Matching features

Computational Photography, 6.882

Prof. Bill Freeman
April 11, 2006

Image and shape descriptors: Harris corner detectors and SIFT features.

Suggested readings: Mikolajczyk and Schmid, David Lowe IJCV.
Matching with Invariant Features

Darya Frolova, Denis Simakov
The Weizmann Institute of Science
March 2004

http://www.wisdom.weizmann.ac.il/~deniss/vision_spring04/files/InvariantFeatures.ppt
Building a Panorama

How do we build panorama?

- We need to match (align) images
Matching with Features

• Detect feature points in both images
Matching with Features

• Detect feature points in both images

• Find corresponding pairs
Matching with Features

• Detect feature points in both images
• Find corresponding pairs
• Use these pairs to align images
Matching with Features

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

no chance to match!

*We need a repeatable detector*
Matching with Features

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

We need a reliable and distinctive descriptor
More motivation...

- Feature points are used also for:
  - Image alignment (homography, fundamental matrix)
  - 3D reconstruction
  - Motion tracking
  - Object recognition
  - Indexing and database retrieval
  - Robot navigation
  - … other
Selecting Good Features

• What’s a “good feature”?
  – Satisfies brightness constancy
  – Has sufficient texture variation
  – Does not have too much texture variation
  – Corresponds to a “real” surface patch
  – Does not deform too much over time
Contents

• Harris Corner Detector
  – Description
  – Analysis
• Detectors
  – Rotation invariant
  – Scale invariant
  – Affine invariant
• Descriptors
  – Rotation invariant
  – Scale invariant
  – Affine invariant
An introductory example:

Harris corner detector

The Basic Idea

• We should easily recognize the point by looking through a small window
• Shifting a window in *any direction* should give *a large change* in intensity
Harris Detector: Basic Idea

“flat” region: no change in all directions

“edge”: no change along the edge direction

“corner”: significant change in all directions
Contents

• Harris Corner Detector
  – Description
  – Analysis
• Detectors
  – Rotation invariant
  – Scale invariant
  – Affine invariant
• Descriptors
  – Rotation invariant
  – Scale invariant
  – Affine invariant
Harris Detector: Mathematics

Window-averaged change of intensity for the shift \([u,v]\):

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

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

or

Gaussian
Go through 2\textsuperscript{nd} order Taylor series expansion on board
Expanding $E(u,v)$ in a 2$^{nd}$ order Taylor series expansion, we have, for small shifts $[u, v]$, a bilinear approximation:

$$E(u, v) \cong [u, v] \begin{bmatrix} u \\ v \end{bmatrix}$$

where $M$ is a 2×2 matrix computed from image derivatives:

$$M = \sum_{x,y} w(x, y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}$$
Harris Detector: Mathematics

Intensity change in shifting window: eigenvalue analysis

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

\[ \lambda_1, \lambda_2 \text{ – eigenvalues of } M \]

Ellipse \( E(u, v) = \text{const} \)
Selecting Good Features

$\lambda_1$ and $\lambda_2$ are large
Selecting Good Features

large $\lambda_1$, small $\lambda_2$
Selecting Good Features

small $\lambda_1$, small $\lambda_2$
Classification of image points using eigenvalues of $M$:

- $\lambda_1$ and $\lambda_2$ are small; $E$ is almost constant in all directions
- $\lambda_1 > \lambda_2$; “Corner”
  - $\lambda_1$ and $\lambda_2$ are large,
  - $\lambda_1 \sim \lambda_2$;
  - $E$ increases in all directions
- $\lambda_2 > \lambda_1$; “Edge”
- $\lambda_1$ and $\lambda_2$ are large,
Harris Detector: Mathematics

Measure of corner response:

$$R = \det M - k (\text{trace } M)^2$$

$$\det M = \lambda_1 \lambda_2$$
$$\text{trace } M = \lambda_1 + \lambda_2$$

($k$ – empirical constant, $k = 0.04$-$0.06$)
Harris Detector: Mathematics

- \( R \) depends only on eigenvalues of \( M \)
- \( R \) is large for a corner
- \( R \) is negative with large magnitude for an edge
- \( |R| \) is small for a flat region
Harris Detector

