DD2423 Image Analysis and Computer Vision

IMAGE FORMATION

Mårten Björkman

Computational Vision and Active Perception
School of Computer Science and Communication

November 8, 2013

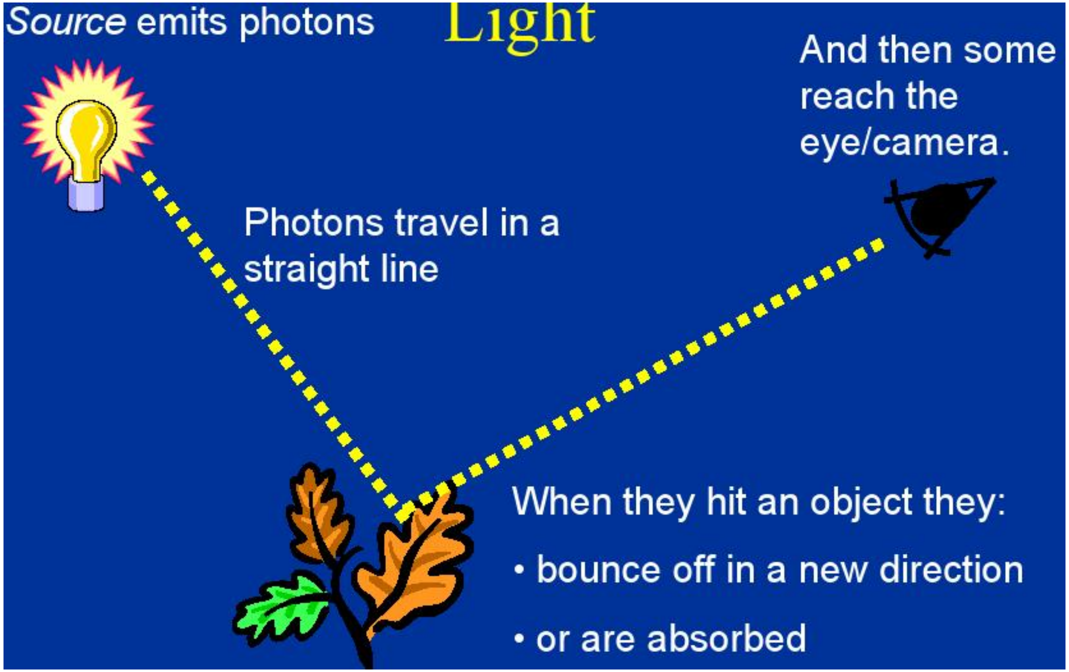
Image formation

Goal: Model the image formation process

- Image acquisition
- Perspective projection
  - properties
  - approximations
- Homogeneous coordinates
- Sampling
- Image warping

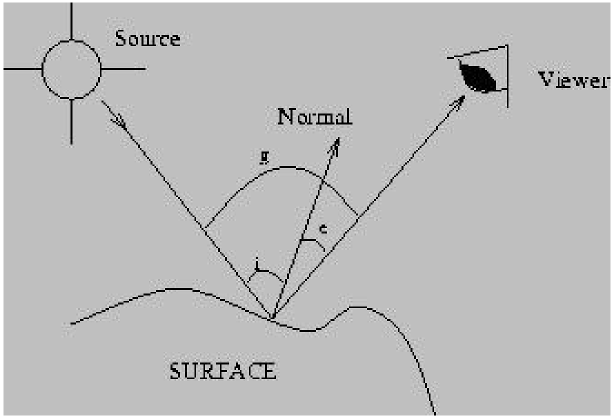
Image formation

Image formation is a physical process that captures scene illumination through a lens system and relates the measured energy to a signal.



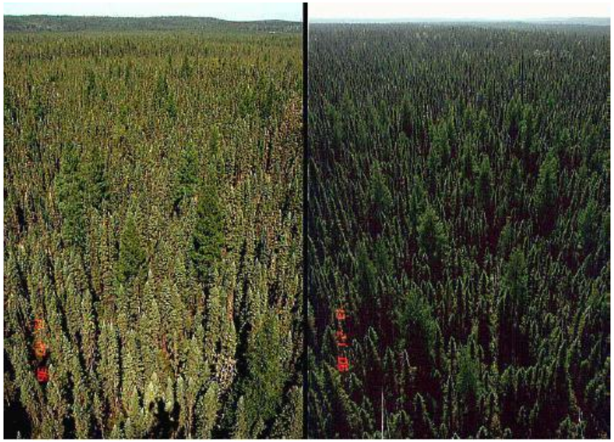
Basic concepts

- Irradiance E: Amount of light falling on a surface, in power per unit area (watts per square meter). If surface tilts away from light, same amount of light strikes bigger surface (foreshortening $\rightarrow$ less irradiance).
- Radiance L: Amount of light radiated from a surface, in power per unit area per unit solid angle. Informally "Brightness".



- Image irradiance E is proportional to scene radiance

Light source examples



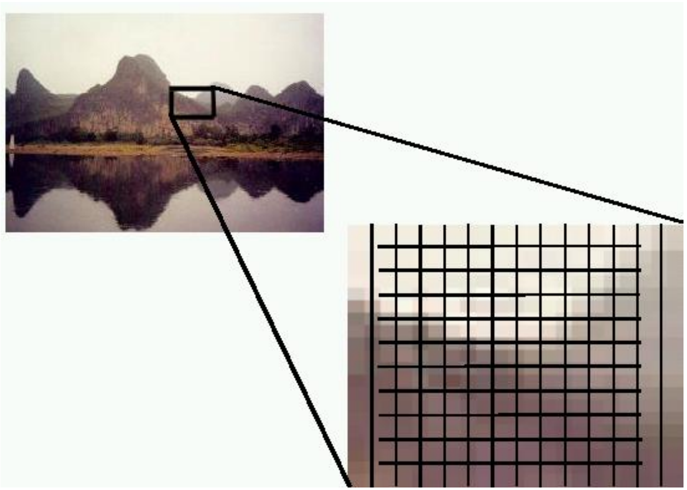
Left: Forest image (left): sun behind observer, (right): sun opposite observer

Right: Field with rough surface (left): sun behind observer, (right): sun opposite observer.

Digital imaging

Image irradiance E $\times$ area $\times$ exposure time $\rightarrow$ Intensity

- Sensors read the light intensity that may be filtered through color filters, and digital memory devices store the digital image information either as RGB color space or as raw data.
- An image is discretized: sampled on a discrete 2D grid $\rightarrow$ array of color values.



