Image Acquisition and Camera Model

Hai Tao
3/29/02
4/1/02
Human visual system

The main components of the eyes (courtesy of Forsyth and Ponce, reprint from Tompson et al., 1966)
Human visual system

Rods and cones. Left: the distribution of rods and cones across the retina. Right: (a) cones in the fovea; (b) rods and cones in the periphery (courtesy of Forsyth and Ponce, reprint from Wandell, 1995). Blind spot - ganglion cell axons exit retina. In the fovea, the density of cones reaches $1.6 \times 10^5$/mm$^2$. Cones – less sensitive to light, but more nerve connection and high spatial resolution. Rods – very sensitive to light, but with low spatial resolution because many rods connection to a single neuron.
Pinhole cameras

- For an ideal pinhole camera, a single ray from any scene point reaches the image plane to form an inverted image. Therefore, the image is sharp.
- The same image can be obtained on a virtual image plane in front of the pinhole.
- The pinhole 3D-to-2D projection is called a perspective projection.
- Small pinholes allow a small amount of light to pass through, but real cameras use sophisticated optics to overcome this problem.
Thin lenses – the fundamental equation

Basic properties

- Any ray parallel to the optical axis on one side pass through the focus on the other side and vice versa
- A thin lens focuses all the rays from one point onto the same point

The Fundamental equation of thin lenses

\[
\frac{1}{z'} - \frac{1}{z} = \frac{1}{f}
\]

Can be treated as a pinhole camera with principal rays passing through the center of the lens
Thin lenses – field of view (FOV)

- Determined by the focal length and the effective area of the retina (or CCD, film, etc)

**Definition**
- $d$ – the effective diameter of the sensor/lens
- $f$ – focal length
- $f/d$ – F number
- FOV - $2\phi$

(courtesy of Forsyth and Ponce)
Camera geometry

3D camera coordinate system
Camera geometry

Perspective camera model - a 3D point $P=[X,Y,Z]^T$ is projected on to the image plane at $Z=f$ as point $p=[x,y,f]^T$, then

$$x = f \frac{X}{Z}$$

$$y = f \frac{Y}{Z}$$
Camera geometry

- **Orthographic camera model**

  \[
  x = X \\
  y = Y
  \]

- **Weak-perspective camera model** - a 3D point \( P = [X, Y, Z]^T \) with \( Z \) close to \( \bar{Z} \) projected on to the image plane at \( Z = f \) as point \( p = [x, y, f]^T \). It is a LINEAR relation! A good approximation of perspective camera when relief of the scene is shallow, i.e., \( |Z - \bar{Z}| / Z \) is small.

  \[
  x = \frac{f}{\bar{Z}} X \\
  y = \frac{f}{\bar{Z}} Y
  \]
Sensing

- **CCD cameras – charge-coupled-device**
  - Rectangular grid of electron-collection sites made on the wafer
  - Photons strike the silicon and the electrons are generated, which are collected over a fixed period of time
  - Charge coupling: change packets are transferred from site to site
  - Output one row at a time

(courtesy of Forsyth and Ponce)
The transformation between $p=[x,y,f]^T$ to $p_{im}=[x_{im},y_{im}]^T$, where $x_{im}$ and $y_{im}$ are image coordinates in pixel units. $S_x$ is the number of pixel per unit length along x direction, $S_y$ is the number of pixel per unit length at y direction. For CCD with square pixel $S_x = S_y$.

\[ x_{im} = xs_x + u_0 \]
\[ y_{im} = ys_y + v_0 \]
Homogeneous coordinate system and the camera model

- Homogeneous coordinates of a 2D points \([x,y]^T\) is \([x,y,1]^T\).
- Homogeneous coordinates \([x,y,1]^T\) and and its scaled version \([\lambda x, \lambda y, \lambda]^T\) represent the same 2D point. This is also regarded as “equal up to a scale”, or \([x,y,1]^T \equiv [\lambda x, \lambda y, \lambda]^T\)
- Homogeneous coordinates of a 3D points \([X,Y,Z]^T\) is \([X,Y,Z,1]^T\).
- Homogeneous coordinates \([X,Y,Z,1]^T\) and and its scaled version \([\lambda X, \lambda Y, \lambda Z, \lambda]^T\) represent the same 3D point, or \([X,Y,Z,1]^T \equiv [\lambda X, \lambda Y, \lambda Z, \lambda]^T\)
Camera model

