CV

Review Questions

Ng Huu Duc

Image Formation

L1) - How images is formed? What are the earliest type of camera?

- What is effected when changing the aperture? What problem there is? How to solve it?

- How do you understand digital image? Resolution of an image?

- Define demosaicing with Bayer grid? Why do we need it?

Image Processing:

L2)

- Name 2 classes of filter (2 categories)
  Give some examples in each.

- What types of noise are there?
  What assumption we usually make regarding the noise in image?

- Formulas of correlation, convolution, box filters (averaging)

- Is averaging filter a correlation filter?

- What is the way to efficiently implement filter & explain why?

Filters

- Formula of Gaussian filter? How to set filter width & explain why? What happens if the width is small?

- What are the ways to deal with boundary issues?

- Compare Correlation filter with Convolution filter. (properties)

- How do you understand the relation between applying a filter and the Fourier Transform? What is Duality in here means?
  Fourier transform of important functions?

- Define Median Filter and what is its characteristics?
  What happens if the filter width too great?

L3) - Explain the Fourier Interpretation of Discrete Sampling Effects & results of sampling with / without gaps.

Multi-Scale Representation

- Is Image Pyramid computationally expensive?
- What & when is there ringing artifact, aliasing artifact?
  Why is there so? How to avoid?

  Nyquist theorem about sampling frequency?
- How to perform the Gaussian pyramid, Laplacian Pyramid?
- What is the different between the two? Why smoothen before resampling?
- What does it mean to imply correlation filtering as template matching?

Image Gradients

- Formula of Prewitt, Sobel, Roberts Filters for image gradients
  Why do we use 3 × 3 instead of 2×2 filters?
  3 × 1 ——— 2 × 1
- What should we do if there is noise?
  The better way to implement image gradients?
- The difference when use 1st grad and 2nd grad of Gaussian (Laplacian) to find edges?

L4) Edge Detection:

- Criterias for an "optimal" edge detector?
- What are the primary steps of an edge detector?
- What is the effect of σ on derivatives?
- Why do we need & how to use threshold here?

+ Formula of edge strength & ~~edge~~ gradient direction.

- The main steps of the Canny Edge Detector.
  What is Hysteresis Thresholding?

- Any comment on the problem of finding edges and other problems of computer vision?

Structure Extraction:

- What is Structure extraction? Name the methods you know.
- Why do we move to use Polar Representation for lines.
- The formulas of Hough transform for lines.
- Explain the steps of Hough Transform Algorithm.
- Extensions of Hough Transform?

L5, Segmentation:

- How do you understand the goal of segmentation?
- What do you know about Gestalt theory?
- What are the two approaches to segmentation? Name some methods for each approach.
- Describe the steps of K-Means. What is its problems?
- What is K-Mean ++?
- To what property can we apply K-Means to group pixels?
- Is K-Means NP-Complete or NP-Hard?

- Describe the steps of Expectation-Maximization! Formula
Compared it with K-Means
- What are the 2 categories in the clustering approach?
- What are the definitions in mean-shift algorithm? Parameter?
- What are the problems of mean-shift? Suggest solutions!
- The energy formula & its components! And of Markov Random Fields.
- Name some inference algorithms!

L6) - Characteristics of Graph-cut algorithm?
- s-t - Min cut problem & its solution?
- When can s-t Graph cut be applied? How can it be extended?
- Graph cut application? & any comment?
- Draw & anotate the graph of Graph-Cut.
- Example unary term with Color potentials.
- Example pairwise term with Contrast sensitive Potts model.
- Is Graph Cut guaranteed to find optimal solution? Is α-Expansion Moves?
- Describe application of Graph Cut (Interactive Image Segmentation and Iterated Graph Cut)

Object Detection

L7) - Differentiate all the terms: Classification, Localization, Object Detection Segmentation, Instance Segmentation, Semantic Segmentation
- Explain the idea of Sliding Window based Object Detection.
How exactly do we search over different scale?
- Explain the idea of Detection via Classification

- For feature extraction, why do we choose gradient-based representation, instead of pixel-based ~~gradie~~ representation?

L8, - For HoG, why do we must use a very good representation? (In general, for which case)

- Formula of SVMs
- What are the steps in HoG Descriptor Processing Chain?
  Explain each in details.
  For which do we need gradient direction?
  Why is it "weighted" vote? How many bins there are, typically?
  Why is it normalize over "overlapping" cells? How big is a cell, a block?

