Lecture 12

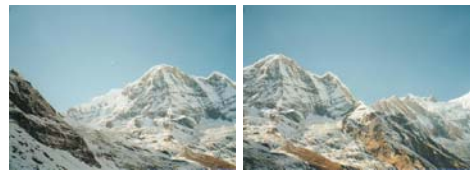
Local Feature Detection

Guest lecturer: Alex Berg

Reading:
- Harris and Stephens
- David Lowe IJCV

Why extract features?

- Motivation: panorama stitching
  - We have two images – how do we combine them?



  - We need to match (align) images

Building a Panorama

M. Brown and D. G. Lowe. Recognising Panoramas. ICCV 2003

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

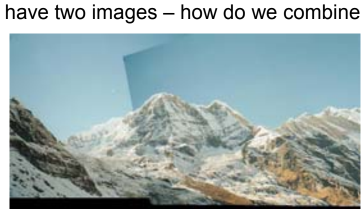
Why extract features?

- Motivation: panorama stitching
  - We have two images – how do we combine them?

Step 1: Detect feature points in both images
Step 2: Find corresponding pairs

Why extract features?

- Motivation: panorama stitching
  - We have two images – how do we combine them?



Step 1: Detect feature points in both images
Step 2: Find corresponding pairs
Step 3: 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

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

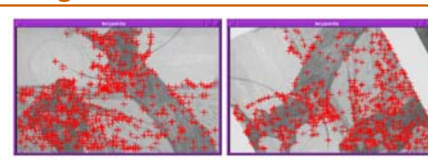
Applications

- Feature points are used for:
  - Motion tracking
  - Image alignment
  - 3D reconstruction
  - Object recognition
  - Indexing and database retrieval
  - Robot navigation

Contents

- Harris Corner Detector
  - Description
  - Analysis
- Detectors
  - Rotation invariant
  - Scale invariant
  - Affine invariant
- Descriptors
  - Rotation invariant
  - Scale invariant
  - Affine invariant

Finding Corners

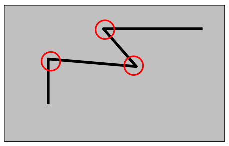


- Key property: in the region around a corner, image gradient has two or more dominant directions
- Corners are repeatable and distinctive

C.Harris and M.Stephens. "A Combined Corner and Edge Detector." *Proceedings of the 4th Alvey Vision Conference*: pages 147--151.

An introductory example:

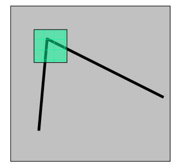
Harris corner detector



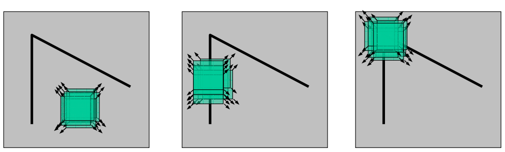
C.Harris, M.Stephens. "A Combined Corner and Edge Detector". 1988

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

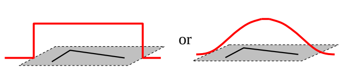
Harris Detector: Mathematics

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 — Shifted intensity — Intensity

Window function $w(x,y)$ =



1 in window, 0 outside — or — Gaussian

Harris Detector: Mathematics

Change of intensity for the shift $[u,v]$:

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

Second-order Taylor expansion of $E(u,v)$ about (0,0) (bilinear approximation for small shifts):

$$E(u,v) \approx E(0,0) + [u \;\; v]\begin{bmatrix} E_u(0,0) \\ E_v(0,0) \end{bmatrix} + \frac{1}{2}[u \;\; v]\begin{bmatrix} E_{uu}(0,0) & E_{uv}(0,0) \\ E_{uv}(0,0) & E_{vv}(0,0) \end{bmatrix}\begin{bmatrix} u \\ v \end{bmatrix}$$

Harris Detector: Mathematics

Expanding E(u,v) in a 2nd order Taylor series expansion, we have,for small shifts $[u,v]$, a *bilinear* approximation:

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

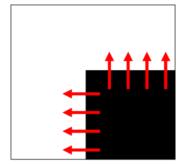
where M is a 2×2 matrix computed from image derivatives:

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

$$M = \begin{bmatrix} \sum I_x I_x & \sum I_x I_y \\ \sum I_x I_y & \sum I_y I_y \end{bmatrix} = \sum \begin{bmatrix} I_x \\ I_y \end{bmatrix} [I_x \; I_y] = \sum \nabla I (\nabla I)^T$$

Interpreting the second moment matrix

First, consider an axis-aligned corner:



Interpreting the second moment matrix

- First, 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 λ is close to 0, then this is **not** a corner, so look for locations where both are large.

