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.

Recap: HOG Descriptor Processing Chain

Object/Non-object

- Linear SVM
- Collect HOGs over detection window
- Contrast normalize over overlapping spatial cells
- Weighted vote in spatial & orientation cells
- Compute gradients
- Gamma compression
- Image Window

Recap: Pedestrian Detection with HOG

- Train a pedestrian template using a linear SVM
- At test time, convolve feature map with template
  \[
  y(x) = w^T x + b
  \]
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.

Resulting weak classifier:

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

For next round, reweight the examples according to errors, choose another filter/threshold combo.

Recap: Viola-Jones Face Detector

- Train with 5K positives, 350M negatives
- Real-time detector using 38 layer cascade
- 6061 features in final layer
- Implementation available in OpenCV: [http://sourceforge.net/projects/opencvlibrary/](http://sourceforge.net/projects/opencvlibrary/)

Limitations: Low Training Resolutions

- Many (older) S/W detectors operate on tiny images
  - Viola&Jones: 24x24 pixels
  - Torralba et al.: 32x32 pixels
  - Dalal&Triggs: 64x96 pixels (notable exception)
- Main reasons
  - Training efficiency (exhaustive feature selection in AdaBoost)
  - Evaluation speed
  - Want to recognize objects at small scales
- But...
  - Limited information content available at those resolutions
  - Not enough support to compensate for occlusions!

Limitations: Changing Aspect Ratios

- Sliding window requires fixed window size
  - Basis for learning efficient cascade classifier
- How to deal with changing aspect ratios?
  - Fixed window size
    - Wastes training dimensions
  - Adapted window size
    - Difficult to share features
  - “Squashed” views [Dalal&Triggs]
    - Need to squash test image, too
Limitations (continued)

• Not all objects are “box” shaped

• Non-rigid, deformable objects not captured well with representations assuming a fixed 2D structure; or must assume fixed viewpoint

• Objects with less-regular textures not captured well with holistic appearance-based descriptions

Limitations (continued)

• If considering windows in isolation, context is lost

• In practice, often entails large, cropped training set (expensive)

• Requiring good match to a global appearance description can lead to sensitivity to partial occlusions

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

by Diva Sian

by swashford

Harder Case

by Diva Sian

by scgbt

Harder Still?

NASA Mars Rover Images

Answer Below (Look for tiny colored squares)

NASA Mars Rover Images

with SIFT feature matches
(Figure by Noah Snavely)

Application: Image Stitching

• Procedure:
  - Detect feature points in both images

Application: Image Stitching

by Darya Frolova, Denis Simakov
**Application: Image Stitching**

- **Procedure:**
  1. Detect feature points in both images
  2. Find corresponding pairs

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

**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.

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)
Harris Detector Formulation

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

\[ E(u, v) = \sum w(x, y) \left( I(x+u, y+v) - I(x, y) \right)^2 \]

Window function

Shifted intensity

Intensity

Window function \(w(x, y)\) = 1 in window, 0 outside or Gaussian

Harris Detector Formulation

- This measure of change can be approximated by:

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

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

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

What Does This Matrix Reveal?

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

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

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 \(M = R^T [\lambda_1 \, 0] \begin{bmatrix} 0 & \lambda_2 \end{bmatrix} R\)

(Eigenvalue decomposition)

- We can visualize \(M\) as an ellipse with axes lengths determined by the eigenvalues and orientation determined by \(R\)

Direction of the fastest change

Direction of the slowest change

\([\lambda_{\text{max}}]^{-1/2} \quad [\lambda_{\text{min}}]^{-1/2}\)
### Interpreting the Eigenvalues

- Classification of image points using eigenvalues of $M$:
  
  $\lambda_1$ and $\lambda_2$ are large; $\lambda_1 \approx \lambda_2$; $E$ increases in all directions
  
  $\lambda_1 > > \lambda_2$; $E$ increases in all directions
  
  $\lambda_1 > > \lambda_2$; $E$ increases in all directions
  
  $\lambda_1 > > \lambda_2$; $E$ increases in all directions

### Corner Response Function

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

### Window Function $w(x,y)$

$M = \sum w(x,y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}$

- Option 1: uniform window
  
  Sum over square window
  
  $M = \sum \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}$
  
  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)
  
