Image Acquisition and Representation

- how digital images are produced
- how digital images are represented
- photometric models-basic radiometry
- image noises and noise suppression methods

Camera

- First photograph was due to Niepce of France in 1827.
- Basic abstraction is the pinhole camera
  - lenses required to ensure image is not too dark
  - various other abstractions can be applied

CCD Camera

- CCD ( Charged Couple Device) camera consists of a lens and an image plane (chip array) containing tiny solid cells that convert light energy into electrical charge. The output is analog image. The key camera parameters include
  - image plane geometries: rectangle, circular, or liner.

Note a digital camera represents a camera system with a built-in digitizer.
- chip array size (e.g. 512 × 512, also referred to as camera resolution, i.e., the number of cells horizontally and vertically).
- cell size (e.g., 16.6 × 12.4 μm, aspect ratio=4:3, not square)
- Spectral response (28%(450nm), 45%(550nm), 62%(650nm))
  visible light: 390-750 nm, IR light 750 nm and higher
- Aperture

Figure 1: CCD camera image plane layout

Other CCD array geometries

Analog Image

An analog image is a 2D image \( F(x, y) \) which has infinite precision in spatial parameters \( x \) and \( y \) and infinite precision in intensity at each point \( (x, y) \).

Usually, \( H \times W/V \times L=4:3 \). This aspect ratio is more suitable for human viewing. For machine vision, aspect ratio of 1:1 is preferred.
CMOS Camera

A CMOS (Complementary Metal Oxide Silicon) camera is an alternative image sensor. It follows the same principle as CCD by converting photons into electrical changes. But it uses different technologies in converting and transporting the electrical charges. Compared to CCD, it's speed is faster and consume less power, and is smaller in size. But its light sensitivity is lower and its image is more noisy. CMOS camera is mainly for low-end consumer applications.

Frame Grabber

An A/D converter that spatially samples the camera image plane and quantizes the voltage into a numerical intensity value.

- Sample frequency (sampling interval) v. image resolution through spatial sampling
- Range of intensity value through amplitude quantization
- On-board memory and processing capabilities

Spatial sampling process

Let \((x,y)\) and \((c,r)\) be the image coordinates before and after sampling. Spatial sampling converts \((x,y)\) to \((c,r)\)

\[
\begin{pmatrix}
c \\
r
\end{pmatrix}
= \begin{pmatrix}
s_x & 0 \\
0 & s_y
\end{pmatrix}
\begin{pmatrix}
x \\
y
\end{pmatrix}
\]

where \(s_x\) and \(s_y\) are sampling frequency (pixels/mm) due to spatial quantization. They are also referred to as scale factors. The sampling frequency determines the image resolution. The higher sampling frequency, the higher image resolution. But the image resolution is limited by camera resolution. Oversampling by the frame grabber requires interpolation and does not necessarily improve image perception.

Amplitude Quantization

Amplitude quantization converts the magnitude of the signal \(F(x,y)\) to produce pixel intensity \(I(c,r)\). The \(I(c,r)\) is obtained by dividing the range of \(F(x,y)\) into intervals and representing each interval with an integer number. The number of intervals to represent \(I(c,r)\) is determined by the number of bits allocated to represent \(F(x,y)\). For example, if 8-bit is used, then \(F(x,y)\) can be divided into 256 intervals with the first interval represented by 0 and the last interval represented by 255. \(I(c,r)\) therefore ranges from 0 to 255.
Computer

Computer (including CPU and monitor): used to access images stored in the frame grabber, process them, and display the results on a monitor.

Digital Image

The result of digitization of an analog image $F(x,y)$ is a digital image $I(c,r)$. $I(c,r)$ is represented by a discrete 2D array of intensity samples, each of which is represented using a limited precision determined by the number of bits for each pixel.

Digital Image (cont’d)

- Image resolution ($W \times H$)
- Intensity range $[0, 2^N-1]$
- Color image (RGB)

Digital Representation

The result of digitization of an analog image $F(x,y)$ is a digital image $I(c,r)$. $I(c,r)$ is represented by a discrete 2D array of intensity samples, each of which is represented using a limited precision determined by the number of bits for each pixel.
Different coordinate systems used for images

(a) Row-column coordinate system with (0,0) at the upper-left corner, (b) Cartesian coordinate system with (0,0) at the lower left corner, and (c) Cartesian coordinate system with (0,0) at the center.

Basic Optics: Pinhole model

Reducing the camera's aperture to a point so that one ray from any given 3D point can enter the camera and create a one-to-one correspondence between visible 3D points and image points.

Pinhole model (cont'd)

Distant objects are smaller due to perspective projection. Larger objects appear larger in the image.
Pinhole model (cont'd)

Parallel lines meet at horizon, where line H is formed by the intersection of the plane parallel to the lines and passing through V, which is referred as vanishing point.

