WST665/CS770A: Web-Scale Image Retrieval

Descriptors

Sung-Eui Yoon
(윤성의)

Course URL:
http://sglab.kaist.ac.kr/~sungeui/IR

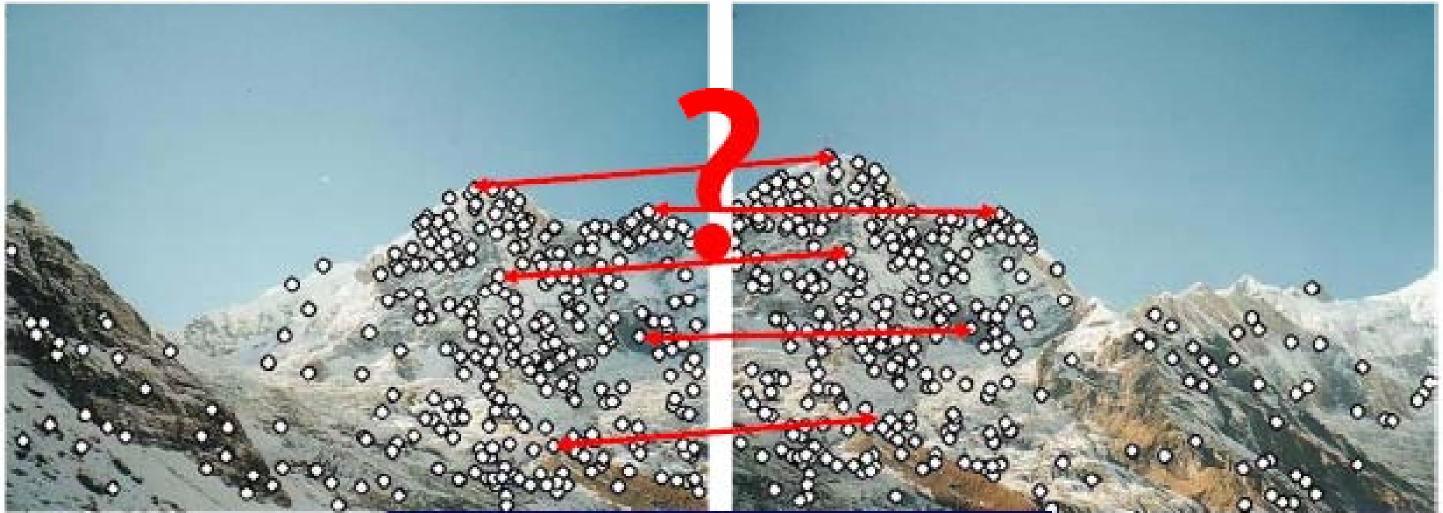
What we will learn today

- Local descriptors
  - SIFT
  - An assortment of other descriptors
  - Applications

Local Descriptors

- We know how to detect points
- Next question:

How to *describe* them for matching?

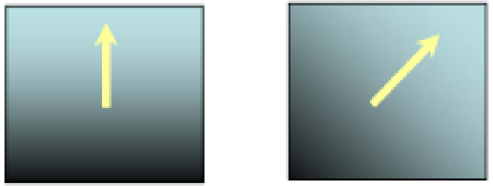


Point descriptor should be:
1. Invariant
2. Distinctive

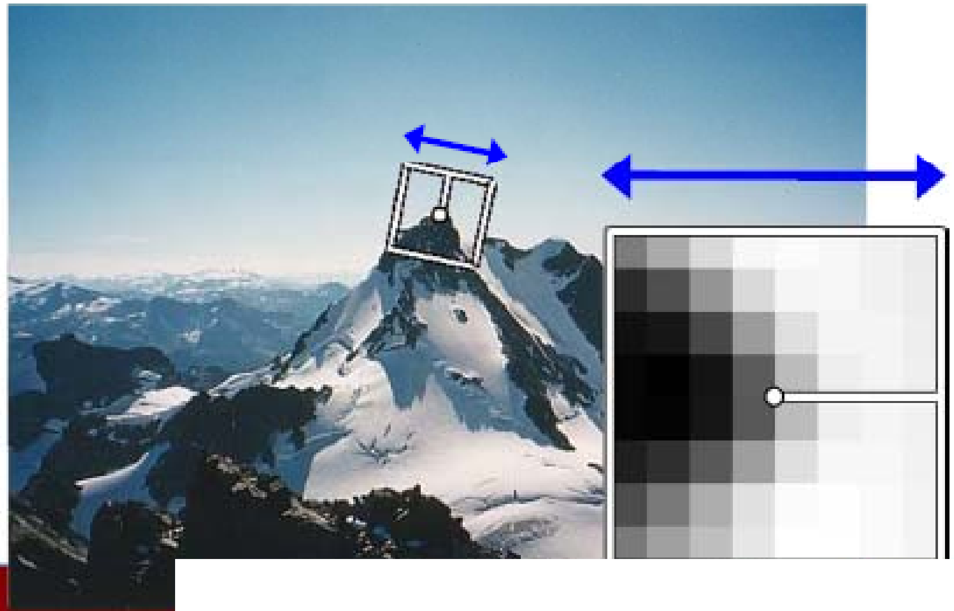
Slide credit: Kristen Grauman

Rotation Invariant Descriptors

- Find local orientation
  - Dominant direction of gradient for the image patch



- Rotate patch according to this angle
  - This puts the patches into a canonical orientation.

