Recap: Sliding-Window Object Detection

- If object may be in a cluttered scene, slide a window around looking for it.
- Essentially, this is a brute-force approach with many local decisions.

Classifier Construction: Many Choices...

- Nearest Neighbor
  - Shakhnarovich, Viola, Darrell 2003
  - Berg, Berg, Malik 2005,
  - Boiman, Shechtman, Irani 2008, ...
- Neural networks
  - LeCun, Bottou, Bengio, Haffner 1998
  - Rowley, Kanade 1998
- Boosting
  - Viola, Jones 2001,
  - Torralba et al. 2004,
  - Opelt et al. 2006,
  - Benenson 2012, ...
- Support Vector Machines
  - Vapnik, Scholkopf 1995,
  - Papageorgiou, Poggio '01,
  - Dalal, Triggs 2005,
  - Vedaldi, Zisserman 2012
- Randomized Forests
  - Amit, Geman 1997,
  - Breiman 2001
  - Lepeit, Fusi 2006,
  - Gall, Lempitsky 2009,...

Recap: AdaBoost

- Want to select the single rectangle feature and threshold that best separates positive (faces) and negative (non-faces) training examples, in terms of weighted error.
- Outputs of a possible rectangle feature on faces and non-faces.

Recap: AdaBoost Feature+Classifier Selection

- Final classifier is combination of the weak classifiers

Resulting weak classifier:

\[ h(x) = \begin{cases} +1 & \text{if } f(x) > \theta_i \\ -1 & \text{otherwise} \end{cases} \]

For next round, reweight the examples according to errors, choose another filter/threshold combo.
Recap: Viola-Jones Face Detector

- Train cascade of classifiers with AdaBoost
- Train with 5K positives, 350M negatives
- Real-time detector using 38 layer cascade
- [Implementation available in OpenCV: http://sourceforge.net/projects/opencvlibrary/]

Topics of This Lecture

- Local Invariant Features
  - Motivation
  - Requirements, Invariances
- Keypoint Localization
  - Harris detector
  - Hessian detector
- Scale Invariant Region Selection
  - Automatic scale selection
  - Laplacian-of-Gaussian detector
  - Difference-of-Gaussian detector
  - Combinations
- Local Descriptors
  - Orientation normalization
  - SIFT

Motivation

- Global representations have major limitations
- Instead, describe and match only local regions
- Increased robustness to
  - Occlusions
  - Articulation
  - Intra-category variations

Application: Image Matching

Slide credit: Steve Seitz

Harder Case

Slide credit: Steve Seitz

Harder Still?

NASA Mars Rover images

Slide credit: Steve Seitz
Application: Image Stitching

• Procedure:
  1. Detect feature points in both images
  2. Find corresponding pairs
  3. Use these pairs to align the images

General Approach

1. Find a set of distinctive keypoints
2. Define a region around each keypoint
3. Extract and normalize the region content
4. Compute a local descriptor from the normalized region
5. Match local descriptors
Common Requirements

• Problem 1:
  - Detect the same point independently in both images

  ![No chance to match!](image)

  **We need a repeatable detector!**

• Problem 2:
  - For each point correctly recognize the corresponding one

  ![?](image)

  **We need a reliable and distinctive descriptor!**

Invariance: Geometric Transformations

Levels of Geometric Invariance

Requirements

• Region extraction needs to be repeatable and accurate
  - Invariant to translation, rotation, scale changes
  - Robust or covariant to out-of-plane (affine) transformations
  - Robust to lighting variations, noise, blur, quantization

• Locality: Features are local, therefore robust to occlusion and clutter.

• Quantity: We need a sufficient number of regions to cover the object.

• Distinctiveness: The regions should contain “interesting” structure.

• Efficiency: Close to real-time performance.

