Digital Image Formation

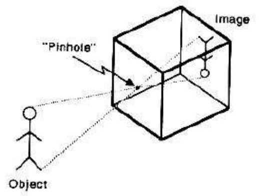
- An (intensity) IMAGE is a 2D array of numbers representing gray level, color, distance or other physical quantities
- We'll usually consider
  - Intensity / Brightness Images $f: \mathbb{R}^2 \to \mathbb{R}^+$
  - Color Images (3-valued)
- 2 KEY Issues:
  1) Where will the image of a scene point appear?
  2) How bright will the image of a scene point appear?

Geometric Image Formation

- 3D scene projects to 2D image
- We'll assume:
  3D scene consists of opaque and reflective objects in a transparent medium (air) with 1 or more light sources.
- Want a sharp image (in focus)
  ⇒ all rays coming from a single scene point P must converge to a single point, p, on the image

Pinhole Camera Model

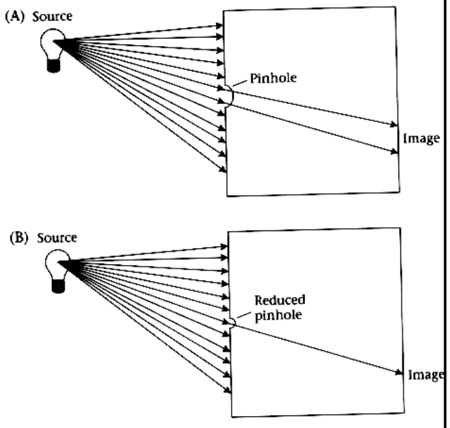
- Aka perspective projection model



- For each scene point only 1 ray can enter camera
- Pinhole = center of projection through which all light passes
- Pinhole too big ⇒ blurring
  Pinhole too small ⇒ diffraction-based blurring
- Long exposure time needed

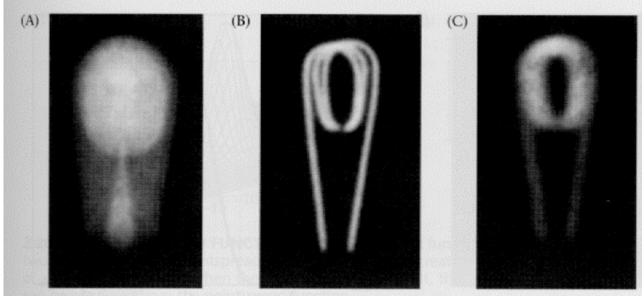
Lensless Imaging Systems: Pinhole Optics

- Pinhole optics focuses images
  - without lens
  - with infinite depth of field
- Smaller the pinhole
  - better the focus
  - less light energy from any single point



2.17 PINHOLE OPTICS. Using ray-tracing, we see that only a small pencil of rays passes through a pinhole. (A) If we use a wide pinhole, light from the source spreads across the image, making it blurry. (B) If we narrow the pinhole, only a small amount of light is let in. The image is sharp; the sharpness is limited by diffraction.

Diffraction and Pinhole Optics

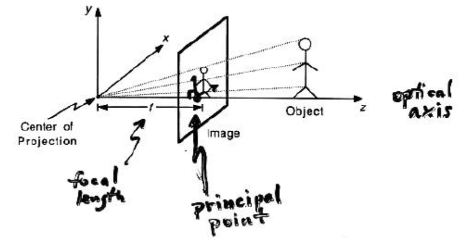


2.18 DIFFRACTION LIMITS THE QUALITY OF PINHOLE OPTICS. These three images of a bulb filament were made using pinholes with decreasing size. (A) When the pinhole is relatively large, the image rays are not properly converged, and the image is blurred. (B) Reducing the size of the pinhole improves the focus. (C) Reducing the size of the pinhole further worsens the focus, due to diffraction. From Ruechardt, 1958.

Diffraction

- Two disadvantages to pinhole systems
  - Light collecting power
  - Diffraction
- Diffraction
  - When light passes through a small aperture it does not travel in a straight line
  - It is scattered in many directions
  - Process is called diffraction and is a quantum effect
- Human vision
  - At high light levels, pupil (aperture) is small and blurring is due to diffraction
  - At low light levels, pupil is open and blurring is due to lens imperfections

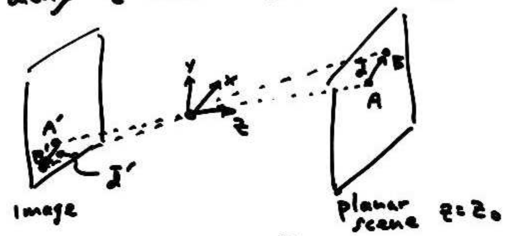
Perspective Projection



- Image plane orthogonal to z axis (called optical axis)
- Camera frame origin at center of projection
- 3D scene point $P = (X, Y, Z)^T$ projects to image point $p' = (x', y', z')^T$ where $z' = f$ (focal length)

Properties of Perspective Projection

- Object size changes as it translates along z axis (scale effect)



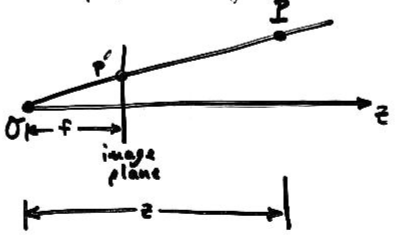
$$\text{Magnification } |m| = \frac{|\bar{d}'|}{|\bar{d}|} = \left|\frac{f}{z_0}\right|$$

⇒ distance b/w points **not** preserved

- As f gets smaller, more world points project onto finite image plane ⇒ more **wide angle** image
- As f gets larger, more **telescopic**
- Lines in 3D project to lines in 2D

Perspective Projection (cont.)