Camera Lens

Lens may be used to focus light so that objects may be viewed brighter. Lens can also increase the size of the objects so that objects in the distance can appear larger.

Without lens in the top figure and with lens in the bottom figure

Basic Optics: Lens Parameters

Lens parameters: focal length (f) and effective diameter (d)

Fundamental equation of thin Lens

\[ \frac{1}{f} = \frac{1}{u} + \frac{1}{v} \]

It is clear that increasing the object distance, while keeping the same focus length, reduces image size. Keeping the object distance, while increasing the focus length, increases the image size.
Angle (Field) of View (AOV)

Angular measure of the portion of 3D space actually seen by the camera. It is defined as

\[ \omega = 2 \arctan \frac{d}{2f} \]

AOV is inversely proportional to focal length and proportional to lens size. Larger lens or smaller focal length give larger AOV.

\( \frac{f}{d} \) is called F-number. AOV is inversely proportional to F-number.

Similar to AOV, Field of View (FOV) determines the portion of an object that is observable in the image. But different from AOV, which is a camera intrinsic parameter and is a function of only lens of parameters, FOV is a camera extrinsic parameter that depend both on lens parameters and object parameters. In fact, FOV is determined by focus length, lens size, object size, and object distance to the camera.

Depth of Field

The allowable distance range such that all points within the range are acceptably (this is subjective!) in focus in the image.

Depth of field is inversely proportional to focus length, proportional to shooting distance, and inversely proportional to the aperture (especially for close-up or with zoom lens).
Since “acceptably in focus” is subjective, as the focus length increases or shooting distance decreases (both make the picture more clear and larger), the tolerance in picture blurriness also decreases, hence a reduction in depth of field.

**Camera and Lens Parameter Summary**
- Camera resolution
- Camera spectral response
- Aperture
- Image resolution
- Lens focus length (f)
- Lens diameter (d)
- Angle of view
- Field of view
- Depth of field

**Other Lens Parameters**
- Fixed focal length v. Zoom lens
- Motorized zoom Lenses—zoom lenses are typically controlled by built-in, variable-speed electric motors. These electric zooms are often referred to as servo-controlled zooms
- Supplementary lens: positive and negative (increase/decrease AOV)
- Digital zoom: a method to digitally change the focus length to focus on certain region of the image typically through interpolation.
Effects of Lens Distortion

Figure 2: Effect of radial distortion. Solid lines: no distortion; dashed lines with distortion. More distortion far away from the center.

Lens Distortion modeling and correction

Radial lens distortion causes image points to be displaced from their proper locations along radial lines from the image center. The distortion can be modeled by

\[
\begin{align*}
  u &= u_d(1 + k_1 r^2 + k_2 r^4) \\
  v &= v_d(1 + k_1 r^2 + k_2 r^4)
\end{align*}
\]

where \( r = \sqrt{(u - u_0)^2 + (v - v_0)^2} \), \((u, v)\) is the ideal and unobserved image coordinates relative to the \((U, V)\) image frame, \((u_d, v_d)\) is the observed and distorted image coordinates, \((u_0, v_0)\) is the center of the image, \(k_1\) and \(k_2\) are coefficients. \(k_2\) is often very small and can be ignored. Besides radial distortion, another type of geometric distortion is tangential distortion. It is however much smaller than radial distortion.

The geometric knowledge of 3D structure (e.g. collinear or coplanar points, parallel lines, angles, and distances) is often used to solve for the distortion coefficients. Refer to [http://www.media.mit.edu/people/sbeck/results/Distortion/distortion.html](http://www.media.mit.edu/people/sbeck/results/Distortion/distortion.html) for lens calibration using parallel lines.

Structure of Eye

- cornea-the front and the transparent part of the coat of the eyeball that reflects and refracts the incoming light
- pupil-the opening in the center of iris that controls the amount of light entering into the eyes
- iris-the colored tiny muscles that surround the pupil. It controls the opening

Figure 3: Radial lens distortion before (a) and after (b) correction

With the modern optics technology and for most computer vision applications, both types of geometric lens distortions are often negligible.
and closing of the pupil
• lens—the crystalline lens located just behind the iris. Its purpose is to focus
  the light on retina.
• retina—the sensory photo-electric sensitive tissue at the back of the eye. It
  captures light and converts it to electrical impulses.
• optic nerve—the optic nerve transmits electrical impulses from the retina to
  the brain.

The question is if it is possible to produce (simulate) the electrical impulses by
other means (e.g., through hearing or other sensing channels) and send the
signals to the brain as if they were from the eyes.

