DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF COPENHAGEN

Object Detection and Recognition I

Kim Steenstrup Pedersen

Plan for today

- General comments on your solutions to assignment 1
- Object detection and recognition?
- Interest point based approaches:
  - SIFT

Comments to your solutions for assignment 1

- Patch extraction: Remember to handle boundary conditions
- Matching: Compare all patches from image A with all patches in image B using NCC.
  Thresholding or best / 2^{nd} best ratio thresholding?
- What was the best patch size?
- Remember to answer the questions in the assignment, include figures and make sure we can run your code (e.g. don’t use absolute file paths – only relative paths)

Remember:

If you have to resubmit, do so at the latest on Friday via the assignment in Absalon

Plan for today

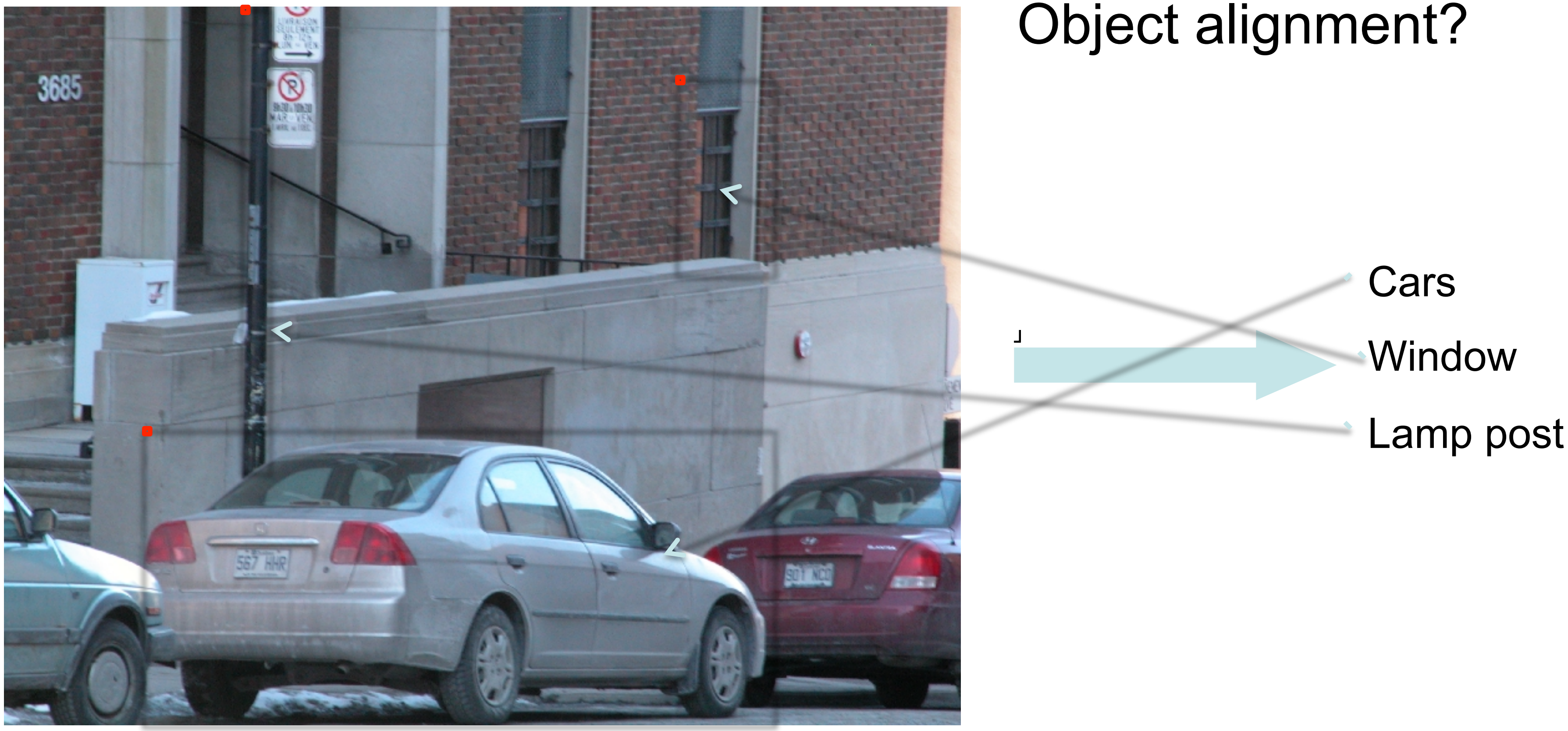
- Object detection and recognition?
- Interest point based approaches:
  - SIFT

Object detection and recognition

What is in this image?

Where?

Object alignment?



