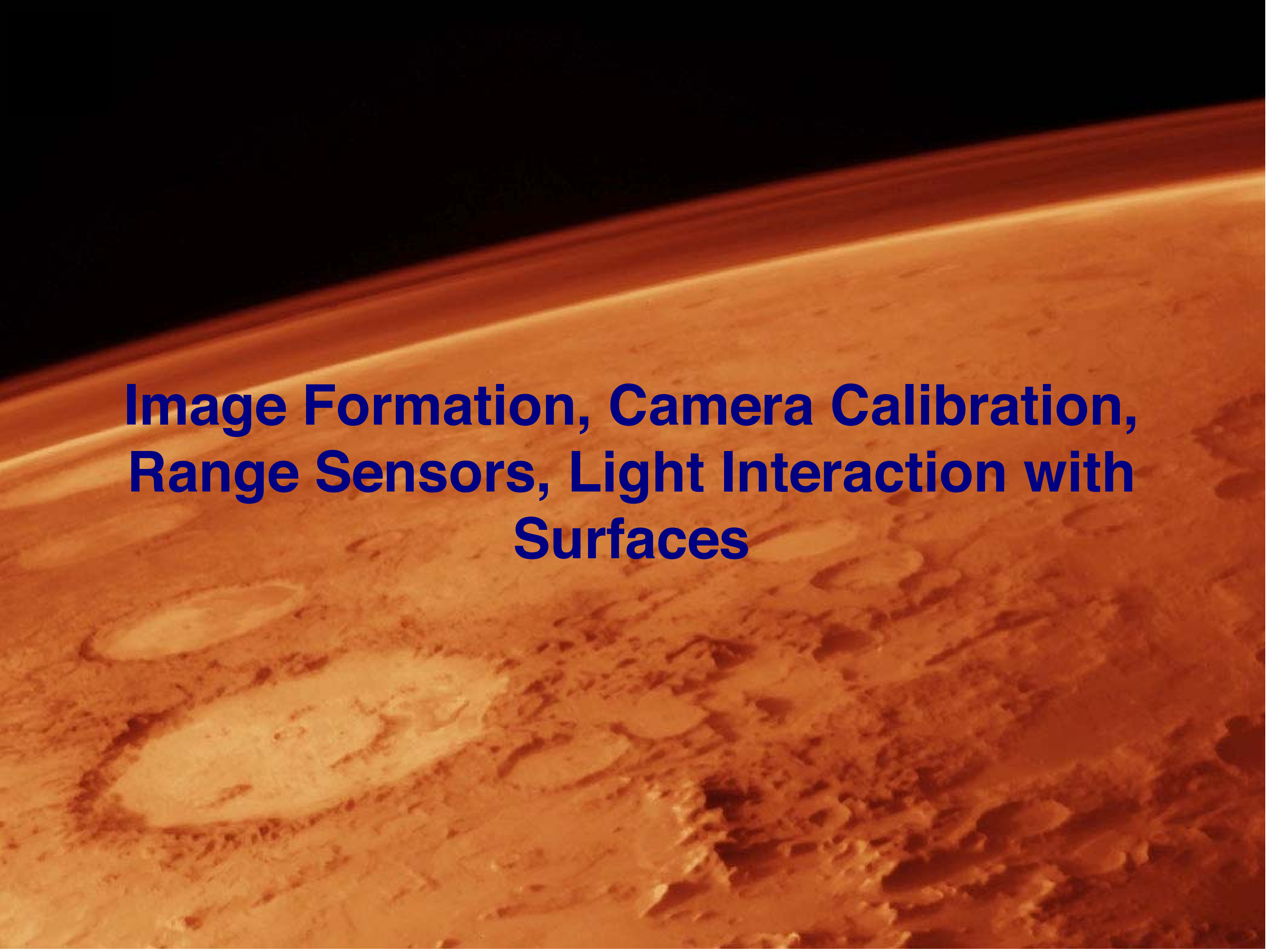
Image Formation, Camera Calibration, Range Sensors, Light Interaction with Surfaces



TAs, Logistics, Prerequisites

- Lectures: Tue/Thu, 1:00-2:30 pm, 105 Annenberg
- Lecturer office hours: after class and by arrangement
- TAs:
    - Lu Li (lli3@caltech.edu) + Zheng Li (zli@caltech.edu), looking for another
- Course wiki:
    - www.its.caltech.edu/~me132
- Course mailing list: me132-students@caltech.edu (sign up on wiki)
- Recommended background:
    - Basic understanding of linear algebra, probability, and statistics
    - Programming experience with Matlab and C; Python optional

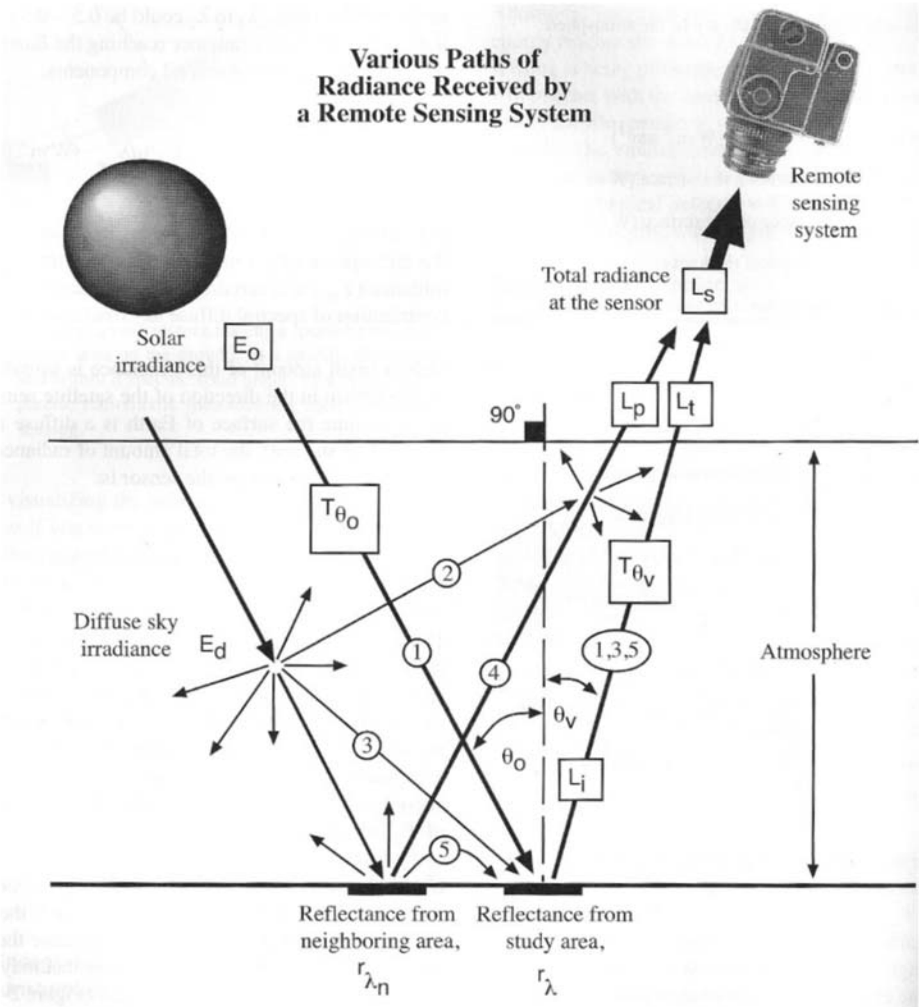
Texts and Supplemental Material

- Not closely following a single text. Good references are:
  - D. Forsyth and J. Ponce, *Computer Vision: A Modern Approach*, 2nd edition!
  - Richard Szeliski, *Computer Vision: Algorithms and Applications*, http://szeliski.org/Book
  - S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*!
  - A. Kelly, *Mobile Robotics: Mathematics, Models, and Methods*!
- Tutorials in appendices on relevant mathematical techniques (prob/stat, optimization algorithms, etc.)
- Reading material for today's lecture:
  - Forsyth chapter 1 and sections 2.1, 3.1, 3.2
- Additional references:
  - Szeliski chapter 2 (image formation)
  - Forsyth sections 22.1 and 22.2, or Szeliski appendix A (appendices on linear algebra and optimization techniques)

Announcements

- First homework is on the wiki today. Due by next Thursday, Jan 15.
  - Hand in at start of class and email to TA (me132-tas@caltech.edu)
- Office hours:
  - In the past, TAs handled all the office hours, using a poll to pick the time.
  - Instructors: easiest for us is after class. Other times are possible by arrangement. If necessary, we'll set up a standard time slot.

What This Lecture is About



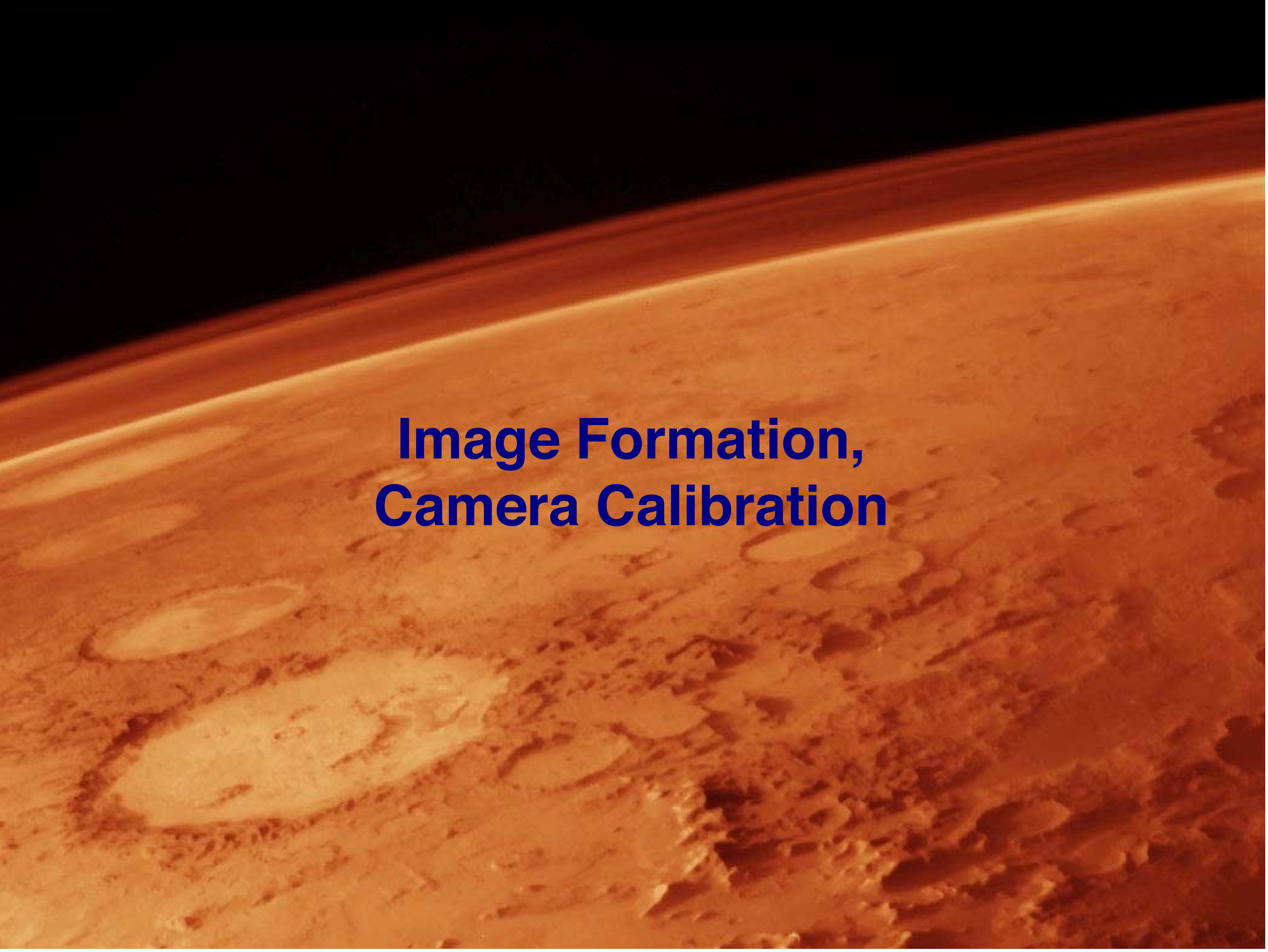


Image Formation,
Camera Calibration

Pinhole Camera

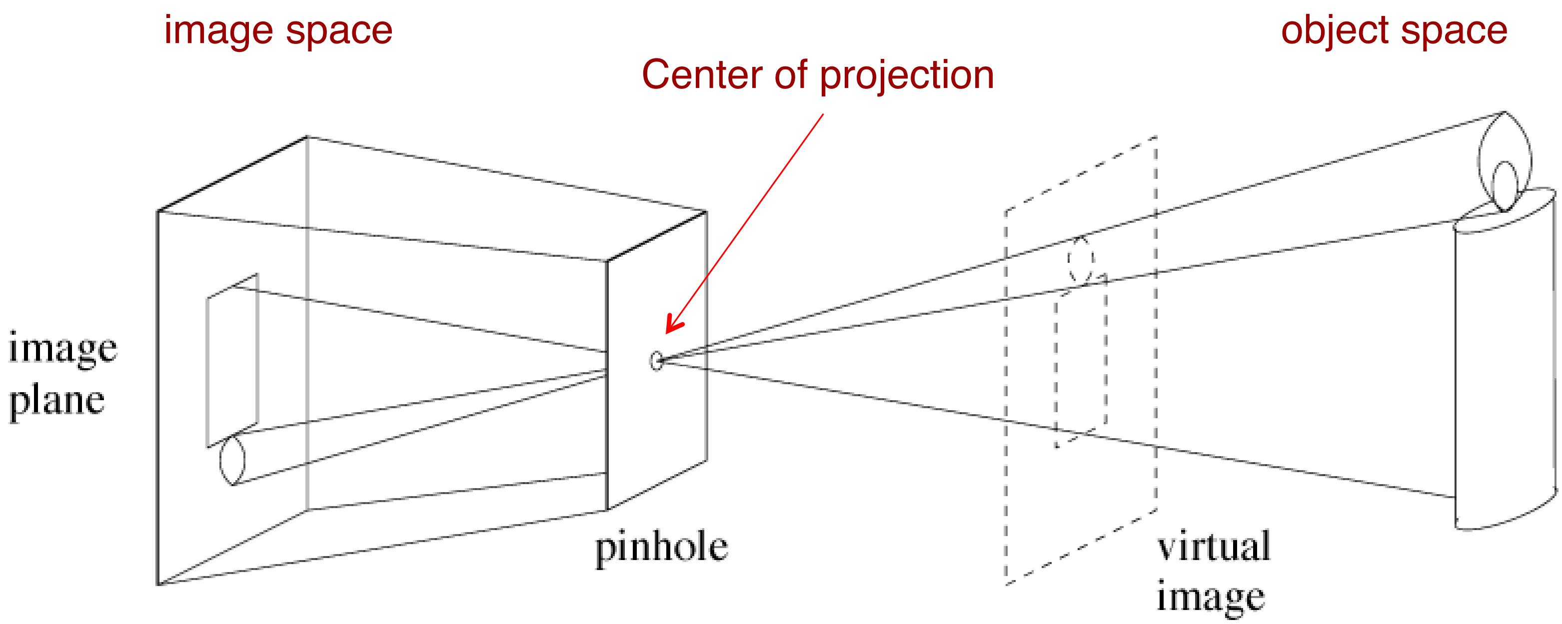
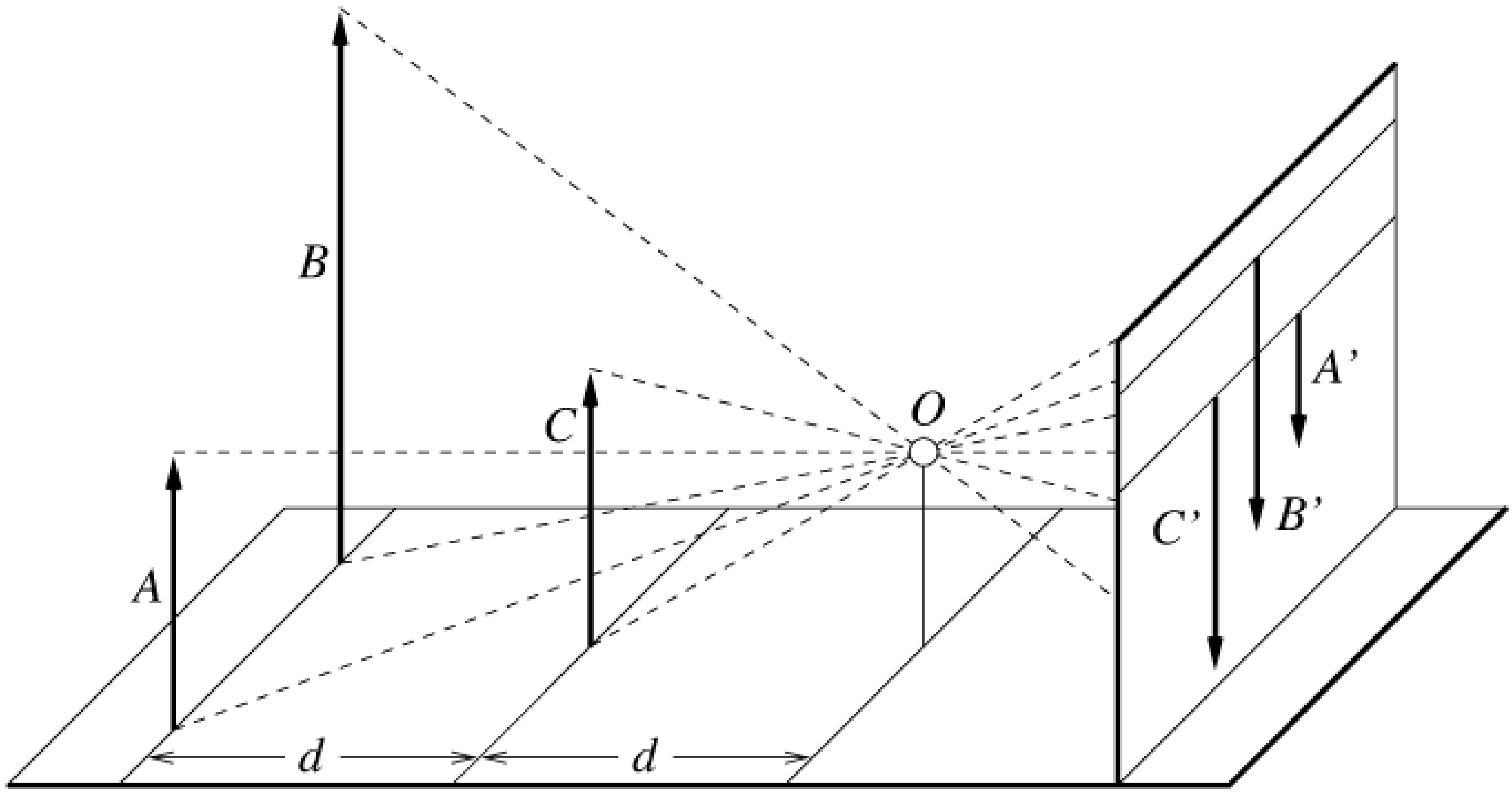
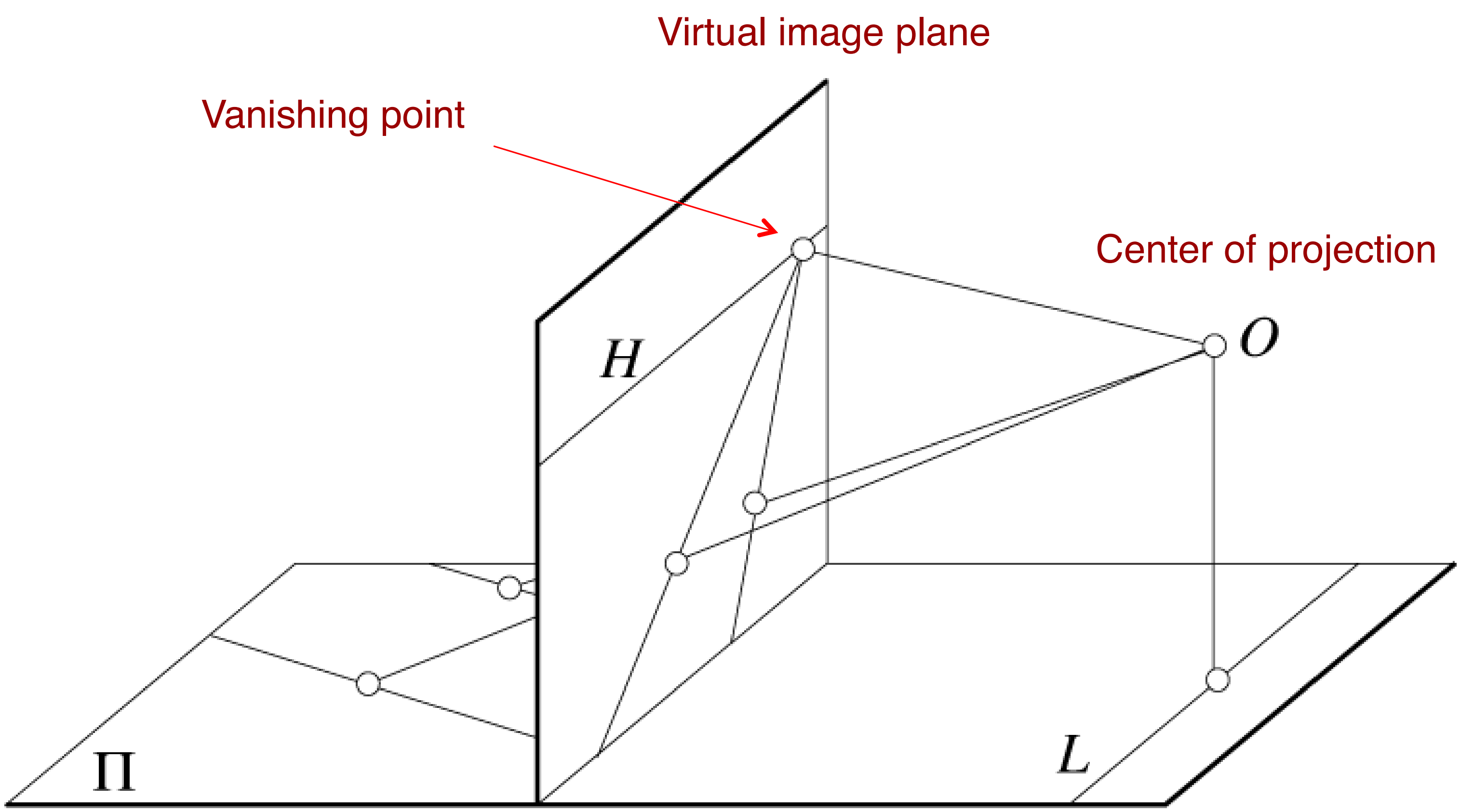