Imaging acqusition - From world point to pixel

- World points are projected onto a camera sensor chip.
- Camera sensors sample the irradiance to compute energy values.
- Positions in camera coordinates (in mm) are converted to image coordinates (in pixels) based on the intrinsic parameters of the camera:

  - size of each sensor element,

  - aspect ratio of the sensor (xsize/ysize),

  - number of sensor elements in total,

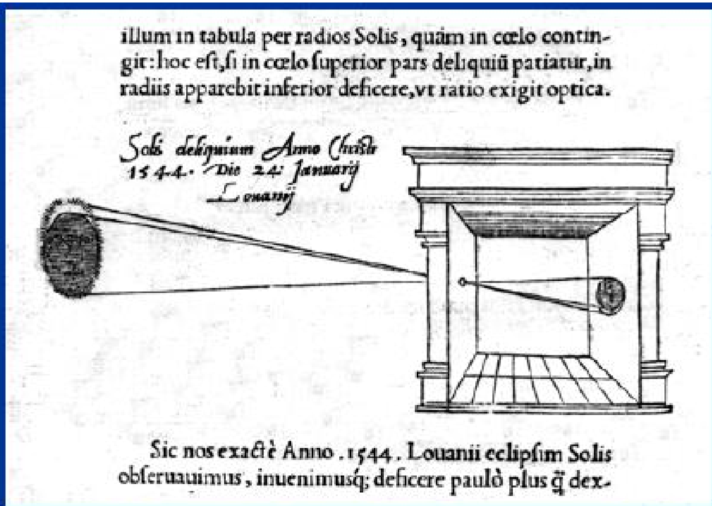
  - image center of sensor chip relative to the lens system.

Steps in a typical image processing system

- Image acquisition: capturing visual data by a vision sensor
- Discretization/digitalization - Quantization - Compression: Convert data into discrete form; compress for efficient storage/transmission
- Image enhancement: Improving image quality (low contrast, blur noise)

- Image segmentation: Partition image into objects or constituent parts.
- Feature detection: Extracting pertinent features from an image that are important for differentiating one class of objects from another.
- Image representation: Assigning labels to an object based on information provided by descriptors.
- Image interpretation: Assigning meaning to image information.

Pinhole camera or “Camera Obscura”



illum in tabula per radios Solis, quàm in cœlo contin-
git: hoc eſt, ſi in cœlo ſuperior pars deliquiũ patiatur, in
radiis apparebit inferior deficere, vt ratio exigit optica.

Sic nos exactè Anno .1544. Louanii eclipſim Solis
obſeruauimus, inuenimusq; deficere paulò plus q̃ dex-

"When images of illuminated objects ... penetrate through a small hole into a very dark room ... you will see [on the opposite wall] these objects in their proper form and color, reduced in size ... in a reversed position, owing to the intersection of the rays".

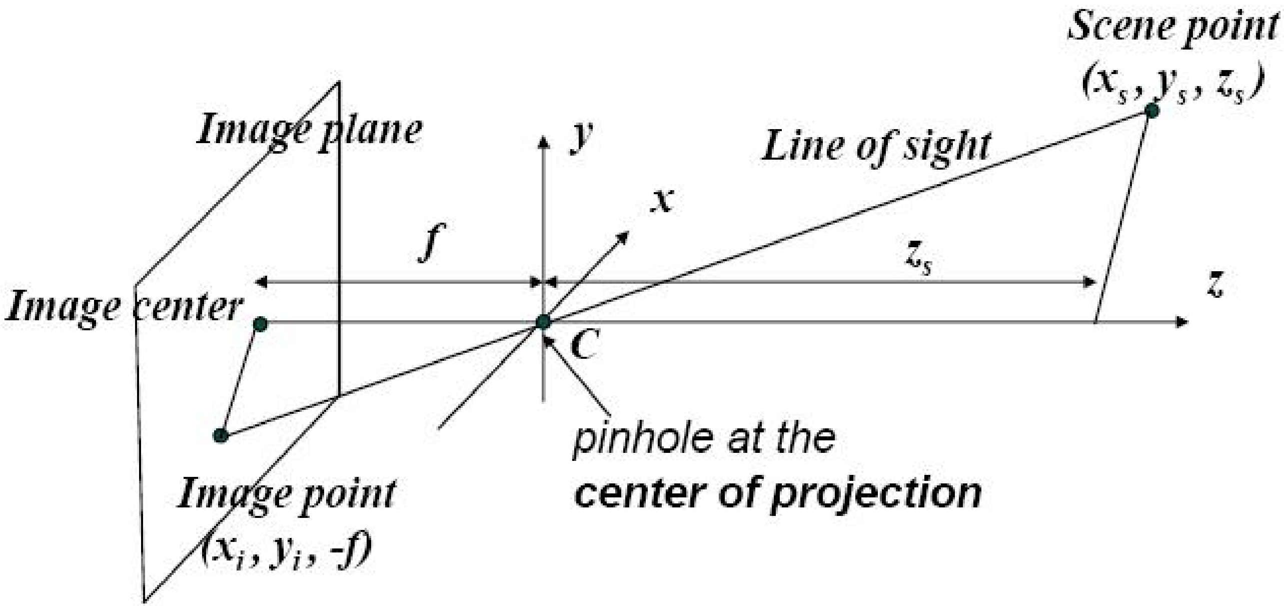
Leonardo Da Vinci

http://www.acmi.net.au/AIC/CAMERA_OBSCURA.html (Russell Naughton)

Pinhole camera and perspective projection

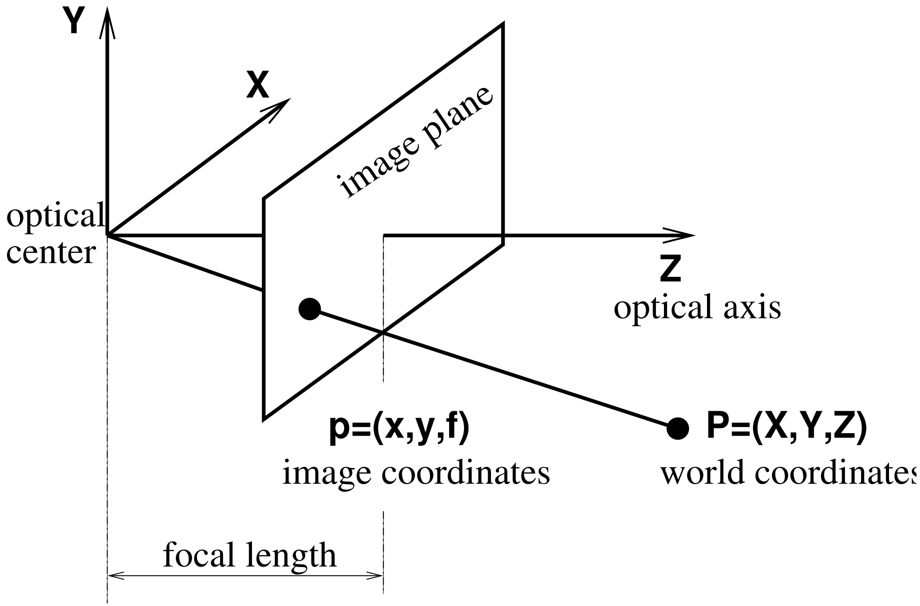
- A mapping from a three dimensionsal (3D) world onto a two dimensional (2D) plane in the previous example is called perspective projection.
- A pinhole camera is the simplest imaging device which captures the geometry of perspective projection.
- Rays of light enter the camera through an infinitesimally small aperture.
- The intersection of light rays with the image plane form the image of the object.