Slide credit: David Jacobs

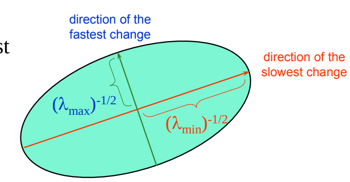
General Case

Since M is symmetric, we have $M = R^{-1} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} R$

We can visualize M as an ellipse with axis lengths determined by the eigenvalues and orientation determined by R

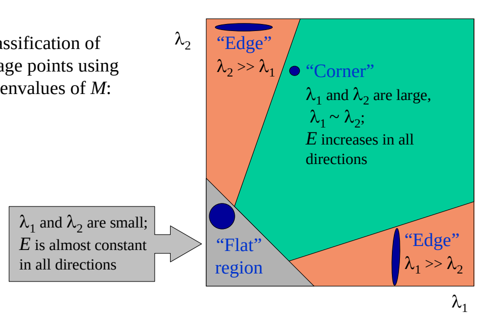
Ellipse $E(u,v)$ = const

$[u \; v] \; M \begin{bmatrix} u \\ v \end{bmatrix}$ = const



Harris Detector: Mathematics

Classification of image points using eigenvalues of M:



Harris Detector: Mathematics

Measure of corner response:

$$R = \det M - k \left(\operatorname{trace} M \right)^2$$

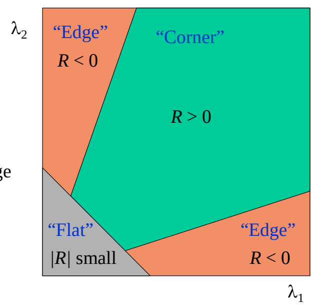
$$\det M = \lambda_1 \lambda_2$$
$$\operatorname{trace} M = \lambda_1 + \lambda_2$$

(k – empirical constant, k = 0.04-0.06)

Harris Detector: Mathematics

$$R = \det(M) - k\,\mathrm{trace}(M)^2 = \lambda_1\lambda_2 - k(\lambda_1 + \lambda_2)^2, \quad 0.04 \le k \le .06$$

- 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: Summary

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

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

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

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

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

Harris Detector

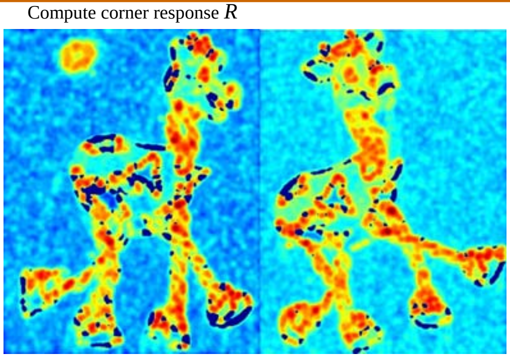
Algorithm:

1. Compute Gaussian derivatives at each pixel
2. Compute second moment matrix M in a Gaussian window around each pixel
3. Compute corner response function R
4. Threshold R
5. Find local maxima of response function (nonmaximum suppression)