image is inverted

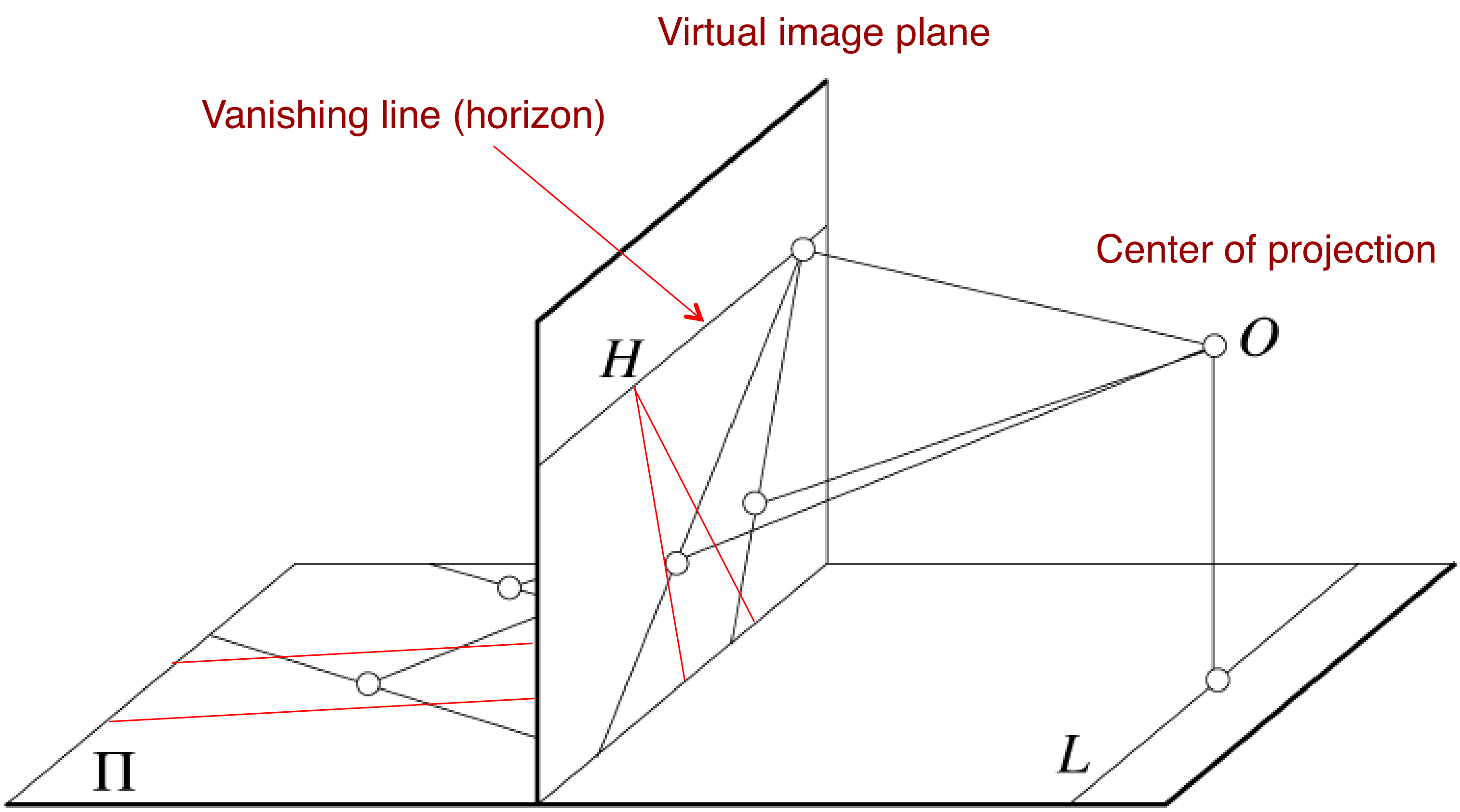
Basic Geometric Properties: (1) Distant Objects are Smaller



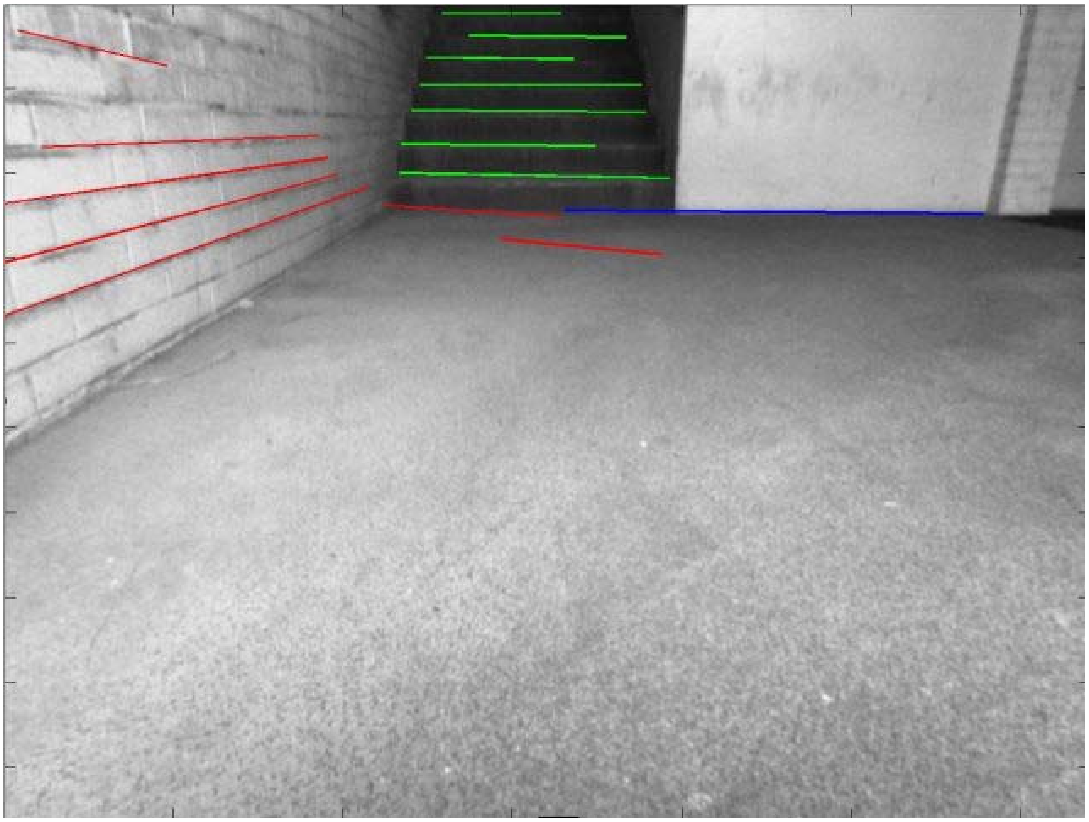
Basic Geometric Properties: (2) Lines Project to Lines and Parallel Lines Meet



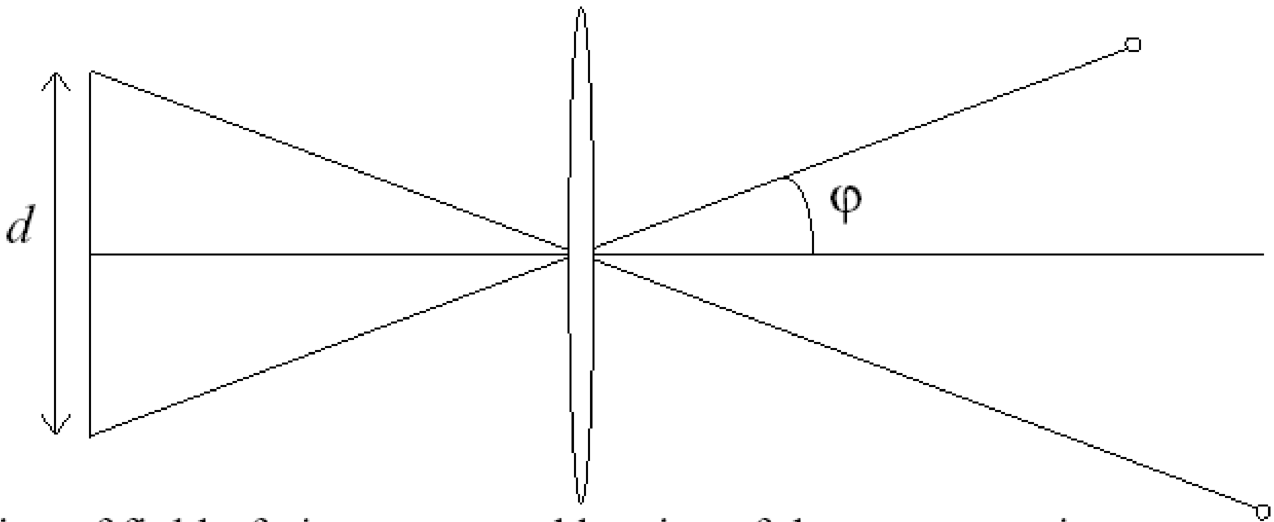
Basic Geometric Properties: (3) Multiple Parallel Lines form a Vanishing Line



Why Care about Vanishing Points and Lines?
Useful for Perception in Man-Made Scenes



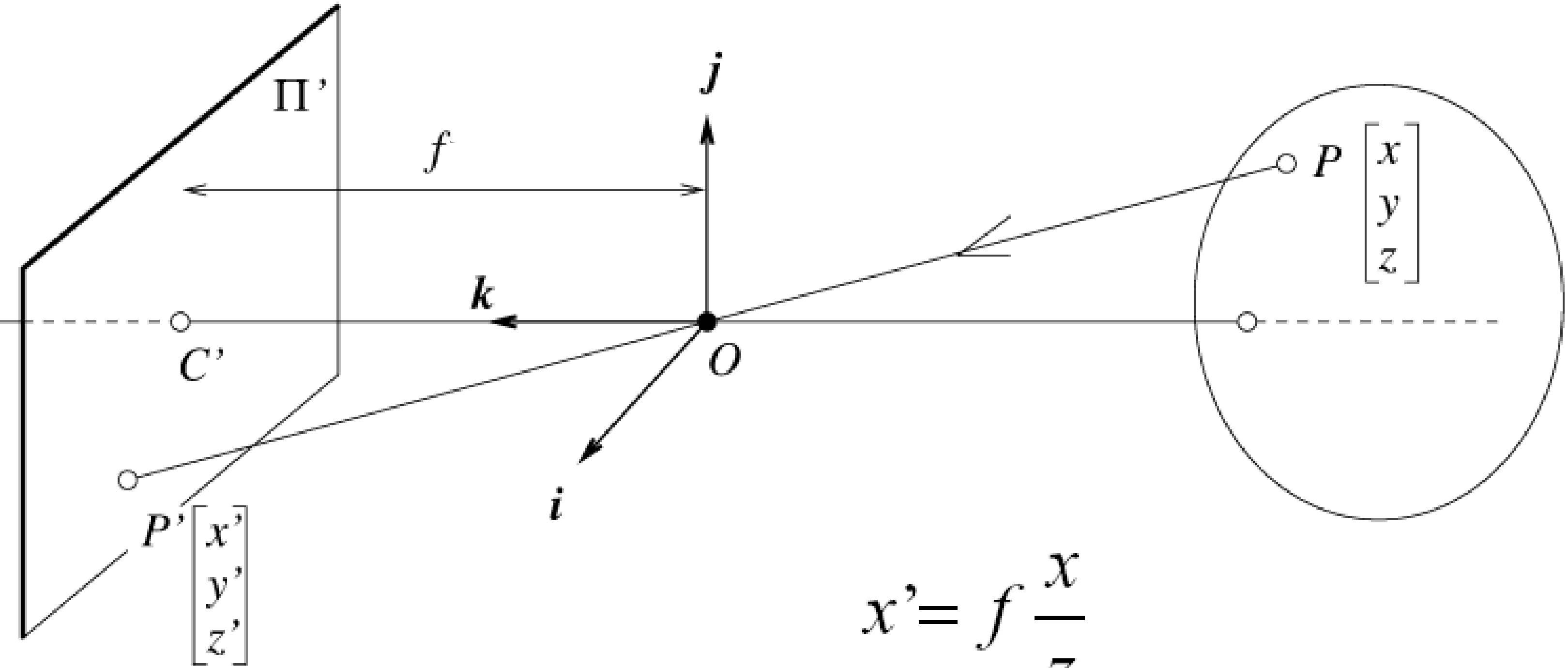
Finite Field of View



Size of field of view governed by size of the camera retina:

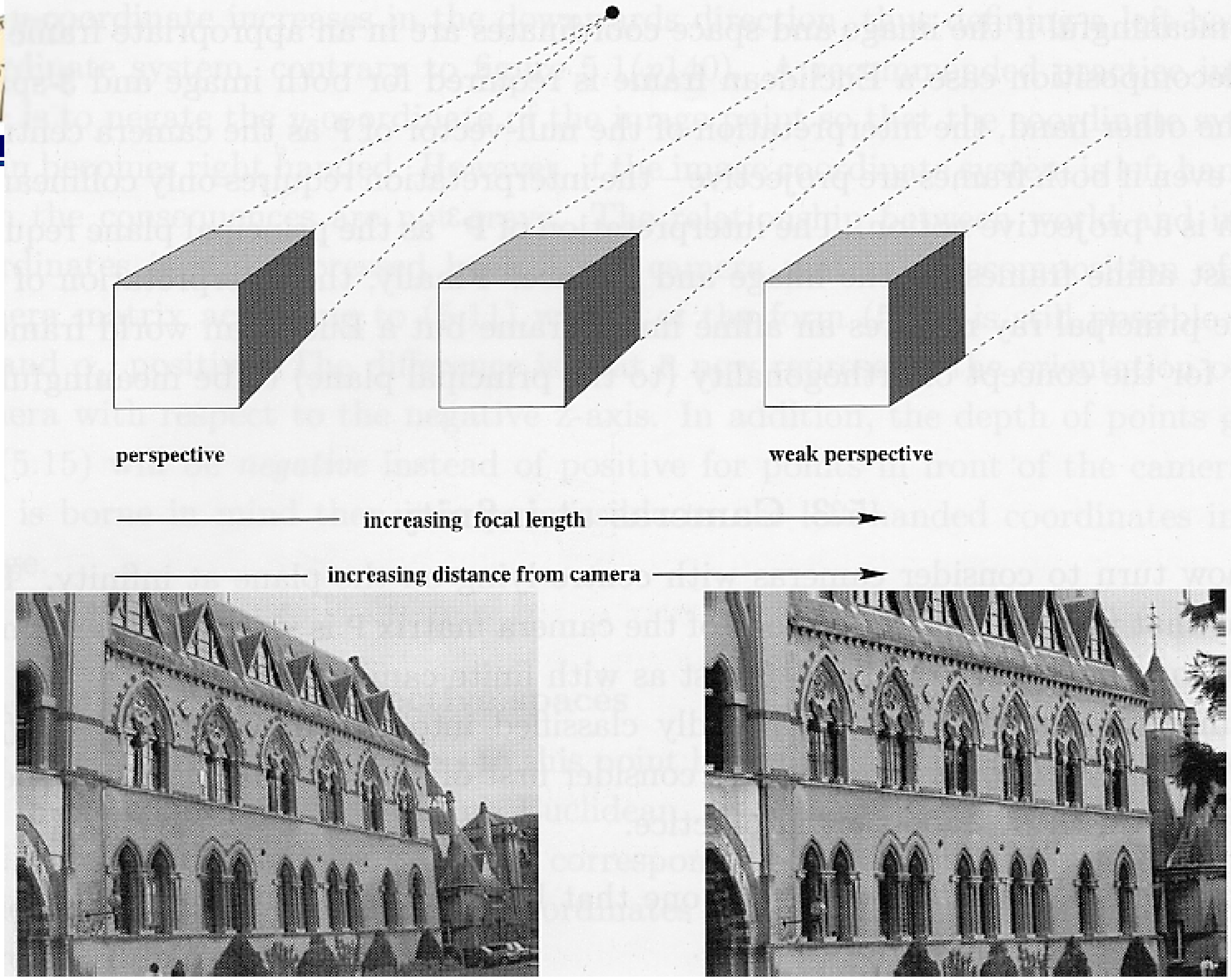
$$\varphi = \tan^{-1}\left(\frac{d}{2f}\right)$$

Perspective Projection



$$x' = f\,\frac{x}{z}$$

$$y' = f\,\frac{y}{z}$$

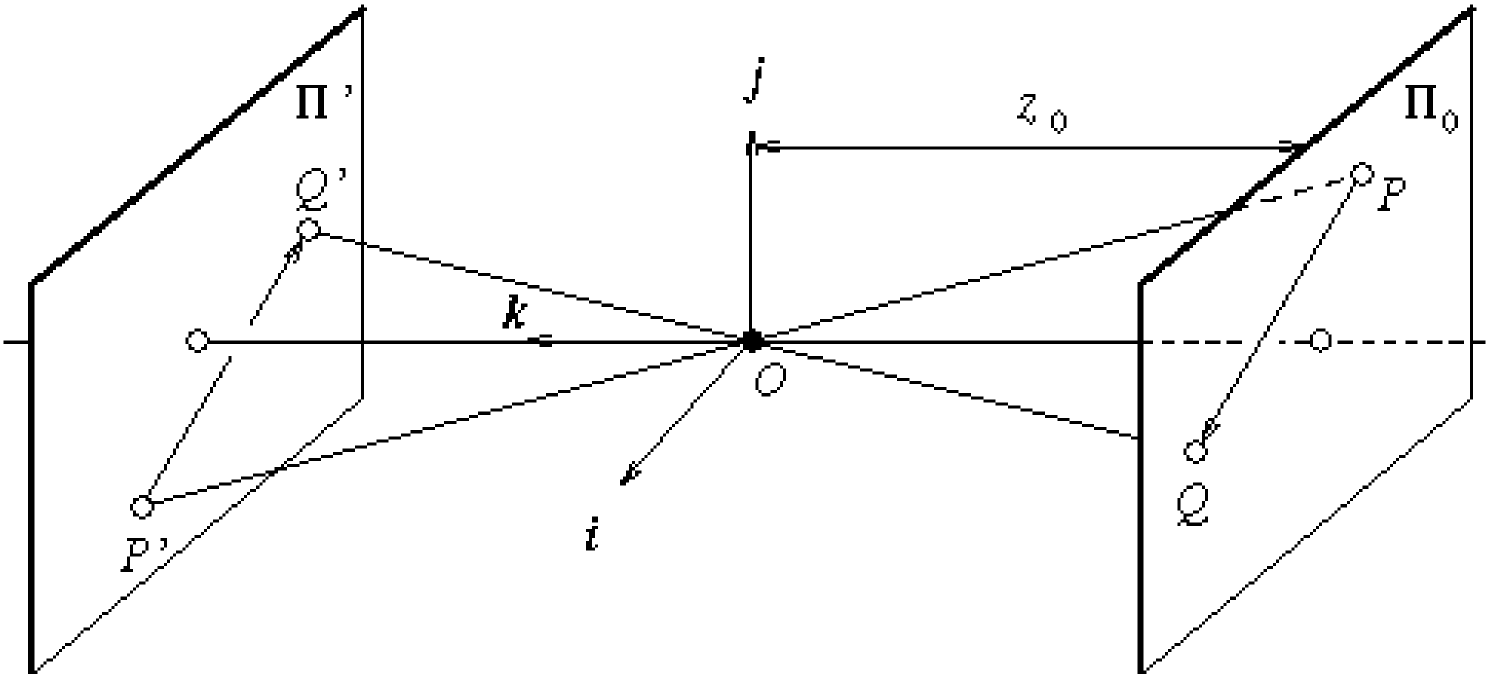

perspective
weak perspective
increasing focal length
increasing distance from camera

Weak Perspective

Strong perspective:
Angles are not preserved
The projections of parallel lines intersect at one point

Weak perspective:
Angles are better preserved
The projections of parallel lines are (almost) parallel

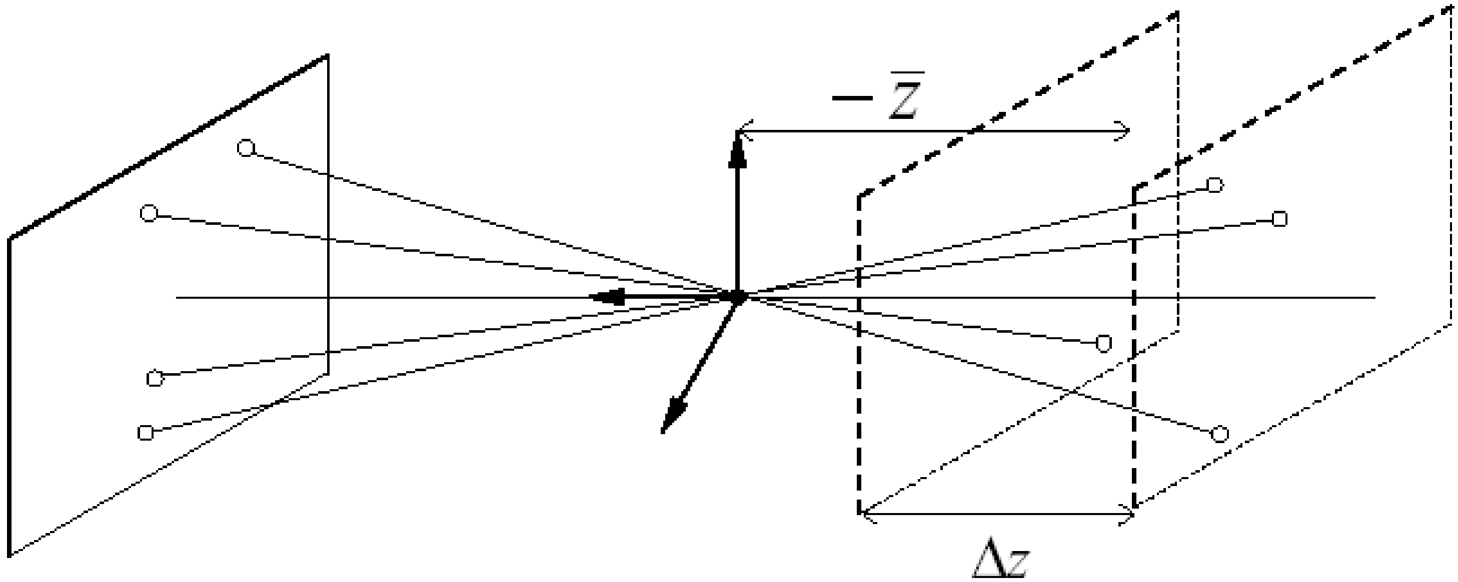
Weak Perspective (Affine Projection)



$$x' \approx -mx \qquad m = -\frac{f}{z_o}$$
$$y' \approx -my$$

If scene points are in a plane, projections are simply magnified by m

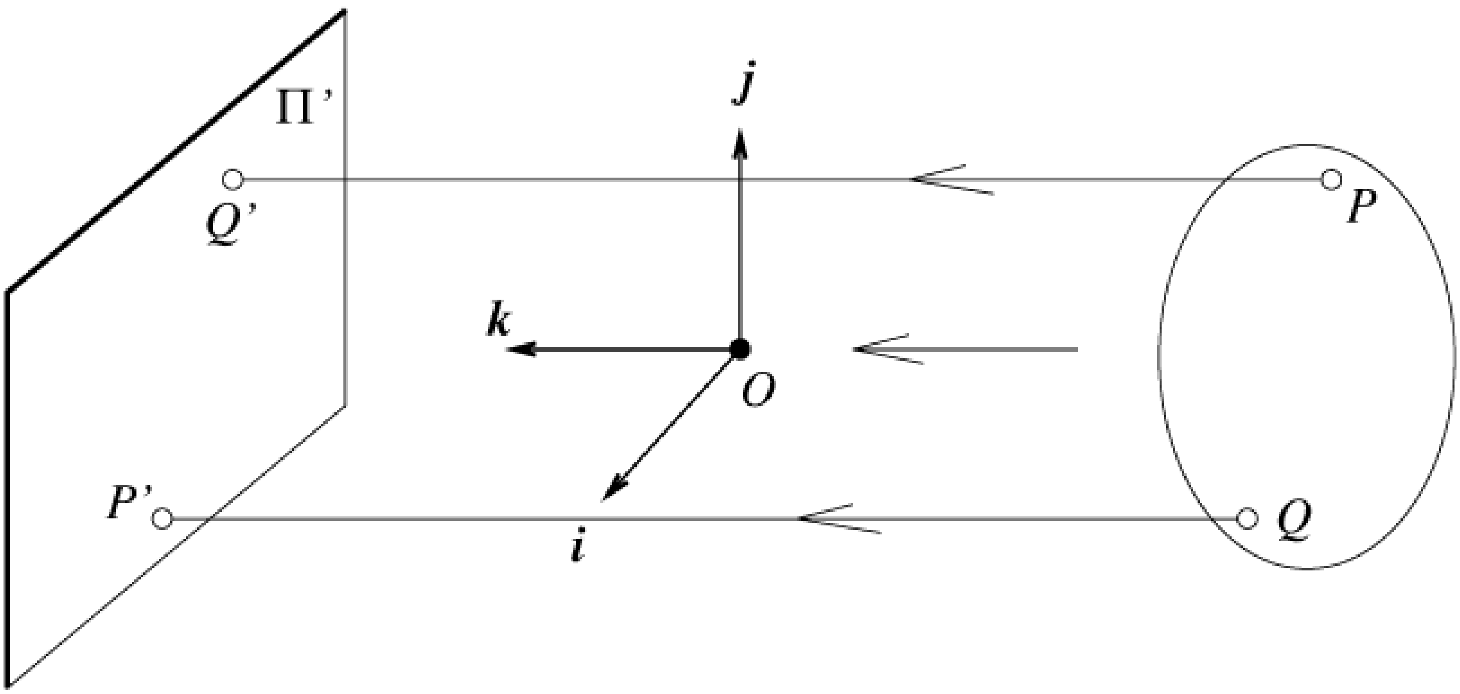
Weak Perspective (Affine Projection)