Perspective projection



- The point on the image plane that corresponds to a particular point in the scene is found by following the line that passes through the scene point and the center of projection

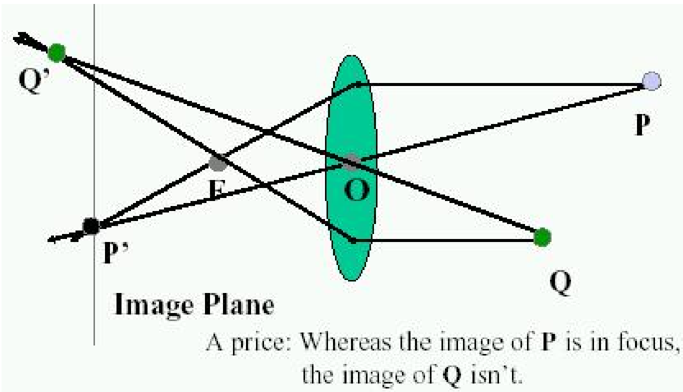
Pinhole camera - Perspective geometry

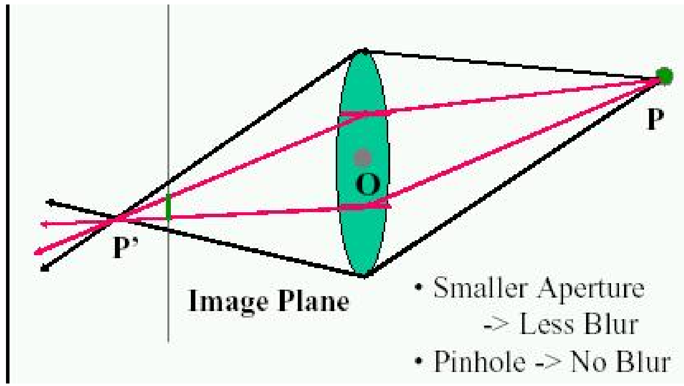


- The image plane is usually modeled in front of the optical center.
- The coordinate systems in the world and in the image domain are parallel. The optical axis is $\perp$ image plane.

Lenses

- Purpose: gather light from from larger opening (aperture)
- Problem: only light rays from points on the focal plane intersect the same point on the image plane
- Result: blurring in-front or behind the focal plane
- Focal depth: the range of distances with acceptable blurring





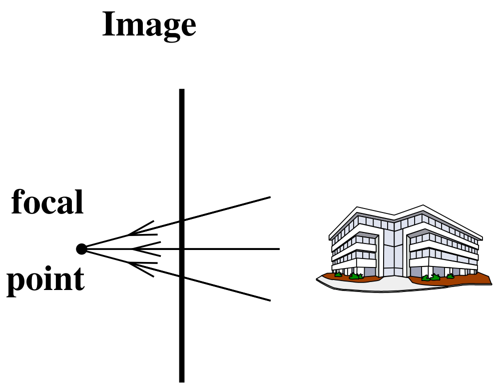
Imaging geometry - Basic camera models

- **Perspective projection** (general camera model)

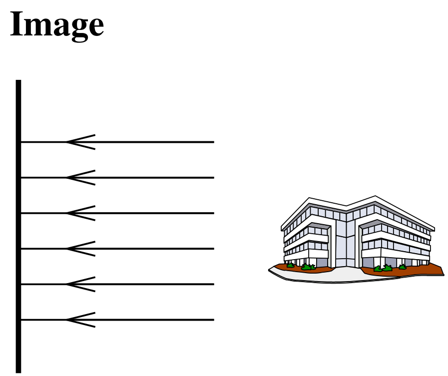
  All visual rays converge to a common point - the focal point

- **Orthographic projection** (approximation: distant objects, center of view)

  All visual rays are perpendicular to the image plane

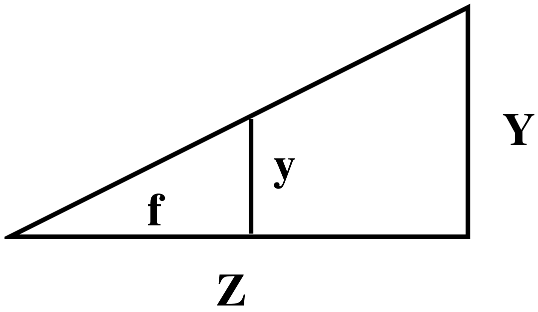


Perspective projection



Orthographic projection

Projection equations



- Perspective mapping

$$\frac{x}{f} = \frac{X}{Z}, \quad \frac{y}{f} = \frac{Y}{Z}$$

- Orthographic projection

$$x = X, \quad y = Y$$

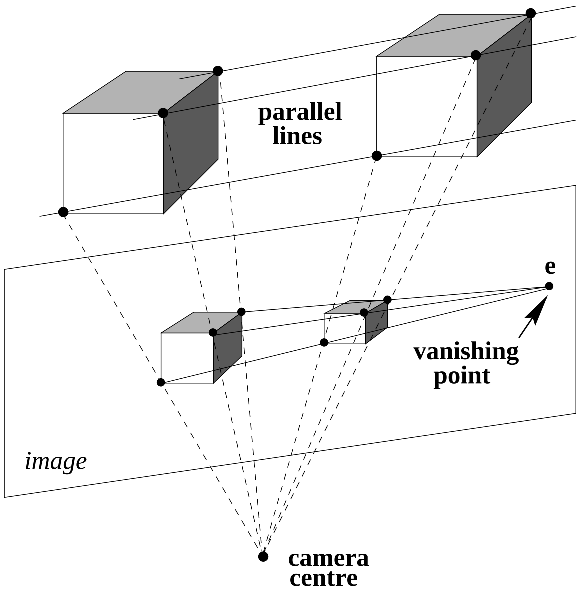
- Scaled orthography - Z_0 constant (representative depth)

$$\frac{x}{f} = \frac{X}{Z_0}, \quad \frac{y}{f} = \frac{Y}{Z_0}$$

Perspective transformation

- A perspective transformation has three components:
  - Rotation - from world to camera coordinate system
  - Translation - from world to camera coordinate system
  - Perspective projection - from camera to image coordinates
- Basic properties which are preserved:
  - lines project to lines,
  - collinear features remain collinear,
  - tangencies,
  - intersections.

Perspective transformation (cont)

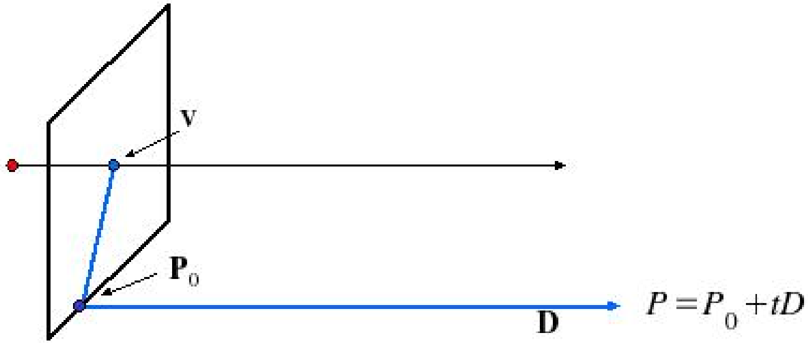


Each set of parallel lines meet at a different vanishing point - vanishing point associated to this direction. Sets of parallel lines on the same plane lead to collinear vanishing points - the line is called the horizon for that plane.