Harris Detector: Workflow

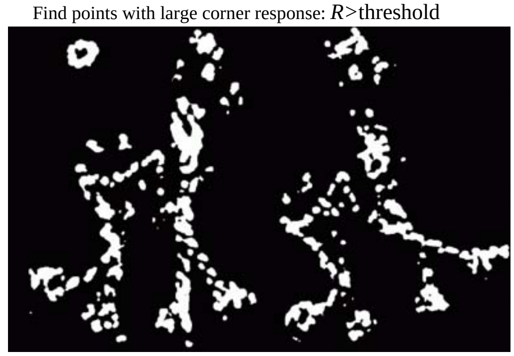
Harris Detector: Workflow

Compute corner response R



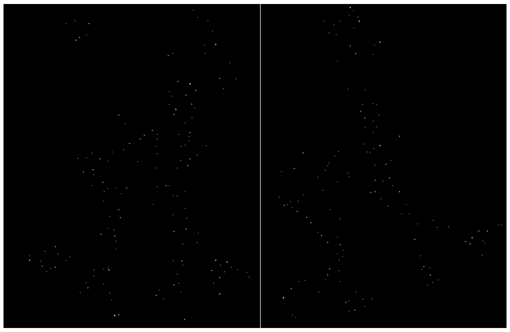
Harris Detector: Workflow

Find points with large corner response: R>threshold



Harris Detector: Workflow

Take only the points of local maxima of R



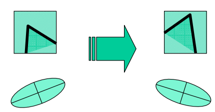
Harris Detector: Workflow

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



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

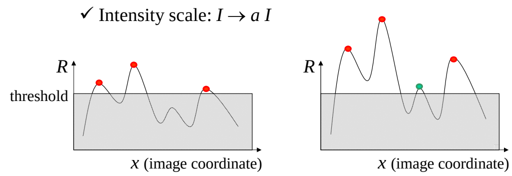
Corner response R is invariant to image rotation

Harris Detector: Some Properties

- Invariance to image intensity change?

Harris Detector: Some Properties

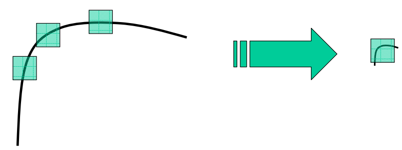
- Partial invariance to additive and multiplicative intensity changes
  - ✓ Only derivatives are used => invariance to intensity shift $I \rightarrow I + b$
  - ✓ Intensity scale: $I \rightarrow a\, I$



Harris Detector: Some Properties

- Invariant to image scale?

Harris Detector: Some Properties



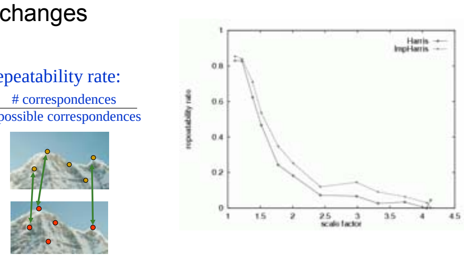
Not invariant to scaling

Harris Detector: Some Properties

- Quality of Harris detector for different scale changes

Repeatability rate:

$$\frac{\text{\# correspondences}}{\text{\# possible correspondences}}$$



C.Schmid et.al. "Evaluation of Interest Point Detectors". IJCV 2000

Contents

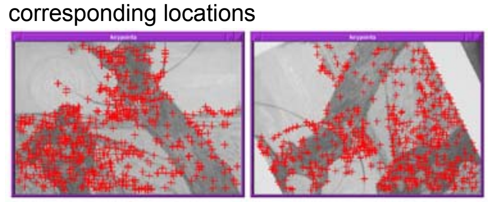
We want to:

detect *the same* interest points regardless of *image changes*

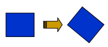
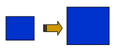
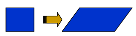
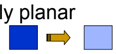
Darya Frolova, Denis Simakov
http://www.wisdom.weizmann.ac.il/~deniss/vision_spring04/files/InvariantFeatures.ppt

Invariance

- We want features to be detected despite geometric or photometric changes in the image: if we have two transformed versions of the same image, features should be detected in corresponding locations



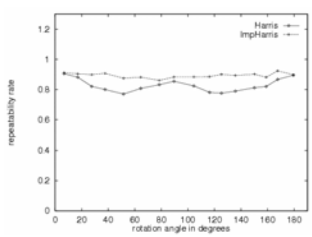
Models of Image Change

- Geometric
  - **Rotation** 
  - **Scale** 
  - 
  - **Affine**
    valid for: orthographic camera, locally planar object 
- Photometric
  - **Affine intensity change** ($I \rightarrow a\,I + b$)