If $\Delta z << -\bar{z}$: $\begin{array}{l} x' \approx -mx \\ y' \approx -my \end{array} \quad m = -\dfrac{f}{\bar{z}}$

Justified if scene depth is small relative to average distance from camera

Orthographic Projection



Pinhole Camera

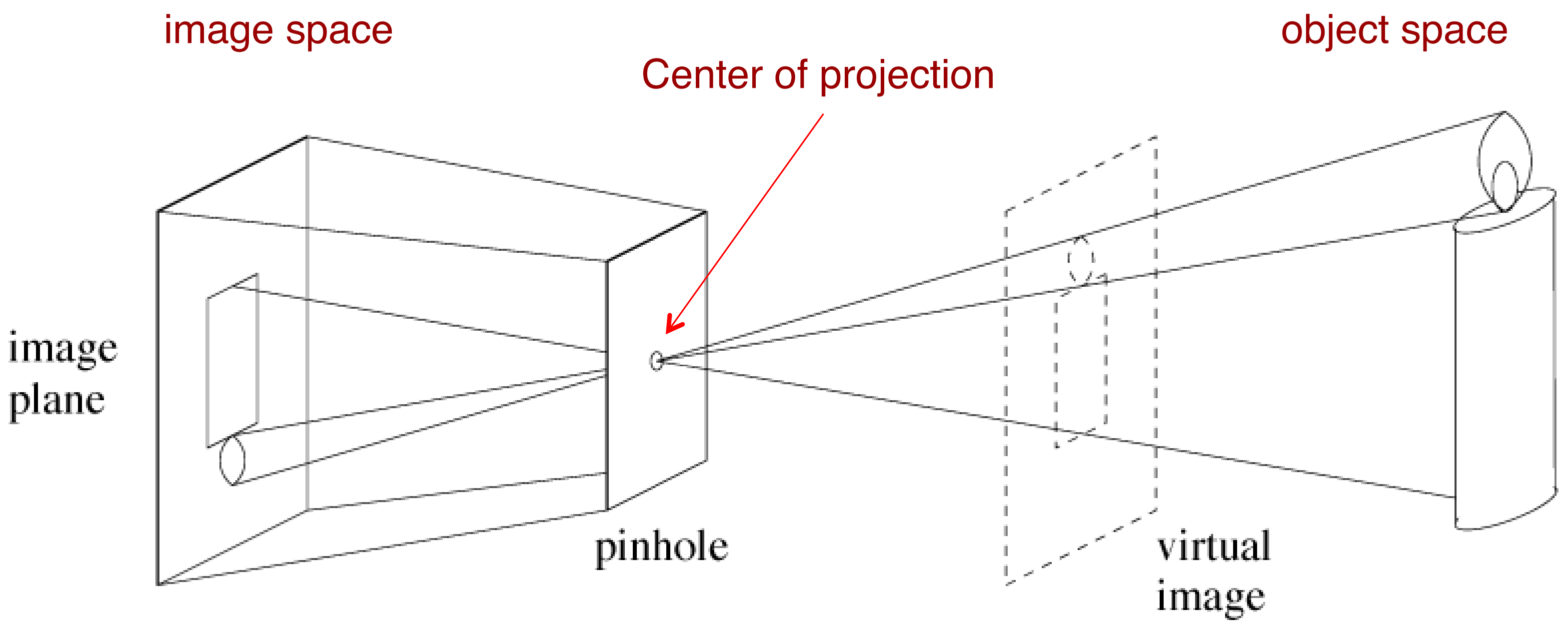
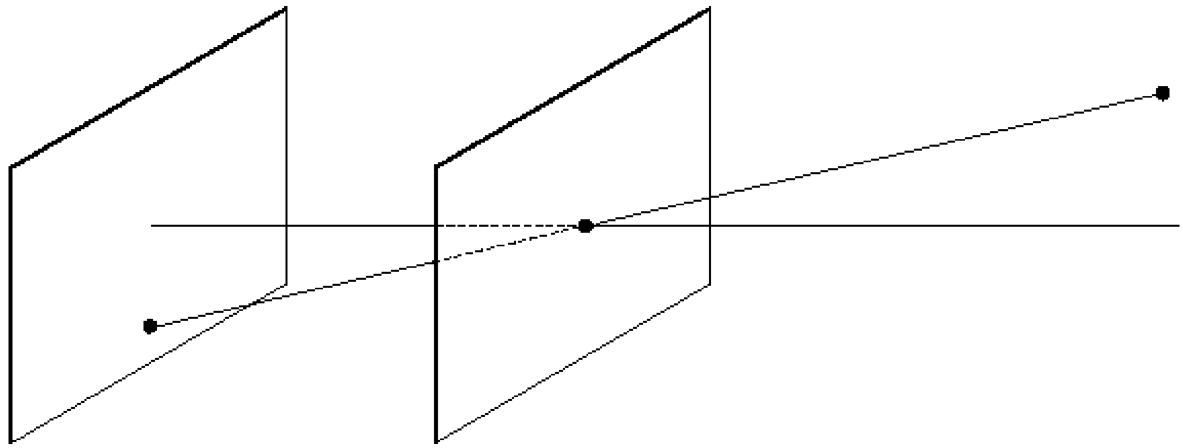


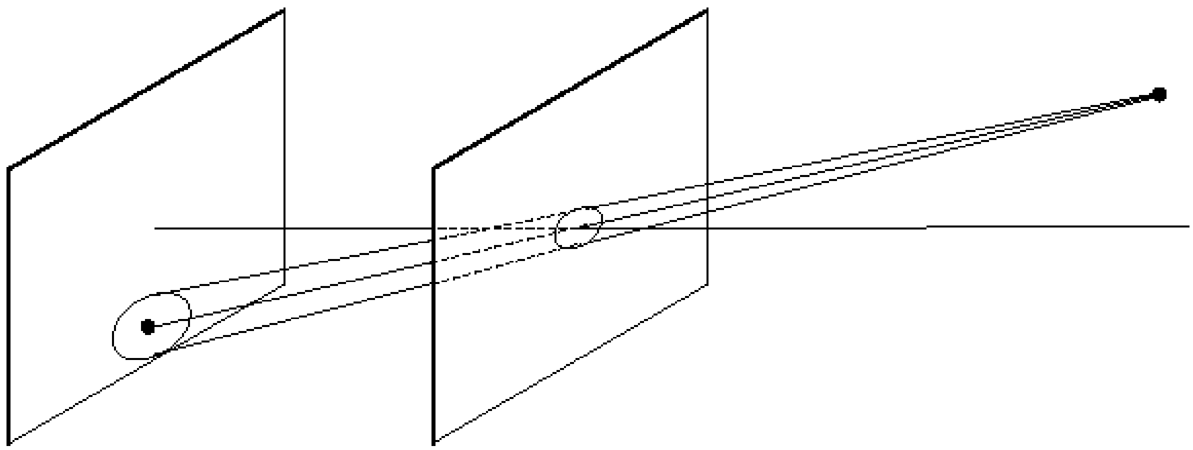
image is inverted

Limitations of Pinhole Cameras/Models

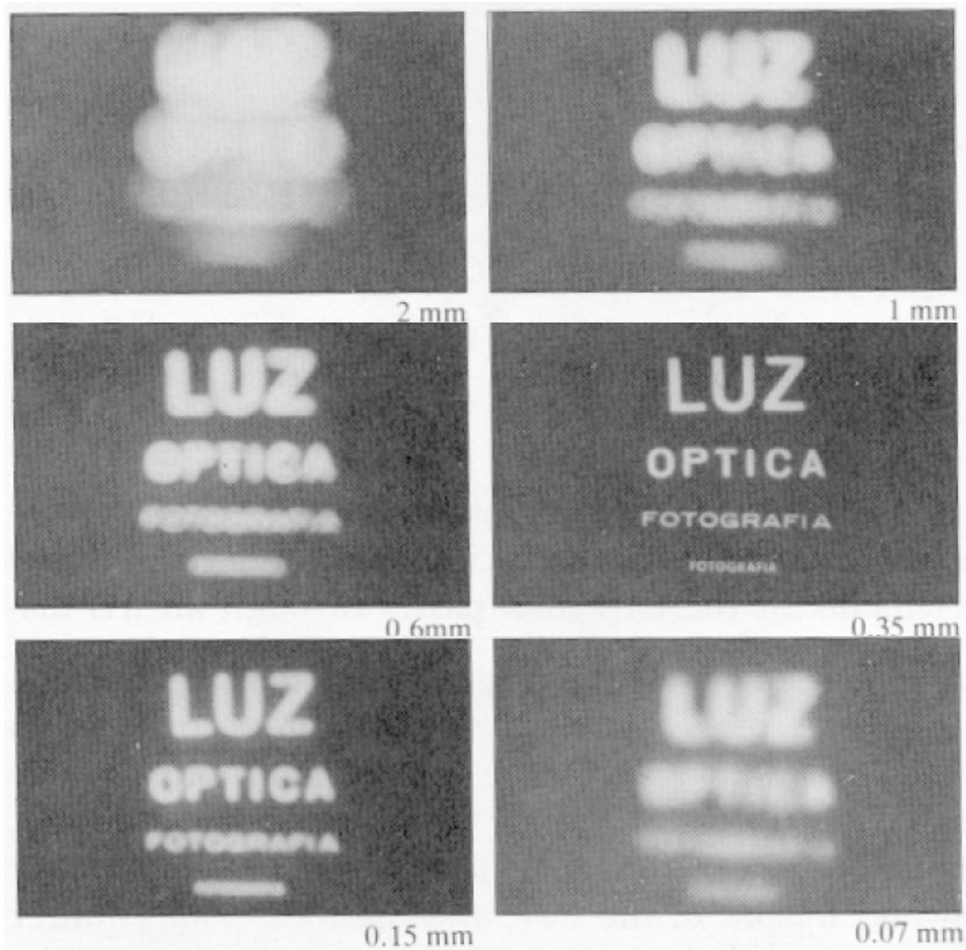
Ideal pinhole:
Single scene point generates single image
but:
Diffraction
Low light level



Finite-size pinhole:
Single scene point generates extended image.
Resulting image is blurry



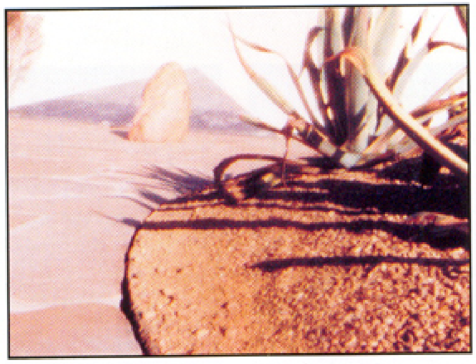
Limitations of Pinhole Cameras/Models

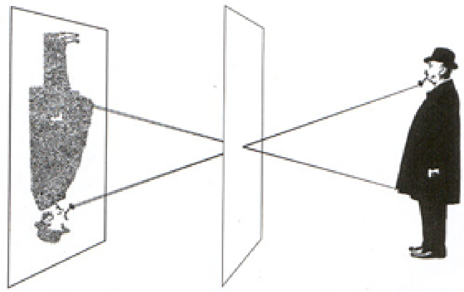


From Hecht, *Optics*

Need to Use Lens: Larger Aperture, But Need to Focus

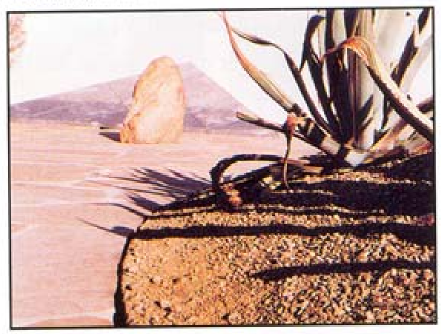
Photograph made with small pinhole

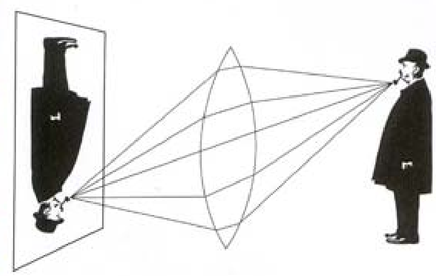




To make this picture, the lens of a camera was replaced with a thin metal disk pierced by a tiny pinhole, *equivalent in size to an aperture of f/182. Only a few rays of light from each point on the subject got through the tiny opening, producing a soft but acceptably clear photograph. Because of the small size of the pinhole, the exposure had to be 6 sec long.*

Photograph made with lens



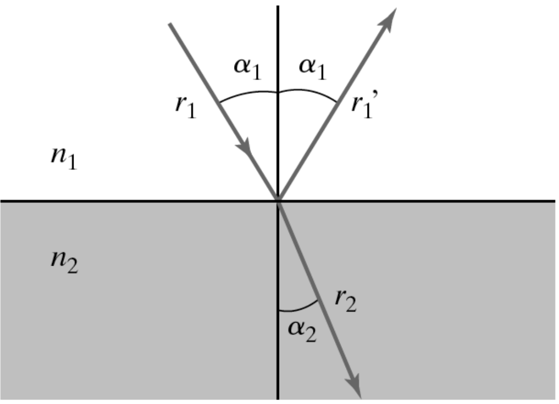


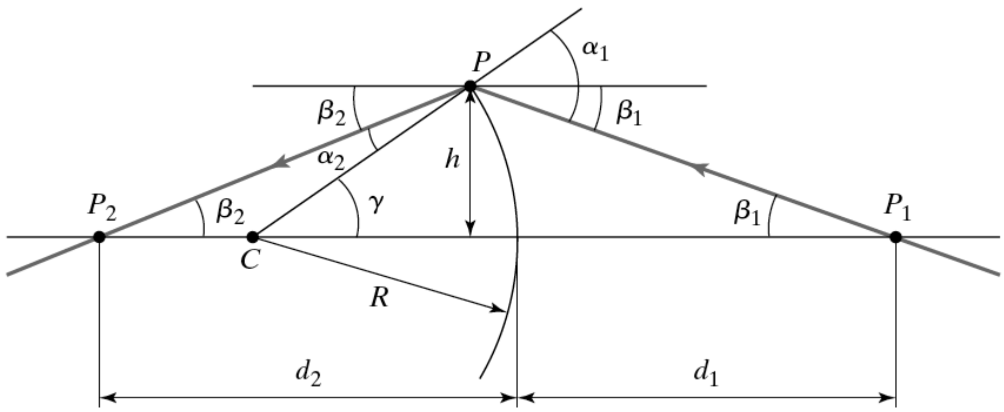
This time, using a simple convex lens with an f/16 aperture, *the scene appeared sharper than the one taken with the smaller pinhole, and the exposure time was much shorter, only 1/100 sec.*

The lens opening was much bigger than the pinhole, letting in far more light, but it focused the rays from each point on the subject precisely so that they were sharp on the film.

From Photography, London et al.

Snell's Law, Spherical Surfaces, and Paraxial Optics



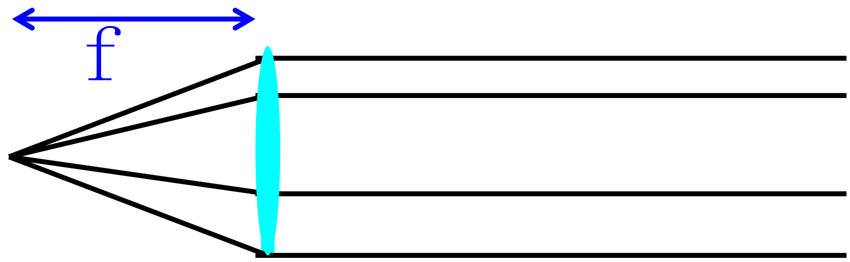


$$n_1 \sin \alpha_1 = n_2 \sin \alpha_2$$

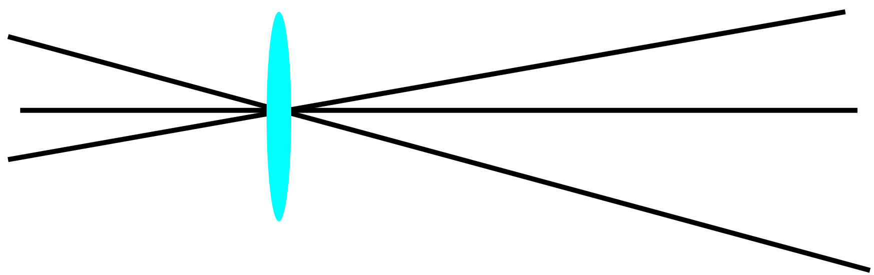
$$n_1 \alpha_1 = n_2 \alpha_2$$

Thin Lens Optics

- Simplification of geometrical optics for well-behaved lenses
- All parallel rays converge to one point on a plane located at the focal length f

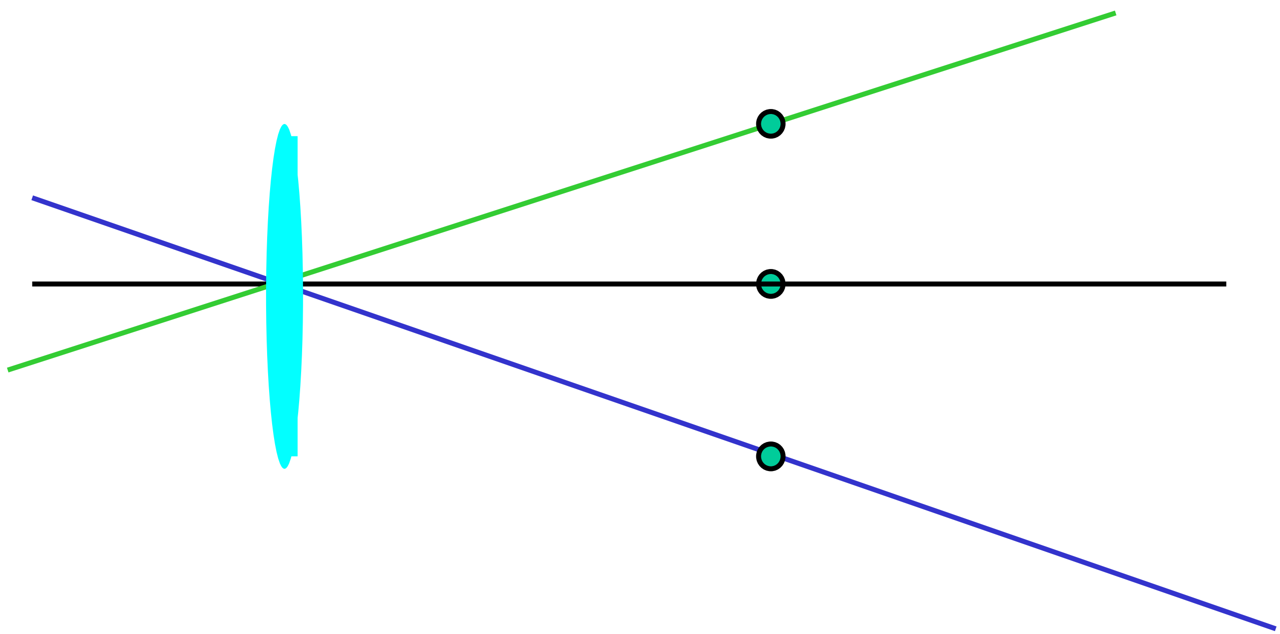


- All rays going through the center are not deviated
  - Hence same perspective as pinhole