Homogeneous coordinates

- Model points (X, Y, Z) in $\mathcal{R}^3$ world by (kX, kY, kZ, k) where k is arbitrary $\neq 0$, and points (x, y) in $\mathcal{R}^2$ image domain by (cx, cy, c) where c is arbitrary $\neq 0$.
- Equivalence relation: (k_1X, k_1Y, k_1Z, k_1) is same as (k_2X, k_2Y, k_2Z, k_2).
- Homogeneous coordinates imply that we regard all points on a ray (cx, cy, c) as equivalent (if we only know the image projection, we do not know the depth).
- Possible to represent "points in infinity" with homogreneous coordinates $(X, Y, Z, 0)$ - intersections of parallel lines.

Computing vanishing points



$$P_t = \begin{bmatrix} P_X + tD_X \\ P_Y + tD_Y \\ P_Z + tD_Z \\ 1 \end{bmatrix} \simeq \begin{bmatrix} P_X / t + D_X \\ P_Y / t + D_Y \\ P_Z / t + D_Z \\ 1/t \end{bmatrix} \quad t \to \infty \quad P_\infty \simeq \begin{bmatrix} D_X \\ D_Y \\ D_Z \\ 0 \end{bmatrix}$$

Properties $\quad v = \mathbf{P}_\infty$

- $\mathbf{P}_\infty$ is a point at *infinity*, **v** is its projection
- They depend only on line *direction*
- Parallel lines $\mathbf{P}_0 + t\mathbf{D}$, $\mathbf{P}_1 + t\mathbf{D}$ intersect at $\mathbf{P}_\infty$

Homogeneous coordinates (cont)

In homogeneous coordinates the projection equations can be written

$$\begin{pmatrix} cx \\ cy \\ c \end{pmatrix} = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} kX \\ kY \\ kZ \\ k \end{pmatrix} = \begin{pmatrix} fkX \\ fkY \\ kZ \end{pmatrix}$$

Image coordinates obtained by normalizing the third component to one (divide by $c = kZ$).

$$x = \frac{xc}{c} = \frac{fkX}{kZ} = f\frac{X}{Z}, \quad y = \frac{yc}{c} = \frac{fkY}{kZ} = f\frac{Y}{Z}$$

Transformations in homogeneous coordinates

- Translation

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \rightarrow \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + \begin{pmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{pmatrix}$$

$$\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 0 & \Delta X \\ 0 & 1 & 0 & \Delta Y \\ 0 & 0 & 1 & \Delta Z \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

- Scaling

$$\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} S_X & 0 & 0 & 0 \\ 0 & S_Y & 0 & 0 \\ 0 & 0 & S_Z & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

Transformations in homogeneous coordinates II

- Rotation around the Z axis

$$\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} \cos\theta & -\sin\theta & 0 & 0 \\ \sin\theta & \cos\theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

- Mirroring in the XY plane

$$\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

Transformations in homogeneous coordinates III

Common case: Rigid body transformations (Euclidean)

$$\begin{pmatrix} X' \\ Y' \\ Z' \end{pmatrix} \rightarrow R \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + \begin{pmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{pmatrix}$$

where R is a rotation matrix ($R^{-1} = R^T$) is written

$$\begin{pmatrix} X' \\ Y' \\ Z' \\ 1 \end{pmatrix} = \begin{pmatrix} & & & \Delta X \\ & R & & \Delta Y \\ & & & \Delta Z \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

Perspective projection - Extrinsic parameters

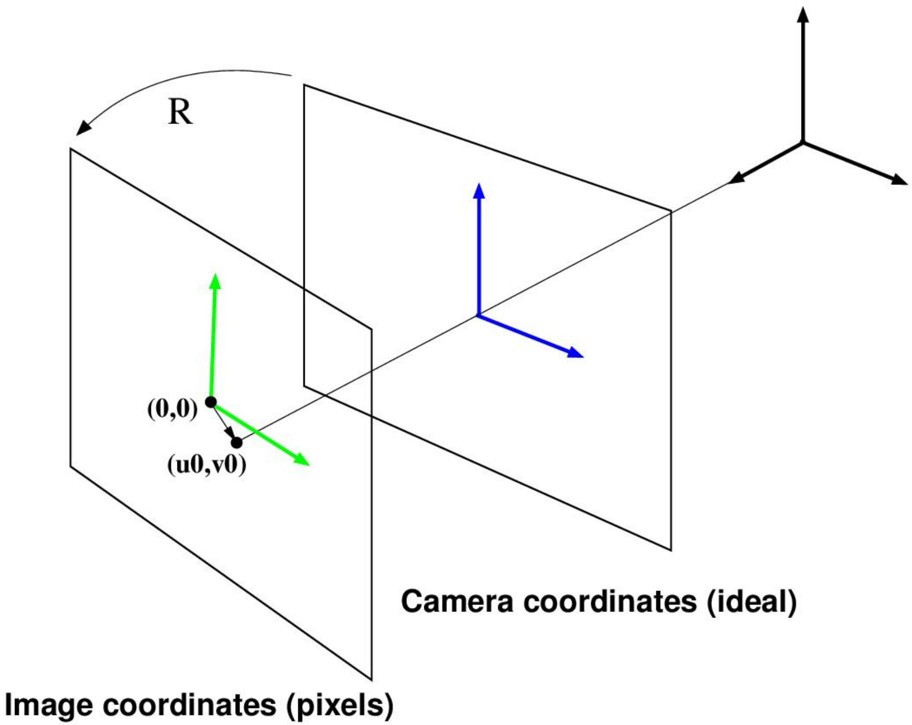
Consider world coordinates $(X',Y',Z',1)$ expressed in a coordinate system not aligned with the camera coordinate system

$$\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = \begin{pmatrix} & & & \Delta X \\ & R & & \Delta Y \\ & & & \Delta Z \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X' \\ Y' \\ Z' \\ 1 \end{pmatrix} = A \begin{pmatrix} X' \\ Y' \\ Z' \\ 1 \end{pmatrix}$$

Perspective projection (more general later)

$$c \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = PA \begin{pmatrix} X' \\ Y' \\ Z' \\ 1 \end{pmatrix} = M \begin{pmatrix} X' \\ Y' \\ Z' \\ 1 \end{pmatrix}$$

Intrinsic camera parameters



Due to imperfect placement of the camera chip relative to the lens system, there is always a small relative rotation and shift of center position.

