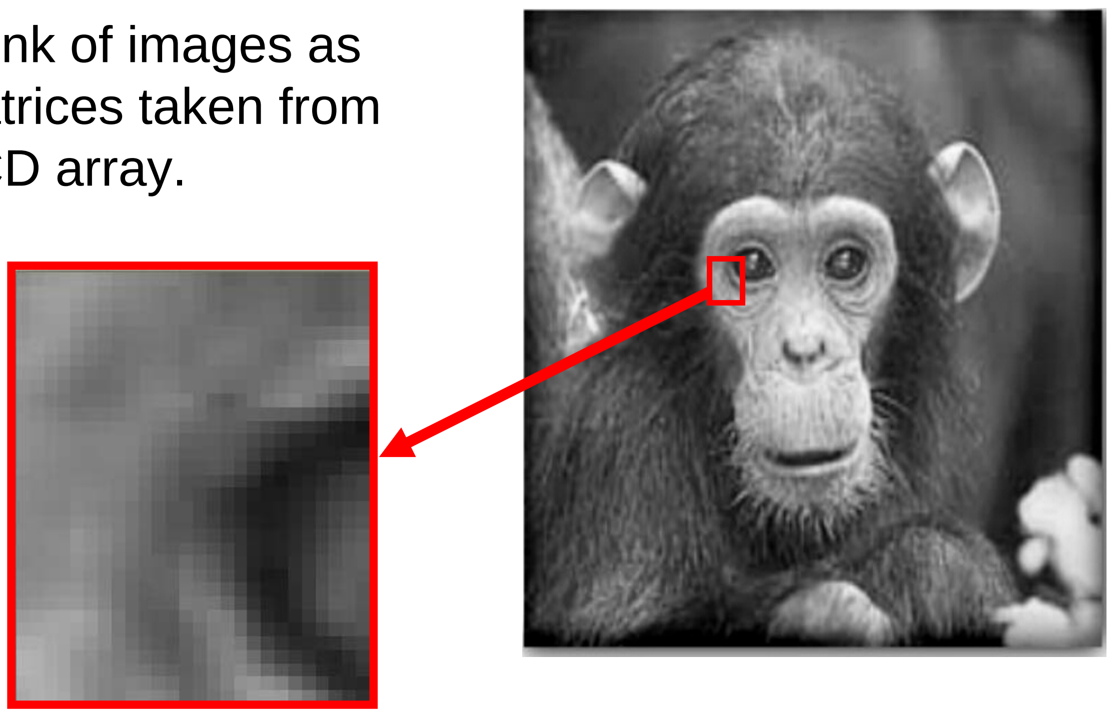
Sensors

- Often CCD camera: charge coupled device
- Record amount of light reaching grid photosensors, which convert light energy into voltage
- Read digital output row-by-row