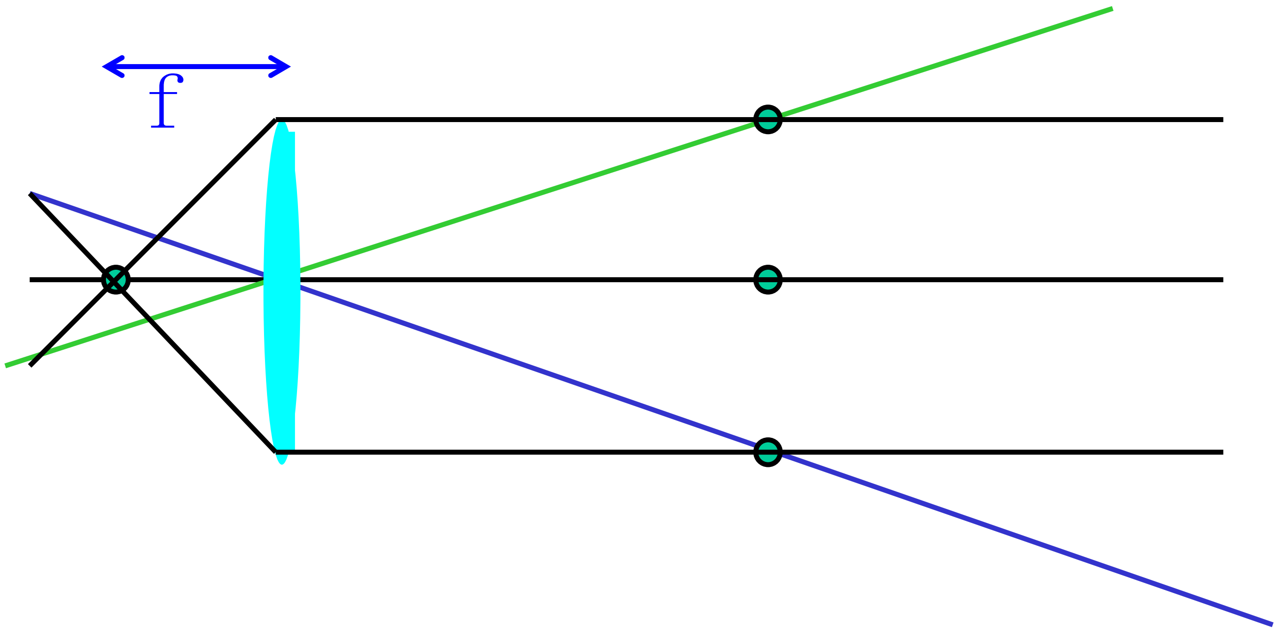
Tracing Rays

- Start by rays through the center



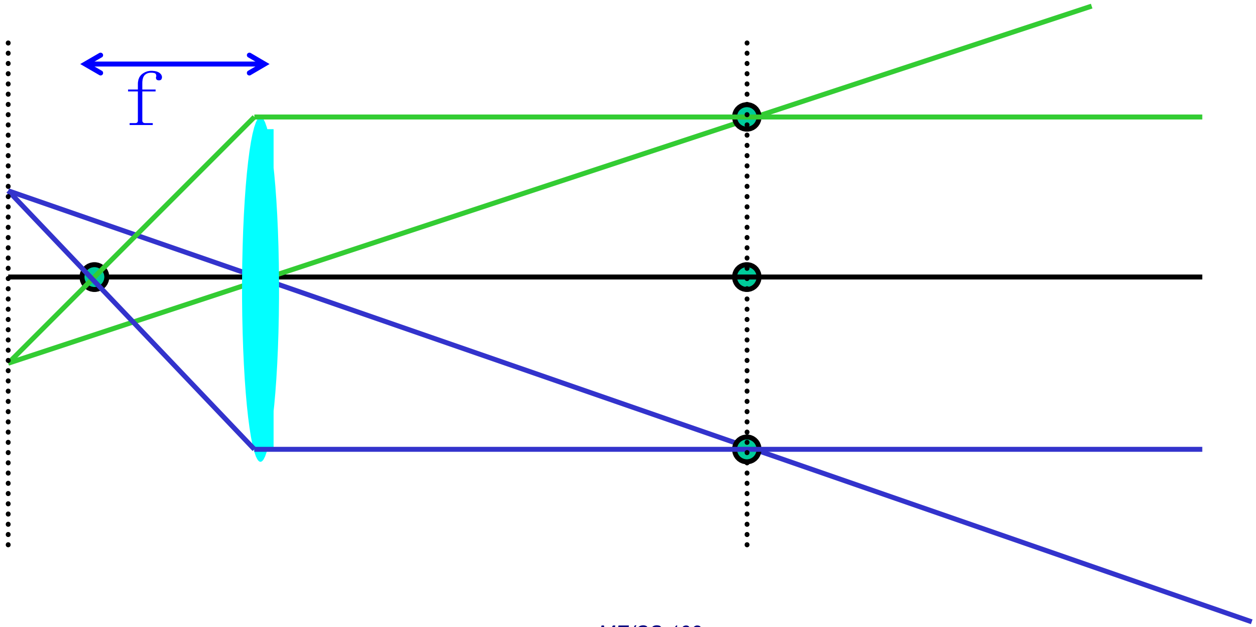
Tracing Rays

- Start by rays through the center
- Choose focal length, trace parallels



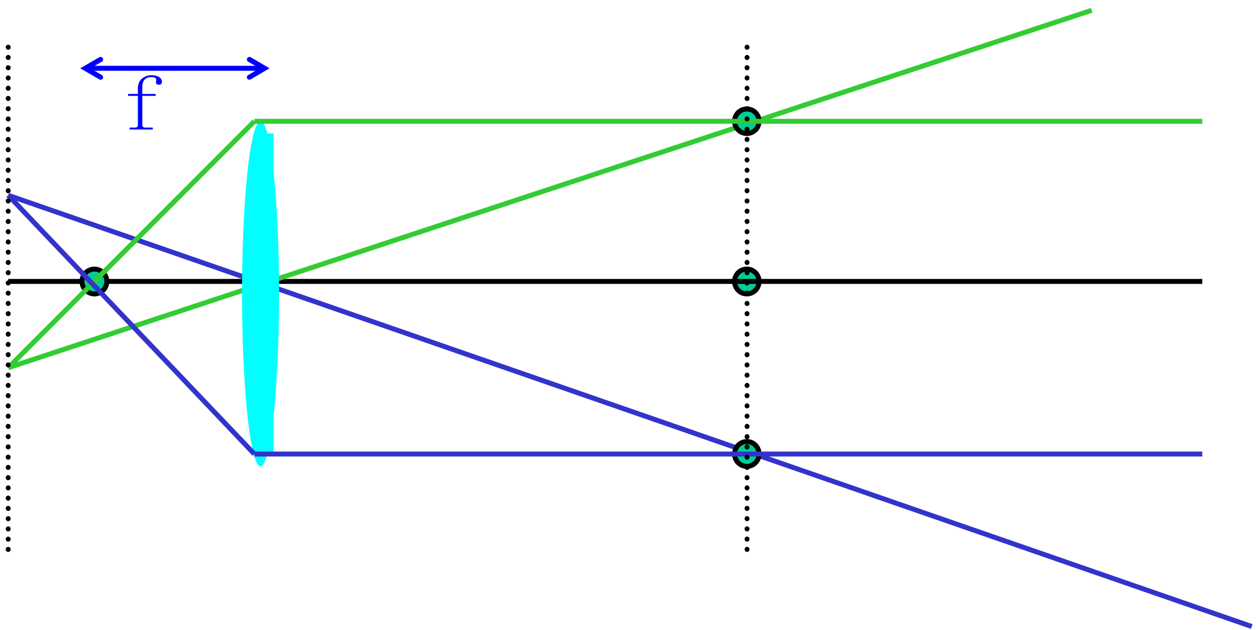
Tracing Rays

- Start by rays through the center
- Choose focal length, trace parallels
- You get the focus plane for a given scene plane
  - All rays coming from points on a plane parallel to the lens are focused on another plane parallel to the lens

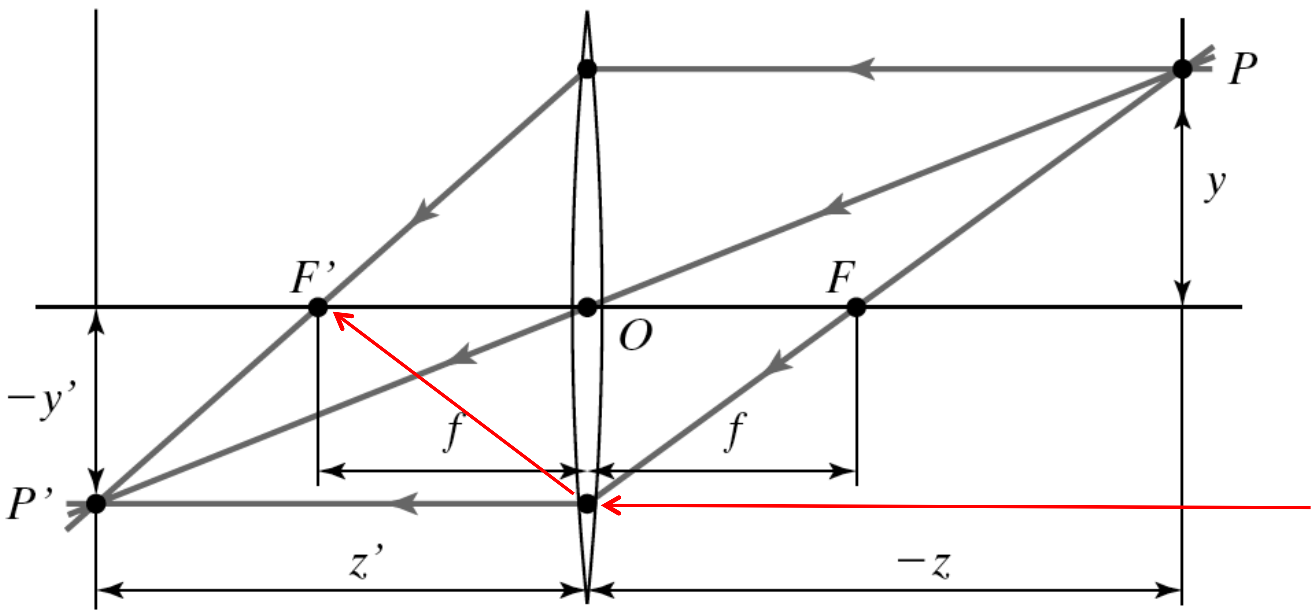


Focusing

- To focus closer than infinity
  - Move the sensor/film further than the focal length

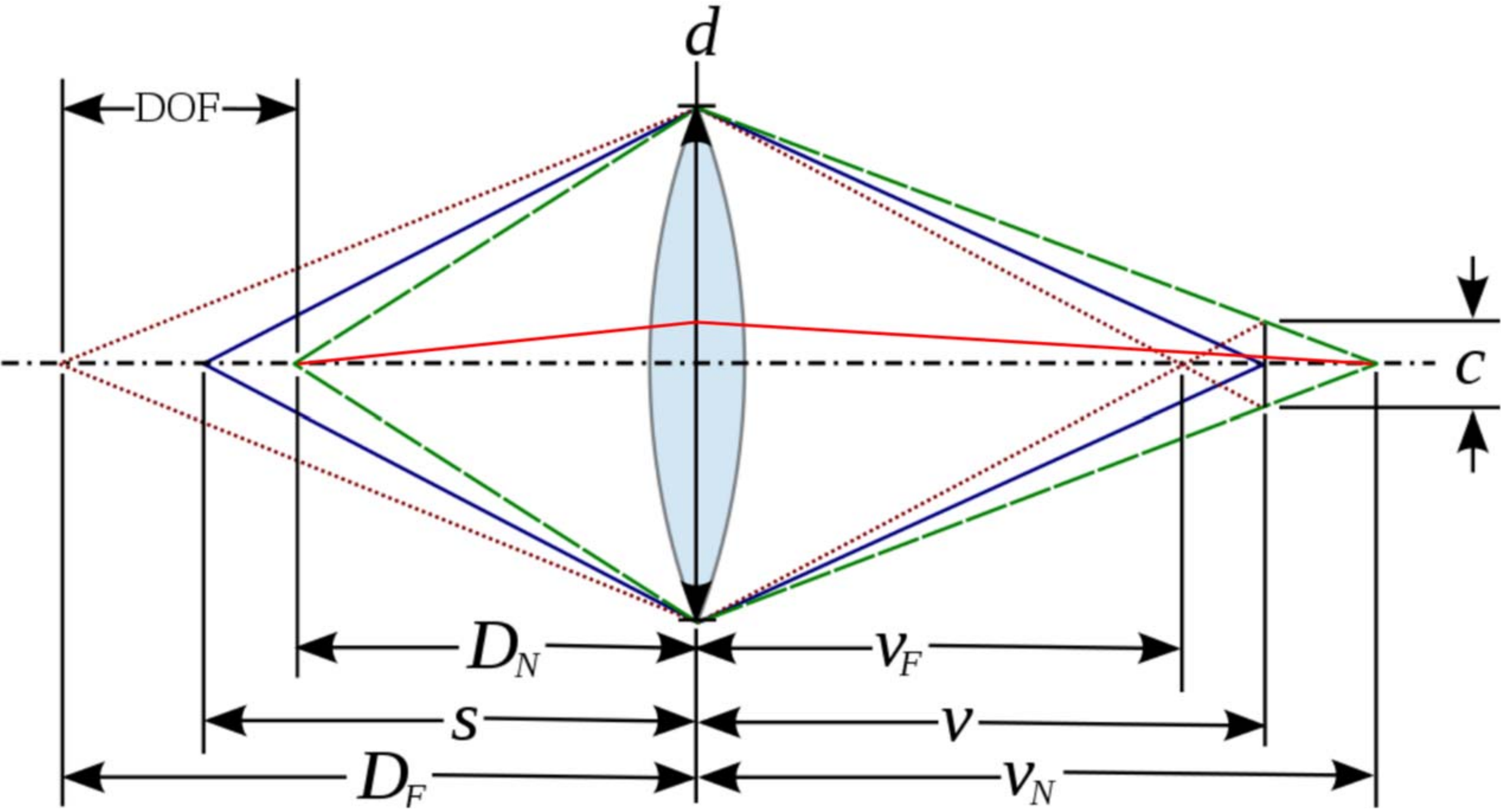


The Thin Lens Model

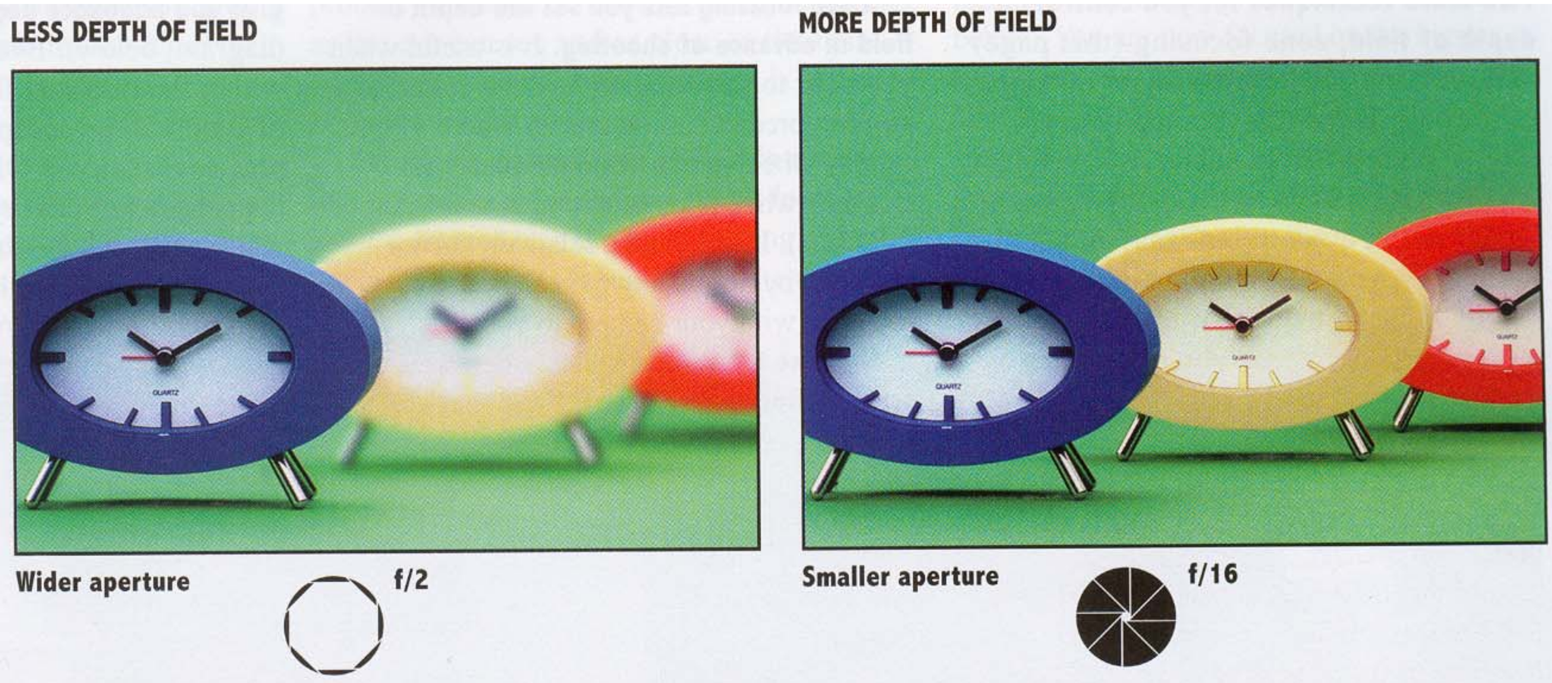


$$\frac{1}{z'} - \frac{1}{z} = \frac{1}{f}$$

Depth of Field and the Effect of Aperture



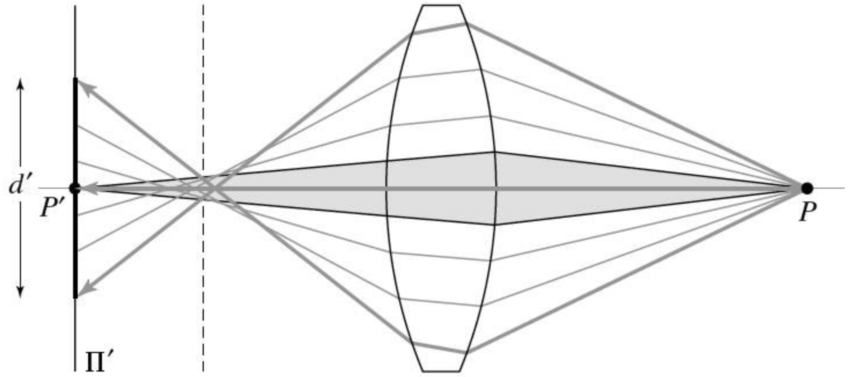
Depth of Field and the Effect of Aperture



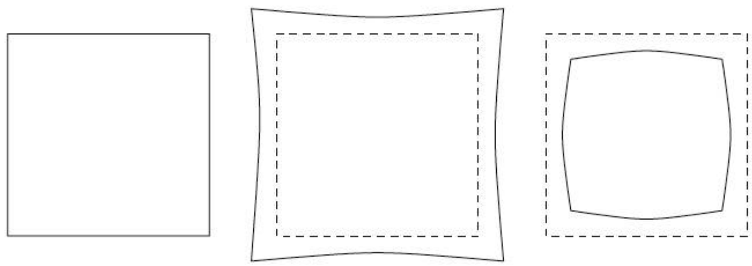
From Photography, London et al.

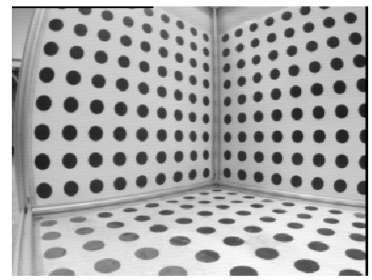
Aberrations

Spherical, coma, astigmatism, field curvature blur the image

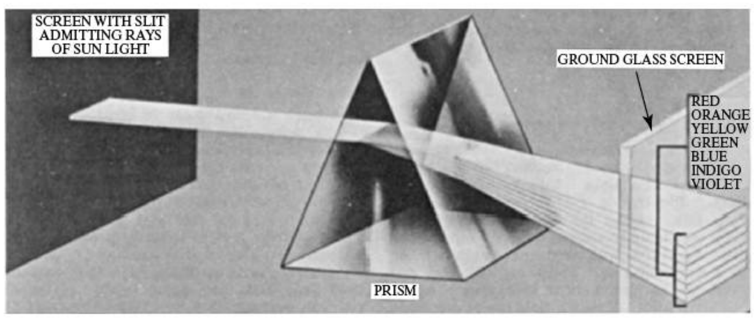


Distortion
(we'll model it)

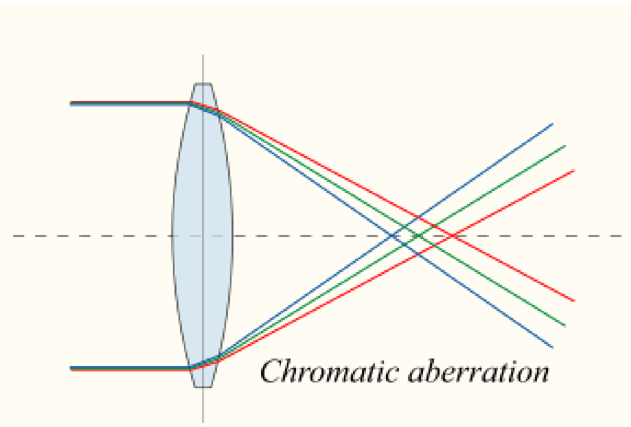




Chromatic

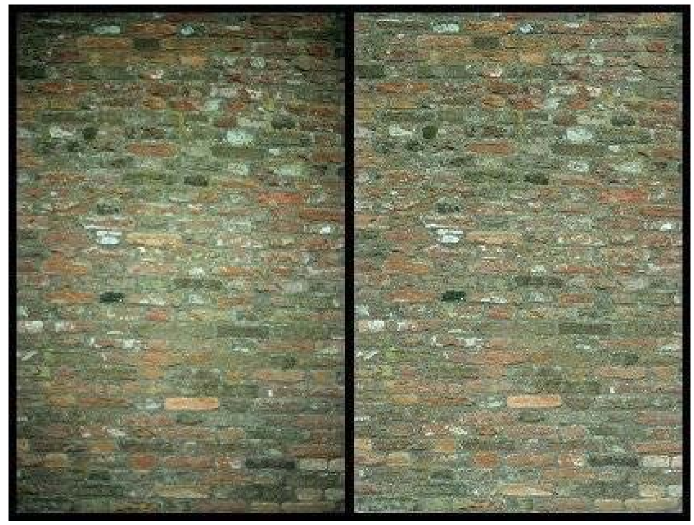


Chromatic Aberration



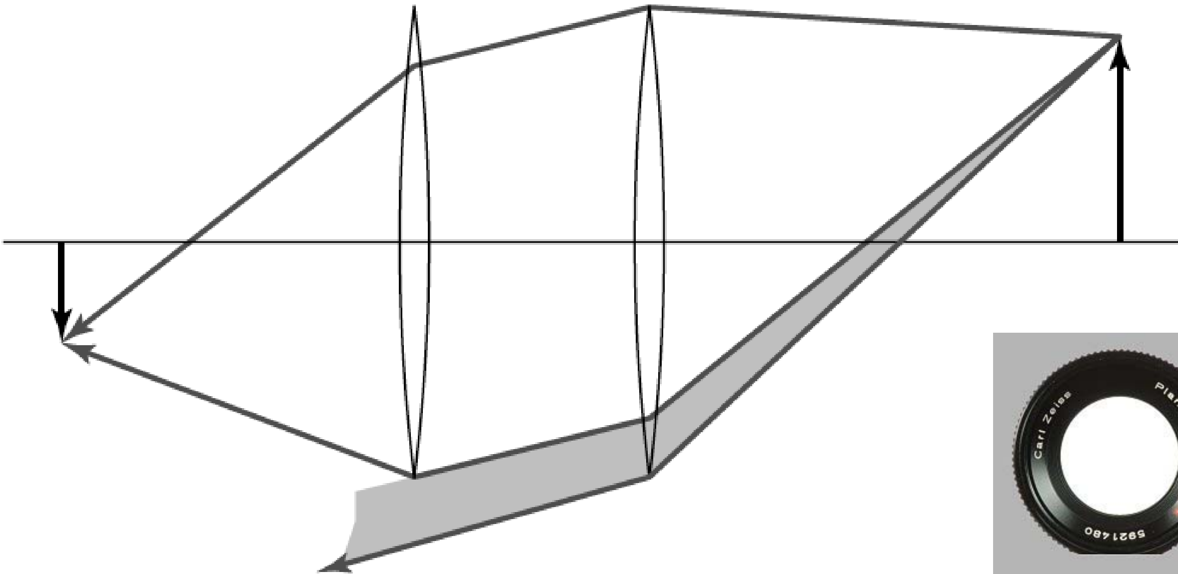
From Wikipedia

Vignetting



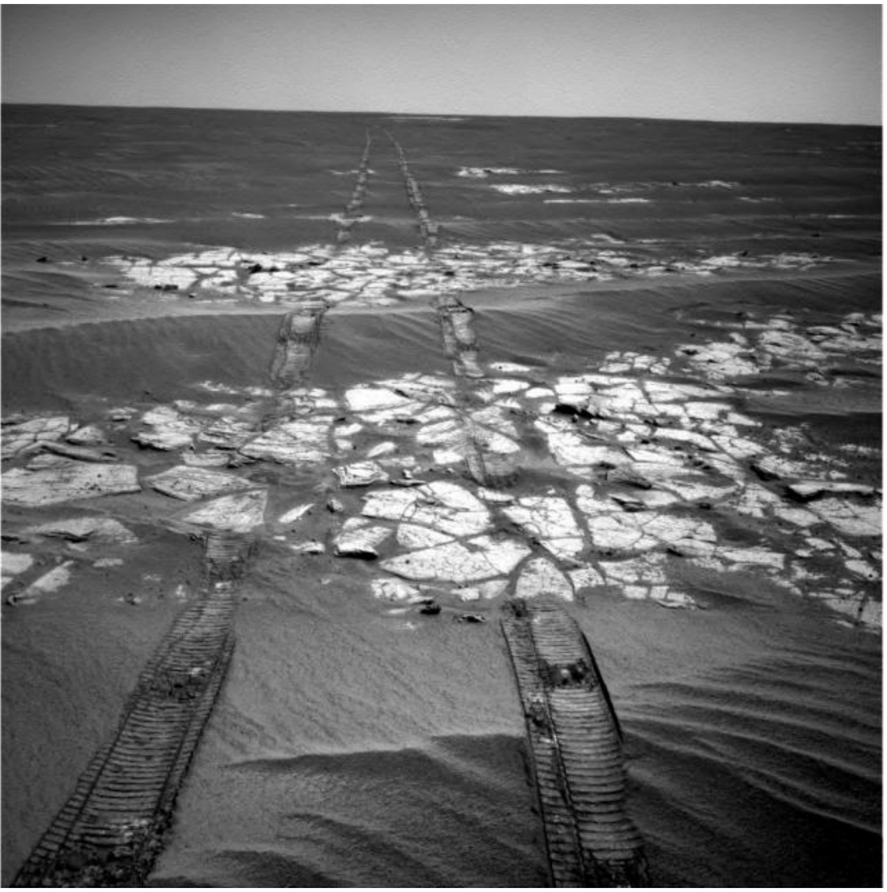
http://toothwalker.org/optics/vignetting.html!