Intrinsic camera parameters

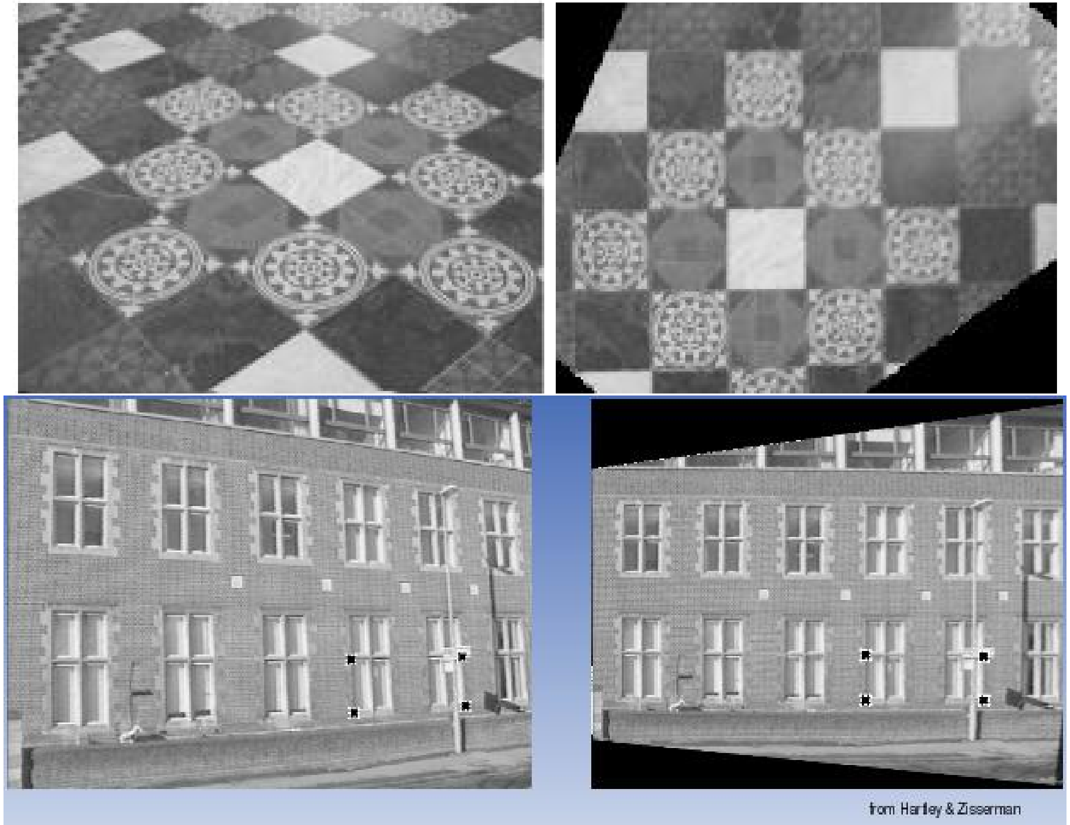
A more general projection matrix allows:

- Image coordinates with an offset origin
- Non-square pixels
- Skewed coordinate axes
- Five variables below are known as the camera's intrinsic parameters

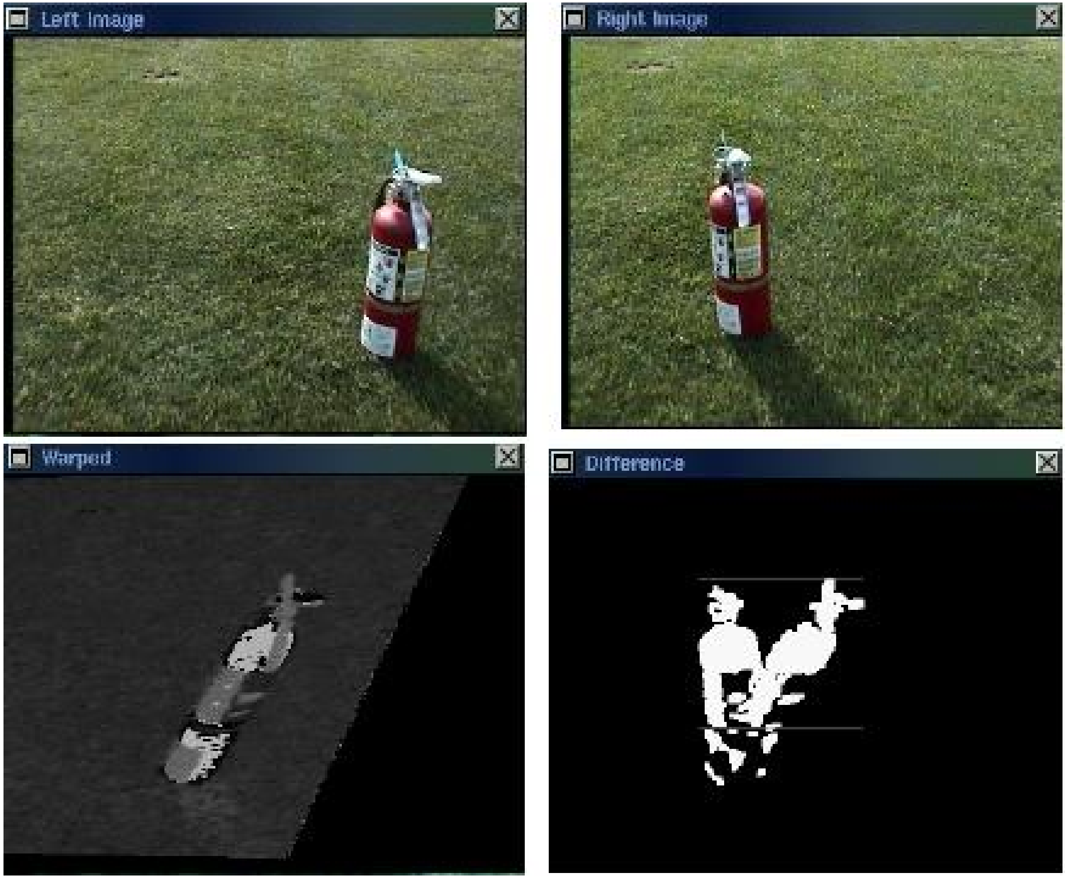
$$K = \begin{pmatrix} f_u & \gamma & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{pmatrix}, \quad P = \begin{pmatrix} K & 0 \end{pmatrix} = \begin{pmatrix} f_u & \gamma & u_0 & 0 \\ 0 & f_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

Most important is the focal length (f_u, f_v). Normally, f_u and f_v are assumed equal and the parameters γ, u_o and v_o close to zero.