Since
\[
\begin{bmatrix}
   x \\
   y \\
   f
\end{bmatrix} \equiv \begin{bmatrix}
   X \\
   Y \\
   Z
\end{bmatrix}
\]
and
\[
\begin{bmatrix}
   x_{im} \\
   y_{im}
\end{bmatrix} = \begin{bmatrix}
   s_x & 0 & u_0/f \\
   0 & s_y & v_0/f \\
   0 & 0 & 1/f
\end{bmatrix}
\begin{bmatrix}
   x \\
   y \\
   f
\end{bmatrix}
\]
therefore
\[
\begin{bmatrix}
   x_{im} \\
   y_{im}
\end{bmatrix} \equiv \begin{bmatrix}
   s_x f & 0 & u_0 \\
   0 & s_y f & v_0 \\
   0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
   X \\
   Y \\
   Z
\end{bmatrix}
\]

\(s_x f\) is often written as \(f_x\), similarly, \(s_y f\) is written as \(f_y\).

\[
K = \begin{bmatrix}
   f_x & 0 & u_0 \\
   0 & f_y & v_0 \\
   0 & 0 & 1
\end{bmatrix}
\]
is called the camera matrix. The intrinsic camera parameters are focal length \(f_x, f_y\), and the principal point \([u_0, v_0]\). \(f_x, f_y\) are measured in pixel.
World coordinate system and camera coordinate system

- The transformation is $P_c = R(P_w - T')$, or in homogeneous coordinate representation

\[
\begin{bmatrix}
X_c \\
Y_c \\
Z_c
\end{bmatrix} = [R, -RT]
\begin{bmatrix}
X_w \\
Y_w \\
Z_w \\
1
\end{bmatrix}
\]

$R$ is an orthonormal matrix with 3 degrees of freedom.
World coordinate system and camera coordinate system

- What is $O_c$ in the world coordinate system?
  
  Since $[0,0,0]^T = R(O_c - T)$, therefore $O_c = T$

- What is $O_w$ in the camera coordinate system?
  
  Since $R(0 - T)$, therefore $O_w = -RT$

- What are the axes of the camera coordinate system in the world coordinate system?
  
  Since $Rx = [1,0,0]^T$, therefore $x = R^T[1,0,0]^T$, which is the first row of $R$, and so on.

- What are the axes of the world coordinate system in the camera coordinate system?
  
  $x = R[1,0,0]^T$, which is the first column of the $R$. Similar for $y$, and $z$. 
Perspective projection

Remember that in the camera coordinate system

\[
\begin{bmatrix}
    x_{im} \\
    y_{im} \\
    1
\end{bmatrix}
\equiv
\begin{bmatrix}
    s_x f & 0 & u_0 \\
    0 & s_y f & v_0 \\
    0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
    X_c \\
    Y_c \\
    Z_c
\end{bmatrix}
\]

Therefore for a 3D point P in the world coordinate system, its image coordinates are

\[
\begin{bmatrix}
    x_{im} \\
    y_{im} \\
    1
\end{bmatrix}
\equiv K[R, -RT]
\begin{bmatrix}
    X_w \\
    Y_w \\
    Z_w \\
    1
\end{bmatrix}
\]

or more concisely \( p \equiv K[R, -RT]P \). R and T are called the camera motion or extrinsic camera parameters. \( M = K[R, -RT] \) is called the projection matrix.
For weak-perspective camera, we have

\[
\begin{bmatrix}
    s_x Z \\
    s_y Z \\
    1
\end{bmatrix}
\begin{bmatrix}
    f_x \\
    f_y \\
    Z
\end{bmatrix}
= 
\begin{bmatrix}
    u_0 \\
    v_0 \\
    1
\end{bmatrix}
\begin{bmatrix}
    X_c \\
    Y_c \\
    1
\end{bmatrix}
\]

where

\[
R = \begin{bmatrix}
    r_1^T \\
    r_2^T \\
    r_3^T
\end{bmatrix}
\]

In addition,

\[
\begin{bmatrix}
    X_c \\
    Y_c \\
    1
\end{bmatrix}
= 
\begin{bmatrix}
    f_x \\
    f_y \\
    Z
\end{bmatrix}
\begin{bmatrix}
    u_0 \\
    v_0 \\
    1
\end{bmatrix}
\begin{bmatrix}
    r_1^T \\
    r_2^T \\
    r_3^T
\end{bmatrix}
\]

Therefore

\[
\begin{bmatrix}
    X_w \\
    Y_w \\
    Z_w \\
    1
\end{bmatrix}
= 
\begin{bmatrix}
    f_x s_y Z \\
    s_y Z \\
    0
\end{bmatrix}
\begin{bmatrix}
    u_0 \\
    v_0 \\
    1
\end{bmatrix}
\begin{bmatrix}
    r_1^T \\
    r_2^T \\
    r_3^T
\end{bmatrix}
\]
Orthographic camera and general affine camera

