Recap

- Analyze the noise
  - Type of noise
    - Spatial invariant
    - Periodic noise
  - How to identify the type of noise?
    - Test pattern
    - Histogram
  - How to evaluate noise level?
    - RMS
    - PSNR
- Noise removal
  - Spatial domain
    - Mean filters
    - Order-statistics filters
    - Adaptive filters
  - Frequency domain
    - Random/Landmark
    - Peak filters
    - Optimal model filters
- Analyze the blur
  - Linear, position-invariant degradation model
    - Modeled by convolution*
  - The point spread function (PSF)
    - Theoretically
    - Deblurring: an ill-posed problem
    - Ill-conditioning of the linear system
- Understand why image restoration is an ill-posed problem and what it means conceptually
- Different restoration approaches
  - Frequency domain
    - Inverse filter
    - Wiener filter
  - Spatial domain
    - Unconstrained approach
    - Constrained approach
    - MAP

Questions

- What is PSF? How to estimate it?
- What is an ill-posed problem? What is an ill-conditioning system?
- Inverse filter and problem?
- Wiener filter and how it solved the problem?
- Unconstrained vs. Constrained approaches (572)
- What is regularization? (572)
Image restoration

- Degradation model

\[ f(x, y) \xrightarrow{H} g(x, y) \]

\[ g(x, y) = H[f(x, y)] + \eta(x, y) \]

Linear vs. Non-linear

- Many types of degradation can be \textit{approximated} by linear, space-invariant processes
- Non-linear and space-variant models are more accurate
  - Difficult to solve
  - Unsolvable

Linear, position-invariant degradation model

Sampling theorem

\[ f(x, y) = \int \int f(\alpha, \beta) |(x - \alpha, y - \beta)| d\alpha d\beta \]

\[ g(x, y) = \int \int f(\alpha, \beta) [h(x, y - \beta) + \eta(x, y)] d\alpha d\beta \]

- Linearity - additivity

\[ \int \int f(\alpha, \beta) [h(x - \alpha, y - \beta) + \eta(x, y)] d\alpha d\beta \]

- Linearity - homogeneity

\[ \int \int f(\alpha, \beta) [\eta(x - \alpha, y - \beta) + \eta(x, y)] d\alpha d\beta \]

- Space invariant

\[ \int \int f(\alpha, \beta) [\eta(x - \alpha, y - \beta) + \eta(x, y)] d\alpha d\beta \]

Convolution integral

\[ \int \int f(\alpha, \beta) |(x - \alpha, y - \beta)| d\alpha d\beta + \eta(x, y) \]
**PSF - Point Spread Function**

- Impulse response of system $H$
  
  $$h(x, \alpha, y, \beta) = H[\delta(x-\alpha, y-\beta)]$$

- Point spread function (PSF)
  - Used in optics - The impulse becomes a point of light $\rightarrow$ impulse response
  - Completely characterize the linear system

**Estimate the degradation**

- By observation
- By experiment
  
  - $g(x, y) = h(x, y)^{*} f(x, y) + \eta(x, y)$
  - $G(u, v) = H(u, v) F(u, v) + N(u, v)$
  - $H(u, v) = G(u, v)$
- By mathematical modeling
  - Sec. 5.6.3

**Image restoration – An ill-posed problem**

- Degradation model
  
  $$G(u, v) = H(u, v) F(u, v) + N(u, v)$$

  $$\tilde{F}(u, v) = \frac{G(u, v)}{H(u, v)} = \frac{F(u, v)}{H(u, v)} + \frac{N(u, v)}{H(u, v)}$$

- $H$ is ill-conditioned which makes image restoration problem an ill-posed problem
  
  - Solution is not stable
**Ill-conditioning**

\[ Ax = b \]

\[
A = \begin{bmatrix} 0.78 & 0.563 \\ 0.913 & 0.659 \end{bmatrix}
\]

\[
b = \begin{bmatrix} 0.217 \\ 0.254 \end{bmatrix}
\]

\[
x = \begin{bmatrix} 1 \\ -1 \end{bmatrix}
\]

\[
E = \begin{bmatrix} 0.001 & 0.001 \\ -0.002 & -0.001 \end{bmatrix}
\]

\[
x = \begin{bmatrix} 7.3085 \\ -5 \end{bmatrix}
\]

\[ \text{cond}(A) = 2.1932 \times 10^6 \]