Contents

Rotation Invariant Detection

- Harris Corner Detector

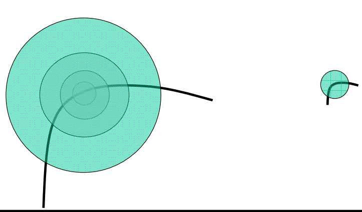


C.Schmid et.al. “Evaluation of Interest Point Detectors”. IJCV 2000

Contents

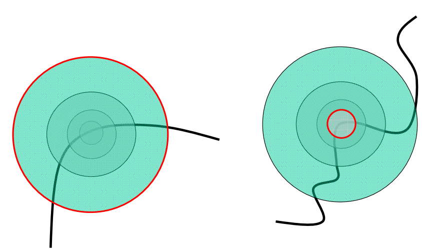
Scale Invariant Detection

- **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?**

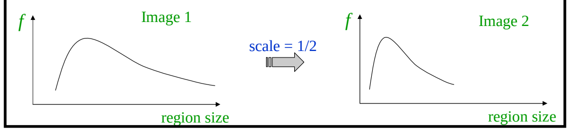


Scale Invariant Detection

- **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)



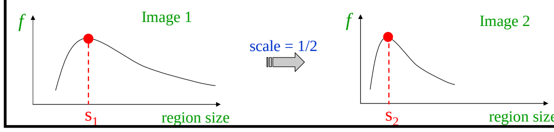
Scale Invariant Detection

- **Common approach:**

  Take a local maximum of this function

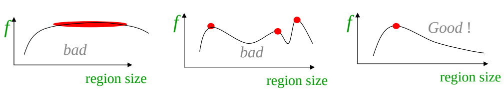
  Observation: region size, for which the maximum is achieved, should be *invariant* to image scale.

  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**

$$f = \text{Kernel} * \text{Image}$$

Kernels:

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

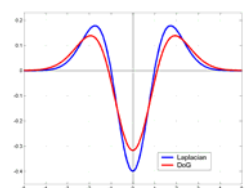
(Laplacian)

$$DoG = G(x, y, k\sigma) - G(x, y, \sigma)$$

(Difference of Gaussians)

where Gaussian

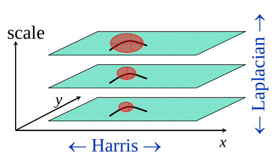
$$G(x, y, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2+y^2}{2\sigma^2}}$$



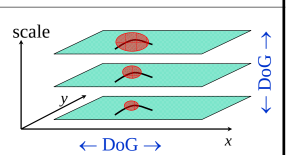
Note: both kernels are invariant to *scale* and *rotation*

Scale Invariant Detectors

- Harris-Laplacian[1]
  Find local maximum of:
  - Harris corner detector in space (image coordinates)
  - Laplacian in scale



- SIFT (Lowe)[2]
  Find local maximum of:
  - Difference of Gaussians in space and scale



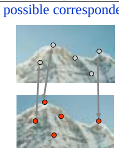
[1] K.Mikolajczyk, C.Schmid. "Indexing Based on Scale Invariant Interest Points". ICCV 2001
[2] D.Lowe. "Distinctive Image Features from Scale-Invariant Keypoints". Accepted to IJCV 2004

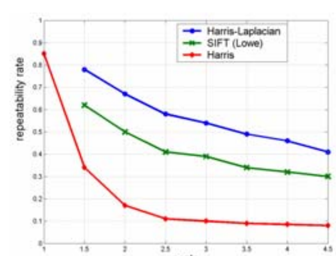
Scale Invariant Detectors

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

Repeatability rate:

$$\frac{\#\ \text{correspondences}}{\#\ \text{possible correspondences}}$$





K.Mikolajczyk, C.Schmid. "Indexing Based on Scale Invariant Interest Points". ICCV 2001

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)**

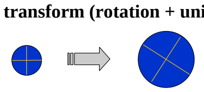
Methods:

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

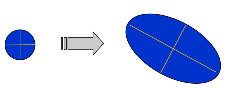
Contents

Affine Invariant Detection

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



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



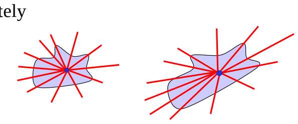
Affine Invariant Detection

- **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**

$$f(t) = \frac{|I(t) - I_0|}{\frac{1}{t}\int_0^t |I(t) - I_0|\,dt}$$

- We will obtain approximately corresponding regions

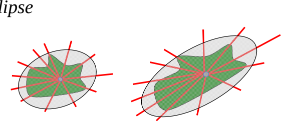
Remark: we search for scale in every direction



T.Tuytelaars, L.V.Gool. "Wide Baseline Stereo Matching Based on Local, Affinely Invariant Regions". BMVC 2000.