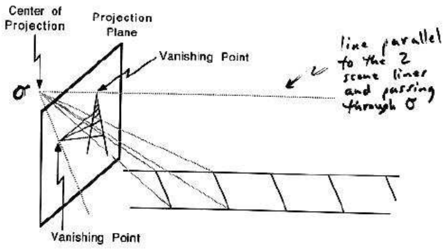
- Perspective projection equations

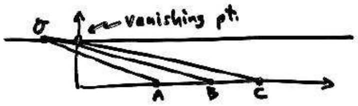


$$\frac{p'}{f} = \frac{P}{z} \quad \Rightarrow$$

$$\begin{cases} x' = \frac{fx}{z} \\ y' = \frac{fy}{z} \\ z' = f \end{cases}$$

- Vanishing point = point in image beyond which projection of straight line cannot extend

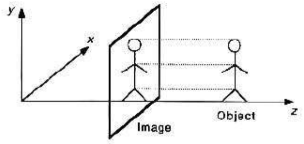




- Focus of Expansion (FOE)

  When camera translates, trajectories of image points appear to move towards or away from a fixed point called FOE which is common vanishing pt. because all pts moving along straight lines relative to camera

Orthographic Projection



- If scene all on a plane $z = z_0$,

  $$x' = \frac{xf}{z_0} = xm$$

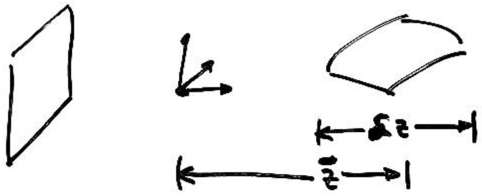
- If range of distances of scene surfaces is small relative to average distance from camera, use constant m as approximation.

  $m = 1 \Rightarrow$ Orthographic projection

  $$\begin{cases} x' = x \\ y' = y \end{cases}$$

- All rays parallel to the optical axis

Orthographic Projection (cont.)

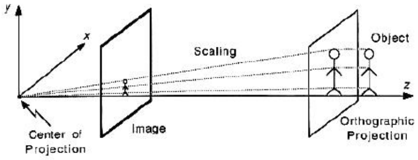


$\bar{z} \gg \Delta z \Rightarrow$ reasonable approximation

Example: Telephoto lens and object(s) are near the principal point and object(s) are small

- Object size in image does <u>not</u> depend on distance from camera

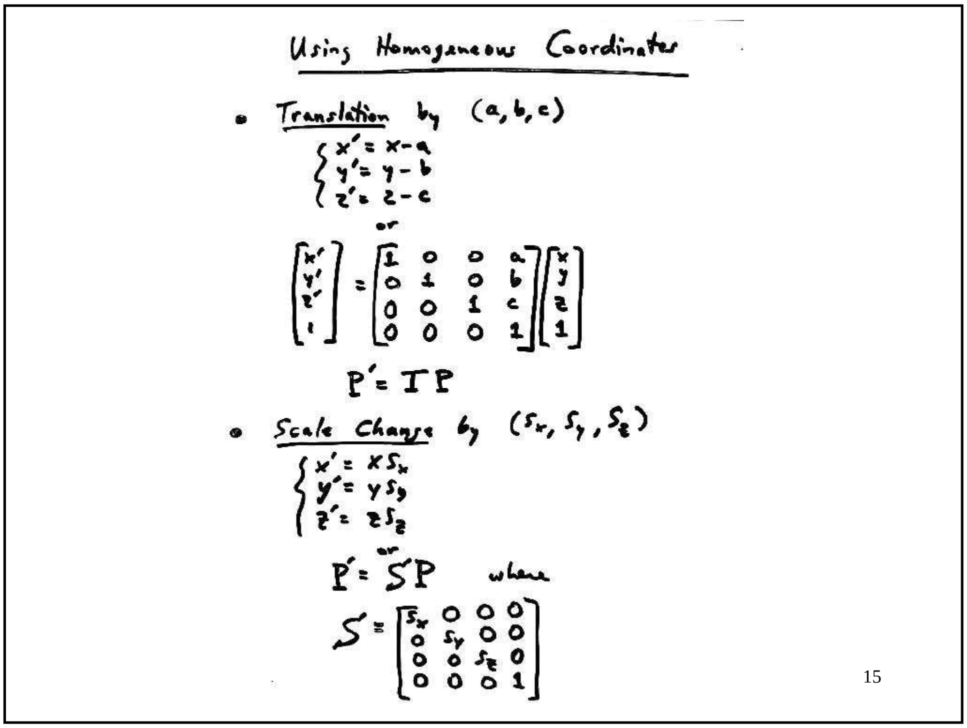
Weak Perspective Projection



- Approximation of perspective proj.
- Orthographic projection + scale:
- 1. Orthographic projection onto a plane $z = \bar{z}$ parallel to image plane at distance $\bar{z}$
  2. Perspective projection onto image plane
     $x' = \left(\frac{f}{\bar{z}}\right)x$, $y' = \left(\frac{f}{\bar{z}}\right)y$
     $\Rightarrow$ scale by $f/\bar{z}$
- Ok approximation when depth of scene points << average distance to camera
  i.e., $\delta z \ll \bar{z}$

Camera Transformations using Homogeneous Coordinates

- Computer vision and computer graphics frequently represent points in Homogeneous Coordinates so that translation, rotation and scale changes are treated in same way
- Cartesian coordinates $P(x, y, z)$ represented as Homogeneous coords $P(wx, wy, wz, w)$ for any scale factor $w \neq 0$.
- Given 4D homogeneous coords (x, y, z, w), the 3D Cartesian coords are $(x/w, y/w, z/w)$
- (x, y, z, w) same as $(x/w, y/w, z/w, 1)$ so usually consider pts of form $(x, y, z, 1)$



Using Homogeneous Coordinates

- Translation by (a, b, c)

$$\begin{cases} x' = x - a \\ y' = y - b \\ z' = z - c \end{cases}$$

or

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & a \\ 0 & 1 & 0 & b \\ 0 & 0 & 1 & c \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