  $M(\sigma_x, \sigma_y) = g(\sigma_x) \begin{bmatrix} I_x(\sigma_x) & I_x I_y(\sigma_x) \\ I_x I_y(\sigma_x) & I_y(\sigma_x) \end{bmatrix}$
  
  1. Image derivatives
  
  2. Square of derivatives
  
  3. Gaussian filter $g(\sigma)$

- 4. Cornerness function - two strong eigenvalues
  
  $R = \det(M(\sigma_x, \sigma_y)) - \alpha \text{trace}(M(\sigma_x, \sigma_y))^2$
  
  $= g(I_x^2) g(I_y^2) - [g(I_x I_y)]^2 - \alpha [g(I_x^2)]^2 + [g(I_y^2)]^2$

- 5. Perform non-maximum suppression

### Harris Detector: Workflow

- Compute corner responses $R$
Harris Detector: Workflow

- Take only the local maxima of $R$, where $R > \text{threshold}$.

Harris Detector: Workflow

- Resulting Harris points

Harris Detector - Responses [Harris88]

Effect: A very precise corner detector.

Harris Detector - Responses [Harris88]

Results are well suited for finding stereo correspondences

Harris Detector: Properties

- Rotation invariance?

Corner response $R$ is invariant to image rotation

Ellipse rotates but its shape (i.e. eigenvalues) remains the same
**Harris Detector: Properties**
- Rotation invariance
- Scale invariance?

All points will be classified as edges!

**Not invariant to image scale!**

**Hessian Detector** [Beaudet78]
- Hessian determinant

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

Effect: Responses mainly on corners and strongly textured areas.

**Hessian Detector - Responses** [Beaudet78]

**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

Naïve Approach: Exhaustive Search

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

Naïve Approach: Exhaustive Search

- Comparing descriptors while varying the patch size
  - Computationally inefficient
  - Inefficient but possible for matching
  - Prohibitive for retrieval in large databases
  - Prohibitive for recognition
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)
  - Example: average intensity. For corresponding regions (even of different sizes) it will be the same.
  - 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.

> \[ f \text{ Region size}_1 \]

> \[ f \text{ Region size}_2 \]

> \[ \text{scale} = \frac{1}{2} \]

- **Important:** this scale invariant region size is found in each image independently!

- **Function responses for increasing scale (scale signature)**

> \[ f \text{ Region size}_1 \]

> \[ f \text{ Region size}_2 \]

> \[ \text{scale} = \frac{1}{2} \]

> \[ \text{scale} = \frac{1}{3} \]

> \[ \text{scale} = \frac{1}{4} \]

- **Common approach:**
  - Take a local maximum of this function.
  - Observation: region size for which the maximum is achieved should be invariant to image scale.
Automatic Scale Selection

- Function responses for increasing scale (scale signature)

Slide credit: Krystian Mikolajczyk

Automatic Scale Selection

- Normalize: Rescale to fixed size

Slide credit: Tinne Tuytelaars

What Is A Useful Signature Function?

- Laplacian-of-Gaussian = "blob" detector

Slide adapted from Krystian Mikolajczyk

Characteristic Scale

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

Slide adapted from Tinne Tuytelaars

Laplacian-of-Gaussian (LoG)

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

Slide adapted from Krystian Mikolajczyk
Laplacian-of-Gaussian (LoG)

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

\[ L_\sigma(\sigma) + L_{\sigma/2}(\sigma) \]

\[ L_\sigma(\sigma) + L_{\sigma/2}(\sigma) \]

\[ L_\sigma(\sigma) + L_{\sigma/2}(\sigma) \]

\[ L_\sigma(\sigma) + L_{\sigma/2}(\sigma) \]

\[ \Rightarrow \text{List of } (x, y, \sigma) \]

LoG Detector: Workflow
Technical Detail

- We can efficiently approximate the Laplacian with a difference of Gaussians:

\[ L = \sigma^2 \left( G_x(x, y, \sigma) + G_y(x, y, \sigma) \right) \] (Laplacian)

\[ \text{DoG} = G(x, y, k\sigma) - G(x, y, \sigma) \] (Difference of Gaussians)

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.

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

Candidate keypoints:
- list of (x,y,σ)

DoG - Efficient Computation

- Computation in Gaussian scale pyramid

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

#### Computing Harris function

![Slide adapted from Krystian Mikolajczyk](image)

Detection of local maxima

![Slide adapted from Krystian Mikolajczyk](image)

### 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).

### You Can Try It At Home...

- 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)