NOT Content-Based Image Retrieval (CBIR)

Query image

Search result

Clarification of terms

- In CBIR we compare images with images
- In object recognition we want to infer whether or not a specific object category is present in an image.
- In object detection we want to localize specific instances from object categories in images
- Object detection is a harder problem than object recognition and CBIR

Recognition and Detection Strategies

- Interest point based approaches:
  - SIFT

- Dense feature template approaches:
  - Histogram of Oriented Gradients (HoG)
  - Deformable parts models (DPM)

Interest point based approaches to object detection

Outline of interest point based approaches

Training:

- Detect interest points and extract features on localized object examples
- For each object, store in a database a representation of interest points found on the object

Query image:

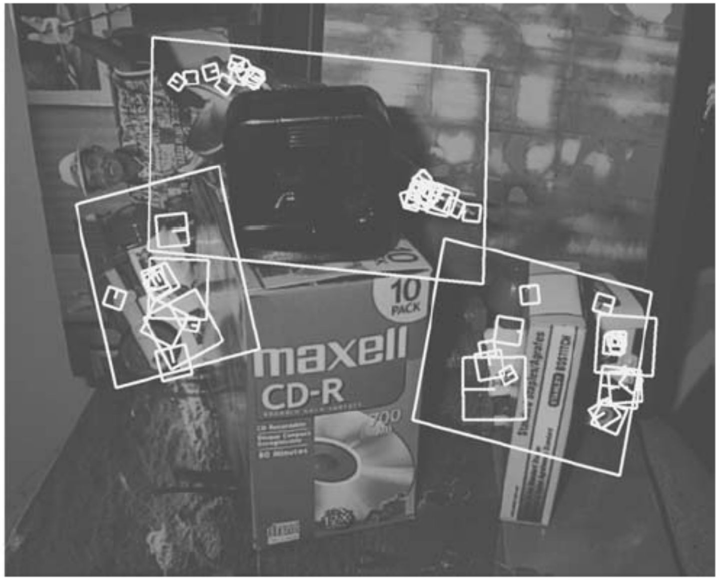
- Detect interest points and extract features
- Look up best matching object in the database by matching the interest points