$$P' = T P$$

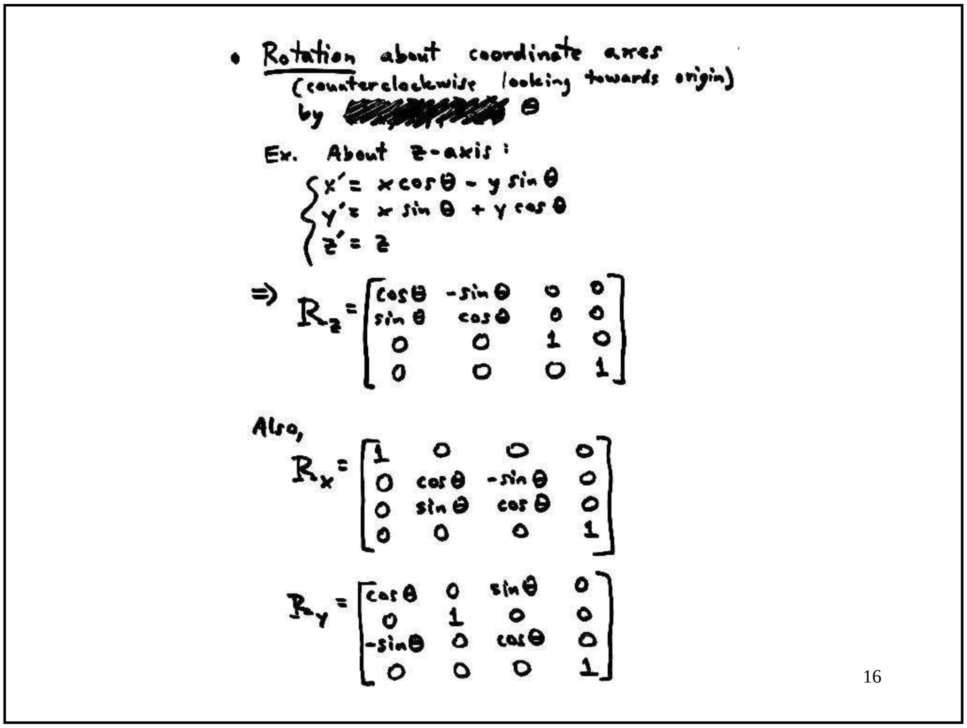
- Scale Change by (S_x, S_y, S_z)

$$\begin{cases} x' = x S_x \\ y' = y S_y \\ z' = z S_z \end{cases}$$

or

$$P' = S P$$ where

$$S = \begin{bmatrix} S_x & 0 & 0 & 0 \\ 0 & S_y & 0 & 0 \\ 0 & 0 & S_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$



- Rotation about coordinate axes (counterclockwise looking towards origin) by [redacted] θ

Ex. About z-axis:

$$\begin{cases} x' = x\cos\theta - y\sin\theta \\ y' = x\sin\theta + y\cos\theta \\ z' = z \end{cases}$$

$$\Rightarrow R_z = \begin{bmatrix} \cos\theta & -\sin\theta & 0 & 0 \\ \sin\theta & \cos\theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Also,

$$R_x = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta & 0 \\ 0 & \sin\theta & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$R_y = \begin{bmatrix} \cos\theta & 0 & \sin\theta & 0 \\ 0 & 1 & 0 & 0 \\ -\sin\theta & 0 & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

- Any transformation involving translation, scale or rotation can be written as

$$\underline{P}' = M\underline{P}$$

where M constructed by composing transformation matrices

Ex.

$$\begin{bmatrix} 1 & 0 & a \\ 0 & 1 & b \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \alpha\cos\theta & \beta\sin\theta & \alpha(a\cos\theta - b\sin\theta) \\ -\alpha\sin\theta & \beta\cos\theta & \beta(a\sin\theta + b\cos\theta) \\ 0 & 0 & 1 \end{bmatrix}$$

- Translations are commutative, rotations are not
- General transformation matrix of form:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & T_x \\ a_{21} & a_{22} & a_{23} & T_y \\ a_{31} & a_{32} & a_{33} & T_z \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Affine Transformation

Projection using Homogeneous Coords

- Perspective Projection

$$x' = \frac{fx}{z}, \quad y' = \frac{fy}{z}$$

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{fx}{z} \\ \frac{fy}{z} \\ f \\ 1 \end{bmatrix} \Leftrightarrow \underset{\substack{\uparrow \\ \text{image pt}}}{\begin{bmatrix} x_h \\ y_h \\ z_h \\ z/f \end{bmatrix}} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1/f & 0 \end{bmatrix}}_{T_{\text{perspective}}} \underset{\substack{\uparrow \\ \text{scene pt}}}{\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}}$$

- Orthographic Projection

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & f \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{T_{\text{orthographic}}} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

- **Weak Perspective Projection**

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \begin{bmatrix} f/\bar{z} & 0 & 0 & 0 \\ 0 & f/\bar{z} & 0 & 0 \\ 0 & 0 & 0 & f \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

- **Note**: Since image plane at $z=f$, perspective projection equation can be written as :

$$\begin{bmatrix} x_h \\ y_h \\ w \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

and $\therefore \begin{cases} x' = x_h / w \\ y' = y_h / w \end{cases}$

$\Rightarrow$ Camera = linear projective transform from 3D projective space to 2D projective plane

- 3 x 4 matrix called <u>camera perspective projection matrix</u>

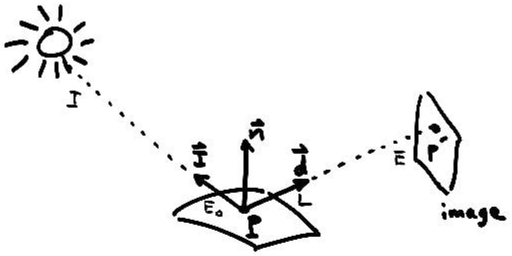
Radiometry

- How bright will an image point appear?
- Assume: Light at a point depends only on the brightness of closest surface point in a given direction.

  (Not met in tomography, X-rays, satellite images w/ atmospherics)
- Irradiance (Illuminance) = power/unit area of radiant energy falling on a surface

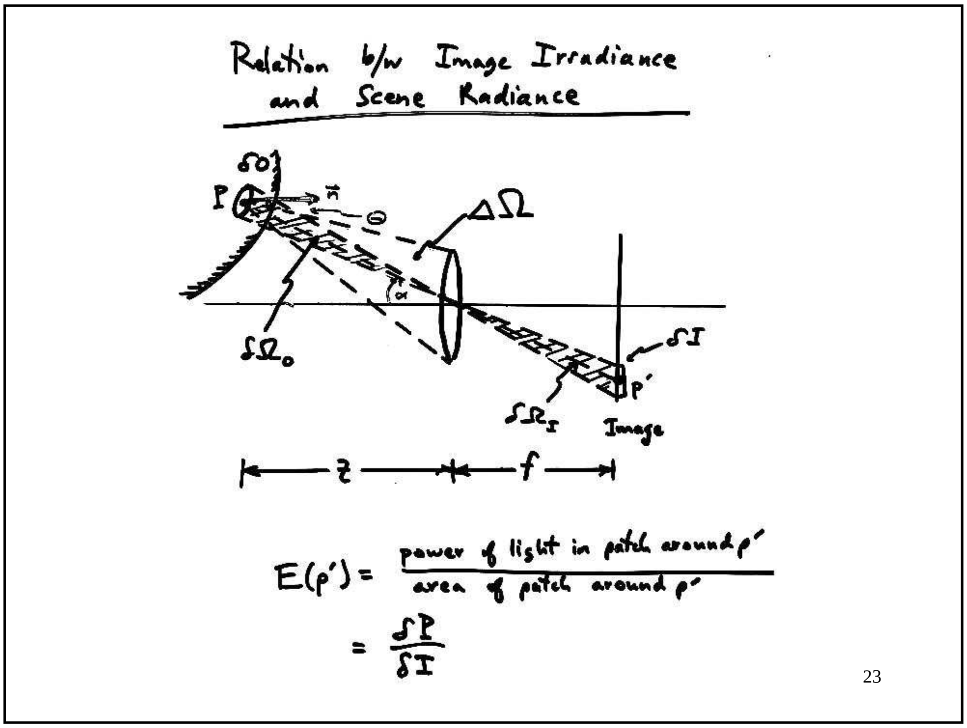
$E = \frac{\delta P}{\delta I}$ (watts/m^2)