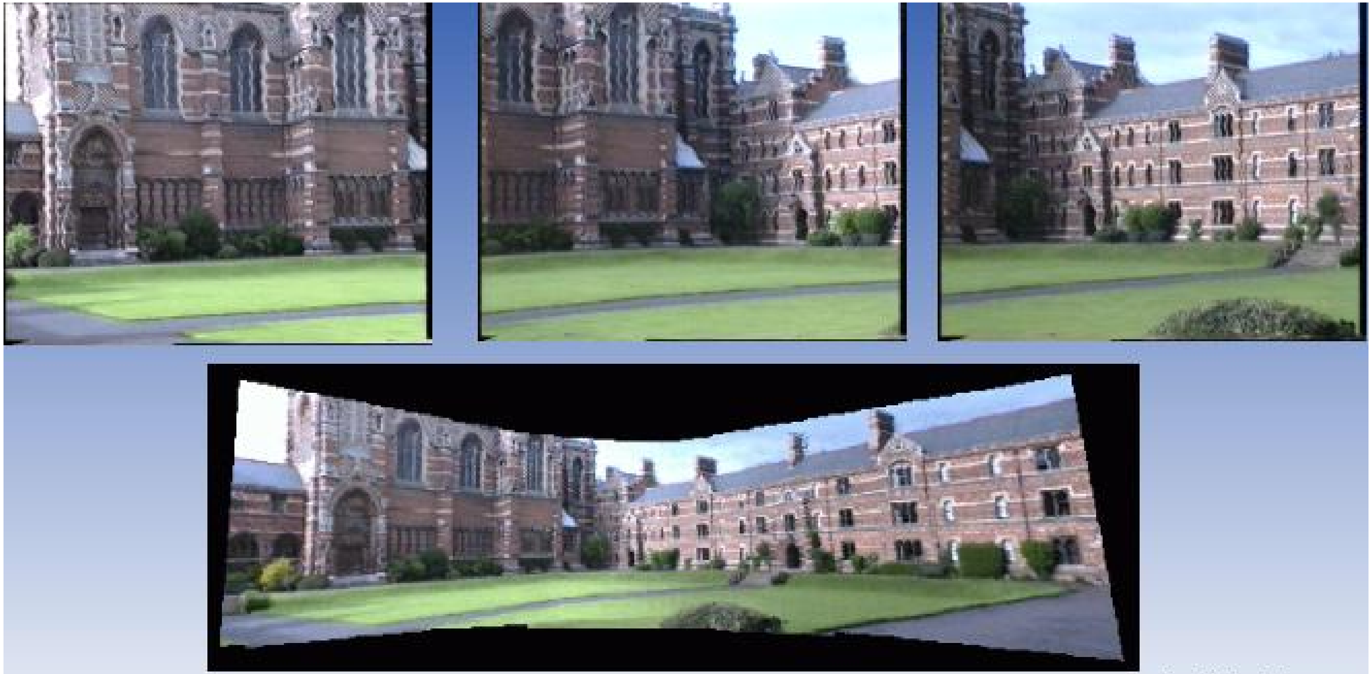
Example: Perspective mapping



Example: Perspective mapping in stereo



Mosaicing



from Hartley & Zisserman

Exercise

Assume you have a point at $(3m, -2m, 8m)$ with respect to the cameras coordinate system. What is the image coordinates, if the image has a size $(w, h) = (640, 480)$ and origin in the upper-left corner, and the focal length is $f = 480$?

Exercise

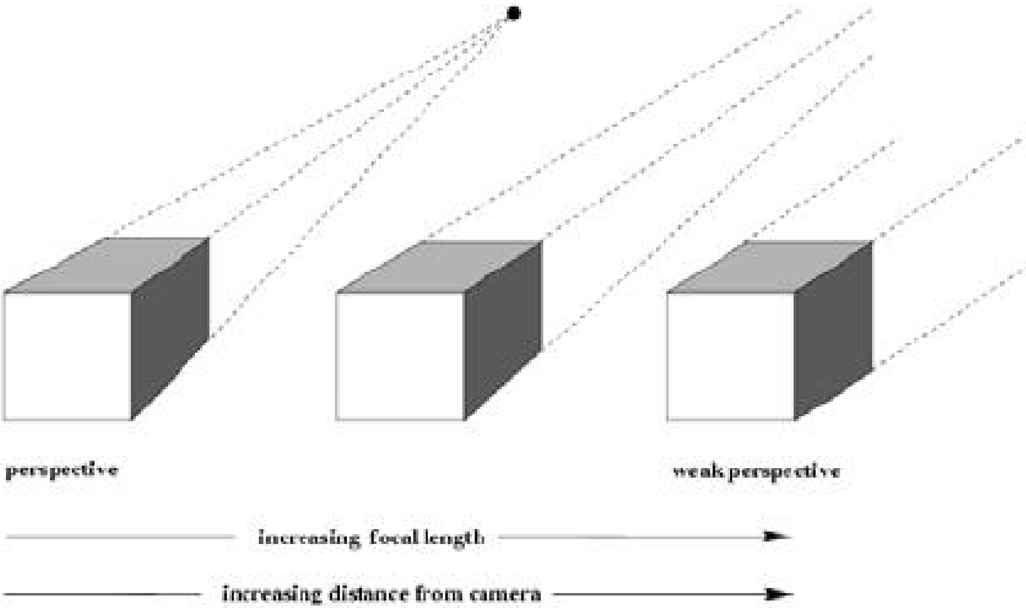
Assume you have a point at $(3m, -2m, 8m)$ with respect to the cameras coordinate system. What is the image coordinates, if the image has a size $(w, h) = (640, 480)$ and origin in the upper-left corner, and the focal length is $f = 480$?

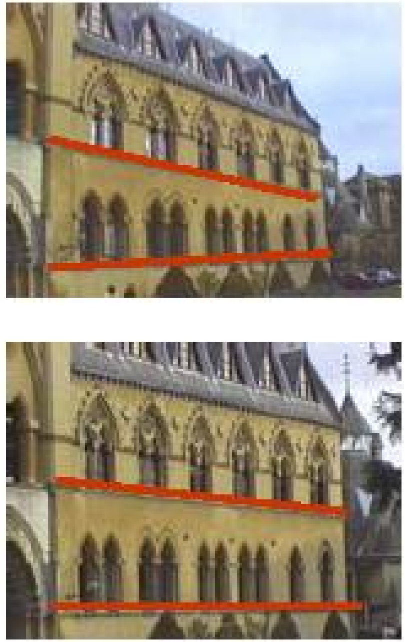
Answer:

$$x = f\frac{X}{Z} + \frac{w}{2} = (480 * 3/8 + 640/2) = 500$$

$$y = f\frac{Y}{Z} + \frac{h}{2} = (-480 * 2/8 + 480/2) = 120$$

Approximation: affine camera





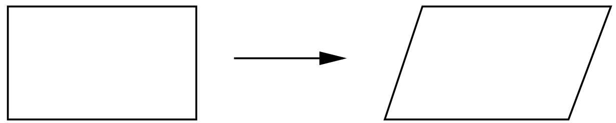
Approximation: affine camera

- A linear approximation of perspective projection

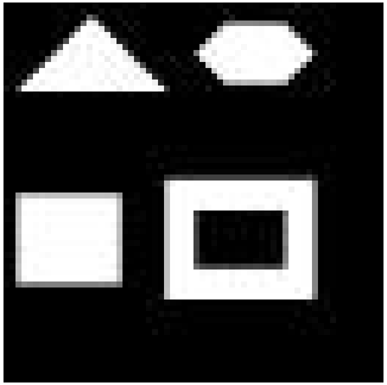
$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

- Basic properties
  - linear transformation (no need to divide at the end)
  - parallel lines in 3D mapped to parallel lines in 2D

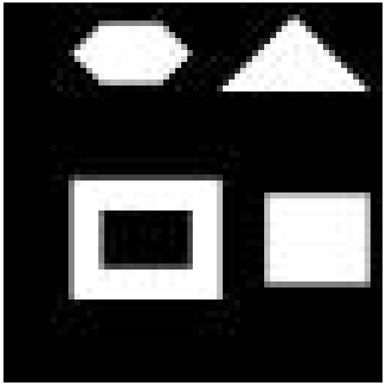
Angles are not preserved!