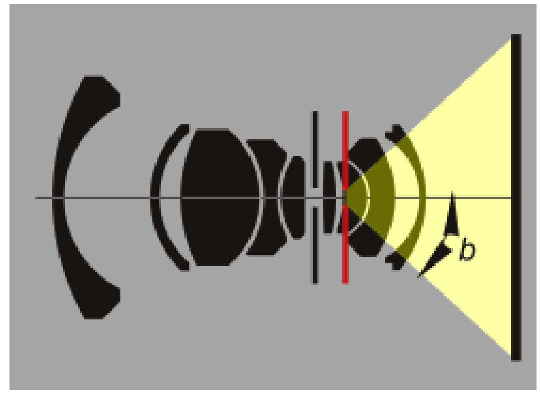
Optical Vignetting

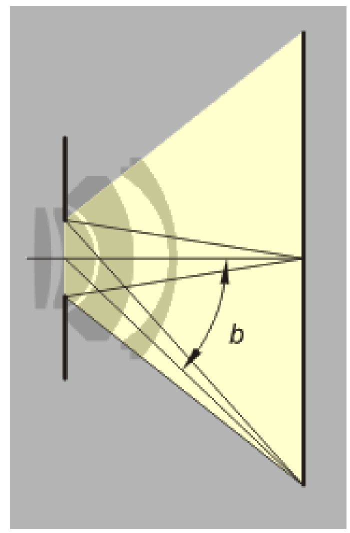


http://toothwalker.org/optics/vignetting.html

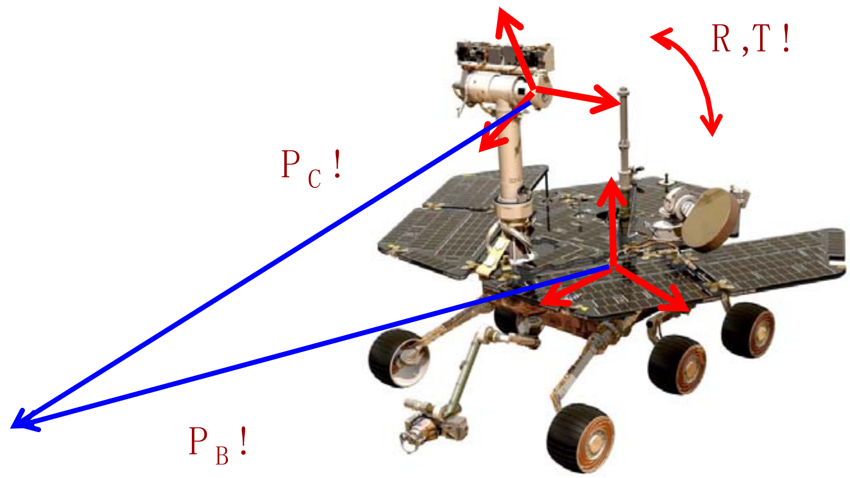
Natural Vignetting ($Cos^4\ \theta$)





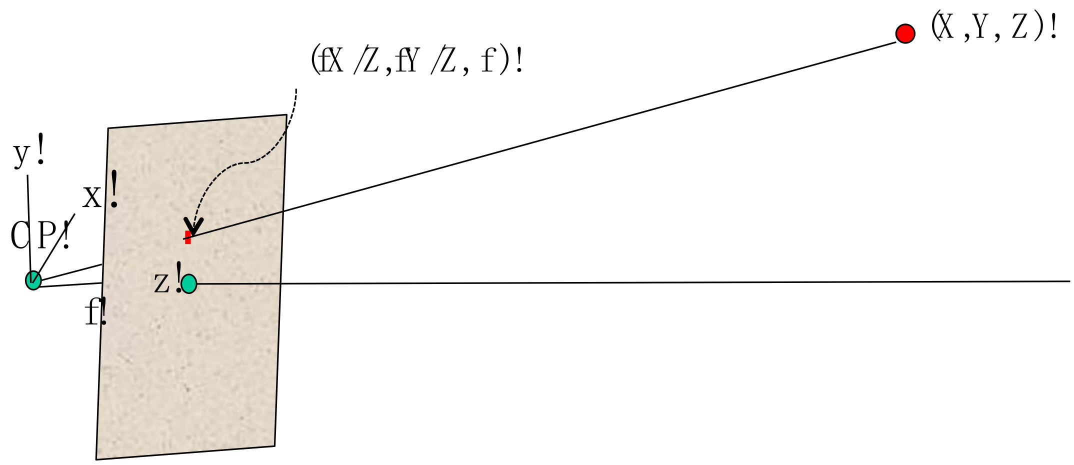


Motivation for Camera Calibration



Szeliski "Computer Vision" text section 2.1 has a tutorial on geometric transformations and homogeneous coordinates

Perspective Projection Revisited



(X, Y, Z) ! (fX/Z, fY/Z, f) ! (fX/Z, fY/Z) in image!

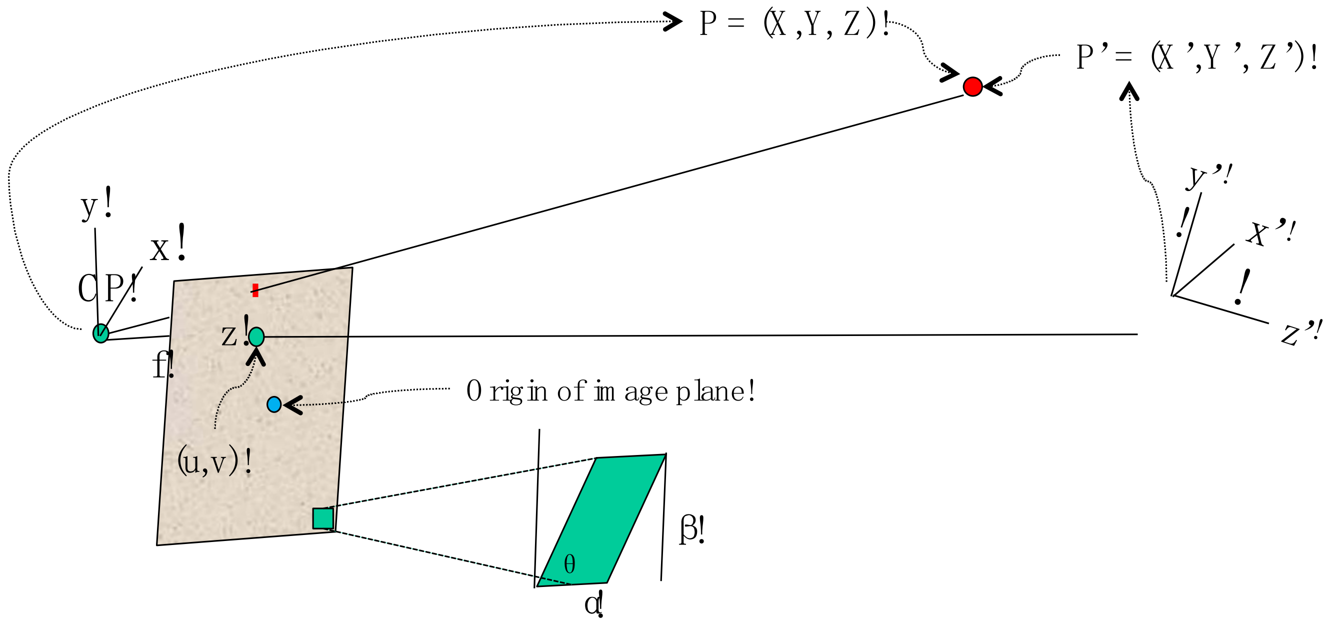
!

Implicit assumptions: (1) Only coordinate frame centered at CP. !

(2) Origin of image plane = intersection point with z-axis. !

(3) Image plane is uniform !

Perspective Projection Generalized



• Let (R,T) = Euclidean transformation from global to camera frame s.t. $R*P' + T = P$ is expression for P' in camera frame.

• (u,v) = coordinates of principal point (intersection of z axis of camera frame with image plane) in image plane.

• Plane has skew θ and scale factors α and β (assume $\theta = 0$ for simplicity).

Perspective Projection Generalized

- Projection from 3D to image plane works as follows:
- P' = (X', Y', Z') ! P = (X, Y, Z) via P = R*P' + T

- P = (X, Y, Z) ! P_n = (X/Z, Y/Z) = normalized coordinates (f = 1 with idealized perspective projection)

- P_n ! I = (x, y) = image coordinates via

$$x = \alpha \frac{X}{Z} + u$$

$$y = \beta \frac{Y}{Z} + v$$

Perspective Projection

$$p = \begin{bmatrix} \lambda x \\ \lambda y \\ \lambda \end{bmatrix} = \begin{bmatrix} \alpha & 0 & u \\ 0 & \beta & v \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} R & T \end{pmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = K_{3x3} E_{3x4} P = M_{3x4} P$$

- Observe $P \approx (X,Y,Z,1)$, and $p \approx (x,y,1)$

Intrinsics vs. Extrinsics

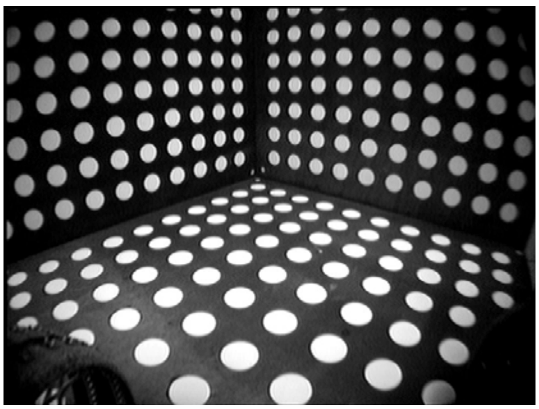
$$p = \begin{bmatrix} \lambda x \\ \lambda y \\ \lambda \end{bmatrix} = \begin{bmatrix} \alpha & 0 & u \\ 0 & \beta & v \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} R & T \end{pmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = K_{3x3} E_{3x4} P = M_{3x4} P$$

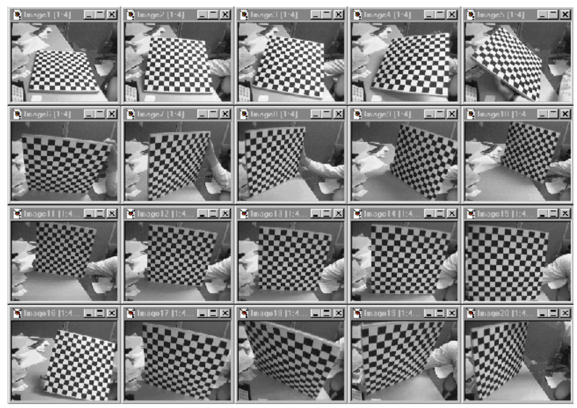
- Extrinsics are the pose (position and orientation) of the camera in some pre-defined coordinate frame. Captured in Euclidean motion matrix E.
- Intrinsics (α, β, u, v) describe the projection geometry of the camera independent of its pose (position and orientation). This is captured in the calibration matrix K.
- The product of the above is called the projection matrix (M in above notation). Given M, we can project any 3D point into image coordinates.
- Observe P≈ (X,Y,Z,1), and p≈(x,y,1)

How to Estimate the Parameters?

$$p = \begin{pmatrix} \lambda x \\ \lambda y \\ \lambda \end{pmatrix} = \begin{pmatrix} \alpha & 0 & u \\ 0 & \beta & v \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} R & T \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = K_{3x3} E_{3x4} P = M_{3x4} P$$

- Typical input derived from pictures of a known object





- Object must have 3-D extent to estimate all parameters

Solving for the Projection Matrix, M

- Suppose we have a set of point correspondences $\{P_i\} \leftrightarrow \{p_i\}$ between 3D and 2D, where the full 3D is known in some coordinate frame. For each point, we have

$$p_i = \begin{bmatrix} \lambda_i x_i \\ \lambda_i y_i \\ \lambda_i \end{bmatrix} = M \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{bmatrix} \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix} = M P_i$$

$$x_i = \frac{m_{11}X_i + m_{12}Y_i + m_{13}Z_i + m_{14}}{m_{31}X_i + m_{32}Y_i + m_{33}Z_i + m_{34}}$$

$$y_i = \frac{m_{21}X_i + m_{22}Y_i + m_{23}Z_i + m_{24}}{m_{31}X_i + m_{32}Y_i + m_{33}Z_i + m_{34}}$$

$$p_i = F(\Phi, P_i)$$

$$\Phi = \arg\min_{\Phi} \sum_i \left\| p_i - F(\Phi, P_i) \right\|^2$$

A Standard Trick

- Rewrite the last set of equations as linear system in m_{ij}

$$A_{2n\times 12}\,\underline{m}_{12\times 1} = \begin{bmatrix} X_1 & Y_1 & Z_1 & 1 & 0 & 0 & 0 & 0 & -x_1X_1 & -x_1Y_1 & -x_1Z_1 & -x_1 \\ 0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 & -y_1X_1 & -y_1X_1 & -y_1X_1 & -y_1 \\ X_2 & Y_2 & Z_2 & 1 & 0 & 0 & 0 & 0 & -x_2X_2 & -x_2Y_2 & -x_2Z_2 & -x_2 \\ 0 & 0 & 0 & 0 & X_2 & X_2 & X_2 & 1 & -y_2X_2 & -y_2Y_2 & -y_2Z_2 & -y_2 \\ & & & & & & & & & & & \\ . & . & . & . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . & . & . & . \\ X_n & Y_n & Z_n & 1 & 0 & 0 & 0 & 0 & -x_nX_n & -x_nY_n & -x_nZ_n & -x_n \\ 0 & 0 & 0 & 0 & X_n & Y_n & Z_n & 1 & -y_nX_n & -y_nY_n & -y_nZ_n & -y_n \end{bmatrix} \begin{bmatrix} m_{11} \\ m_{12} \\ m_{13} \\ . \\ \\ . \\ . \\ m_{33} \\ m_{34} \end{bmatrix} = 0_{6\times 1}$$

- For 6 non-coplanar points, matrix A has rank 11, hence 1-dimensional kernel. Let $A = UDV^T$ be SVD. It follows that $\underline{m}$ is column of V corresponding to smallest singular value in D.

Solving for Intrinsics and Extrinsics

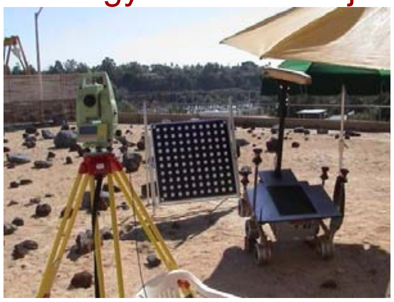
- Having solved for M up to scale, write it explicitly in terms of K, E

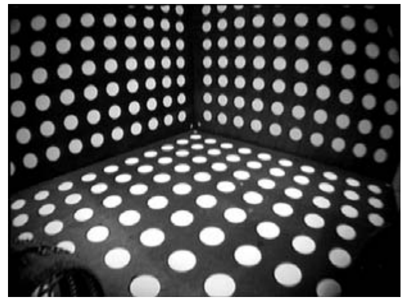
$$\begin{pmatrix} \alpha & 0 & u \\ 0 & \beta & v \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} R & T \end{pmatrix} = K \begin{pmatrix} R & T \end{pmatrix} = \begin{pmatrix} KR & KT \end{pmatrix} = \lambda M_{3\times 4}$$

- Let N = left 3x3 block of M, and S = last column of M
- Observe that $KR(KR)^T = KK^T = \lambda^2 NN^T$. Since $(KK^T)_{3,3}=1$, we can solve for λ up to a sign. Remaining solutions for K is straightforward algebra.
- Given K and λ, $R = \lambda K^{-1}N$. $T = \lambda K^{-1}S$.
  - Use SVD (orthonormal Procrustes) to refine R.

About Extrinsics

- 3D coordinates $\{P_i\}$ obtained from
  - Explicit metrology or known object geometry



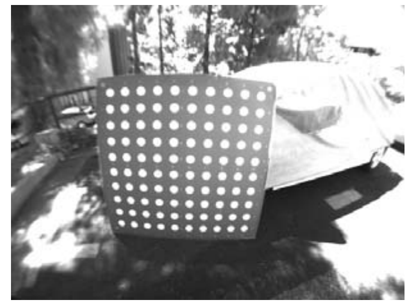


  - For target planes, only per-plane geometry is known. Full 3D (i.e. orientation of plane in fixed coordinate frame is not known)

Real Cameras

Strong perspective:
Angles are not preserved
The projections of parallel lines intersect at one point

Weak perspective:
Angles are better preserved
The projections of parallel lines are (almost) parallel



- But what about this?

Real Cameras

- Perspective projection is idealized. No real camera works like this.
  - Often a good approximation for narrow FOV $\rightarrow$ close to pinhole camera
- More sophisticated models are needed for more complex cases
  - Radial distortion is a radially symmetric function of distance from center of distortion.
  - Tangential distortion: arises from lens decentering, also commonly modeled
- We will apply non-linear distortion after Euclidean motion but before the camera matrix: i.e. in normalized coordinates. Any advantage to this ordering?

- Consider:

$$Q = \begin{bmatrix} Q_x \\ Q_y \\ Q_z \end{bmatrix} = \begin{pmatrix} R & T \end{pmatrix} P$$

$$p_n = \begin{bmatrix} x_n \\ y_n \end{bmatrix} = \begin{bmatrix} Q_x / Q_z \\ Q_y / Q_z \end{bmatrix}$$

Radial and Tangential Distortion

- Let (x_u, y_u) be the result of applying radial and tangential corrections to (x_n, y_n)

$$x_u = x_n + x_n\left(k_1 r^2 + k_2 r^4 + k_3 r^6\right) + 2 p_1 x_n y_n + p_2\left(r^2 + 2 x_n^2\right)$$
$$y_u = y_n + y_n\left(k_1 r^2 + k_2 r^4 + k_3 r^6\right) + 2 p_2 x_n y_n + p_1\left(r^2 + 2 y_n^2\right)$$

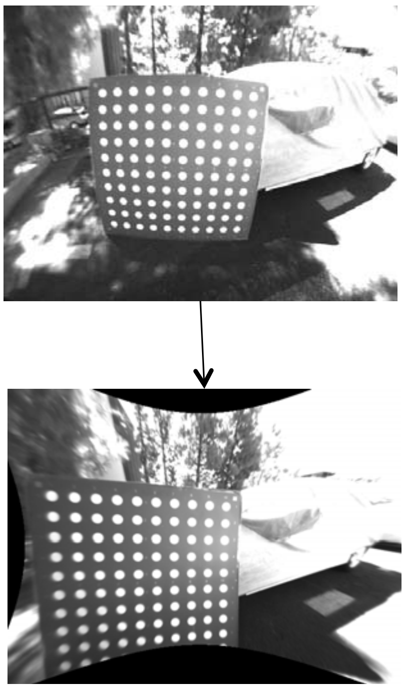
- Then the image coordinates corresponding to P are given by

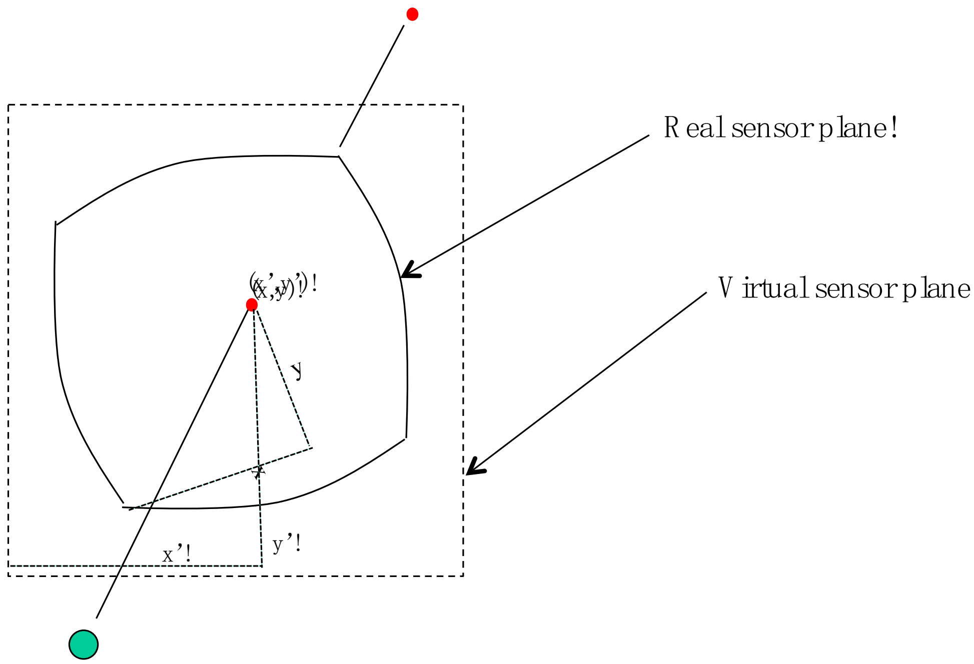
$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = K \begin{bmatrix} x_u \\ y_u \\ 1 \end{bmatrix}$$

Where K is the calibration matrix discussed above.