Many Existing Detectors Available

• Hessian & Harris
  - [Beaudet ‘78], [Harris ‘88]

• Laplacian, DoG
  - [Lindeberg ‘98], [Lowe ‘99]

• Harris-/Hessian-Laplace
  - [Mikolajczyk & Schmid ‘01]

• Harris-/Hessian-Affine
  - [Mikolajczyk & Schmid ‘04]

• EBR and IBR
  - [Tuytelaars & Van Gool ‘04]

• MSER
  - [Matas ‘02]

• Salient Regions
  - [Kadir & Brady ‘01]

• Others...

• Those detectors have become a basic building block for many recent applications in Computer Vision.
Keypoint Localization

- Goals:
  - Repeatable detection
  - Precise localization
  - Interesting content

⇒ Look for two-dimensional signal changes

Finding Corners

- Key property:
  - In the region around a corner, image gradient has two or more dominant directions
- Corners are repeatable and distinctive


Corners as Distinctive Interest Points

- Design criteria
  - We should easily recognize the point by looking through a small window (locality)
  - Shifting the window in any direction should give a large change in intensity (good localization)

"flat" region: no change in all directions
"edge": no change along the edge direction
"corner": significant change in all directions

Corners as Distinctive Interest Points

Harris Detector Formulation

- Change of intensity for the shift \([u,v]\):

\[
E(u,v) = \sum_{x,y} w(x,y) \left[ I_x(x+u,y+v) - I_x(x,y) \right]^2 + \left[ I_y(x+u,y+v) - I_y(x,y) \right]^2
\]

- This measure of change can be approximated by:

\[
E(u,v) \approx [u \ v] M \begin{bmatrix} u \\ v \end{bmatrix}
\]

where \(M\) is a 2x2 matrix computed from image derivatives:

\[
M = \sum_{x,y} w(x,y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}
\]

Sum over image region - the area we are checking for corner

\[
M = \sum_{x,y} \begin{bmatrix} I_x I_x & I_x I_y \\ I_x I_y & I_y I_y \end{bmatrix} = \sum_{x,y} \begin{bmatrix} I_x \\ I_y \end{bmatrix} \begin{bmatrix} I_x & I_y \end{bmatrix}
\]
What Does This Matrix Reveal?

- First, let’s consider an axis-aligned corner:

First, let’s consider an axis-aligned corner:

\[
M = \begin{bmatrix}
\sum I_x^2 & \sum I_x I_y \\
\sum I_x I_y & \sum I_y^2
\end{bmatrix} = \begin{bmatrix}
\lambda_1 & 0 \\
0 & \lambda_2
\end{bmatrix}
\]

This means:
- Dominant gradient directions align with \( x \) or \( y \) axis
- If either \( \lambda \) is close to 0, then this is not a corner, so look for locations where both are large.

What if we have a corner that is not aligned with the image axes?

General Case

- Since \( M \) is symmetric, we have \( M = R \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} R \) (Eigenvalue decomposition)
- We can visualize \( M \) as an ellipse with axis lengths determined by the eigenvalues and orientation determined by \( R \)

Interpreting the Eigenvalues

- Classification of image points using eigenvalues of \( M \):

\[
\lambda_1, \lambda_2 \text{ are large, } \lambda_1 \approx \lambda_2; \quad \xi \text{ increases in all directions}
\]

\[
\lambda_1 > \lambda_2 > 0; \quad \xi \text{ is almost constant in all directions}
\]

\[
\lambda_2 > 0; \quad \text{“Edge” region}
\]

\[
\lambda_1 > 0; \quad \text{“Corner”}
\]

\[
\lambda_1 < 0; \quad \text{“Flat” region}
\]

\[
\lambda_2 < 0; \quad \text{“Edge”}
\]

Corner Response Function

\[
R = \det(M) - \alpha \text{ trace}(M)^2 = \lambda_1 \lambda_2 - \alpha(\lambda_1 + \lambda_2)^2
\]

- Fast approximation
  - Avoid computing the eigenvalues
  - \( \alpha \): constant (0.04 to 0.06)