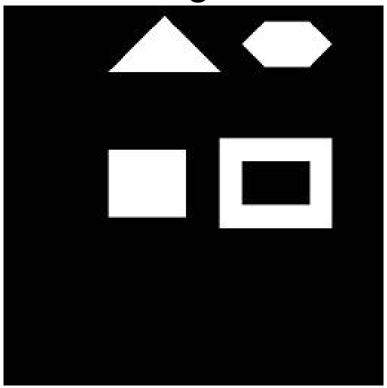
Planar Affine Transformation



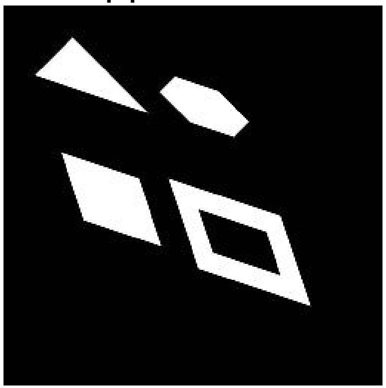
Original



Flipped x-size



Shifted and scaled



Sheared

Summary of models

Projective (11 degrees of freedom):

$$M = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{pmatrix}$$

Affine (8 degrees of freedom):

$$M = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Scaled orthographic (6 degrees of freedom):

$$M = \begin{pmatrix} r_{11} & r_{12} & r_{13} & \Delta X \\ r_{21} & r_{22} & r_{23} & \Delta Y \\ 0 & 0 & 0 & Z_0 \end{pmatrix}$$

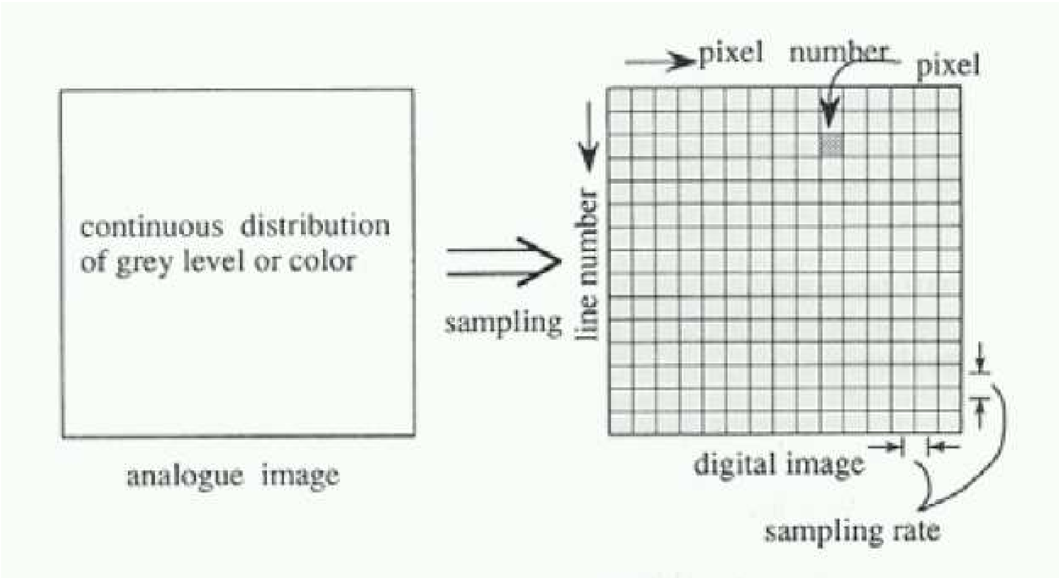
Orthographic (5 degrees of freedom):

$$M = \begin{pmatrix} r_{11} & r_{12} & r_{13} & \Delta X \\ r_{21} & r_{22} & r_{23} & \Delta Y \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

All these are just approximations, since they all assume a pin-hole, which is supposed to be infinitesimally small.

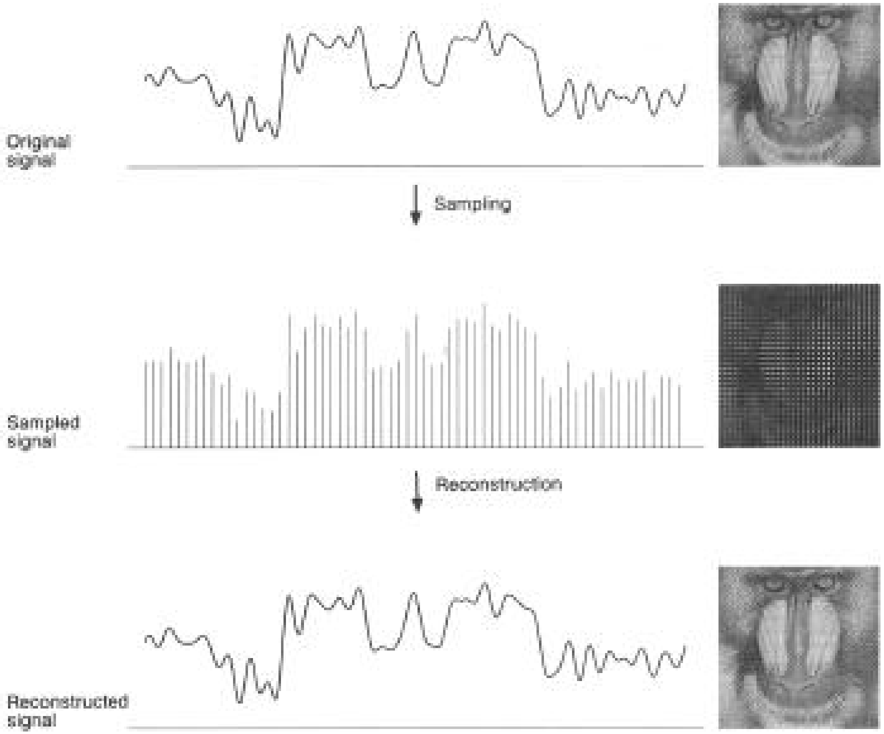
Sampling and quantization

- Sample the continuous signal at a finite set of points and quantize the registered values into a finite number of levels.
- Sampling distances $\Delta x, \Delta y$ and Δt determine how rapid spatial and temporal variations can be captured.



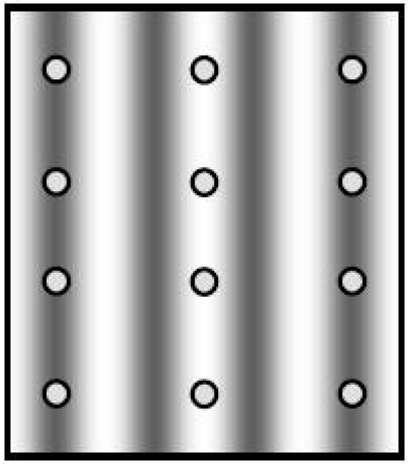
Sampling and quantization

- Sampling due to limited spatial and temporal resolution.
- Quantization due to limited intensity resolution.