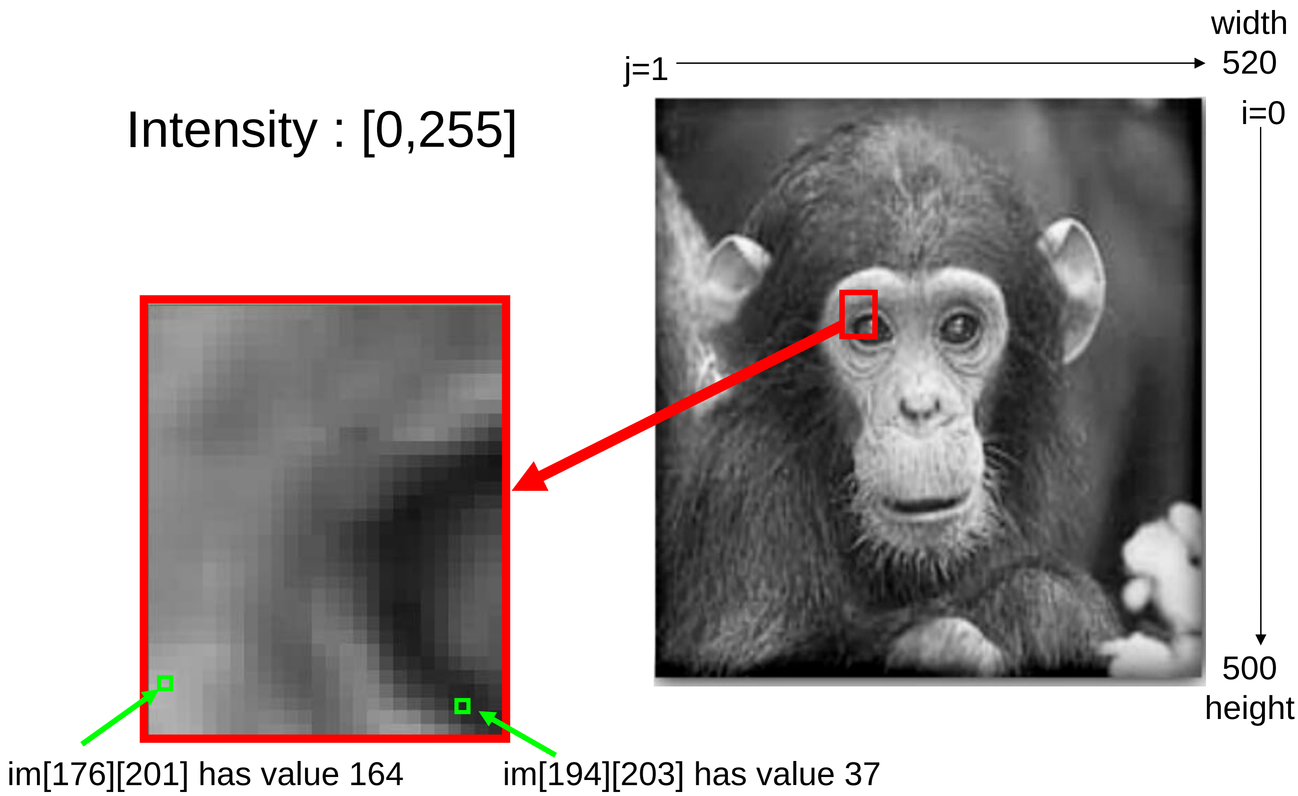
Digital images

Think of images as matrices taken from CCD array.

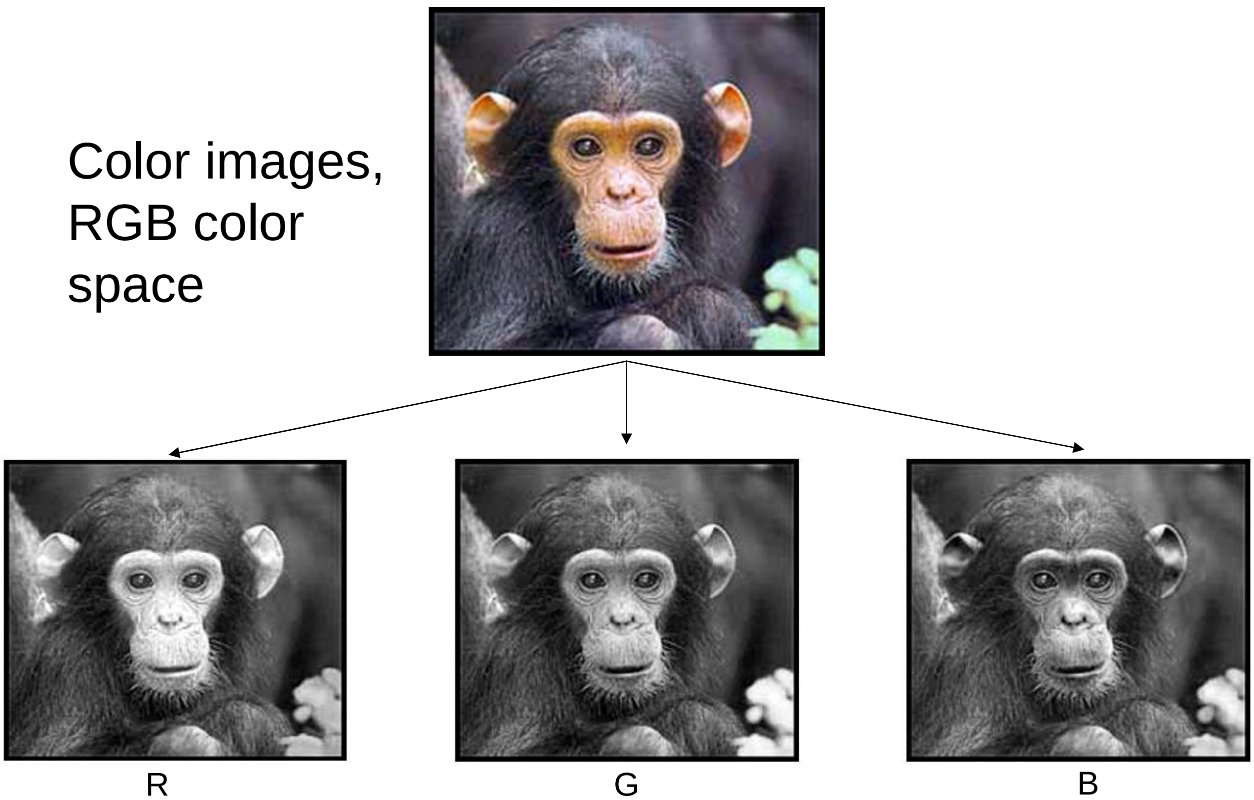


Digital images

Intensity : [0,255]



Color images, RGB color space



Resolution

- sensor: size of real world scene element a that images to a single pixel
- image: number of pixels
- Influences what analysis is feasible, affects best representation choice.

[Mori et al]

Resolution

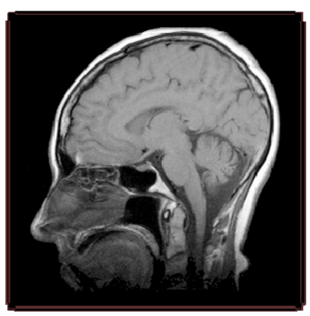
…though not necessarily for the human visual system with familiar faces…

[Sinha et al]

Other sensors

- Stereo cameras
- MRI scans
- Xray
- LIDAR devices…

[Jim Gasperini]



geospatial-online.com

Summary

- Image formation affected by geometry, photometry, and optics.
- Projection equations express how world points mapped to 2d image.
- Lenses make pinhole model practical.
- Imaged points related to real world coordinates via calibrated cameras.

Next

Problem set 0 due Sept 6

- Matlab warmup
- Image formation questions
- Read F&P Chapter 1

Reading for next lecture:

- F&P Chapter 6