• The Algorithm:
  – Find points with large corner response function $R$ ($R >$ threshold)
  – Take the points of local maxima of $R$
Harris Detector: Workflow
Harris Detector: Workflow

Compute corner response $R$
Harris Detector: Workflow

Find points with large corner response: $R > \text{threshold}$
Harris Detector: Workflow

Take only the points of local maxima of $R$
Harris Detector: Workflow
Harris Detector: Summary

- Average intensity change in direction $[u,v]$ can be expressed as a bilinear form:

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

- Describe a point in terms of eigenvalues of $M$: measure of corner response

$$R = \lambda_1 \lambda_2 - k(\lambda_1 + \lambda_2)^2$$

- A good (corner) point should have a large intensity change in all directions, i.e. $R$ should be large positive
Contents

• Harris Corner Detector
  – Description
  – Analysis
• Detectors
  – Rotation invariant
  – Scale invariant
  – Affine invariant
• Descriptors
  – Rotation invariant
  – Scale invariant
  – Affine invariant
Harris Detector: Some Properties

• Rotation invariance?
Harris Detector: Some Properties

• Rotation invariance

Corner response $R$ is invariant to image rotation
Harris Detector: Some Properties

• Invariance to image intensity change?
Harris Detector: Some Properties

• Partial invariance to additive and multiplicative intensity changes
  ✓ Only derivatives are used => invariance to intensity shift $I \rightarrow I + b$
  ✓ Intensity scale: $I \rightarrow a I$
Harris Detector: Some Properties

• Invariant to image scale?
Harris Detector: Some Properties

- Not invariant to \textit{image scale}!
Harris Detector: Some Properties

- Quality of Harris detector for different scale changes

Repeatability rate:
\[
\frac{\text{# correspondences}}{\text{# possible correspondences}}
\]

Evaluation plots are from this paper

Evaluation of Interest Point Detectors

CORDELIA SCHMID, ROGER MOHR AND CHRISTIAN BAUCKHAGE

INRIA Rhône-Alpes, 655 av. de l’Europe, 38330 Montbonnot, France

Cordelia.Schmid@inrialpes.fr

Abstract. Many different low-level feature detectors exist and it is widely agreed that the evaluation of detectors is important. In this paper we introduce two evaluation criteria for interest points: repeatability rate and information content. Repeatability rate evaluates the geometric stability under different transformations. Information content measures the distinctiveness of features. Different interest point detectors are compared using these two criteria. We determine which detector gives the best results and show that it satisfies the criteria well.
Contents

• Harris Corner Detector
  – Description
  – Analysis
  • Detectors
    – Rotation invariant
    – Scale invariant
    – Affine invariant
• Descriptors
  – Rotation invariant
  – Scale invariant
  – Affine invariant
We want to:

detect *the same* interest points regardless of *image changes*
Models of Image Change

• Geometry
  – Rotation
  – Similarity (rotation + uniform scale)
  – Affine (scale dependent on direction)
    valid for: orthographic camera, locally planar object

• Photometry
  – Affine intensity change \((I \rightarrow a I + b)\)
Contents

• Harris Corner Detector
  – Description
  – Analysis
• Detectors
  – Rotation invariant
  – Scale invariant
  – Affine invariant
• Descriptors
  – Rotation invariant
  – Scale invariant
  – Affine invariant
Rotation Invariant Detection

- Harris Corner Detector

Contents

• Harris Corner Detector
  – Description
  – Analysis

• Detectors
  – Rotation invariant
    – Scale invariant
    – Affine invariant

• Descriptors
  – Rotation invariant
  – Scale invariant
  – Affine invariant
Scale Invariant Detection

• Consider regions (e.g. circles) of different sizes around a point
• Regions of corresponding sizes will look the same in both images
Scale Invariant Detection

• The problem: how do we choose corresponding circles *independently* in each image?
Scale Invariant Detection

• Solution:
  – Design a function on the region (circle), 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 (circle radius)
Scale Invariant Detection

• Common approach:

Take a local maximum of this function

Observation: region size, for which the maximum is achieved, should be *invariant* to image scale.