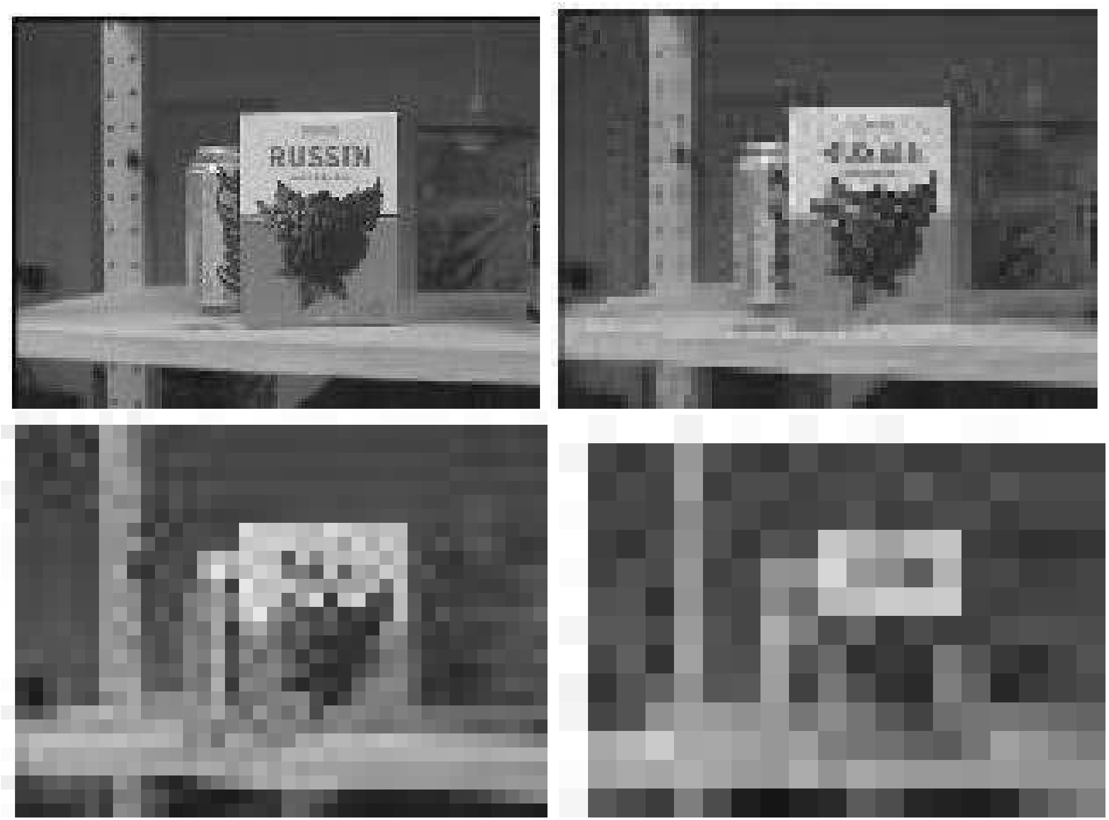


Factors that affect quality

- Quantization: Assigning, usually integer, values to pixels (sampling an amplitude of a function).
- Quantization error: Difference between the real value and assigned one.
- Saturation: When the physical value moves outside the allocated range, then it is represented by the end of range value.



Different image resolutions



Different number of bits per pixel

Image warping

Resample image $f(x,y)$ to get a new image $g(u,v)$, using a coordinate transformation: $u = u(x,y)$, $v = v(x,y)$.

Examples of transformations:

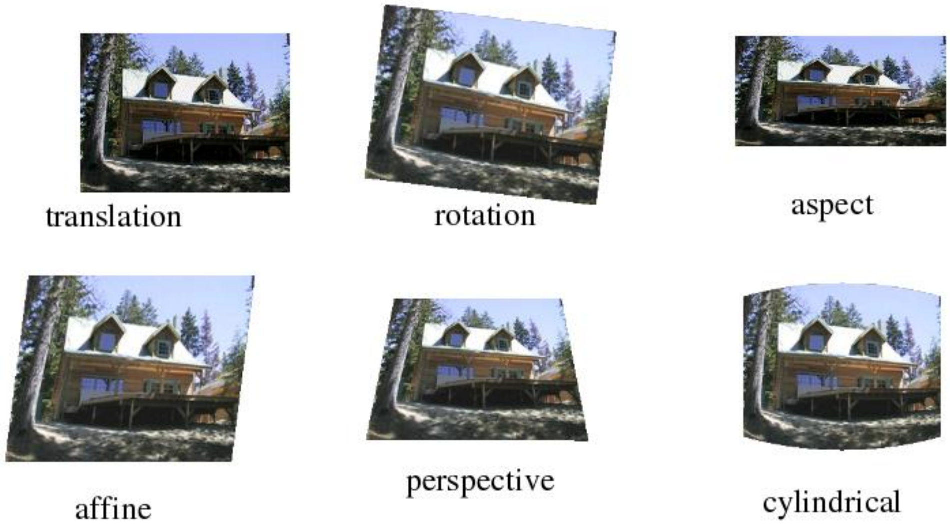
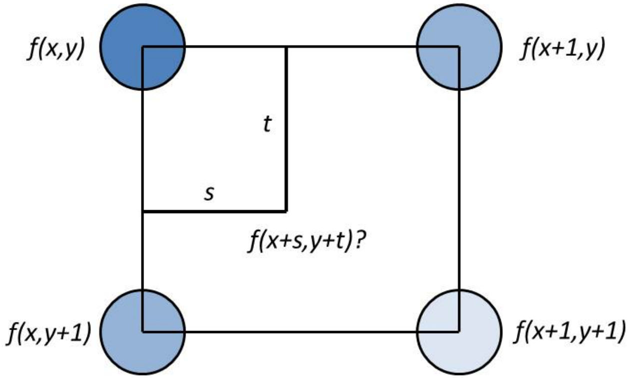


Image Warping

- For each grid point in (u, v) domain compute corresponding (x, y) values.

  Note: transformation is inverted to avoid holes in result.
- Create $g(u, v)$ by sampling from $f(x, y)$ either by:
  - Nearest neighbour look-up (noisy result)
  - Bilinear interpolation (blurry result)



$$\begin{aligned} f(x+s, y+t) &= (1-t)\cdot((1-s)\cdot f(x,y) + s\cdot f(x+1,y)) + \\ &+ t\cdot((1-s)\cdot f(x,y+1) + s\cdot f(x+1,y+1)) \end{aligned}$$

Nearest Neighbor vs. Bilinear Interpolation

Summary of good questions

- What parameters affects the quality in the acquisition process?
- What is a pinhole camera model?
- What is the difference between intrinsic and extrinsic camera parameters?
- How does a 3D point get projected to a pixel with a perspective projection?
- What are homogeneous coordinates and what are they good for?
- What is a vanishing point and how do you find it?
- What is an affine camera model?
- What is sampling and quantization?

Readings

- Gonzalez and Woods: Chapter 2
- Szeliski: Chapters 2.1 and 2.3.1