Yes, this is can be done!. Research about bionic eyes is doing this. See the
video at http://www.youtube.com/watch?v=696dxY6BYBM

---

Basic Radiometry

We introduce the basic photometric image model.

\[
\begin{align*}
L & : \text{Light source} \\
N & : \text{Illumination vector} \\
R & : \text{Scene radiance} \\
E & : \text{Image irradiance} \\
I & : \text{Image intensity} \\
\end{align*}
\]

Scene radiance \( R \): is the power of the light, per unit area, ideally emitted by
a 3D point

- Image irradiance \( E \): the power of the light per unit area a CCD array
  element receives from the 3D point
- Image intensity \( I \): the intensity of the corresponding image point

Lambertian Surface Reflectance Model

\[
R = \rho L \cdot N
\]

where \( L \) represents the incident light, \( N \) surface normal, and \( \rho \) surface albedo.
The object looks equally bright from all view directions.
Surface Radiance and Image Irradiance

The fundamental radiometric equation:

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

For small angular aperture (pin-hole) or object far from camera, \( \alpha \) is small, the \( \cos^4 \alpha \) can be ignored. The image irradiance is uniformly proportional to scene radiance. Large \( d \) or small F number produces more image irradiance and hence brighter image.

Image Irradiance and Image Intensity

\[ I = \beta E \]

where \( \beta \) is a coefficient dependent on camera and frame grabber settings.

The Fundamental Image Radiometric Equation

\[ I = \beta \rho \frac{\pi}{4} \left( \frac{d}{f} \right)^2 \cos^4 \alpha L \cdot N \]

Image Formats

Images are usually stored in computer in different formats. There are two image formats: Raster and Vector.
A Raster image consists of a grid of colored dots called pixels. The number of bits used to represent the gray levels (or colors) denotes the depth of each pixel. Raster files store the location and color of every pixel in the image in a sequential format.

There are many different Raster image formats such as TIFF, PGM, JPEG, GIF, and PNG. They all can be organized as follows:

- image header (in ASCII, image size, depth, date, creator, etc.)
- image data (in binary either compressed or uncompressed) arranged in sequential order.

PGM stands for Portable Greyscale Map. Its header consists of:

- P5
- number of columns number of rows
- Max intensity (determine the no of bits)
- Raw image data (in binary, pixels are arranged sequentially)

Some software may add additional information to the header. For example, the PGM header created by XV looks like:

```
P5
320 240
255
```
**PPM**

PPM (Portable PixMap) format is for color image. Use the same format.

P6
640 480
255
raw image data (each pixel consists of 3 bytes data in binary)

---

**Vector Format**

A Vector image is composed of lines, not pixels. Pixel information is not stored; instead, formulas that describe what the graphic looks like are stored. They're actual vectors of data stored in mathematical formats rather than bits of colored dots. Vector format is good for image cropping, scaling, shrinking, and enlarging but is not good for displaying continuous-tone images.

---

**Image noise**

- intensity noise
- positional error

Note image noise is the intrinsic property of the camera or sensor, independent of the scene being observed. It may be used to identify the imaging sensors/cameras.

---

**Intensity Noise Model**

Let \( \hat{I} \) be the observed image intensity at an image point and \( I \) be the ideal image intensity, then

\[
\hat{I}(c, r) = I(c, r) + \epsilon(c, r)
\]

where \( \epsilon \) is white image noise, following a distribution of \( \epsilon \sim N(0, \sigma^2(c, r)) \).

Note we do not assume each pixel is identically and independently perturbed.
**Estimate \( \sigma \) from Multiple Images**

Given \( N \) images of the same scene \( \hat{I}_0, \hat{I}_1, ..., \hat{I}_{N-1} \), for each pixel \( (c, r) \),

\[
\hat{I}(c, r) = \frac{1}{N} \sum_{i=0}^{N-1} \hat{I}_i(c, r)
\]

\[
\sigma(c, r) = \left\{ \frac{1}{N - 1} \sum_{i=0}^{N-1} [\hat{I}_i(c, r) - \bar{I}(c, r)]^2 \right\}^{\frac{1}{2}}
\]

see figure 2.11 (Trucco’s book). Note noise averaging can reduce the noise of \( \bar{I}(c, r) \) to \( \frac{\sigma^2}{N} \).