- HoG in Testing Phase:
  What should be noticed when doing Non-Maximum Suppression? 3
  with Sliding-Window approach? 2

- Formula of Adaboost. What property has to be fulfilled by the weak classifiers? (for Ensembles)

- What are the limitations of Sliding-Window approach? 2

- Describe Viola-Jones Face Detector?
  How can the "rectangular" filters be efficiently computed?
  Explain the idea of Cascading Classifiers, why do we need it?

Local Feature

L9, - What is the motivation to find, search for Local feature?
  What applications use local feature?

- What is the general procedure of a local feature ~~detector~~?

- What are the requirements for the distinctive key points detector?

- Describe Harris Detector & Hessian Detector.
  Give formulas and explain. Why we have σ^4?
  Write pseudo code.
- Compare Harris & Hessian Detector! What are their pros & cons?
  In practice, what should we use?

40, - What to look for from a region selection?
- Explain & describe Laplacian of Gaussian!
  How can we efficiently compute it?
- What is octave, how should we choose no. of scale per octave, σ? Draw the Scale Pyramid! Draw the scale space!
  How to chose interest points from the scale space of LoG?
- What kind of local descriptors do we want? (characteristics)
- What are the procedure to extract SIFT feature/descriptor?
- For SIFT descriptor, how many sub-patches are in the region of interest, how many bins for the histogram?
  What is the minimum size of the region?
- What to notice when conducting key-point matching?
- What are the 2 type of problems regarding (finding) configurations?
  Differentiate them!
- What kinds of 2D (Affine) Transformation there are?
  Write the formula for each. What is Euclidean Transf., Similarity, Affine, Projective Transformation?
- Describe Homography, formula?
- How to deal with outliers?
  Explain the idea of RANSAC.

Deep Learning:

L11 ↓ 16/

- Why do we use Max Pooling?
- If we use Batch Norm, do we need bias? What should we do then, & why?
- Instead of Max pooling, what can we do with CONV? Why?
- What is Global Average Pooling Layer? What is the benefit?
- Why do we use augmentation? Which result will it increase?
- What is the different between VGG Net and AlexNet?
- ~~How many FCs layer is in GoogleNet?~~
- What is top-1 & top-5 accuracy rate? Which one will be higher?
- While conducting transfer learning, what should we notice regarding different data set size?
- Describe the naive version of an inception modul? What is the problem here? How to solve it? Where to put the 1x1 conv layer?
- How many inception modules are there in the GoogLeNet? How many auxiliary branches are there? Why do we have it?
- What is IoU, how to calculate it?
- What is landmark detection & what are its applications?
- Describe YOLO, R-CNN, Fast R-CNN, Faster R-CNN!
- How is Semantic Segmentation conducted?
- Why do we have downsampling in FCN?
- What are the methods for up-sampling? 4

- Describe the combination of Inception with ResNet
  Skip connection with FCN
- Describe Masked R-CNN.
- Compare ROI Align with ROI pooling
- Describe Atrous Conv, SPP, Atrous SPP? The benefit?
- Describe Siamese Network. What are the 3 types of model for Matching Tasks?
- Formullar of Triplet Loss, Batch all & Batch hard
  The idea behind it? OHM?

3D Reconstruction

- What is Stereo Vision? Epipolar Geometry?
  Extrinsic & Intrinsic Parameters
- Formula to estimate the depth
- What is the Stereo correspondence constraint? Formula?
  What is baseline, epipolar lines, epipoles, epipolar plane?
- What is Stereo Image Rectification?
- Compare Dense & Sparse Correspondence Search.
- What is Essential Matrix & give formula?
- Describe and compare Dense & Sparse Correspondence Search
- What are all parameters of Camera calibration?
  Give formula.
- How many DoF there are for general pinhole camera, CCD camera with square pixels, general camera?

- How to calibrate / Find calibration of a camera?
- How many points we need at least & how many should we have?

- Compared to the ideal case of Stereo Reconstruction, in practice, what are the problems we will face?
- What are the 2 approach to deal with non-ideal triangulation? Describe them.
- What happens with non ideal epipolar geometry? Describe the properties of E, F
- How to estimate F? 3 Methods

- How to conduct 3D Reconstruction with Weak Calibration?

- What is epipolar transfer & in what way can we utilize it?