“Linearizing” an Image

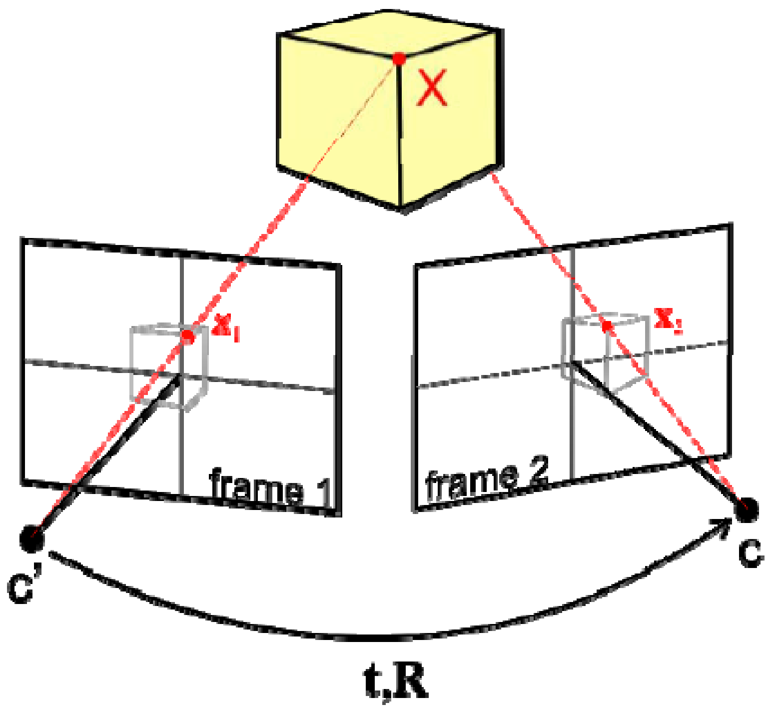
- Given an image with non-linear distortion, we generate a virtual image that is pure perspective. Imagine a virtual perspective camera with projection center identical to the real camera’s. What would this virtual camera see?





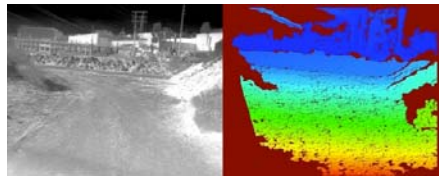
Overview of 3-D Sensors

Passive Triangulation-based Ranging

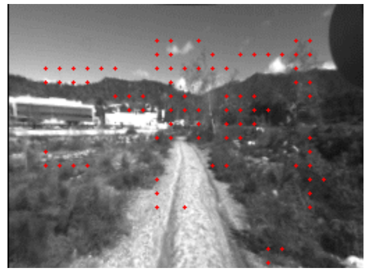


Passive Triangulation
(and Optional Pattern Projection)

- Passive stereo
  - Works reasonably well with visible and thermal images
  - Range error quadratic in range
- Optical flow
  - Essential part of egomotion estimation (visual odometry)
  - As a range sensor, poor resolution near focus of expansion
- Active stereo (pattern projection)
  - Now commercially available, e.g. Ensenso
  - Quite limited range in sunlight

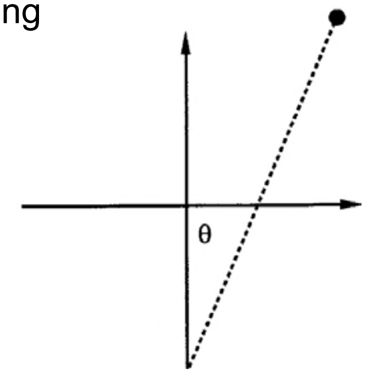


SAD5 local block matching

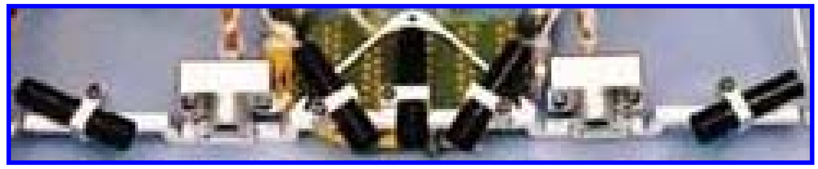


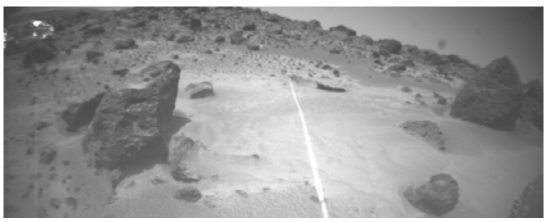
Real-time SAD optical flow

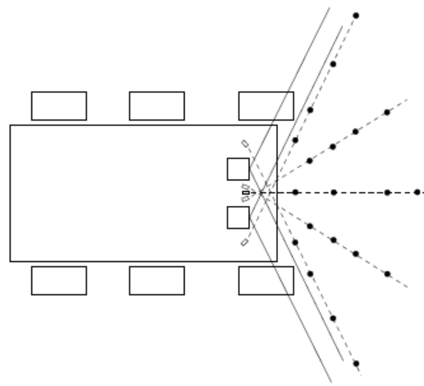
$\sigma_R \sim R^2$



$\sigma_R \sim 1/\tan\ \theta$

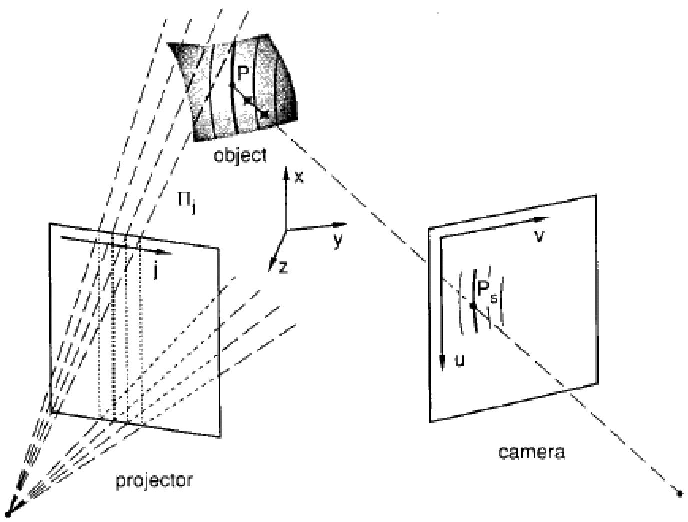


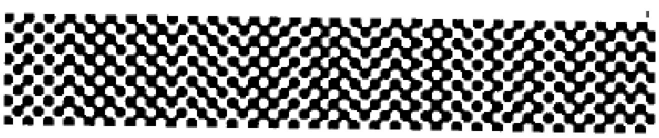




Active Triangulation: Coded Light

P. Vuylsteke and A. Oosterlinck, "Range image acquisition with a single binary-encoded light pattern", IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Volume 12, Issue 2, Pages 148-164, February 1990



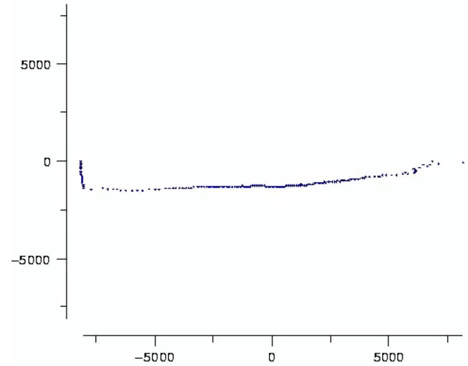


- Principle used by Primesense, now part of Apple, and the Kinect RGBD sensor. Indoor only.
- Also used in handheld 3-D mapping devices by Mantis Vision; useful at several meters range in direct sunlight

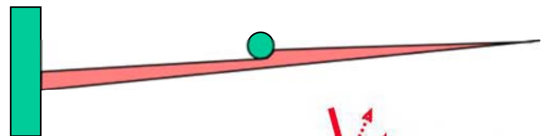
Source: Mantis Vision

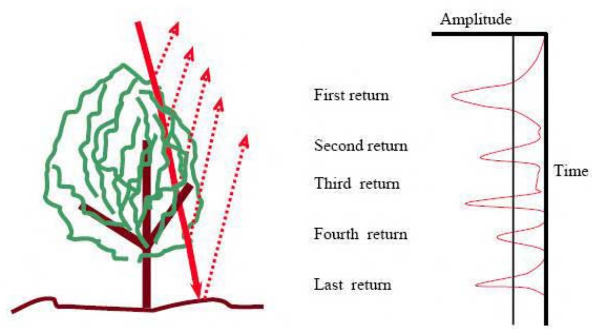
Direct (Pulsed) Time of Flight Lidar

- Principle: measure flight time of a short pulse of light
  - 1-axis and 2-axis scanners
  - Usually get NIR reflectance channel as well
  - SNR depends on albedo and angle of incidence
  - Single echo models can get false readings from dust
  - Multi-echo models useful in dust, foliage removal, etc.
  - “Flash” lidar gets full 2-D array of XYZ measurements in one pulse; needs more power, so trade-offs with max range and field of view



Dust hits with single echo lidar





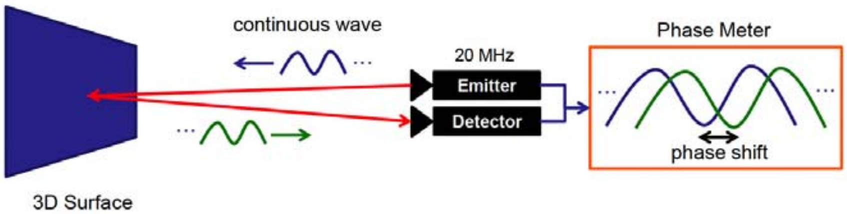
Source: Lohani and Barat

Hokuyo UTM-30LX

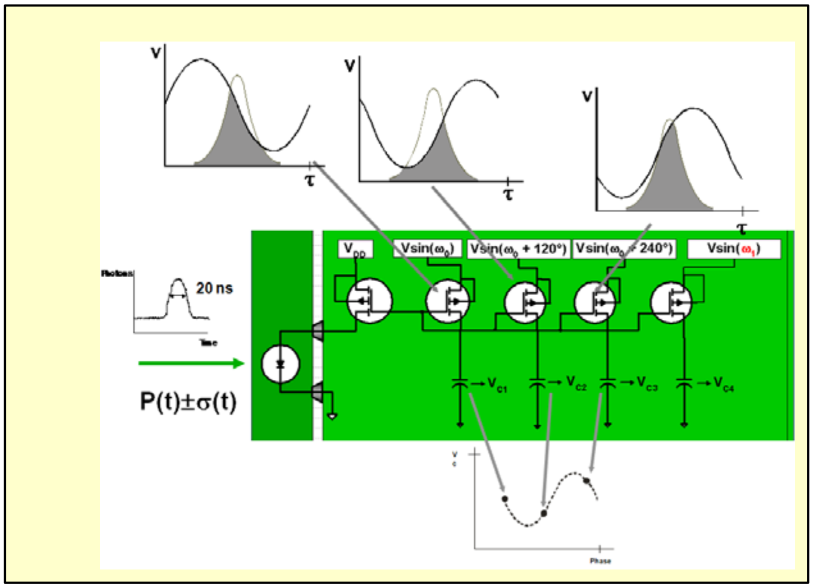
Velodyne HDL32E

Indirect (Phase Measurement) Time of Flight Lidar

- Multiple variants of this idea exist

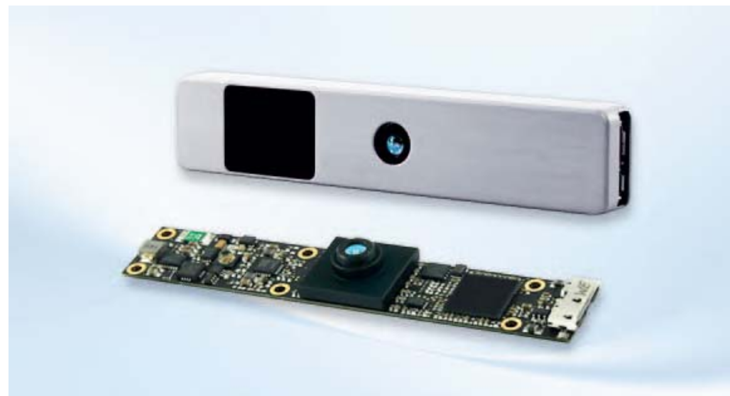


Source: Castaneda and Navab, TU Munchen

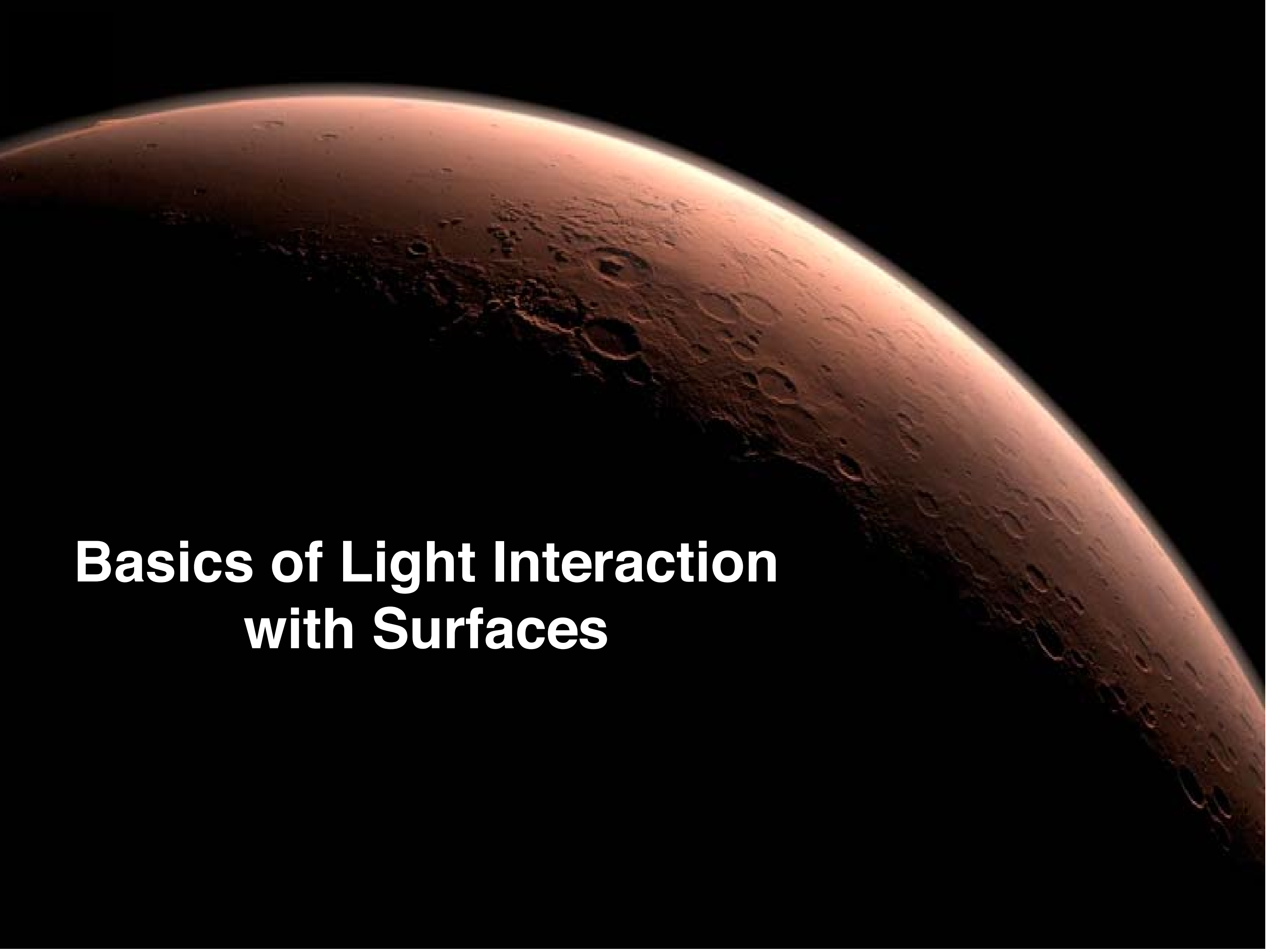


Source: JPL design

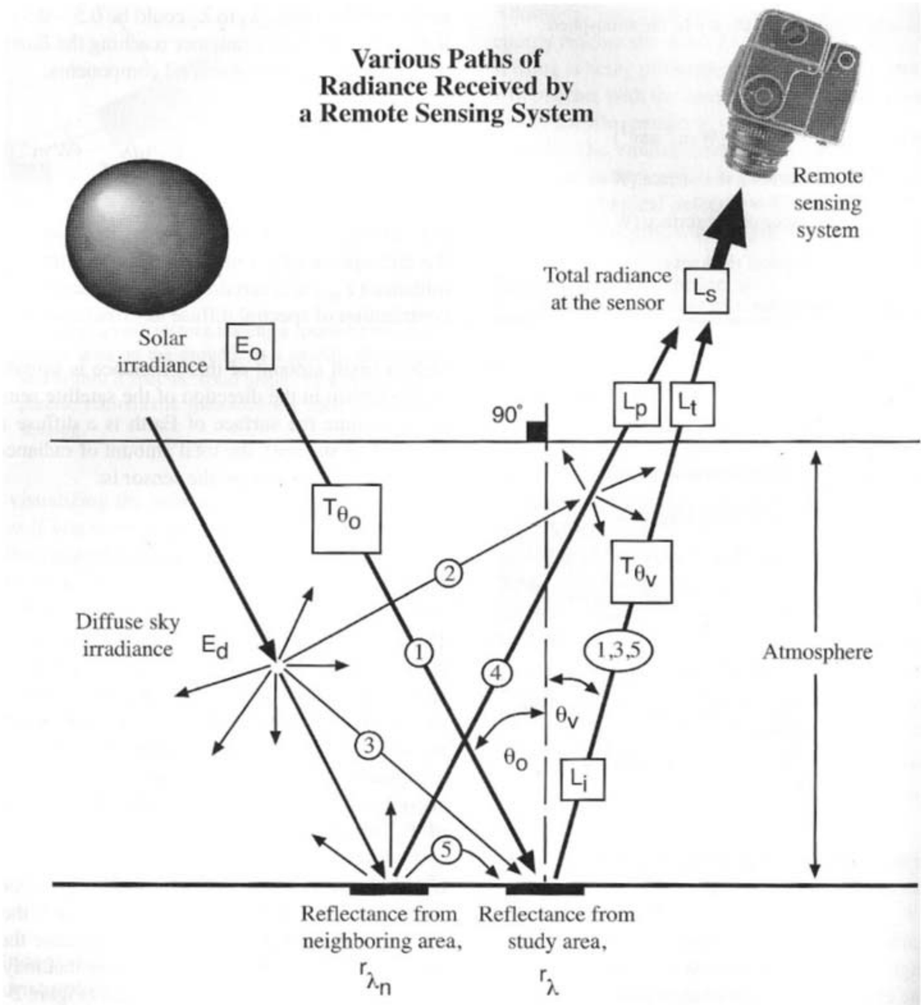
- E.g. the Camboard Pico from PMDTec/Infineon, 8.5 cm long, range of 1 m for 160x120 pixels, 90° FOV. Direct sunlight may be possible at short range



Source: Infineon



Basics of Light Interaction with Surfaces


Various Paths of
Radiance Received by
a Remote Sensing System
Remote
sensing
system
Total radiance
at the sensor
L_s
Solar
irradiance
E_o
90°
L_p
L_t
T_{θ_o}
T_{θ_v}
Diffuse sky
irradiance
E_d
1
2
3
4
5
1,3,5
Atmosphere
θ_v
θ_o
L_i
Reflectance from
neighboring area,
r_{λ_n}
Reflectance from
study area,
r_λ

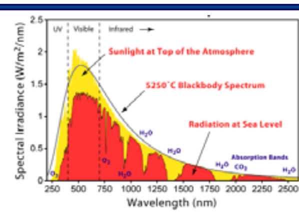
Color Response in Cameras

$$p = \int \sigma(\lambda)\rho(\lambda)S(\lambda)d\lambda$$

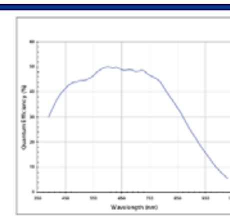
Photoreceptor sensitivity

Spectral albedo ("object color")

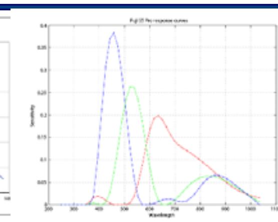
Illuminant color

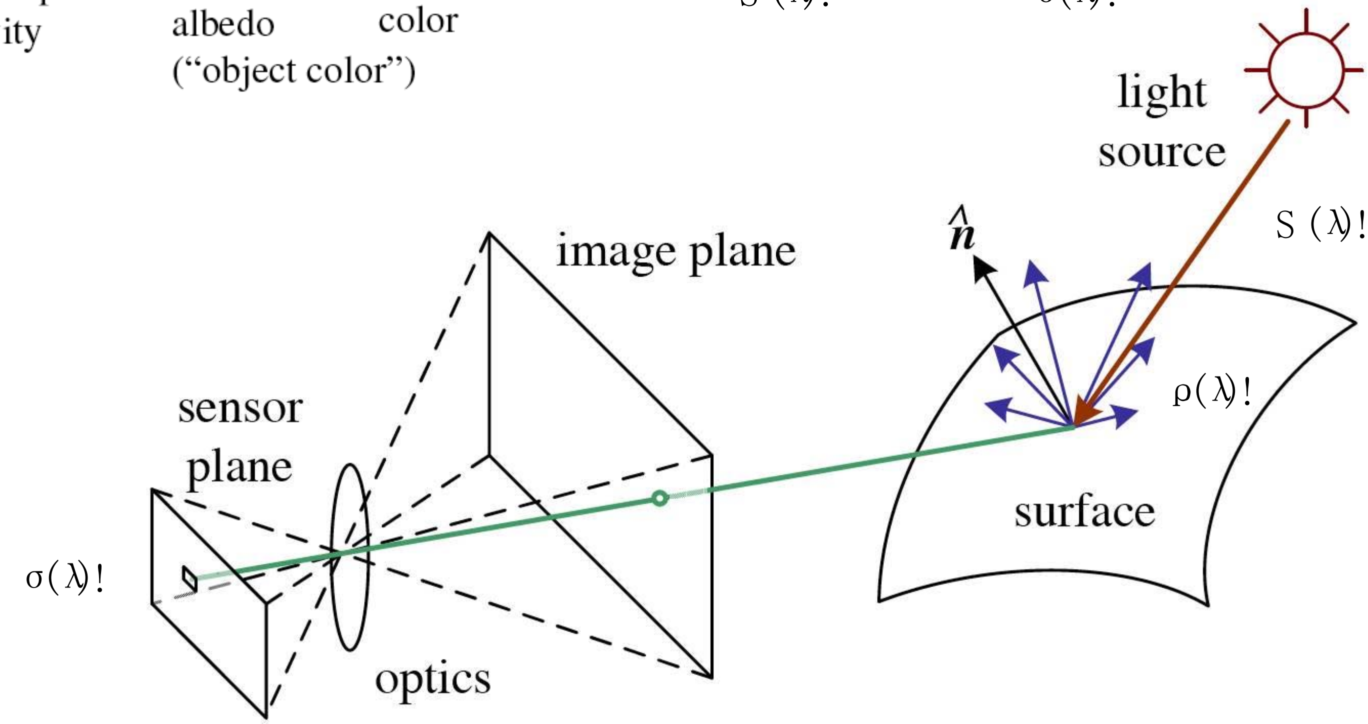


S (λ)!