**Important:** this scale invariant region size is found in each image *independently!*
Scale Invariant Detection

• A “good” function for scale detection: has one stable sharp peak

• For usual images: a good function would be a one which responds to contrast (sharp local intensity change)
Scale Invariant Detection

- Functions for determining scale

Kernels:

\[ L = \sigma^2 \left( G_{xx}(x, y, \sigma) + G_{yy}(x, y, \sigma) \right) \]

(Laplacian)

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

(Difference of Gaussians)

where Gaussian

\[ G(x, y, \sigma) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{x^2+y^2}{2\sigma^2}} \]

Note: both kernels are invariant to scale and rotation
Scale Invariant Detection

• Compare to human vision: eye’s response
**Scale Invariant Detectors**

- **Harris-Laplacian**
  
  *Find local maximum of:*
  
  - Harris corner detector in space (image coordinates)
  - Laplacian in scale

\[ \leftarrow \text{Harris} \rightarrow \]

- **SIFT (Lowe)**
  
  *Find local maximum of:*
  
  - Difference of Gaussians in space and scale

\[ \leftarrow \text{DoG} \rightarrow \]

---


Scale Invariant Detectors

- Experimental evaluation of detectors w.r.t. scale change

Repeatability rate:

\[
\frac{\text{# correspondences}}{\text{# possible correspondences}}
\]
Scale Invariant Detection: Summary

• **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)

Methods:

1. **Harris-Laplacian** [Mikolajczyk, Schmid]: maximize Laplacian over scale, Harris’ measure of corner response over the image
2. **SIFT** [Lowe]: maximize Difference of Gaussians over scale and space
Contents

• Harris Corner Detector
  – Description
  – Analysis
• Detectors
  – Rotation invariant
  – Scale invariant
  – Affine invariant
• Descriptors
  – Rotation invariant
  – Scale invariant
  – Affine invariant
Affine Invariant Detection

• Above we considered:
  Similarity transform (rotation + uniform scale)

• Now we go on to:
  Affine transform (rotation + non-uniform scale)
Affine Invariant Detection

- Take a local intensity extremum as initial point
- Go along every ray starting from this point and stop when extremum of function $f$ is reached

\[ f(t) = \frac{|I(t) - I_0|}{\int_0^t |I(t) - I_0| \, dt} \]

- We will obtain approximately corresponding regions

Remark: we search for scale in every direction

Affine Invariant Detection

• The regions found may not exactly correspond, so we approximate them with ellipses

• Geometric Moments:

\[ m_{pq} = \int_{\mathbb{R}^2} x^p y^q f(x, y) \, dx \, dy \]

Fact: moments \( m_{pq} \) uniquely determine the function \( f \)

Taking \( f \) to be the characteristic function of a region (1 inside, 0 outside), moments of orders up to 2 allow to approximate the region by an ellipse

This ellipse will have the same moments of orders up to 2 as the original region
**Affine Invariant Detection**

- **Covariance matrix** of region points defines an ellipse:

\[
p^T \Sigma_1^{-1} p = 1
\]

\[
\Sigma_1 = \langle pp^T \rangle_{\text{region 1}}
\]

\[
( p = [x, y]^T \text{ is relative to the center of mass})
\]

\[
\Sigma_2 = A \Sigma_1 A^T
\]

\[
q^T \Sigma_2^{-1} q = 1
\]

\[
\Sigma_2 = \langle qq^T \rangle_{\text{region 2}}
\]

Ellipses, computed for corresponding regions, also correspond!
Affine Invariant Detection