Window Function \( w(x, y) \)

\[
M = \sum_{x,y} w(x, y) \begin{bmatrix} I_x^2 & I_x I_y \\
I_x I_y & I_y^2 \end{bmatrix}
\]

- Option 1: uniform window
  - Sum over square window

\[
M = \sum_{x,y} \begin{bmatrix} I_x^2 & I_x I_y \\
I_x I_y & I_y^2 \end{bmatrix}
\]

- Problem: not rotation invariant

- Option 2: Smooth with Gaussian
  - Gaussian already performs weighted sum

\[
M = g(\sigma) \begin{bmatrix} I_x^2 & I_x I_y \\
I_x I_y & I_y^2 \end{bmatrix}
\]

- Result is rotation invariant
Summary: Harris Detector [Harris88]

- Compute second moment matrix (autocorrelation matrix)
  \[
  M(\sigma_x, \sigma_y) = \sigma_x \cdot \sigma_y \cdot \begin{bmatrix}
    I_x^2(\sigma_x) & I_x I_y(\sigma_x) \\
    I_x I_y(\sigma_x) & I_y^2(\sigma_y)
  \end{bmatrix}
  \]

1. Image derivatives
2. Square of derivatives
3. Gaussian filter \(g(\alpha)\)

4. Cornerness function - two strong eigenvalues

\[
R = \det(M(\sigma_x, \sigma_y)) - \alpha \text{trace}(M(\sigma_x, \sigma_y))^2
\]

\[
= g(I_x^2)g(I_y^2) - [g(I_x I_y)]^2 - \alpha [g(I_x^2) + g(I_y^2)]^2
\]

5. Perform non-maximum suppression

Harris Detector: Workflow

- Compute corner responses \(R\)
- Take only the local maxima of \(R\), where \(R > \) threshold.

Effect: A very precise corner detector.
Harris Detector - Responses [Harris88]

Results are well suited for finding stereo correspondences

Harris Detector: Properties

- Rotation invariance?

Ellipse rotates but its shape (i.e., eigenvalues) remains the same

Corner response \( R \) is invariant to image rotation

Hessian Detector [Beaudet78]

- Hessian determinant

Note: these are 2nd derivatives!

Intuition: Search for strong derivatives in two orthogonal directions

Hessian determinant

\[
\text{Hessian}(I) = \begin{bmatrix}
I_{xx} & I_{xy} \\
I_{xy} & I_{yy}
\end{bmatrix}
\]

\[
\text{det}(\text{Hessian}(I)) = I_{xx}I_{yy} - I_{xy}^2
\]

In Matlab:

\[
I_{xx} \cdot I_{yy} - (I_{xy})^2
\]
Topics of This Lecture

- **Local Invariant Features**
  - Motivation
  - Requirements, Invariances
- **Keypoint Localization**
  - Harris detector
  - Hessian detector
- **Scale Invariant Region Selection**
  - Automatic scale selection
  - Laplacian-of-Gaussian detector
  - Difference-of-Gaussian detector
  - Combinations
- **Local Descriptors**
  - Orientation normalization
  - SIFT

From Points to Regions...

- The Harris and Hessian operators define interest points.
  - Precise localization
  - High repeatability
- In order to compare those points, we need to compute a descriptor over a region.
- How can we define such a region in a scale invariant manner?
- *I.e. how can we detect scale invariant interest regions?*

Naïve Approach: Exhaustive Search

- Multi-scale procedure
  - Compare descriptors while varying the patch size
Naïve Approach: Exhaustive Search

- Multi-scale procedure
  - Compare descriptors while varying the patch size

Automatic Scale Selection

- Solution:
  - Design a function on the region, which is “scale invariant” (the same for corresponding regions, even if they are at different scales)
  - For a point in one image, we can consider it as a function of region size (patch width)

**Example:** average intensity. For corresponding regions (even of different sizes) it will be the same.

Important: this scale invariant region size is found in each image independently!
Automatic Scale Selection