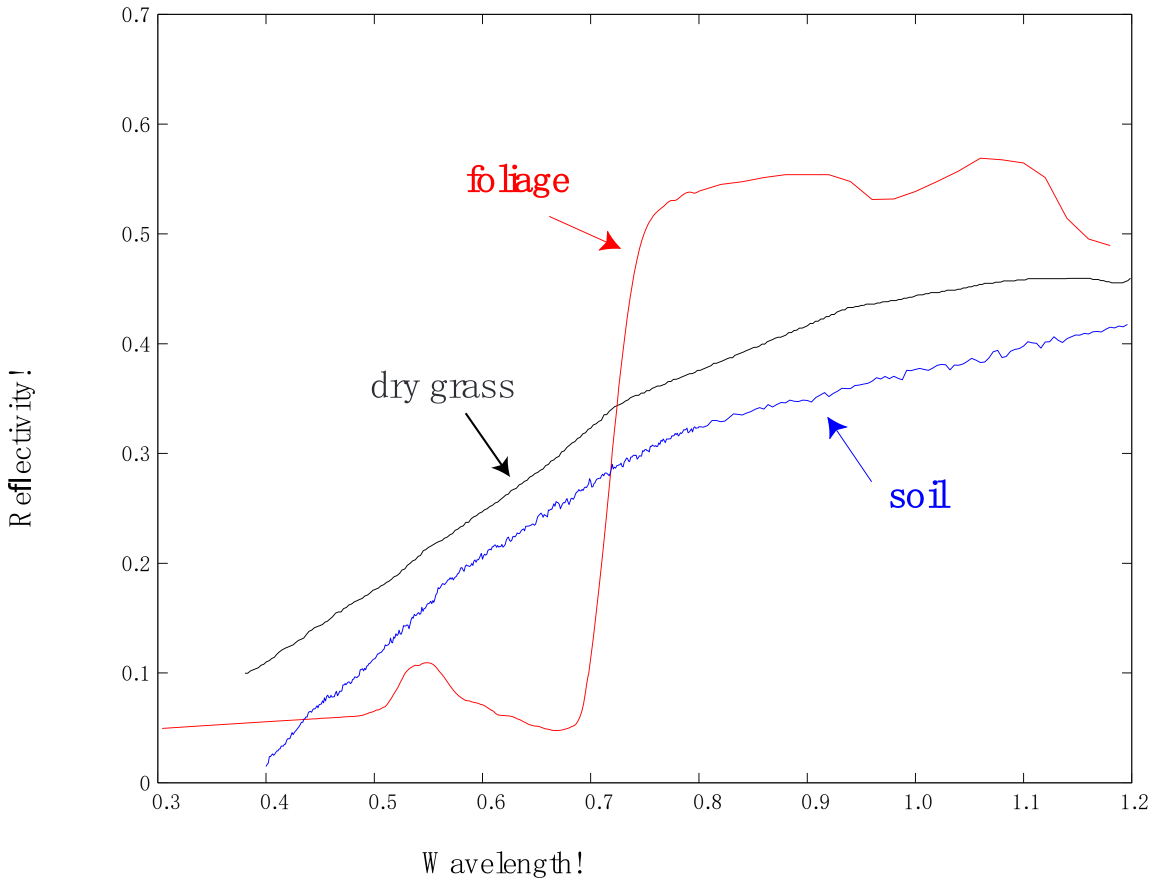


σ(λ)!





Spectral Reflectance Example: Visible and Near Infrared (VNIR)



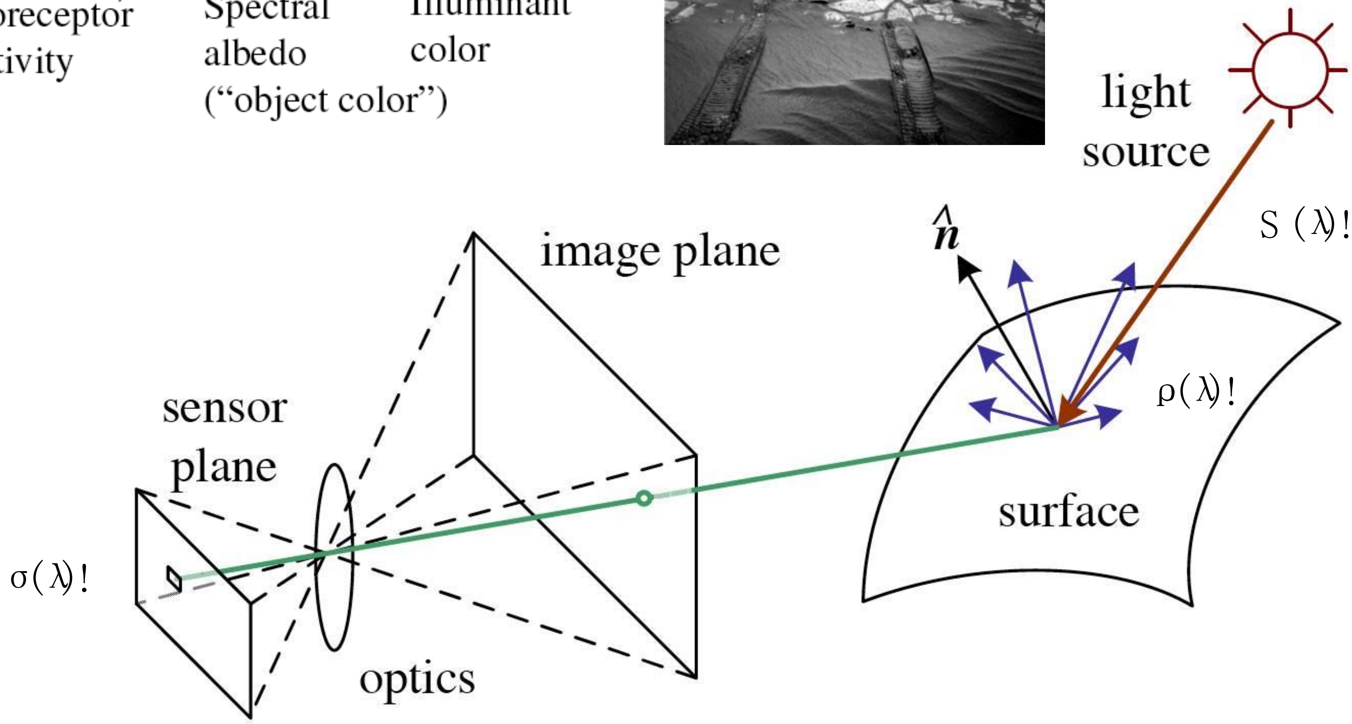
Color Response in Cameras

$$p = \int \sigma(\lambda)\rho(\lambda)S(\lambda)d\lambda$$

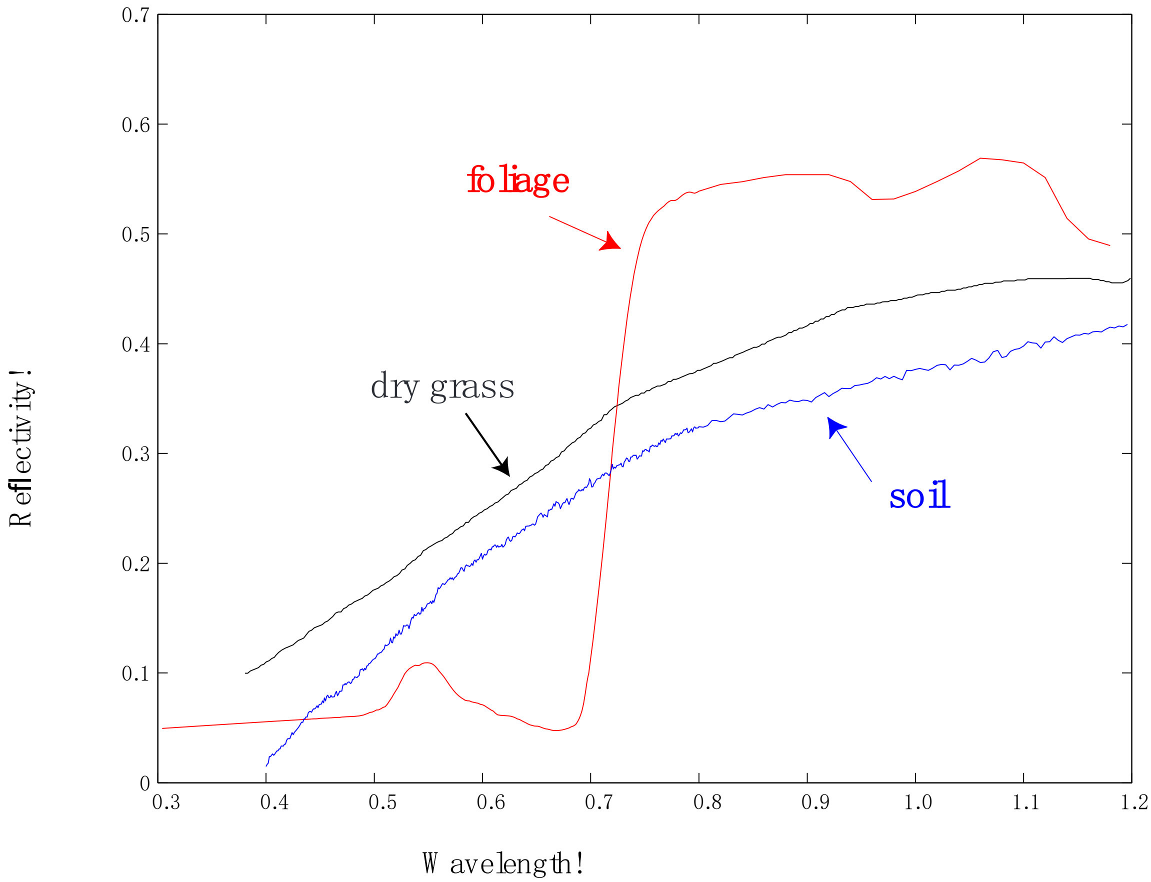
Photoreceptor sensitivity

Spectral albedo ("object color")

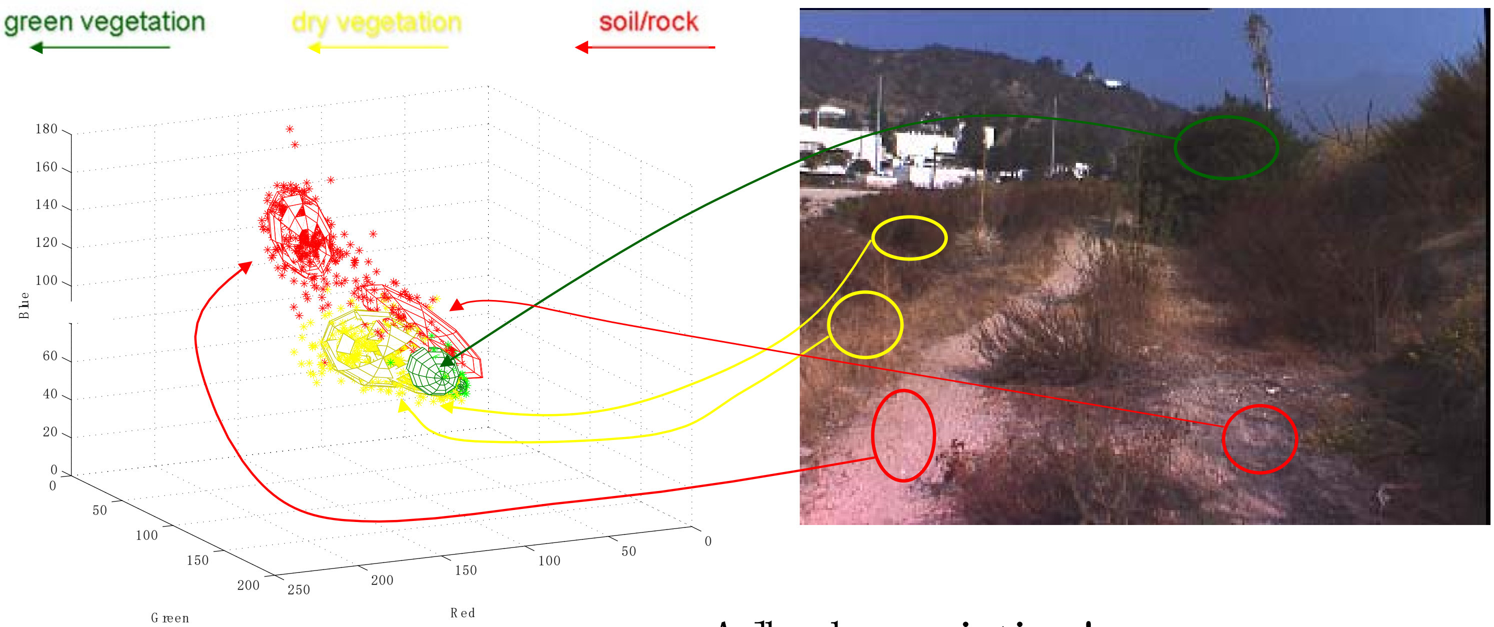
Illuminant color



Spectral Reflectance Example: Visible and Near Infrared (VNIR)

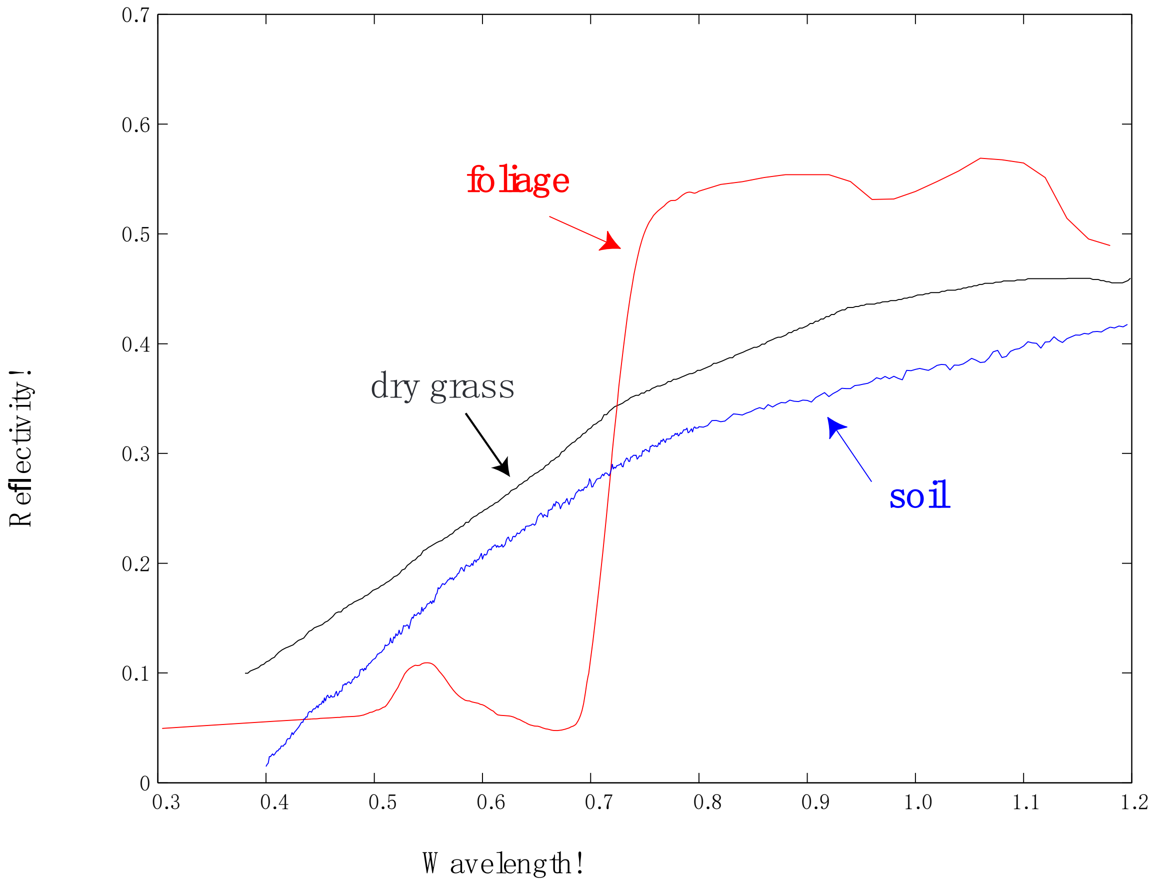


RGB Color Space



- Classes may overlap and have large variance!
- Causes of variation:!
- Albedo variation!
- Power of illumination!
- Spectral content of illumination!
- Illumination and viewing geometry!
- Noise!

Spectral Reflectance Example: Visible and Near Infrared (VNIR)



Spectral Reflectance: Visible and Near Infrared (VNIR)

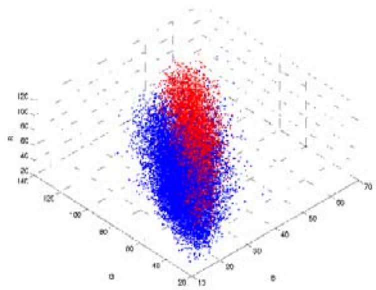
450nm image

550nm image

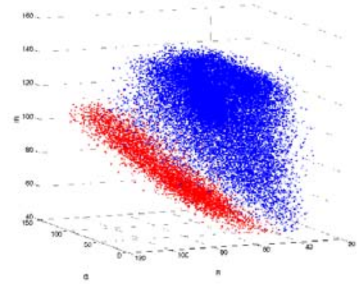
650nm image

800nm image

NIR Channel Better Separates the Clusters



B-G-R !



G-R-IR !

Vegetation Classification with R:NIR Ratio

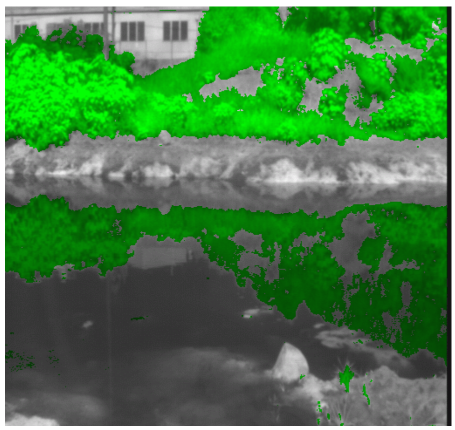
:

=

650nm image!

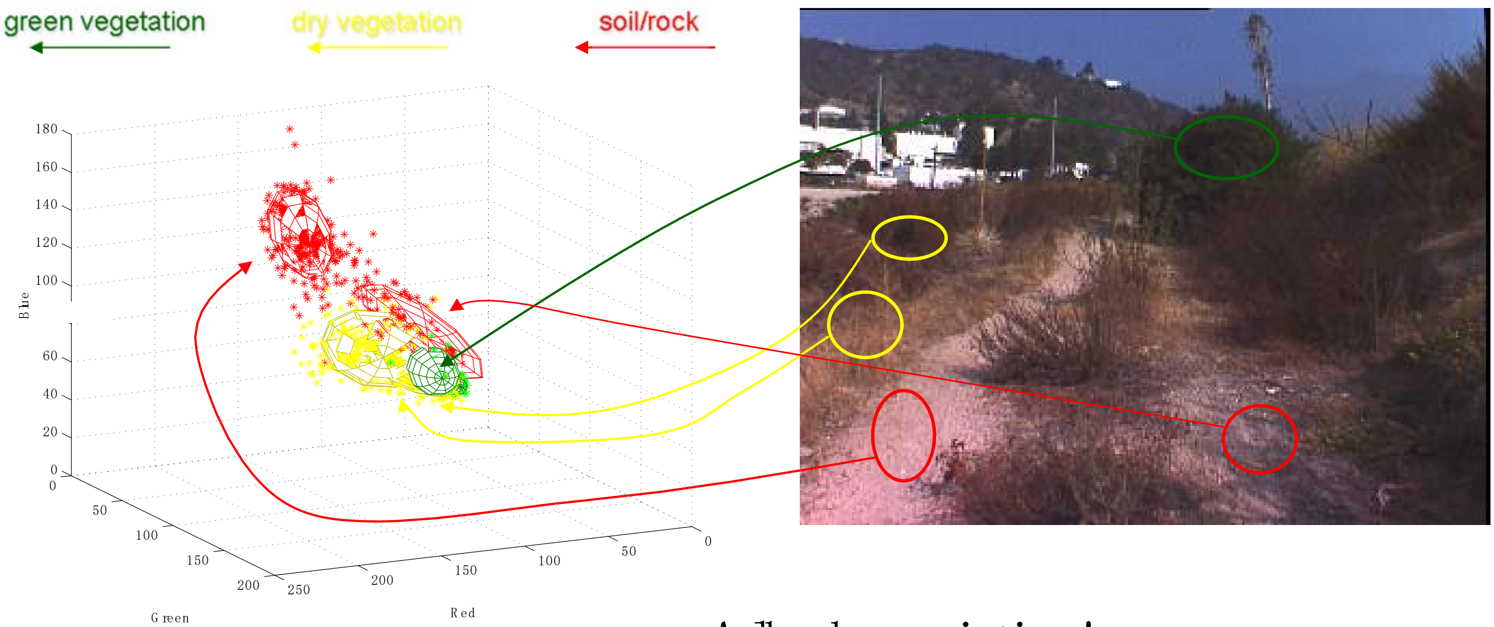
800nm image!

650:800 ratio image!



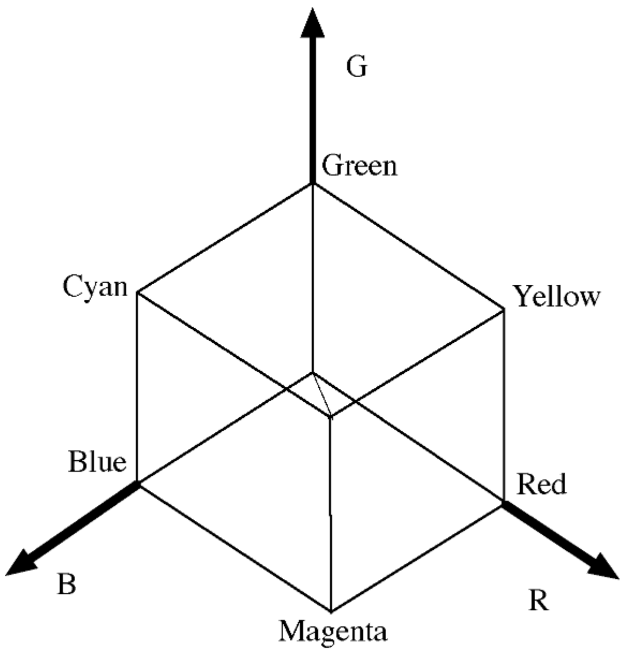
Classified image!