- Algorithm summary (detection of affine invariant region):
  - Start from a local intensity extremum point
  - Go in every direction until the point of extremum of some function $f$
  - Curve connecting the points is the region boundary
  - Compute geometric moments of orders up to 2 for this region
  - Replace the region with ellipse

---

Affine Invariant Detection

• Maximally Stable Extremal Regions
  – Threshold image intensities: \( I > I_0 \)
  – Extract connected components ("Extremal Regions")
  – Find a threshold when an extremal region is "Maximally Stable", i.e. local minimum of the relative growth of its square
  – Approximate a region with an ellipse

Affine Invariant Detection: Summary

- Under affine transformation, we do not know in advance shapes of the corresponding regions.
- Ellipse given by geometric covariance matrix of a region robustly approximates this region.
- For corresponding regions ellipses also correspond.

Methods:
1. Search for extremum along rays [Tuytelaars, Van Gool]:
2. Maximally Stable Extremal Regions [Matas et.al.]
Contents

• Harris Corner Detector
  – Description
  – Analysis

• Detectors
  – Rotation invariant
  – Scale invariant
  – Affine invariant

• Descriptors
  – Rotation invariant
  – Scale invariant
  – Affine invariant
Point Descriptors

• We know how to detect points
• Next question:

How to match them?

Point descriptor should be:
1. Invariant
2. Distinctive
Contents

• Harris Corner Detector
  – Description
  – Analysis

• Detectors
  – Rotation invariant
  – Scale invariant
  – Affine invariant

• Descriptors
  – Rotation invariant
  – Scale invariant
  – Affine invariant
Descriptors Invariant to Rotation

- **Harris corner response measure:**
  - depends only on the eigenvalues of the matrix $M$

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

Descriptors Invariant to Rotation

- Image moments in polar coordinates

\[
m_{kl} = \int \int r^k e^{-i\theta l} I(r, \theta) dr d\theta
\]

Rotation in polar coordinates is translation of the angle:
\[
\theta \rightarrow \theta + \theta_0
\]
This transformation changes only the phase of the moments, but not its magnitude

Rotation invariant descriptor consists of magnitudes of moments:
\[
|m_{kl}|
\]

Matching is done by comparing vectors \([|m_{kl}|]_{k,l}\)

Descriptors Invariant to Rotation

• Find local orientation
  Dominant direction of gradient

• Compute image derivatives relative to this orientation

2 D.Lowe. “Distinctive Image Features from Scale-Invariant Keypoints”. Accepted to IJCV 2004
Contents

• Harris Corner Detector
  – Description
  – Analysis
• Detectors
  – Rotation invariant
  – Scale invariant
  – Affine invariant
• Descriptors
  – Rotation invariant
  – Scale invariant
  – Affine invariant
Descriptors Invariant to Scale

• Use the scale determined by detector to compute descriptor in a normalized frame

For example:
  • moments integrated over an adapted window
  • derivatives adapted to scale: $sI_x$
Contents

• Harris Corner Detector
  – Description
  – Analysis
• Detectors
  – Rotation invariant
  – Scale invariant
  – Affine invariant
• Descriptors
  – Rotation invariant
  – Scale invariant
  – Affine invariant
Affine Invariant Descriptors

- Affine invariant color moments

\[ m_{pq}^{abc} = \int_{\text{region}} x^p y^q R^a(x, y) G^b(x, y) B^c(x, y) dx dy \]

Different combinations of these moments are fully affine invariant

Also invariant to affine transformation of intensity \( I \rightarrow a I + b \)

F. Mindru et al. “Recognizing Color Patterns Irrespective of Viewpoint and Illumination”. CVPR99
Affine Invariant Descriptors