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*



T.Tuytelaars, L.V.Gool. "Wide Baseline Stereo Matching Based on Local, Affinely Invariant Regions". BMVC 2000.

Affine Invariant Detection

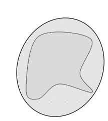
- **The regions found may not exactly correspond, so we approximate them with ellipses**

- Geometric Moments:

$$m_{pq} = \int_{\mathbb{R}^2} x^p y^q f(x,y)\,dxdy$$

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

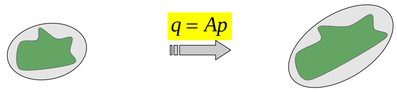


This ellipse will have the same moments of orders up to 2 as the original region

Affine Invariant Detection

- Covariance matrix of region points defines an ellipse:

$q = Ap$



$p^T \Sigma_1^{-1} p = 1$

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

$q^T \Sigma_2^{-1} q = 1$

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

($p = [x, y]^T$ is relative to the center of mass)

$\Sigma_2 = A \Sigma_1 A^T$

Ellipses, computed for corresponding regions, also correspond!

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.]

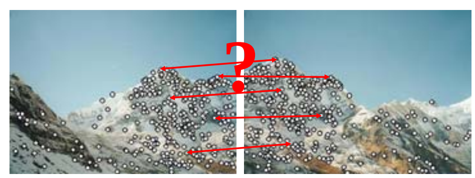
Contents

- 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

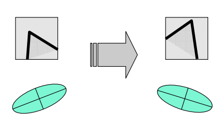
Contents

- 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*

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



C.Harris, M.Stephens. “A Combined Corner and Edge Detector”. 1988

Descriptors Invariant to Rotation

- **Image moments in polar coordinates**

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

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: $|m_{kl}|$

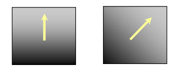
Matching is done by comparing vectors $[|m_{kl}|]_{k,l}$

J.Matas et.al. "Rotational Invariants for Wide-baseline Stereo". Research Report of CMP, 2003

Descriptors Invariant to Rotation

- **Find local orientation**

Dominant direction of gradient



- Compute image derivatives relative to this orientation

[1] K.Mikolajczyk, C.Schmid. "Indexing Based on Scale Invariant Interest Points". ICCV 2001
[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**

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

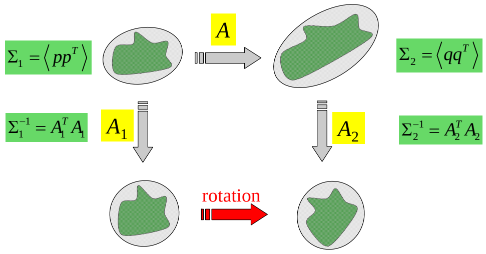
Different combinations of these moments are fully affine invariant

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**



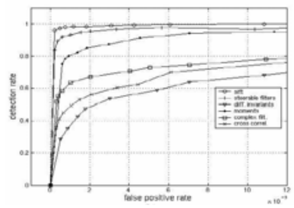
- Compute rotational invariant descriptor in this normalized frame

J.Matas et.al. "Rotational Invariants for Wide-baseline Stereo". Research Report of CMP, 2003

SIFT – Scale Invariant Feature Transform[1]

- **Empirically found[2] to show very good performance, invariant to *image rotation*, *scale*, *intensity change*, and to moderate *affine* transformations**

Scale = 2.5
Rotation = 45^0



[1] D.Lowe. "Distinctive Image Features from Scale-Invariant Keypoints". Accepted to IJCV 2004
[2] K.Mikolajczyk, C.Schmid. "A Performance Evaluation of Local Descriptors". CVPR 2003