Outline of interest point based approaches

Query image

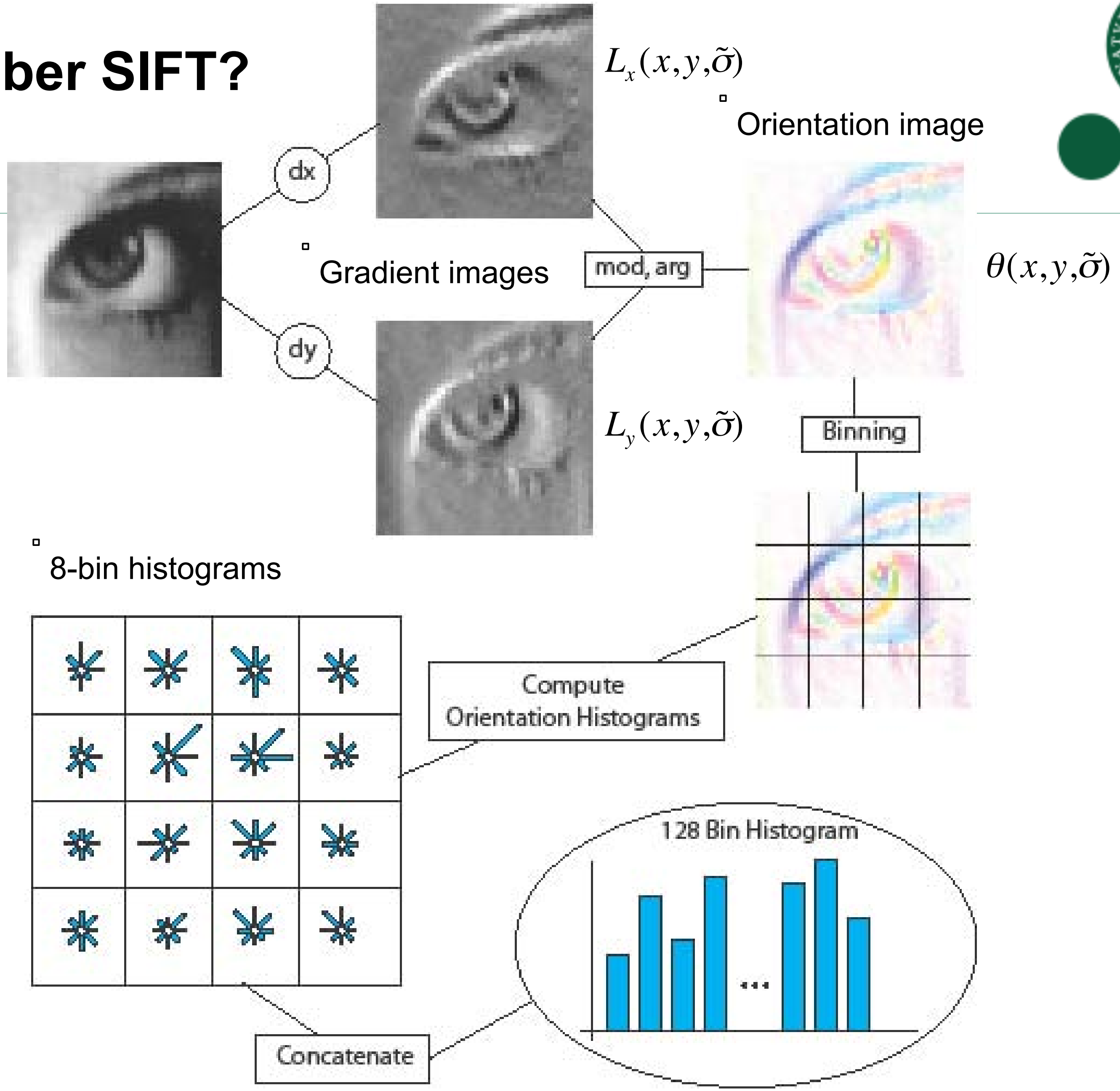
Recognition and detection

Objects



We will look at object detection by Lowe’04:
Can detect specific objects under heavy occlusion and scene clutter

Remember SIFT?



Interest point based approach using SIFT

- Building the object database:
  - For each object, collect one or more training images of the object.
  - Training images: One view of object in the image center without background clutter.
  - For each training image:
    - Extract DoG interest points and SIFT descriptors.
    - Per interest point store the location relative to object, scale, orientation (x, y, σ, θ), SIFT descriptor vector, and object label.

Interest point based approach using SIFT

- Recognition given a query image:
  - Detect DoG interest points and extract SIFT descriptors in query image.
  - For all interest points in query image, match SIFT descriptors with descriptors in the object database. This is performed by a nearest neighbor search.
  - NOTICE: We might match to interest points from different objects in the database.

Recall SIFT matching:

- SIFT features are compared with Euclidean distance (L2-norm):

$$d_2(F_1,F_2) = \sqrt{\sum_{i=1}^{128}(F_1(i) - F_2(i))^2}$$

- Matching SIFT features:
  - A match if this is true

$$\frac{\text{Best}}{\text{2nd Best}} \leq 0.8$$

  - Best refers to the distance for the pair of features with smallest distance.
  - 2nd Best refers to the distance for the pair of features with second smallest distance.

Interest point based approach using SIFT

- Recognition given a query image:
  - Detect DoG interest points and extract SIFT descriptors in query image.
  - For all interest points in query image, match SIFT descriptors with descriptors in the object database. This is performed by a nearest neighbor search.
  - NOTICE: We might match to interest points from different objects in the database.
- For all matched SIFT descriptors:
  - Find object categories by majority vote among descriptors.
  - Cluster matching descriptors using Hough voting to handle object occlusion and determine rough estimate of position, scale and orientation.
  - Improve estimate of position, orientation, scale and shear of object using an affine model.

Voting using Hough transform
(For each object, estimate position, scale and orientation)

- For each potential object category:
  - For all matching interest points have this object category:
    - Apply Hough transform to the 4 parameters (x, y, σ, θ)
  - Find all maxima in Hough space with more than 3 points in the bin. These are potential object candidates (one or more detections).
  - For each object candidate (x, y, σ, θ) , refine the estimate using an affine transformation model.

Hough transform:
A 4 dimensional histogram tilling the parameter space.

Improving the estimate of location, orientation and scale

- Using all interest points that vote for candidate (x,y,σ,θ), improve the estimate by:
  - Solving for affine transformation parameters – model points (x,y) and query points (u,v)

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} m_1 & m_2 \\ m_3 & m_4 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix}$$

  - Using all matching points then improve by least squares solution:

$$\begin{pmatrix} x_1 & y_1 & 0 & 0 & 1 & 0 \\ 0 & 0 & x_1 & y_1 & 0 & 1 \\ & & \cdots & & & \\ & & \cdots & & & \end{pmatrix}\begin{pmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \\ t_x \\ t_y \end{pmatrix} = \begin{pmatrix} u_1 \\ v_1 \\ \vdots \end{pmatrix} \quad \Leftrightarrow \mathbf{Az} = \mathbf{b} \Rightarrow \mathbf{z} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{b}$$

Interest point based approach using SIFT
(Quick overview)

- Post-processing:
  Check if estimated (x, y, σ, θ) and interest points in query and database fit closely, otherwise discard.

- Decision (scoring):
  Use a probability model and only accept object detections when we are really sure according to the model.

SIFT Results

SIFT Results

Problems

This approach:

- Only works well for detecting specific objects or well-defined object categories with low intra-class variability

In general:

- Inter-class versus intra-class variability

Examples of inter-class proximity
(Flowers dataset, Oxford Visual Geometry Group)

English Marigold | Tree Poppy | Hibiscus



Barbeton Daisy | Japanese Anemone | Azalea



Examples of large intra-class variation
(Caltech-UCSD Birds 200 dataset – North American birds)

Black capped Vireo

Black footed Albatross

Black Tern

Problems

This approach:

- Only works well for detecting specific objects or well defined object categories with low intra-class variability

In general:

- Inter-class versus intra-class variability
- Large variation in view-point and illumination (intra-class variation)

Summary

- Object detection and recognition is harder than CBIR
- Interest point based approaches:
  - SIFT

Literature

Reading material:

- Lowe IJCV 2004 (Sec. 7 – 9)