- Radiance (Luminance) = power emitted per unit area into a cone of unit diameter



3 relations of interest:

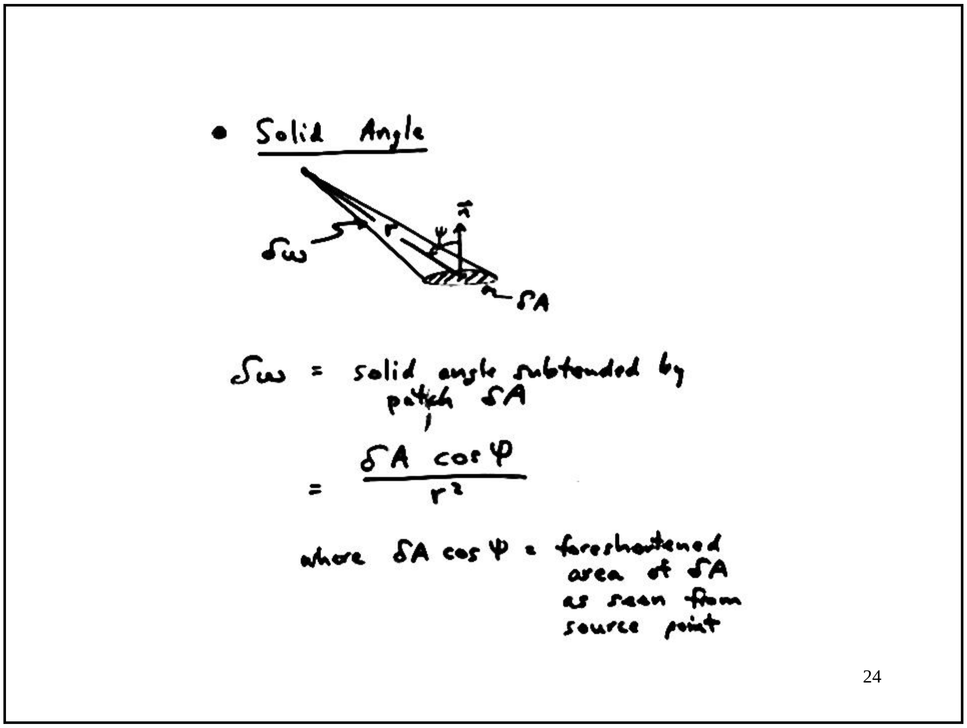
1. L to E : surface radiance to image irradiance
2. E_o to L : surface reflectance function
3. I to E_o : surface brightness function



Relation b/w Image Irradiance and Scene Radiance

$$E(p') = \frac{\text{power of light in patch around } p'}{\text{area of patch around } p'}$$

$$= \frac{\delta P}{\delta I}$$

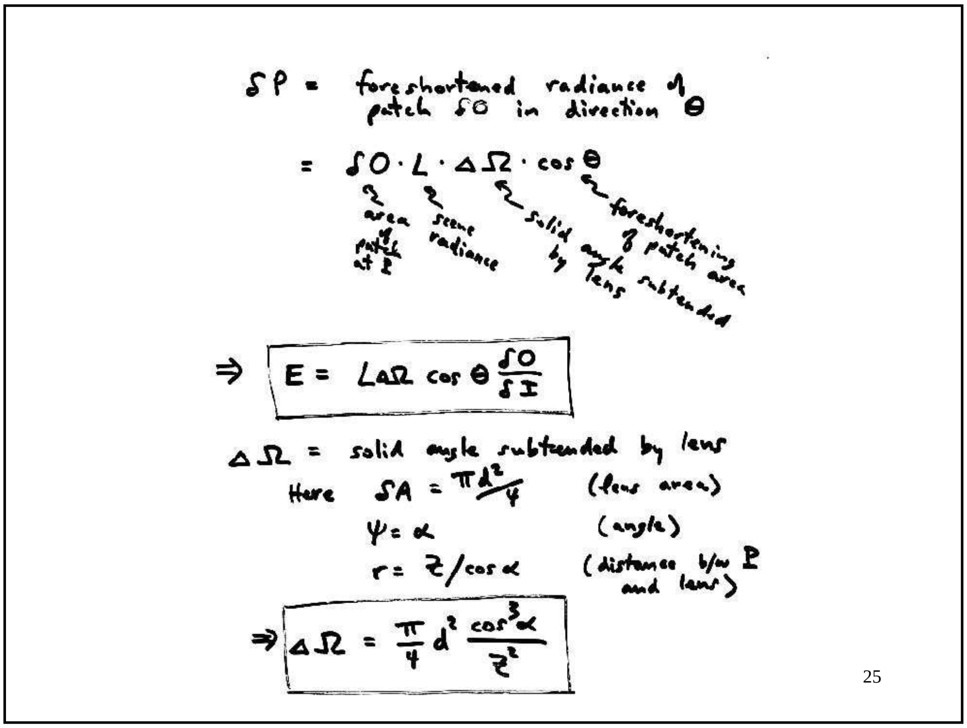


- Solid Angle

δω = solid angle subtended by patch δA

$$= \frac{\delta A \cos\psi}{r^2}$$

where $\delta A \cos\psi$ = foreshortened area of δA as seen from source point



δP = foreshortened radiance of patch δO in direction θ

$$= \underbrace{\delta O}_{\text{area of patch at P}} \cdot \underbrace{L}_{\text{scene radiance}} \cdot \underbrace{\Delta \Omega}_{\text{solid angle subtended by lens}} \cdot \underbrace{\cos\theta}_{\text{foreshortening of patch area}}$$