The first two rows orthogonal to each other (except for the last column)

\[
M = \begin{bmatrix}
  s_x \frac{f}{Z} & 0 & u_0 \\
  0 & s_y \frac{f}{Z} & v_0 \\
  0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
  r_1^T & -r_1^T T \\
  r_2^T & -r_2^T T \\
  0 & 1
\end{bmatrix}
= \begin{bmatrix}
  s_x \frac{f}{Z} r_1^T & -s_x \frac{f}{Z} r_1^T T + u_0 \\
  s_y \frac{f}{Z} r_2^T & -s_y \frac{f}{Z} r_2^T T + v_0 \\
  0 & 1
\end{bmatrix}
\]

Orthographic camera

\[
M = \begin{bmatrix}
  s_x r_1^T & -s_x r_1^T T + u_0 \\
  s_y r_2^T & -s_y r_2^T T + v_0 \\
  0 & 1
\end{bmatrix}
\]

General affine camera

\[
M = \begin{bmatrix}
  m_{11} & m_{12} & m_{13} & t_1 \\
  m_{21} & m_{22} & m_{23} & t_2 \\
  0 & 0 & 0 & 1
\end{bmatrix}
\]
Thin lenses - radiometry

The fundamental equation of radiometric image formation

\[ E = \left[ \frac{\pi}{4} \left( \frac{d}{f} \right)^2 \cos^4 \alpha \right] L \]

- \( L \) – scene radiance at a point (power of light per unit area, per time unit, per unit solid angle)
- \( E \) – image irradiance (power of light per unit area, per time unit, on the image plane)
Thin lenses - radiometry

Remarks

- Image is “darker” around the border
- Smaller the F number, the brighter the image.
Chromatic aberration

The index of refraction depends on the wavelength of the light – light rays of different colors emitted from a single scene point follow different paths and will not meet at the same point.

(courtesy of Forsyth and Ponce, reprint from Navy 1969)
Lens distortion

- Real length can be better modeled using thick lens model
- Geometric distortion caused by difference in focal length at different parts of the lens
- Barrel distortion can be modeled using simple radial function with the relation
  \[ x = x_d (1 + k_1 r^2 + k_2 r^4) \]
  \[ y = y_d (1 + k_1 r^2 + k_2 r^4) \]
  where \( r^2 = x_d^2 + y_d^2 \)

Nominal image

Barrel distortion
Lens distortion

Example

Original image with lens distortion
Corrected with $k_1=0.11$, $k_2=0.019$

(courtesy of Shawn Becker)
Still cameras and video cameras

- **Still cameras**
  - Film cameras – effective resolution is approximately 6,000 by 4,000 pixels
  - Digital cameras – 1M pixels (low end), 3-3.5 M pixels (midrange), better optics and resolution (high end), price is dropping

- **Video cameras**
  - Analog video camcorders (VHS, Hi8, etc), need frame grabber (internal PCI board or external converter) for digitization. (resolution 640x480 or 720x486 depends on frame grabbers)
  - Digital video camcorders (resolution 640x480)
    - Store on digital media such as Mini DV
    - Output digital video - USB 1.1 12 Mbps, FireWire 400 Mbps, USB 2.0 480 Mbps
  - Web cams – low image quality, low price
  - Industry video cameras – various specs
Acquiring high quality images and videos

Main elements of a camera
- Lens – controlled by focusing ring and zoom
- Aperture – controlled by aperture ring
- Shutter – speed control

Goals
- Find the combination that achieves sharpest images for the largest range of scene with correct exposure

(reprint from The Complete Kodak Book of Photography)
Focus

Manual focus and auto focus

If focused at close distance, only a shallow part of the scene appear sharp.

(reprint from The Complete Kodak Book of Photography)
Shutter speed

- The amount of time a film/CCD is exposed to the light
- A slow shutter speed causes motion blur
- 1/60 or 1/100 is good for common tasks

1/30 sec  1/125  1/500

(reprint from The Complete Kodak Book of Photography)
Aperture

- Opening of the length, usually has the scale of f/1.4, f/1.4, f/2, f/2.8, f/4, f/5.6, f/8, f/11, f/16, f/22.
- The numbers get bigger, the aperture opening gets smaller
- Smaller aperture means sharper image (pinhole camera)

(reprint from The Complete Kodak Book of Photography)
Tips

- Unwanted camera motion during the exposure can be reduced by holding the camera still or using a tripod
- Using lighting equipment to achieve small aperture+fast shutter speed
Homework

A camera has the following intrinsic camera matrix. If it is also known that the size of the output images is 640x480 pixels, what is the focal length (in pixels)? What is the field of view of the camera?

\[
K = \begin{bmatrix}
1000 & 0 & 320 \\
0 & 1000 & 240 \\
0 & 0 & 1
\end{bmatrix}
\]