• Find affine normalized frame

\[ \Sigma_1 = \langle pp^T \rangle \]

\[ A \]

\[ \Sigma_1^{-1} = A_1^T A_1 \]

\[ A_1 \]

\[ \Sigma_2 = \langle qq^T \rangle \]

\[ A_2 \]

\[ \Sigma_2^{-1} = A_2^T A_2 \]

rotation

• Compute rotational invariant descriptor in this normalized frame

SIFT – Scale Invariant Feature Transform\textsuperscript{1}

- Empirically found\textsuperscript{2} to show very good performance, invariant to *image rotation, scale, intensity change*, and to moderate *affine* transformations

Scale = 2.5  
Rotation = 45\degree

\textsuperscript{1} D.Lowe. “Distinctive Image Features from Scale-Invariant Keypoints”. Accepted to IJCV 2004  
CVPR 2003 Tutorial

Recognition and Matching Based on Local Invariant Features

David Lowe
Computer Science Department
University of British Columbia
Invariant Local Features

- Image content is transformed into local feature coordinates that are invariant to translation, rotation, scale, and other imaging parameters

SIFT Features
Advantages of invariant local features

• **Locality:** features are local, so robust to occlusion and clutter (no prior segmentation)

• **Distinctiveness:** individual features can be matched to a large database of objects

• **Quantity:** many features can be generated for even small objects

• **Efficiency:** close to real-time performance

• **Extensibility:** can easily be extended to wide range of differing feature types, with each adding robustness
Scale invariance

Requires a method to repeatably select points in location and scale:

• The only reasonable scale-space kernel is a Gaussian (Koenderink, 1984; Lindeberg, 1994)
• An efficient choice is to detect peaks in the difference of Gaussian pyramid (Burt & Adelson, 1983; Crowley & Parker, 1984 – but examining more scales)
• Difference-of-Gaussian with constant ratio of scales is a close approximation to Lindeberg’s scale-normalized Laplacian (can be shown from the heat diffusion equation)
Scale space processed one octave at a time
Key point localization

- Detect maxima and minima of difference-of-Gaussian in scale space
- Fit a quadratic to surrounding values for sub-pixel and sub-scale interpolation (Brown & Lowe, 2002)
- Taylor expansion around point:
  \[ D(x) = D + \frac{\partial D^T}{\partial x} x + \frac{1}{2} x^T \frac{\partial^2 D}{\partial x^2} x \]
- Offset of extremum (use finite differences for derivatives):
  \[ \hat{x} = -\frac{\partial^2 D^{-1}}{\partial x^2} \frac{\partial D}{\partial x} \]
Select canonical orientation

- Create histogram of local gradient directions computed at selected scale
- Assign canonical orientation at peak of smoothed histogram
- Each key specifies stable 2D coordinates (x, y, scale, orientation)
Example of keypoint detection

Threshold on value at DOG peak and on ratio of principle curvatures (Harris approach)

(a) 233x189 image
(b) 832 DOG extrema
(c) 729 left after peak value threshold
(d) 536 left after testing ratio of principle curvatures
SIFT vector formation

- Thresholded image gradients are sampled over 16x16 array of locations in scale space
- Create array of orientation histograms
- 8 orientations x 4x4 histogram array = 128 dimensions

![Image showing SIFT vector formation](image.png)
Feature stability to noise

- Match features after random change in image scale & orientation, with differing levels of image noise
- Find nearest neighbor in database of 30,000 features
Feature stability to affine change

- Match features after random change in image scale & orientation, with 2% image noise, and affine distortion
- Find nearest neighbor in database of 30,000 features
Distinctiveness of features

- Vary size of database of features, with 30 degree affine change, 2% image noise
- Measure % correct for single nearest neighbor match
Figure 12: The training images for two objects are shown on the left. These can be recognized in a cluttered image with extensive occlusion, shown in the middle. The results of recognition are shown on the right. A parallelogram is drawn around each recognized object showing the boundaries of the original training image under the affine transformation solved for during recognition. Smaller squares indicate the keypoints that were used for recognition.
Figure 13: This example shows location recognition within a complex scene. The training images for locations are shown at the upper left and the 640x315 pixel test image taken from a different viewpoint is on the upper right. The recognized regions are shown on the lower image, with keypoints shown as squares and an outer parallelogram showing the boundaries of the training images under the affine transform used for recognition.
A good SIFT features tutorial