$$\Rightarrow \boxed{E = L \Delta\Omega \cos\theta \frac{\delta O}{\delta I}}$$

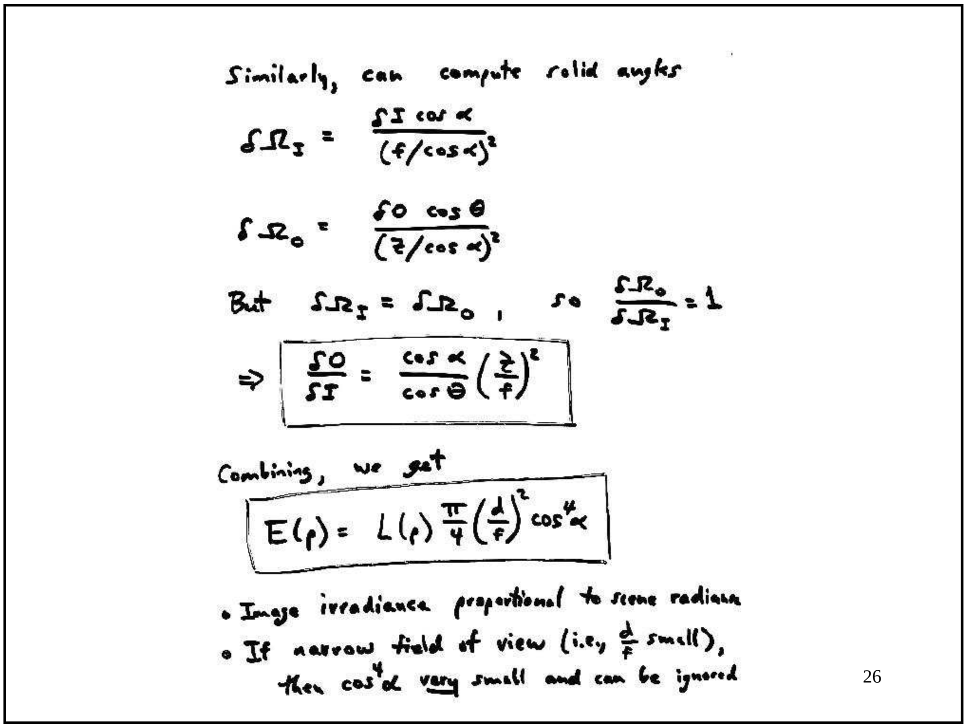
$\Delta\Omega$ = solid angle subtended by lens

Here $\delta A = \pi d^2 / 4$ (lens area)

$\psi = \alpha$ (angle)

$r = z / \cos\alpha$ (distance b/w P and lens)

$$\Rightarrow \boxed{\Delta\Omega = \frac{\pi}{4} d^2 \frac{\cos^3\alpha}{z^2}}$$



Similarly, can compute solid angles

$$\delta\Omega_I = \frac{\delta I \cos\alpha}{(f/\cos\alpha)^2}$$

$$\delta\Omega_o = \frac{\delta O \cos\theta}{(z/\cos\alpha)^2}$$

But $\delta\Omega_I = \delta\Omega_o$, so $\frac{\delta\Omega_o}{\delta\Omega_I} = 1$

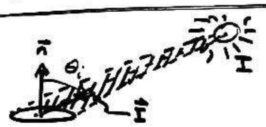
$$\Rightarrow \boxed{\frac{\delta O}{\delta I} = \frac{\cos\alpha}{\cos\theta}\left(\frac{z}{f}\right)^2}$$

Combining, we get

$$\boxed{E(p) = L(p) \frac{\pi}{4} \left(\frac{d}{f}\right)^2 \cos^4\alpha}$$

- Image irradiance proportional to scene radiance
- If narrow field of view (i.e., $\frac{d}{f}$ small), then $\cos^4\alpha$ very small and can be ignored

Relation b/w Illumination and Surface Irradiance



Assuming point light source at infinity

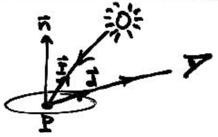
$$E_o = \frac{I \cos \theta_i}{r^2} \qquad (\theta_i < \frac{\pi}{2})$$

$$= I \cos \theta_i$$

But $\cos \theta_i = \bar{I}^T \cdot \bar{n}$
where $\bar{I}$ and $\bar{n}$ are unit vectors

$\Rightarrow \quad E_o = \vec{I}^T \cdot \bar{n}$
where $\vec{I}$ represents direction <u>and</u> amount of incident light

Relation b/w Surface Irradiance and Surface Radiance



$$L(p) = f(\vec{n}, \vec{I}, \vec{d})$$

- f depends on characteristics of surface at P and light properties
- f called <u>Bidirectional Reflectance Distribution Function</u> (BRDF)
- Example: Lambertian surfaces (perfect matte)
  - Surface appears equally bright in <u>all</u> directions

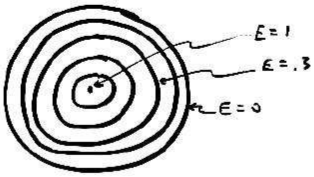
$\Rightarrow \quad L = \frac{\rho}{\pi} E_o \qquad \rho > 0$ called <u>albedo</u>

So, for Lambertian surfaces

$$L = \rho \bar{I}^T \bar{n}$$

and, when camera is viewed from a long distance

$$E = \rho \bar{I}^T \bar{n}$$

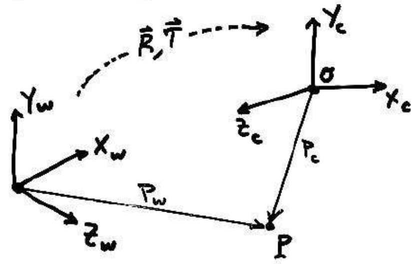


Camera Parameters

- 3 reference frames of interest:
  1. Camera frame
     origin = optical center
     aka center of projection
     z-axis = optical axis
  2. Image frame
     I(i,j) are <u>pixel</u> coordinates in digital image
  3. World frame
     Fixed reference frame in the 3D scene
- Camera parameters used to relate these 3 frames

Camera Parameters (cont.)