Formulas

+) **Filter**: - Correlation

$$G[i,j] = \frac{1}{(2k+1)^2} \sum_{u=-k}^{k} \sum_{v=-k}^{k} F[i+u, j+v]$$

$$G[i,j] = \sum_{u=-k}^{k} \sum_{v=-k}^{k} H[u,v]\, F[i+u, j+v]$$

$$G = H \otimes F$$

Convolution

$$G[i,j] = \sum_{u=-k}^{k} \sum_{v=-k}^{k} H[u,v]\, F[i-u, j-v]$$

$$G = H * F$$

- Gaussian: $$\frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) = g_\sigma(x)$$

$$G_\sigma(x,y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)$$

- Prewitt ; Sobel ; Roberts

$$\begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} ; \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} \quad \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} ; \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \quad \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} ; \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

+) **Edge Detection**: $$\|\nabla f\| = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2} ; \quad \theta = \tan^{-1}\left(\frac{\partial f}{\partial y} \Big/ \frac{\partial f}{\partial x}\right)$$

+) **Hough Transform**: $$y = m_0 x_0 + b_0 \Leftrightarrow b = -x_0 m + y_0$$

$x \cdot \cos\theta + y \cdot \sin\theta = d$ (for all θ or θ from gradient)

+) **EM algorithm**:

N datas

K gaussians

$$p(x|\theta) = \sum_{k=1}^{K} \pi_k\, p(x|\theta_k)$$

E-Step: $$\gamma_j(x_n) = \frac{\pi_j \mathcal{N}(x_n | \mu_j, \Sigma_j)}{\sum_{k=1}^{K} \pi_k \mathcal{N}(x_n | \mu_k, \Sigma_k)} \qquad \forall j = 1..K,\ n = 1..N$$

M-Step: $$\hat{N}_j = \sum_{n=1}^{N} \gamma_j(x_n) ; \quad \hat{\pi}_j = \frac{\hat{N}_j}{N}$$

$$\hat{\mu}_j = \frac{1}{\hat{N}_j} \sum_{n=1}^{N} \gamma_j(x_n) . x_n ; \quad \hat{\Sigma}_j = \frac{1}{\hat{N}_j} \sum_{n=1}^{N} \gamma_j(x_n)(x_n - \hat{\mu}_j)(x_n - \hat{\mu}_j)^T$$

+) **Segmentation as Energy Minimization**:

maximize $$p(x,y) = \prod_i \Phi(x_i, y_i) . \prod_{i,j} \Psi(x_i, x_j)$$

$$\phi(x_i, y_i) = -\log \Phi(x_i, y_i) ; \quad \psi(x_i, x_j) = -\log \Psi(x_i, x_j)$$

$$\Rightarrow E(x,y) = \sum_i \phi(x_i, y_i) + \sum_{i,j} \psi(x_i, x_j)$$

$$\begin{cases} \phi_i(s) \propto \exp(-\|I_i - I^s\|^2 / 2\sigma^2) \\ \phi_i(t) \propto \exp(-\|I_i - I^t\|^2 / 2\sigma^2) \end{cases}$$

$$w_{ij} = \exp\left(\frac{-\Delta I_{ij}^2}{2\sigma^2}\right)$$

$$\psi(x_i, x_j) = w_{ij} . \delta(x_i \neq x_j)$$

- Color Potentials: unary potentials

$$\phi(x_i, y_i ; \theta_\phi) = -\log \sum_k \theta_\phi(x_i, k) . p(k|x_i) . \mathcal{N}(y_i ; \mu_k, \Sigma_k)$$

- Edge Potentials - Contrast Sensitive Potts model: pairwise potentials

$$\psi(x_i, x_j, g_{ij}(y); \theta_\psi) = \theta_\psi\, g_{ij}(y) . \delta(x_i \neq x_j)$$

$$g_{ij}(y) = e^{-\beta \| y_i - y_j \|^2} ; \quad \beta = \frac{1}{2}\left[avg(\| y_i - y_j \|^2)\right]^{-1}$$

+) SVM : Minimize $\frac{1}{2} w^T w$, subject to: $t_n(w^T x_n + b) \geq 1$

$$w = \sum_{n=1}^{N} a_n t_n x_n$$

+) Adaboost: $H(x) = sign\left(\sum_{m=1}^{M} \alpha_m . h_m(x)\right)$

(N sample datas
M models, iterations

- Init: $w_n^{(1)} = \frac{1}{N}$ for $n = 1 .. N$

- For m iterations:

- Loss func: $J_m = \sum_{n=1}^{N} w_n^{(m)} . I(h_m(x_n) \neq t_n)$

- Weighted error: $\varepsilon_m = \dfrac{J_m}{\sum_{n=1}^{N} w_n^{(m)}}$

- Weight coefficient: $\alpha_m = \ln\left(\dfrac{1-\varepsilon_m}{\varepsilon_m}\right)$

- Update weights: $w_n^{(m+1)} = w_n^{(m)} . \exp\{\alpha_m . I(h_m(x_n) \neq t_n)\}$

+) Harris Detector: I_x, I_y with σ_1

$\alpha = 0{,}04 \to 0{,}06$

$\Rightarrow I_x^2, I_xI_y, I_y^2$

$\Rightarrow g(I_x^2), g(I_xI_y), g(I_y^2)$ with $\sigma_2 = 2\sigma_1$

$\Rightarrow R = det(M) - \alpha . trace(M)^2$

$= g(I_x^2).g(I_y^2) - [g(I_xI_y)]^2 - \alpha\left[g(I_x^2) + g(I_y^2)\right]^2$

+) Hessian Detector: I_{xx}, I_{xy}, I_{yy}

$$\text{scores} = \det(\text{Hessian}(I)) = I_{xx} I_{yy} - I_{xy}^2$$
$$= \sigma^4 \left[I_{xx}(\sigma) \cdot I_{yy}(\sigma) - I_{xy}^2(\sigma) \right] > t$$

+) 2D Affine Transformation:

$$\begin{bmatrix} x' \\ y' \\ w' \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ w \end{bmatrix} ; \underbrace{\begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix}}_{\text{translation}} ; \underbrace{\begin{bmatrix} c\theta & -s\theta & 0 \\ s\theta & c\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{\text{rotation}} ; \underbrace{\begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{\text{scaling}} ; \underbrace{\begin{bmatrix} 1 & sh_x & 0 \\ sh_y & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{\text{shearing}}$$

+) Hard Triplet Mining:

- Batch All: $d_{j,a,n}^{i,a,p} = D(f_\theta(x_a^i), f_\theta(x_p^i)) - D(f_\theta(x_a^i), f_\theta(x_n^j))$

$$\mathcal{L}_{BA}(\theta; X) = \sum_{i=1}^{P} \sum_{a=1}^{K} \sum_{\substack{p=1 \\ p\neq a}}^{K} \sum_{\substack{j=1 \\ j\neq i}}^{P} \sum_{n=1}^{K} \left[m - d_{j,a,n}^{i,a,p} \right]_+$$

P classes
each class
K images

- Batch Hard:

$$\mathcal{L}_{BH}(\theta; X) = \sum_{i=1}^{P} \sum_{a=1}^{K} \left[m + \max_{p=1..K} D(f_\theta(x_a^i), f_\theta(x_p^i)) - \min_{\substack{j=1..P \\ j\neq a \\ n=1..K}} D(f_\theta(x_a^i) - f_\theta(x_n^j)) \right]_+$$

- Lifted Embedding:

$$\mathcal{L}_L(\theta; X) = \sum_{(a,p)\in X} \left[D_{a,p} + \log \sum_{\substack{n\in X \\ n\neq a \\ n\neq p}} \left(e^{m-D_{a,n}} + e^{m-D_{p,n}} \right) \right]$$

+) Epipolar Geometry:

$x^T E x' = 0$

$Ee' = E^T e = 0$; $K = \begin{bmatrix} m_x & & \\ & m_y & \\ & & 1 \end{bmatrix} \cdot \begin{bmatrix} f & s & p_x \\ & f & p_y \\ & & 1 \end{bmatrix} = \begin{bmatrix} \alpha_x & s & x_0 \\ & \alpha_y & y_0 \\ & & 1 \end{bmatrix} \Rightarrow P = K \cdot [R \mid -R\tilde{C}]$

lines: Ex' & $E^T x$

$E = T_x R$

$x_i \times P X_i = 0$

$[x_{i\times}] P X = 0$

$$\arg\min_X \left[d^2(x_1, P_1 X) + d^2(x_2, P_2 X) \right]$$