• Function responses for increasing scale (scale signature)

Slide credit: Krystian Mikolajczyk

Automatic Scale Selection

• Function responses for increasing scale (scale signature)

Slide credit: Krystian Mikolajczyk

Automatic Scale Selection

• Function responses for increasing scale (scale signature)

Slide credit: Krystian Mikolajczyk

Automatic Scale Selection

• Function responses for increasing scale (scale signature)

Slide credit: Krystian Mikolajczyk

Automatic Scale Selection

• Function responses for increasing scale (scale signature)

Slide credit: Tinne Tuytelaars

Automatic Scale Selection

• Normalize: Rescale to fixed size

Slide credit: Tinne Tuytelaars
What Is A Useful Signature Function?

- Laplacian-of-Gaussian = “blob” detector

Characteristic Scale

- We define the characteristic scale as the scale that produces peak of Laplacian response

Laplacian-of-Gaussian (LoG)

- Interest points:
  - Local maxima in scale space of Laplacian-of-Gaussian

Laplacian-of-Gaussian (LoG)

- Interest points:
  - Local maxima in scale space of Laplacian-of-Gaussian

=> List of (x, y, σ)
Technical Detail

- We can efficiently approximate the Laplacian with a difference of Gaussians:
  \[ L = \sigma^2 \left( G_n^x(x, y, \sigma) + G_n^y(x, y, \sigma) \right) \]  
  (Laplacian)
  \[ \text{DoG} = G(x, y, \kappa \sigma) - G(x, y, \sigma) \]  
  (Difference of Gaussians)

Key point localization with DoG

- Detect maxima of difference-of-Gaussian (DoG) in scale space
- Then reject points with low contrast (threshold)
- Eliminate edge responses

Difference-of-Gaussian (DoG)

- Difference of Gaussians as approximation of the LoG
  - This is used e.g. in Lowe’s SIFT pipeline for feature detection.
- Advantages
  - No need to compute 2nd derivatives
  - Gaussians are computed anyway, e.g. in a Gaussian pyramid.
**DoG - Efficient Computation**

- Computation in Gaussian scale pyramid

- Sampling with step $\sigma^2$

**Results: Lowe’s DoG**

**Example of Keypoint Detection**

- (a) 233x189 image
- (b) 832 DoG extrema
- (c) 729 left after peak value threshold
- (d) 536 left after testing ratio of principle curvatures (removing edge responses)

**Harris-Laplace [Mikolajczyk '01]**

1. Initialization: Multiscale Harris corner detection

2. Scale selection based on Laplacian
   (same procedure with Hessian $\Rightarrow$ Hessian-Laplace)

**Summary: Scale Invariant Detection**

- **Given:** Two images of the same scene with a large scale difference between them.
- **Goal:** Find the same interest points independently in each image.
- **Solution:** Search for maxima of suitable functions in scale and in space (over the image).

- Two strategies
  - Laplacian-of-Gaussian (LoG)
  - Difference-of-Gaussian (DoG) as a fast approximation
  - These can be used either on their own, or in combinations with single-scale keypoint detectors (Harris, Hessian).
For most local feature detectors, executables are available online:

- [http://robots.ox.ac.uk/~vgg/research/affine](http://robots.ox.ac.uk/~vgg/research/affine)
- [http://www.vision.ee.ethz.ch/~surf](http://www.vision.ee.ethz.ch/~surf)

References and Further Reading

- Read David Lowe’s SIFT paper
  
  - D. Lowe, *Distinctive image features from scale-invariant keypoints*, IJCV 60(2), pp. 91-110, 2004

- Good survey paper on Int. Pt. detectors and descriptors
  

- Try the example code, binaries, and Matlab wrappers
  
  - Good starting point: Oxford interest point page  
    [http://www.robots.ox.ac.uk/~vgg/research/affine/detectors.html#binaries](http://www.robots.ox.ac.uk/~vgg/research/affine/detectors.html#binaries)