- Extrinsic Parameters relate camera frame wrt world frame



Unique transformation defined by 6 parameters:

3D translation vector, $\vec{T}$

3×3 rotation matrix R (orthogonal matrix ⇒ only 3 dofs)

- Intrinsic Parameters relate camera frame wrt image frame

⇒ for perspective projection camera model, specify:

- focal length, f
- coords of image origin relative to principal point
  ⇒ 2 params (o_x, o_y)
- scale factors in x and y directions
  ⇒ 2 params (s_x, s_y)

$$x' = -(x_{im} - o_x)s_x$$

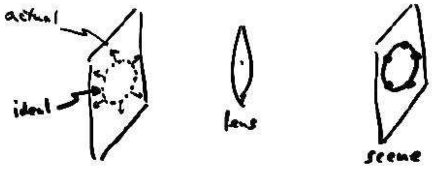
$$y' = -(y_{im} - o_y)s_y$$

↑ coords in camera frame

↑ coords in pixels in image frame

- If camera optics introduce significant distortions, then add parameters to correct. Most common type:

  - Radial distortion



$$x = x_d(1 + k_1 r^2 + k_2 r^4)$$

$$y = y_d(1 + k_1 r^2 + k_2 r^4)$$

coords of distorted pt in image frame

where
$r^2 = x_d^2 + y_d^2$

k_1, k_2 parameters

Camera Parameters

In summary, need

- 6 extrinsic camera parameters
- 6 (usually) intrinsic parameters
  $(f, o_x, o_y, s_x, s_y, k_1)$

Problem of estimating these parameters called Calibration Problem

IMAGE Digitization

- The image falling on the image plane can be thought of as a continuous function of 2 variables, $f(x,y)$
- 2 issues is creating digital image:

1. SAMPLING

   Quantize domain of f to a discrete set of real numbers called samples or pixels. Usually, regularly spaced on a rectangular grid.

2. QUANTIZATION

   Quantize the range of f to a discrete set of real numbers called gray levels. Usually represented as non-negative integers.

Image Sampling

- What grid size is sufficient to encode an image? I.e., from a set of samples, reconstruct exactly the continuous image function?

Answer — It depends on image content.

- If image is nearly constant, then few samples needed.
- If images is highly textured, fine detail, then fine sampling required.

* Sample rate determined by spatial frequency content of image

Fourier Transform

- A continuous image function $f(x,y)$ can be represented as a sum of infinite number of sinusoids of form $e^{i(ux+vy)}$ where

  $$e^{iux} = \cos ux + i \sin ux$$

- $$F(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\, e^{-i(ux+vy)}\, dx\, dy$$

  Fourier Transform

- $$f(x,y) = \iint F(u,v)\, e^{i(ux+vy)}\, du\, dv$$

- ~~[illegible] spatial~~
  ~~[illegible]~~
  (u,v) small for low spatial frequencies
  (u,v) big for high spatial frequencies

Shannon Sampling Theorem

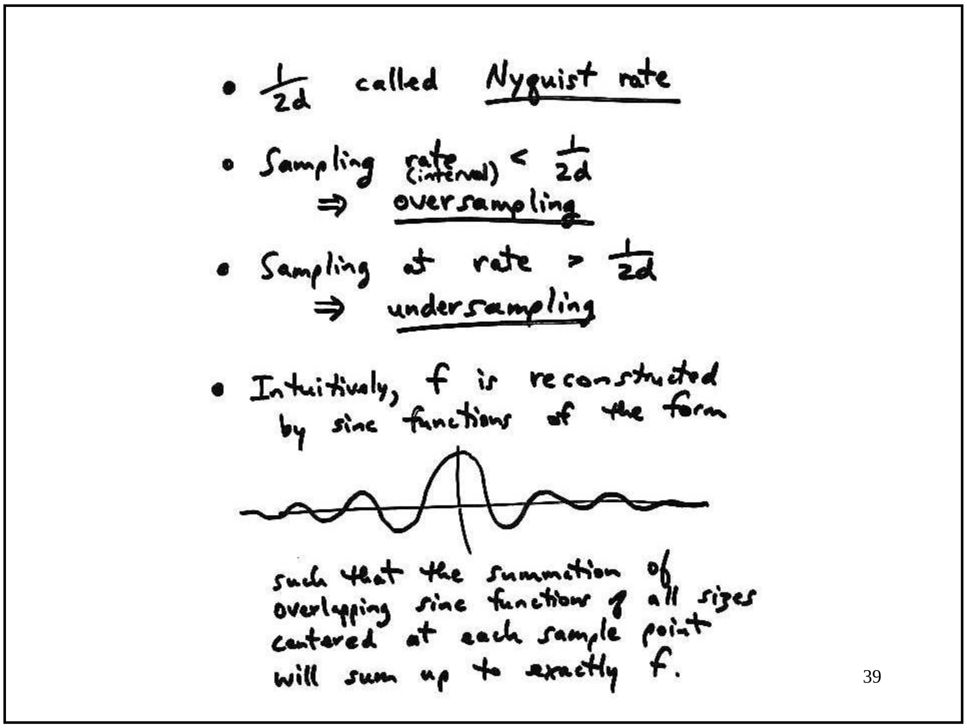
If f is bandlimited,
i.e., $F(u,v) = 0 \quad \forall\ |u| > \pi/w,$
$|v| > \pi/h$

then f can be reconstructed exactly from a sampled version
$f_{kl} = f(kw, lh)$ $\quad k,l$ integers

by:

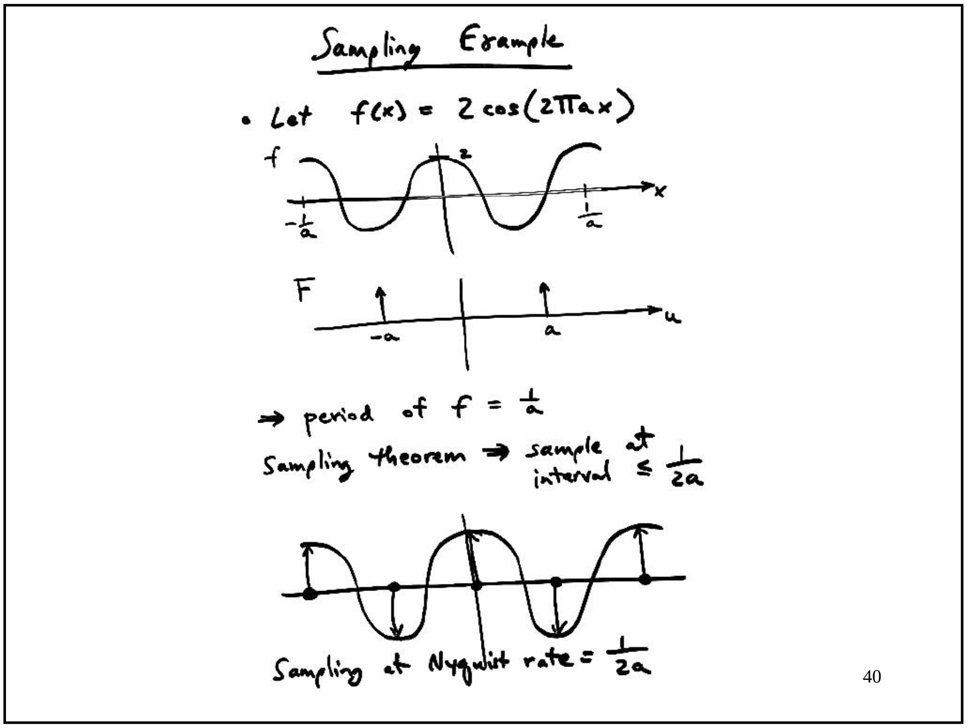
$$f(x,y) = \sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} f_{kl} \cdot \frac{\sin(\pi(x/w-k))}{\pi(x/w-k)} \cdot \underbrace{\frac{\sin(\pi(y/h-l))}{\pi(y/h-l)}}_{\text{"sinc" interpolation function}}$$

If only spatial frequencies of size $< d$ occur, then sampling at interval $\leq \frac{1}{2d}$ ok.



- $\frac{1}{2d}$ called Nyquist rate
- Sampling rate (interval) $< \frac{1}{2d}$
  $\Rightarrow$ oversampling
- Sampling at rate $> \frac{1}{2d}$
  $\Rightarrow$ undersampling
- Intuitively, f is reconstructed by sinc functions of the form

such that the summation of overlapping sinc functions of all sizes centered at each sample point will sum up to exactly f.



Sampling Example

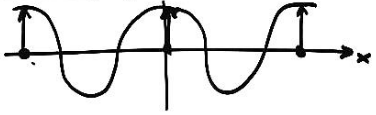
- Let $f(x) = 2\cos(2\pi a x)$

$\Rightarrow$ period of $f = \frac{1}{a}$

Sampling theorem $\Rightarrow$ sample at interval $\leq \frac{1}{2a}$

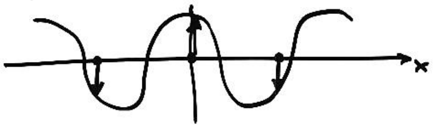
Sampling at Nyquist rate $= \frac{1}{2a}$

- Undersampling at rate = $\frac{1}{a}$

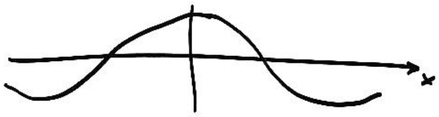


⇒ interpolated function
(i.e. reconstruction of f)
is a constant

- Undersampling at rate = $\frac{2}{3a}$



⇒ reconstructed function at $\frac{1}{2}$ frequency of f:

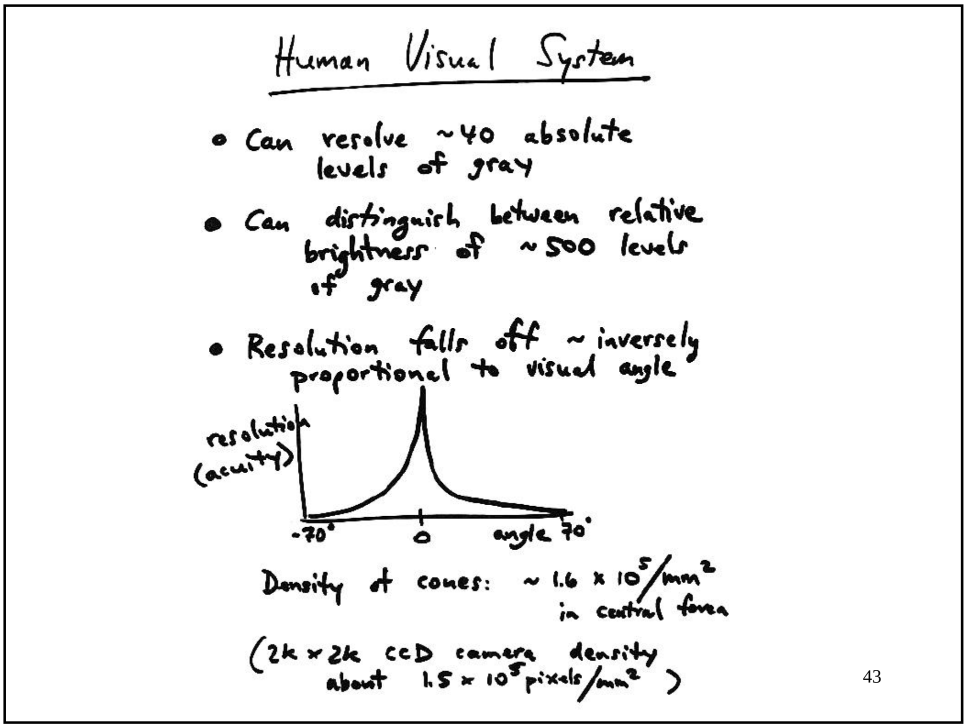


- Undersampling causes change in frequency for 1D images, and change in frequency and orientation for 2D images

  Called Aliasing

  - Most visible in high frequency, periodic structures (bar patterns, fine textured regions).
  - Moiré patterns and jagged edges are visible effects

- If spacing between sensor elements is d, then aliasing will occur if image f has spatial frequencies $> \frac{1}{2d}$ (i.e., period of finest detail $< 2d$)