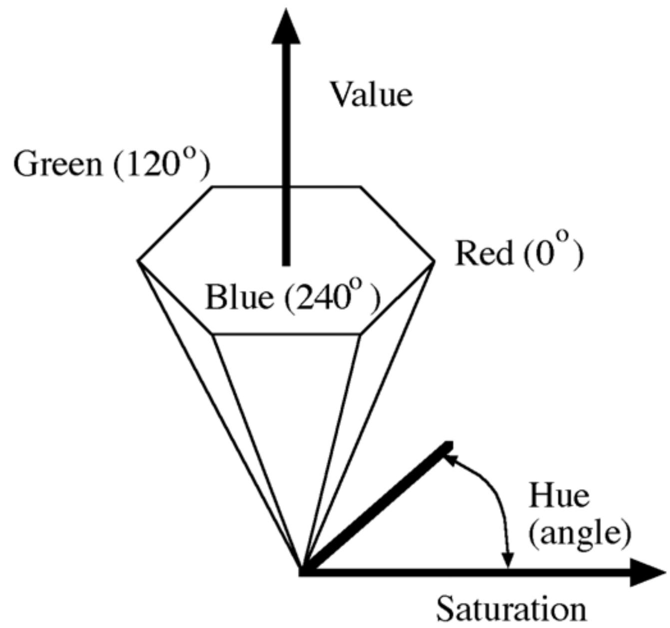
Color



- Classes may overlap and have large variance!
- Causes of variation:!
- Albedo variation!
- Power of illumination!
- Spectral content of illumination!
- Illumination and viewing geometry!
- Noise!

Color Space Conversion (e.g. RGB to HSV)



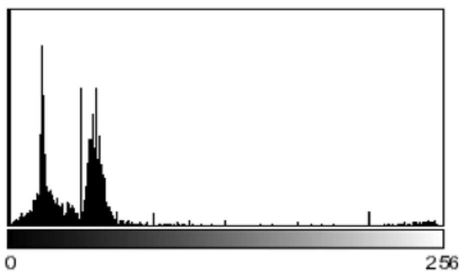


$$HSV(i,j) = f(RGB(i,j))!$$

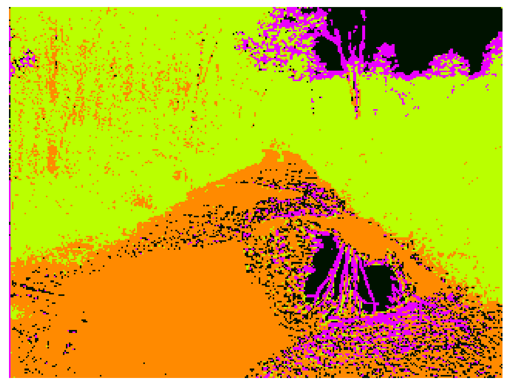
Application of RGB to HSV Color Space Conversion: Simple Segmentation

(a)

(b)



(c)



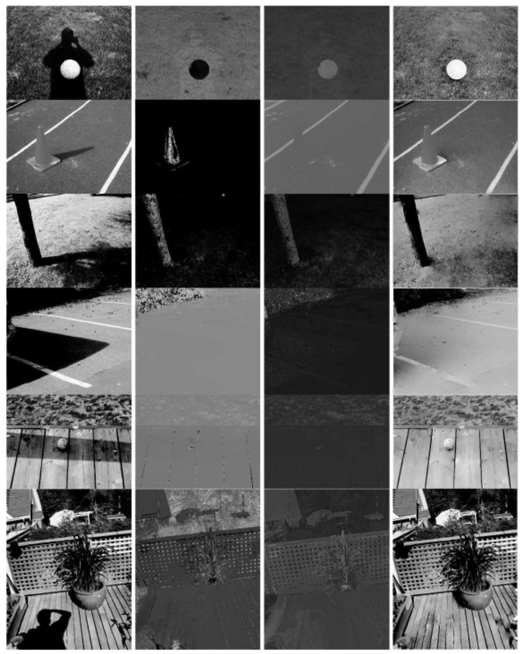
(d)

Color Space Conversion

- There are several other color spaces in use
- CIE and related ones are used in modeling human color perception
- We won't use or discuss these in this course

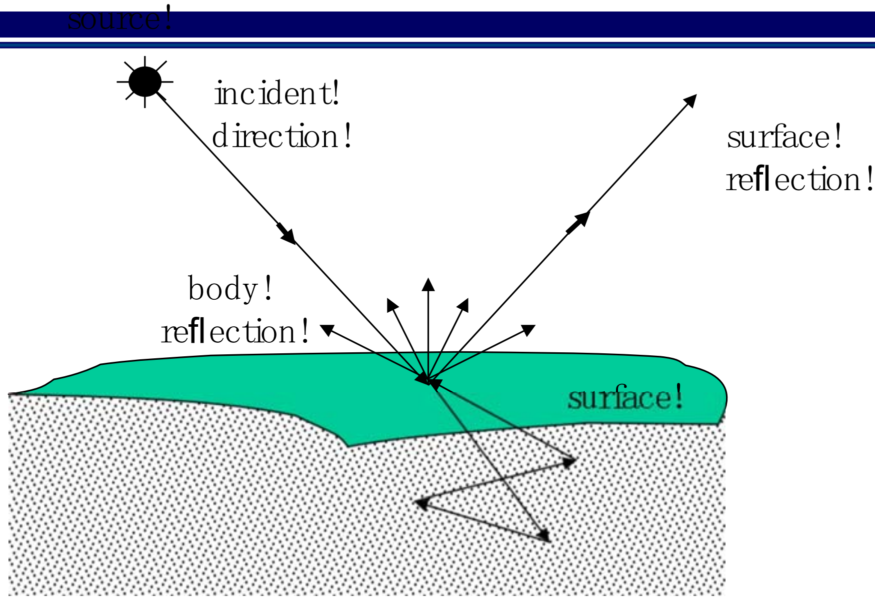
Other Important Applications of Color Analysis

- Shadow removal
- Color constancy
- Both continue to be active research areas
- Both are covered briefly in Forsyth; not covered further in the lectures



Finlayson et al, "On the removal of shadows from images", IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(1), January 2006!

Mechanisms of Reflection



Body Reflection:
- Diffuse Reflection
- Matte Appearance
- Non-Homogeneous Medium
- Clay, paper, etc

Surface Reflection:
- Specular Reflection
- Glossy Appearance
- Highlights
- Dominant for Metals

Image Intensity = Body Reflection + Surface Reflection

Mechanisms of Reflection

source!

Body Reflection:!
!

- !Diffuse Reflection!
- !Matte Appearance!
- !Non-Homogeneous Medium!
- !Clay, paper, etc!

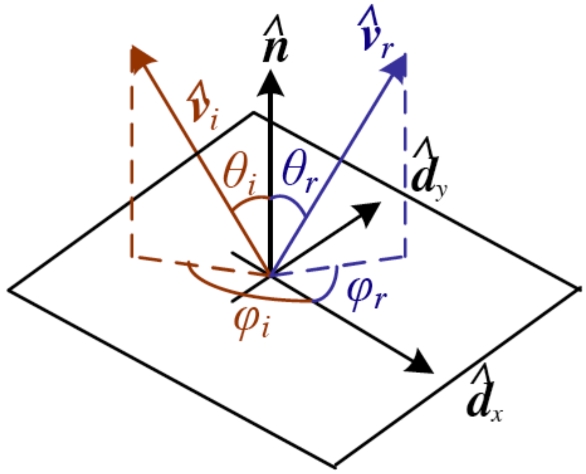
Many materials exhibit both Reflections:!

Surface Reflection:!
!

- !Specular Reflection!
- !Glossy Appearance!
- !Highlights!
- !Dominant for Metals!

Bidirectional Reflectance Distribution Function

Given an incoming ray (θ_i, ϕ_i) and outgoing ray (θ_r, ϕ_r)
what proportion of the incoming light is reflected along outgoing ray?



Answer given by the BRDF:

$$f(\theta_i, \phi_i, \theta_r, \phi_r)$$

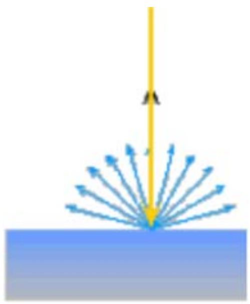
Units: sr^{-1}!

Isotropic case:

$$f(\theta_i, \theta_r, |\phi_i - \phi_r|)$$

Diffuse Reflection

- Diffuse reflection is governed by **Lambert's law**
  - Viewed brightness does not depend on viewing direction
  - Brightness *does* depend on direction of illumination
  - This is the model most often used in computer vision

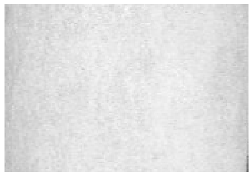
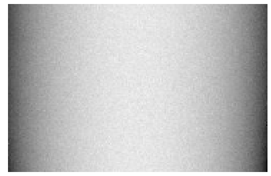


$$f(\theta_i, \phi_i, \theta_r, \phi_r) = k$$

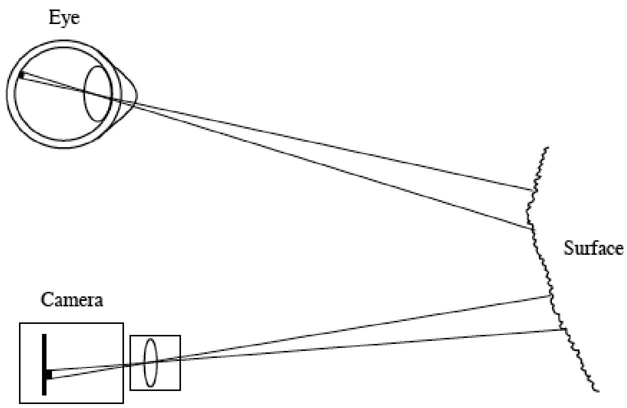
$$E_i = E \cos \theta_i$$

$$L_r = kE \cos \theta_i$$

Lambert's Law is an Idealization; the Real World is More Complex

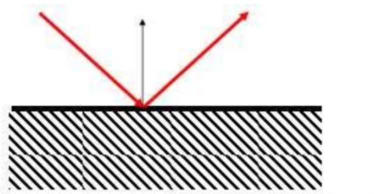
Rough cylinder   Smooth cylinder

(a)

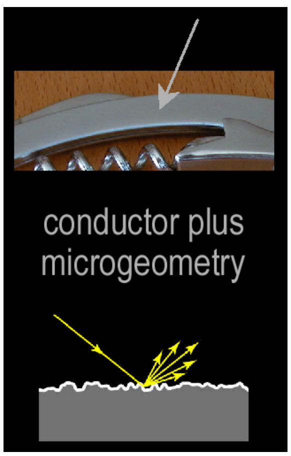


S. Nayar and M. Oren, "Visual Appearance of Matte Surfaces", Science, Vol. 267, pp. 1153-1156, 1995

Other Reflectance Models



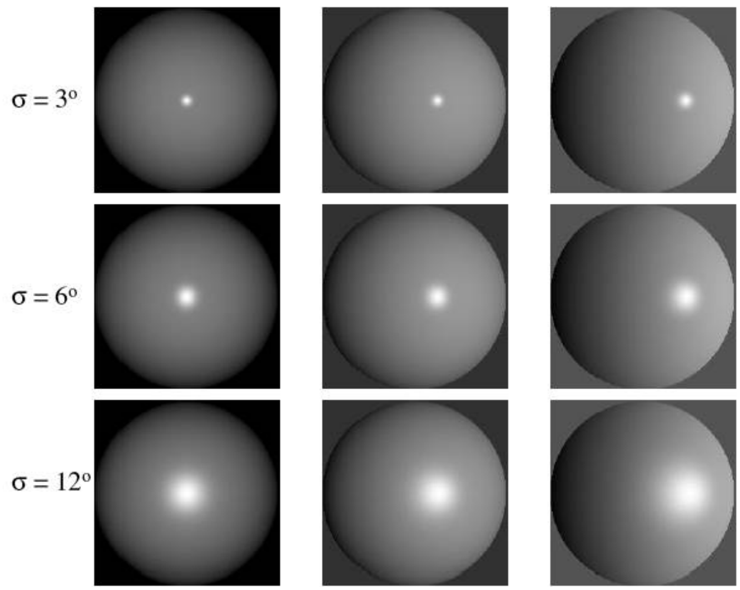
Pure specular



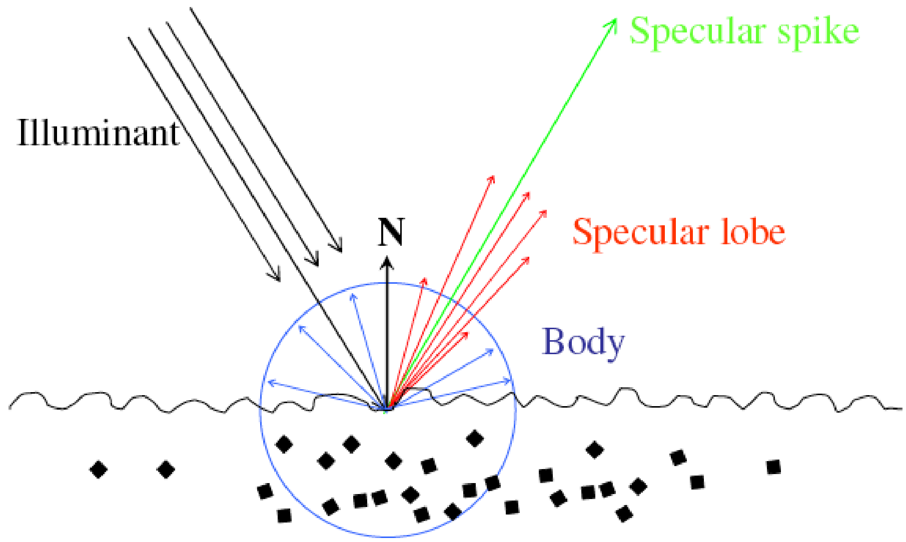
Specular lobe

Varying incidence angle

Varying roughness



Other Reflectance Models



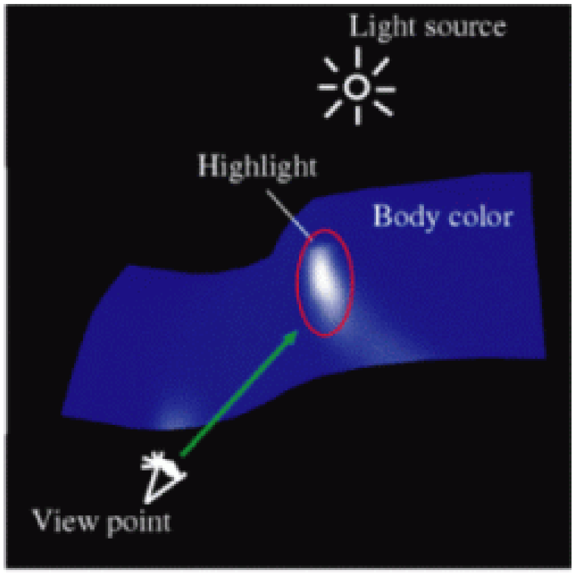
$$a\cos\theta_i + \frac{b}{\cos\theta_e} e^{-\frac{\beta^2}{2\sigma_\alpha^2}} + c\,\delta(\theta_i - \theta_e)\,\delta(\varphi_i - \varphi_e - \pi)$$

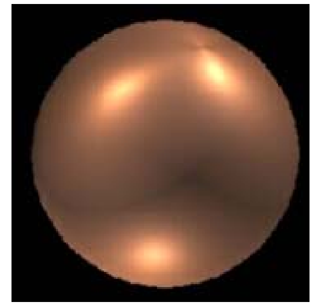
Diffuse lobe — Specular lobe — Specular spike

S. Nayar, K. Ikeuchi, T. Kanade, "Surface Reflection: Physical and Geometrical Perspectives", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 13, No. 7, 1991

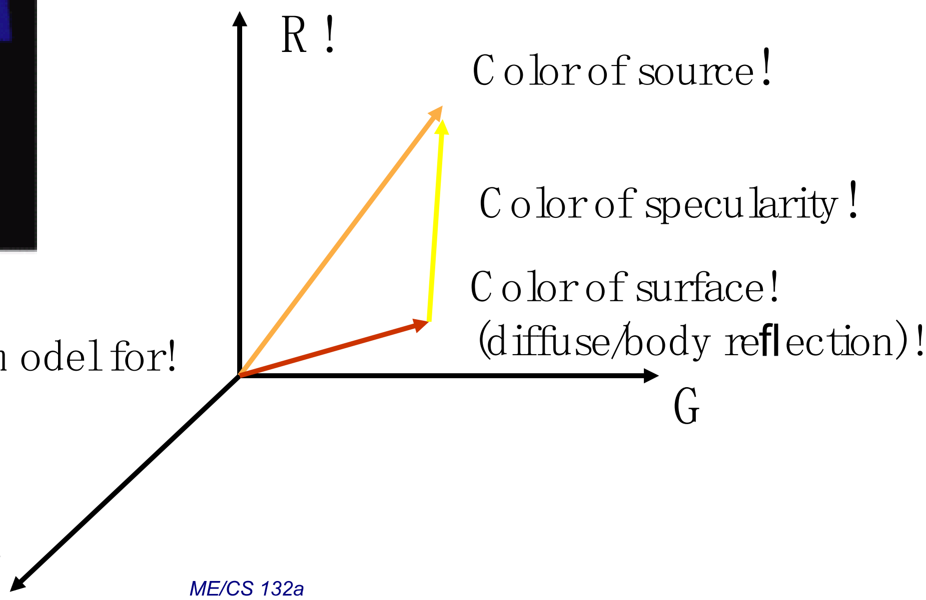
Dichromatic Reflection Model

Observed Image Color = a x Body Color + b x Specular Reflection Color!

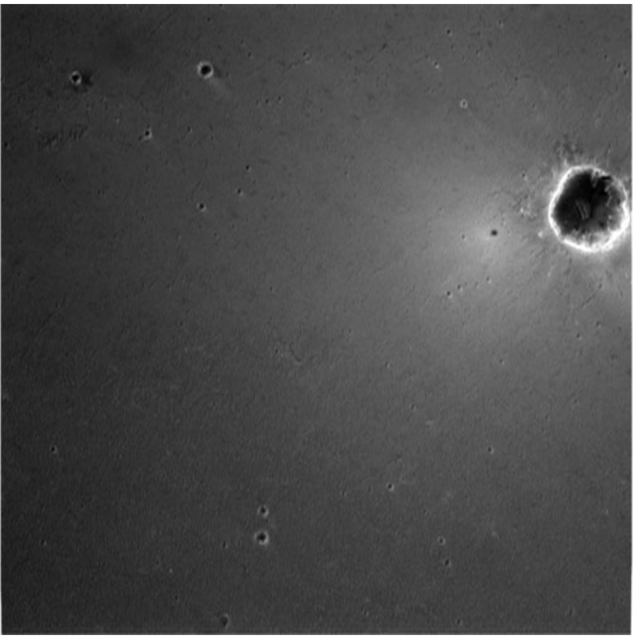




Does not specify any specific model for!
diffuse/specular reflection!



Other Reflectance Models: Opposition Effect



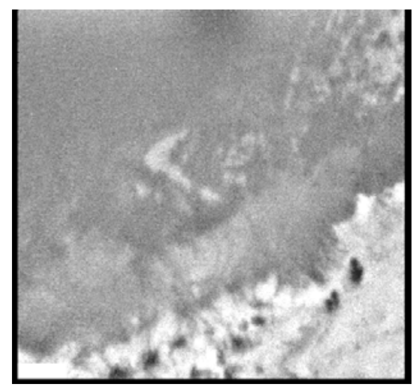
B. Hapke, *Theory of Reflectance and Emittance Spectroscopy*, 1993

Thermal Infrared

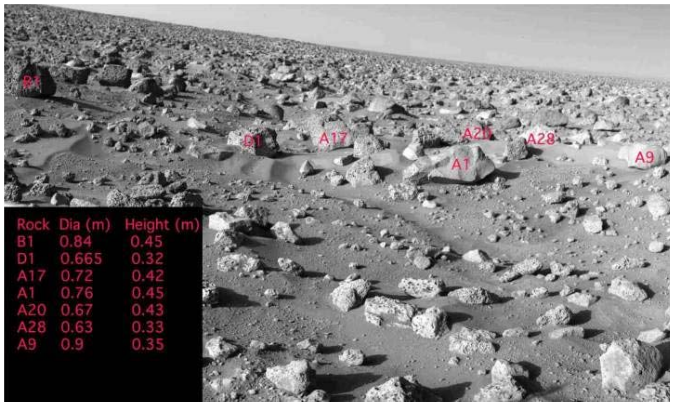
- Seeing in the dark
- Seeing through atmospheric obscurants
- Recognizing things due to:
  - Intrinsic heat
  - Heat transfer characteristics
  - Thermal “color”

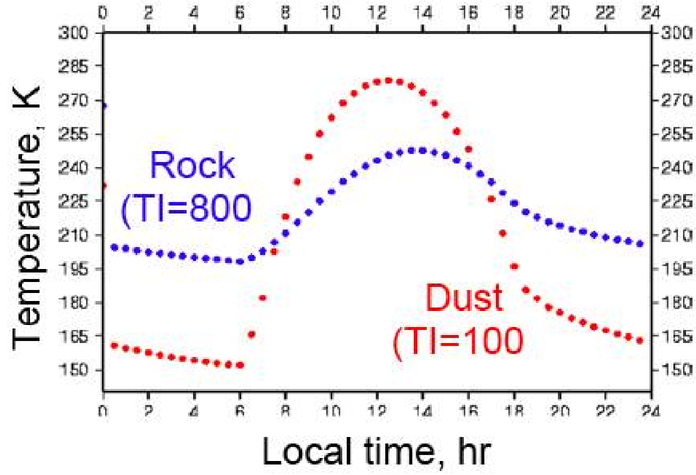
Visible!

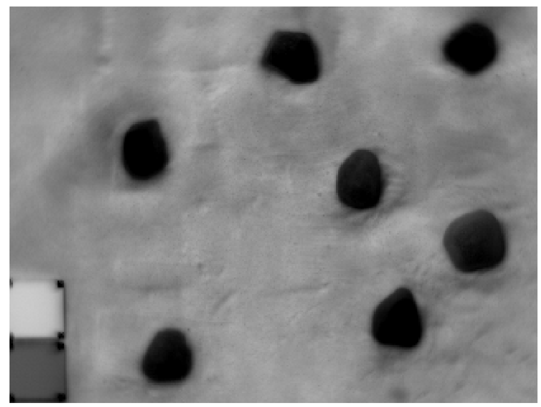
LWIR!



Heat Transfer Characteristics: Thermal Inertia for Hazard Detection on Mars







(U. Arizona)