By Estrada, Jepson, and Fleet.
Talk Resume

• Stable (repeatable) feature points can be detected regardless of image changes
  – **Scale**: search for correct scale as *maximum* of appropriate function
  – **Affine**: approximate regions with *ellipses* (this operation is affine invariant)

• Invariant and distinctive descriptors can be computed
  – Invariant *moments*
  – *Normalizing* with respect to scale and affine transformation
Evaluation of interest points and descriptors

Cordelia Schmid
CVPR’03 Tutorial
Introduction

• Quantitative evaluation of interest point detectors
  – points / regions at the same relative location

  => repeatability rate

• Quantitative evaluation of descriptors
  – distinctiveness

  => detection rate with respect to false positives
Quantitative evaluation of detectors

- Repeatability rate: percentage of corresponding points

  - Two points are corresponding if
    1. The location error is less than 1.5 pixel
    2. The intersection error is less than 20%

  ![homography](image)
Comparison of different detectors

repeatability - image rotation

[Comparing and Evaluating Interest Points, Schmid, Mohr & Bauckhage, ICCV 98]
Comparison of different detectors

repeatability – perspective transformation

[Comparing and Evaluating Interest Points, Schmid, Mohr & Bauckhage, ICCV 98]
Harris detector + scale changes
Harris detector – adaptation to scale
Evaluation of scale invariant detectors

repeatability – scale changes
Evaluation of affine invariant detectors

repeatability – perspective transformation

Repeatability of detectors

- Harris–Laplacian
- Harris–Affine
- Harris–Affine Region

Repeatability % vs. viewpoint angle °
Quantitative evaluation of descriptors

• Evaluation of different local features
  – SIFT, steerable filters, differential invariants, moment invariants, cross-correlation

• Measure: distinctiveness
  – receiver operating characteristics of detection rate with respect to false positives
    – detection rate = correct matches / possible matches
    – false positives = false matches / (database points * query points)

[A performance evaluation of local descriptors, Mikolajczyk & Schmid, CVPR’03]
Experimental evaluation
Scale change (factor 2.5)

Harris-Laplace

DoG
Viewpoint change (60 degrees)

Harris-Affine (Harris-Laplace)
Descriptors - conclusion

• SIFT + steerable perform best

• Performance of the descriptor independent of the detector

• Errors due to imprecision in region estimation, localization
end
SIFT – Scale Invariant Feature Transform

• Descriptor overview:
  – Determine **scale** (by maximizing DoG in scale and in space), **local orientation** as the dominant gradient direction. Use this scale and orientation to make all further computations invariant to scale and rotation.
  – Compute **gradient orientation histograms** of several small windows (128 values for each point)
  – Normalize the descriptor to make it invariant to intensity change

D.Lowe. “Distinctive Image Features from Scale-Invariant Keypoints”. Accepted to IJCV 2004
Affine Invariant Texture Descriptor

- Segment the image into regions of different textures (by a non-invariant method)
- Compute matrix $M$ (the same as in Harris detector) over these regions
  
  $$M = \sum_{x,y} w(x, y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}$$

- This matrix defines the ellipse
  
  $$\begin{bmatrix} x \\ y \end{bmatrix} M \begin{bmatrix} x \\ y \end{bmatrix} = 1$$

- Regions described by these ellipses are invariant under affine transformations
- Find affine normalized frame
- Compute rotation invariant descriptor

Invariance to Intensity Change

• Detectors
  – mostly invariant to affine (linear) change in image intensity, because we are searching for maxima

• Descriptors
  – Some are based on derivatives => invariant to intensity shift
  – Some are normalized to tolerate intensity scale
  – Generic method: pre-normalize intensity of a region (eliminate shift and scale)