---

**Estimate \( \sigma \) from a Single Image**

Assume the pixel noise in the neighborhood is IID distributed, i.e.,

\[
\hat{I}(c, r) = I(c, r) + \epsilon
\]

where \( (c, r) \in R \). \( \sigma \) can then be estimated by sample variance of the pixels inside \( R \)

\[
\hat{\sigma}(c, r) = \sqrt{\frac{\sum_{(c,r) \in R} (I(c, r) - \bar{I})^2}{N - 1}}
\]

(2)

As a result, we can obtain \( \hat{\sigma}^2 \), an estimate of \( \sigma^2 \), as follows

\[
\hat{\sigma}^2 = \frac{\epsilon^2}{M \times N - 2}
\]

Let \( \hat{\sigma}^2_k \) be an estimate of \( \sigma^2 \) from the k-th neighborhood. Given a total of \( K \) neighborhoods across the image, we can obtain

\[
\hat{\sigma}^2 = \frac{1}{K} \sum_{k=1}^{K} \hat{\sigma}^2_k
\]

Note here we assume each pixel is identically and independently perturbed.

---

*we can obtain the same estimate by using the samples in the neighborhood and assumes each sample is IID distributed.*
Independence Assumption Test

We want to study the validity of the independence assumption among pixel values. To do so, we compute correlation between neighboring pixel intensities. Figure 2.12 (Trucco's book) plot the results. We can conclude that neighboring pixel intensities correlate with each other and the independence assumption basically holds for pixels that are far away from each other.

Consequences of Image Noise

- image degradation
- errors in the subsequent computations e.g., derivatives

Types of Image Noise

Gaussian Noise and impulsive (salt and pepper) noise.
Noise Removal

In image processing, intensity noise is attenuated via filtering. It is often true that image noise is contained in the high frequency components of an image, a low-pass filter can therefore reduce noise. The disadvantage of using a low-pass filter is that image is blurred in the regions with sharp intensity variations, e.g., near edges.

Noise Filtering

\[ I_f(x, y) = I * F = \sum_{h=-m}^{m} \sum_{k=-m}^{m} F(h, k)I(x-h, y-k) \]

where \( m \) is the window size of filter \( F \) and \( * \) indicates discrete convolution. The filtering process replaces the intensity of a pixel with a linear combination of neighborhood pixel intensities.

Noise Filtering (cont'd)

- Filtering by averaging

\[ F = \frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \]

- Gaussian filtering

\[ g(x, y) = \frac{1}{2\pi} e^{-\frac{1}{2} \left( \frac{x^2+y^2}{\sigma^2} \right)} \]

window size \( w = 5\sigma \).

An example of \( 5 \times 5 \) Gaussian filter

Figure 3.2 (a) Results of applying Gaussian filtering kernel with \( \sigma = 1 \) to the "Lena.png" image corrupted by Gaussian white noise and impulse noise. (b) Results of applying the "Lena.png" image corrupted by salt-and-pepper noise.
Noise Filtering (cont’d)

Gaussian filtering has two advantages:

- no secondary lobes in the frequency domain (see figure 3.3 (Trucco’s book)).
- can be implemented efficiently by using two 1D Gaussian filters.

Non-linear Filtering

Median filtering is a filter that replaces each pixel value by the median values found in a local neighborhood. It performs better than the low pass filter in that it does not smear the edges as much and is especially effective for salt and pepper noise.

Signal to Noise Ratio

$$SNR = 10 \log_{10} \frac{S}{N} dB$$

For image, SNR can be estimated from

$$SNR = 10 \log_{10} \frac{I}{\sigma}$$

where \(I\) is the unperturbed image intensity.
Quantization Error

Let \((c, r)\) be the pixel position of an image point resulted from spatial quantization of \((x, y)\), the actual position of the image point. Assume the width and length of each pixel (pixel/mm), i.e., the scale factors, are \(s_x\) and \(s_y\) respectively, then \((x, y)\) and \((r, c)\) are related via

\[
\begin{align*}
c &= s_x x + \xi_x \\
r &= s_y y + \xi_y
\end{align*}
\]

where \(\xi_x\) and \(\xi_y\) represent the spatial quantization errors in \(x\) and \(y\) directions respectively.

Quantization Error (cont’d)

Assume \(\xi_x\) and \(\xi_y\) are uniformly distributed over the range determined by \([-0.5s_x, 0.5s_x]\) and \([-0.5s_y, 0.5s_y]\), i.e.,

\[
\begin{align*}
f(\xi_x) &= \begin{cases} 
\frac{1}{s_x} & -0.5s_x \leq \xi_x \leq 0.5s_x \\
0 & \text{otherwise}
\end{cases} \\
f(\xi_y) &= \begin{cases} 
\frac{1}{s_y} & -0.5s_y \leq \xi_y \leq 0.5s_y \\
0 & \text{otherwise}
\end{cases}
\end{align*}
\]

Now let’s estimate variance of row and column coordinates \(c\) and \(r\).

\[
\begin{align*}
Var(c) &= Var(\xi_x) = \frac{s_x^2}{12} \\
Var(r) &= Var(\xi_y) = \frac{s_y^2}{12}
\end{align*}
\]