---

**Example**

Noise-free

- Exact H

Sinusoidal noise

- Exact H

- Not exact H

---

**Different restoration approaches**

- **Frequency domain**
  - Inverse filter
  - Wiener (minimum mean square error) filter

- **Algebraic approaches**
  - Unconstrained optimization
  - Constrained optimization
  - The regularization theory
The block-circulant matrix

- Stacking rows of image \( f, g, n \) to make \( MN \times 1 \) column vectors \( f, g, n \). (Also called lexicographic representation of the original image). Correspondingly, \( H \) should be a \( MN \times MN \) matrix
- \( H \) is called block-circulant matrix

\[
H = \begin{bmatrix}
H_0 & H_{-1} & \cdots & H_{N-2} \\
H_1 & H_0 & \cdots & H_{N-3} \\
\vdots & \vdots & \ddots & \vdots \\
H_{N-2} & H_{N-3} & \cdots & H_0
\end{bmatrix}
\]

\[ H = \begin{bmatrix}
h(j,0) & h(j,N-1) & \cdots & h(j,1) \\
h(j,1) & h(j,0) & \cdots & h(j,2) \\
\vdots & \vdots & \ddots & \vdots \\
h(j,N-1) & h(j,N-2) & \cdots & h(j,0)
\end{bmatrix} \]

Inverse filter

- In most images, adjacent pixels are highly correlated, while the gray levels of widely separated pixels are only loosely correlated.
- Therefore, the autocorrelation function of typical images generally decreases away from the origin.
- Power spectrum of an image is the Fourier transform of its autocorrelation function, therefore, we can argue that the power spectrum of an image generally decreases with frequency.
- Typical noise sources have either a flat power spectrum or one that decreases with frequency more slowly than typical image power spectra.
- Therefore, the expected situation is for the signal to dominate the spectrum at low frequencies while the noise dominates at high frequencies.

Wiener filter (1942)

- Objective function: find an estimate \( \hat{f} \) of \( f \) such that the mean square error between them is minimized

\[
e^2 = E(\{ f - \hat{f} \}^2)
\]

\[
\hat{f}(u,v) = \frac{1}{|H(u,v)|^2 + S(u,v)} \cdot \left[ H(u,v) F(u,v) + S(u,v) \cdot \hat{F}(u,v) \right]
\]

- Potential problems:
  - Weights all errors equally regardless of their location in the image, while the eye is considerably more tolerant of errors in dark areas and high-gradient areas in the image.
  - In minimizing the mean square error, Wiener filter also smooth the image more than the eye would prefer
Algebraic approach – Unconstrained restoration vs. Inverse filter

\[ g = Hf + n \]

Seek \( f \) such that \( HF \) approximates \( g \) in a least squares sense.

\[ n = g - Hf \]

Differentiate right hand side with respect to \( f \).

\[ \frac{\partial J(f)}{\partial f} = 0 = -2H^T(g - Hf) \Rightarrow H^THf = H^Tg \]

\[ f = (H^TH)^+H^Tg = H^Tg \]

Compared to the inverse filter:

\[ \hat{F}(u,v) = \frac{G(u,v)}{H(u,v)} \]

Algebraic approach – Constrained restoration vs. Wiener filter

Minimizing \( \|Qf\| \) where \( Q \) is a linear operator on \( f \), subject to the constraint \( f - Hf = H \).

Model this problem using Lagrange optimization method.

We seek \( f \) that minimizes the criterion (or objective) function

\[ \mathcal{J}(f) = \|Qf\| + \alpha(\|f - H\|^2) \]

\( \alpha \) is a constant, called the Lagrange multiplier.

\[ \frac{\partial \mathcal{J}(f)}{\partial f} = 0 = -2Q^TQ(f - H) \]

\[ f = (H^TQH + \alpha Q^TQ)^{-1}H^TQg \]

Compared to:

\[ \hat{F}(u,v) = \frac{1}{H(u,v)} \frac{G(u,v)}{\|H(u,v)\|^2} \]

Regularization theory

* Generally speaking, any regularization method tries to analyze a related well-posed problem whose solution approximates the original ill-posed problem.

* The well-posedness is achieved by implementing one or more of the following basic ideas:
  - restriction of the data;
  - change of the space and/or topologies;
  - modification of the operator itself;
  - the concept of regularization operators; and
  - well-posed stochastic extensions of ill-posed problems.
**Solution formulation**

* For \( g = Hf + \eta \), the regularization method constructs the solution as

\[
\arg\min_{\hat{f}} \left[ \ell(g, \hat{f}) + \beta \varepsilon \right]
\]

* \( \ell(g, \hat{f}) \) describes how the real image data is related to the degraded data. In other words, this term models the characteristic of the imaging system.