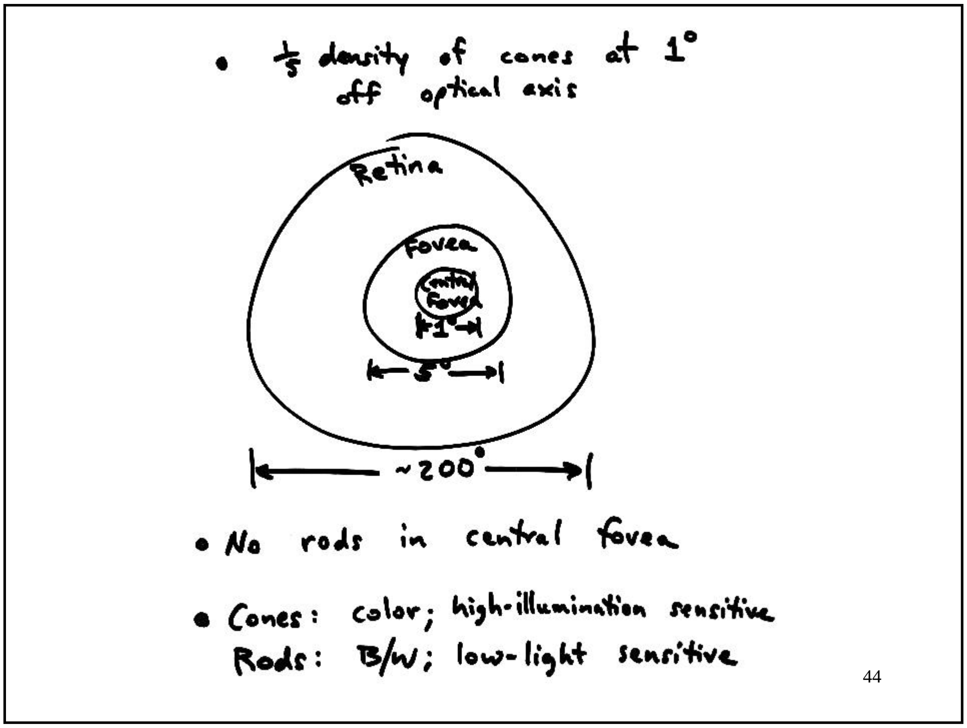


Human Visual System

- Can resolve ~40 absolute levels of gray
- Can distinguish between relative brightness of ~500 levels of gray
- Resolution falls off ~ inversely proportional to visual angle

Density of cones: ~ $1.6 \times 10^5/\text{mm}^2$ in central fovea

(2k × 2k CCD camera density about 1.5×10^5 pixels/mm^2)

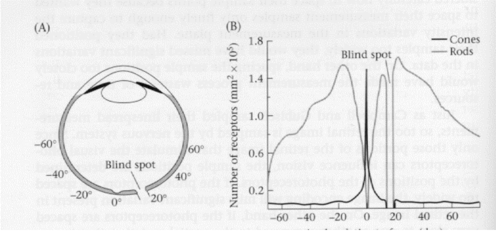


- 1/5 density of cones at 1° off optical axis

- No rods in central fovea
- Cones: color; high-illumination sensitive
  Rods: B/W; low-light sensitive

The Human Eye

- Limitations of human vision
  - the image is upside-down!
  - high resolution vision only in the fovea
    - only one small fovea in man
    - other animals (birds, cheetas) have different foveal organizations
  - blind spot



Quantization

- 2 issues:
  1. How many gray levels?
  2. How are intervals of intensity associated with gray levels?
- In most cameras, number of gray levels = 256 (8 bits/pixel) for B/W images, and 256 * 3 for color images
- Too few gray levels ⇒ <u>false contours</u> in areas of image where intensity changes <u>slowly</u>

- How to assign gray levels?

  Given dynamic range of image is: $z_0 \le f(x,y) \le z_k$ and desired # quantization levels is k, find a set of decision levels $z_1, \ldots, z_{K-1}$ such that

  if $z_i \le f(x_0, y_0) < z_{i+1}$ for $0 \le i < k$

  then assign gray level i representing discrete value $z_i \le g_i < z_{i+1}$

- Human visual system uses approx. log-spaced intervals

- Uniform-sized intervals easy, but not optimal

- To minimize MSE quantize finely where most intensity values occur, coarsely elsewhere => tapered quantization

---

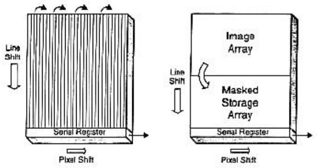
- Minimize $$\sum_{i=0}^{K-1} \int_{z=z_i}^{z_{i+1}} (z - g_i)^2 p(z)\, dz$$

  where $p(z)$ = histogram (prob. density function) of image

- If $p(z)$ is Gaussian distribution, result is a set of intervals with equal # of pixels, i.e., flat histogram.

CCD Cameras

- Photosensitive area divided into 2D array of MOS capacitors that act as charge storage elements
- Charge proportional to number of photons during integration period
- Square vs. rectangular elements
- Interlaced vs. progressive scan
- Interline vs. frame transfer



Interline Transfer CCD   Frame Transfer CCD

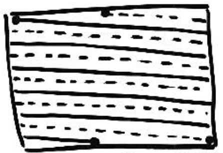
Color Cameras

- Two types of color cameras
  - three built in filters
    - three images are collected through red, green and blue filters
    - such cameras are 3x slower than comparable black and white cameras
  - 3 CCD arrays packed together, each sensitive to different wavelengths of light
    - more similar to human vision

- Broadcast TV

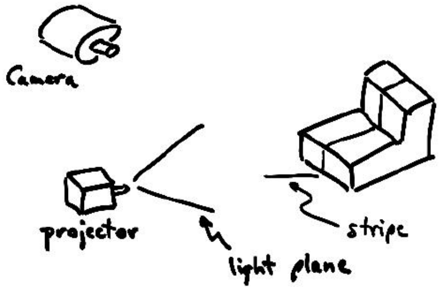
RS-170 (Monochrome)
NTSC (Color)

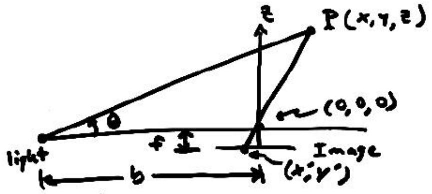
525 lines/frame @ 30 frames/sec
(480 visible)
2 interlaced fields/frame
(even field, odd field)



Frame grabber samples 640 pts/line $\left(\begin{smallmatrix}4/3\\ \text{aspect}\\ \text{ratio}\end{smallmatrix}\right)$

- Light Stripe Camera





$$\begin{cases} x = bx'/(f\cot\theta - x') \\ y = by'/(f\cot\theta - x') \\ z = bf/(f\cot\theta - x') \end{cases}$$