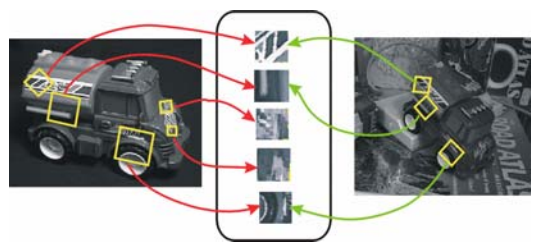
CVPR 2003 Tutorial

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

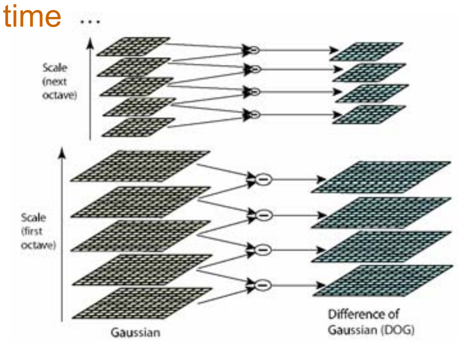
- Locality: **features are local, so robust to occlusion and clutter (no prior segmentation)**
- Distinctiveness: **individual features can be matched to a large database of objects**
- Quantity: **many features can be generated for even small objects**
- Efficiency: **close to real-time performance**
- 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 …



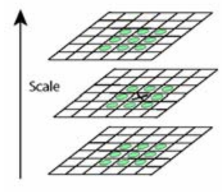
Key point localization

- 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:

$$D(\mathbf{x}) = D + \frac{\partial D^T}{\partial \mathbf{x}}\mathbf{x} + \frac{1}{2}\mathbf{x}^T \frac{\partial^2 D}{\partial \mathbf{x}^2}\mathbf{x}$$

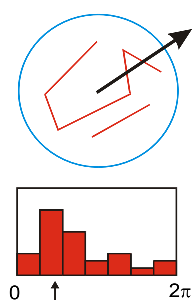
- Offset of extremum (use finite differences for derivatives):

$$\hat{\mathbf{x}} = -\frac{\partial^2 D}{\partial \mathbf{x}^2}^{-1}\frac{\partial D}{\partial \mathbf{x}}$$



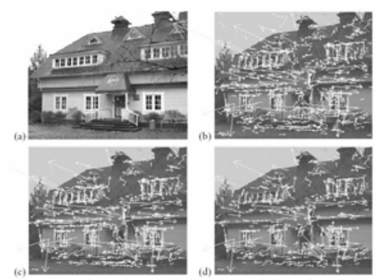
Select canonical orientation

- 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

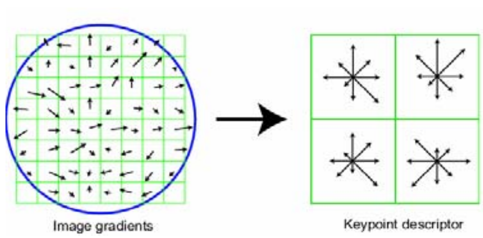
Threshold on value at DOG peak and on ratio of principle curvatures (Harris approach)



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

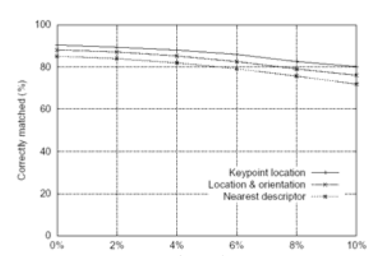
SIFT vector formation

- 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



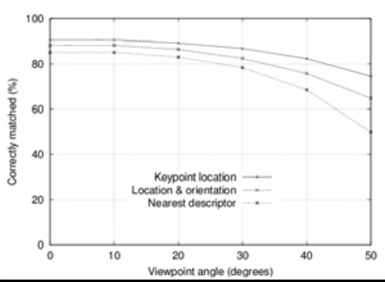
Feature stability to noise

- 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

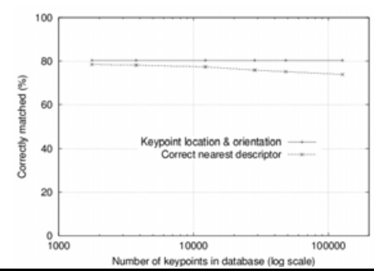


Figure 12: The training images for two objects are shown on the left. These can be recognized in a cluttered image with extensive occlusion, shown in the middle. The results of recognition are shown on the right. A parallelogram is drawn around each recognized object showing the boundaries of the original training image under the affi ne 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 affi ne transform used for recognition.

A good SIFT features tutorial

http://www.cs.toronto.edu/~jepson/csc2503/tutSIFT04.pdf

By Estrada, Jepson, and Fleet.