* \( \beta \varepsilon \) is the regularization term with the regularization operator \( \varepsilon \) operating on the original image \( f \), and the regularization parameter \( \beta \) used to tune up the weight of the regularization term.

* By adding the regularization term, the original ill-posed problem turns into a well-posed one, that is, the insertion of the regularization operator puts some constraints on what \( f \) might be, which makes the solution more stable.

**MAP (maximum a-posteriori probability)**

* Formulate solution from statistical point of view: MAP approach tries to find an estimate of image \( f \) that maximizes the a-posteriori probability \( p(f | g) \) as

\[
\hat{f} = \arg\max_f p(f | g)
\]

* According to Bayes' rule,

\[
p(f | g) = \frac{p(g | f) p(f)}{p(g)}
\]

  - \( p(f) \) is the a-priori probability of the unknown image \( f \). We call it the prior model
  - \( p(g) \) is the probability of \( g \) which is a constant when \( g \) is given
  - \( p(g | f) \) is the conditional probability density function (pdf) of \( g \). We call it the sensor model, which is a description of the noisy or stochastic processes that relate the original unknown image \( f \) to the measured image \( g \).

**MAP - Derivation**

* Bayes interpretation of regularization theory

\[
\hat{f} = \arg\max_f p(f | g) = \arg\max_f \left[ p(g | f) p(f) \right]
\]

Let \( \Omega = -\ln[p(g | f) p(f)] = -\ln[p(g | f)] - \ln[p(f)] \)

\( \Omega_x = -\ln[p(g | f)] \) Noise term

\( \Omega_y = -\ln[p(f)] \) Prior term
The noise term

* Assume Gaussian noise of zero mean, $\sigma$ the standard deviation

$$\Omega = -\ln[p(g|f)] = -\ln\left[\frac{1}{\sqrt{2\pi\sigma}}\exp\left(-\frac{(Hf-g)^2}{2\sigma^2}\right)\right]$$

$$= \frac{1}{2\sigma^2}\sum_{i}(h \otimes h - g)$$

The prior model

* The a-priori probability of an image by a Gibbs distribution is defined as

$$p(f) \propto \frac{\exp(-U(f)/T)}{Z}$$

- $U(f)$ is the energy function
- $T$ is the temperature of the model
- $Z$ is a normalization constant

$$\Omega = -\ln[p(f)] = -\ln\left[\frac{\exp(-U(f)/T)}{Z}\right] = \frac{U(f)}{T}$$

The prior model (cont’)

* $U(f)$, the prior energy function, is usually formulated based on the smoothness property of the original image. Therefore, $U(f)$ should measure the extent to which the smoothness is violated.
The prior model (cont')

\[ \frac{U(f)}{\mathcal{T}} = \left[ -\beta \frac{1}{\sqrt{2\pi T}} \exp\left( -\frac{\|f - \mu\|^2}{2\sigma^2} \right) \right] \mathcal{T} \]

\[ = \sum_{k=1}^{m} \left[ -\beta \frac{1}{\sqrt{2\pi T}} \exp\left( -\frac{(f \otimes r)^2}{2r^2} \right) \right] \mathcal{T} \]

* \( \beta \) is the parameter that adjusts how smooth the image goes
* The \( k \)-th derivative models the difference between neighbor pixels. It can also be approximated by convolution with the right kernel

The prior model – Kernel \( r \)

* Laplacian kernel

\[ \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \]

\[ r = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix} \]

The objective function

\[ \Omega = \Omega_s + \Omega_r \]

\[ = \frac{1}{2\sigma^2} \sum_{i}^{M} (f \otimes h - g_i)^2 - \beta \frac{1}{\sqrt{2\pi T}} \sum_{m}^{M} \exp\left( -\frac{(f \otimes r)^2}{2r^2} \right) \]

* Use gradient descent to solve \( f \)

\[ f^{k+1} = f^k - \alpha \frac{\partial \Omega}{\partial f} \]

\[ \frac{\partial \Omega}{\partial f} = \frac{1}{\sigma^2} \sum_{i}^{M} [(f \otimes h - g_i) \otimes h_i] + \sum_{m}^{M} \left( \frac{\beta (f \otimes r)}{\sqrt{2\pi T}} \exp\left( -\frac{(f \otimes r)^2}{2r^2} \right) \right) \otimes r_m \]