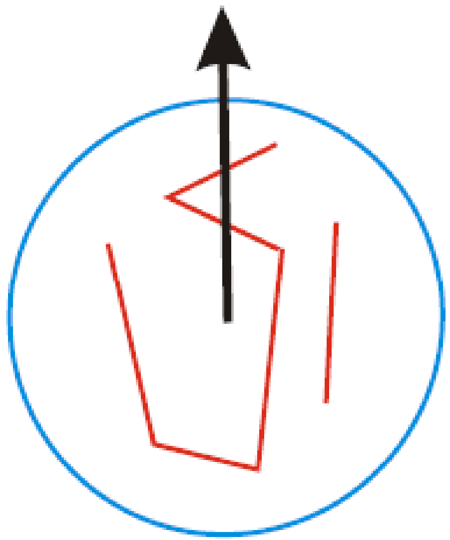


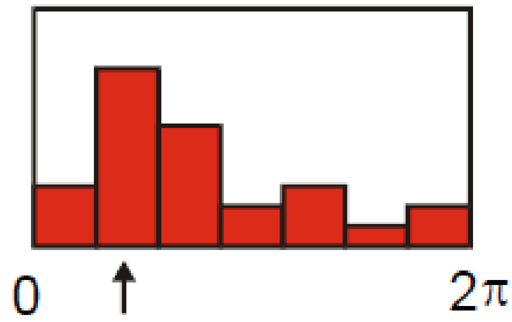
Slide credit: Svetlana Lazebnik, Matthew Brown

Orientation Normalization: Computation

[Lowe, SIFT, 1999]

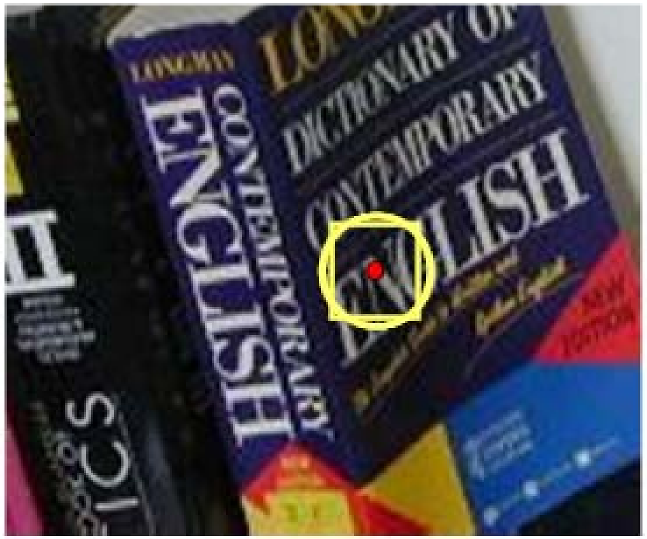
- Compute orientation histogram
- Select dominant orientation
- Normalize: rotate to fixed orientation

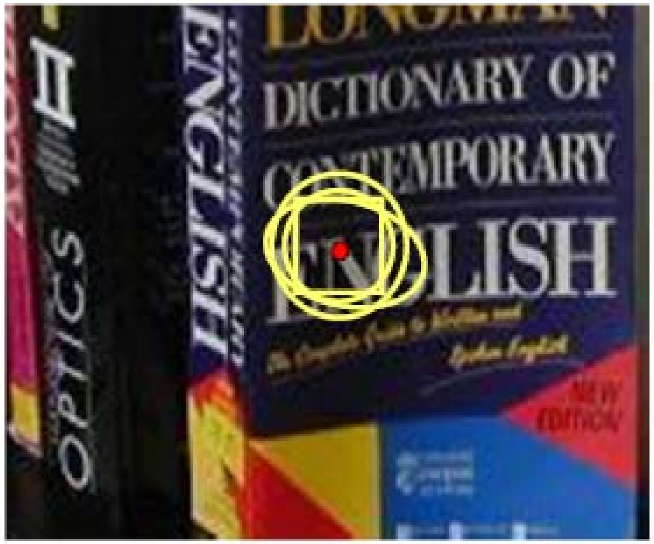




Slide adapted from David Lowe

The Need for Invariance



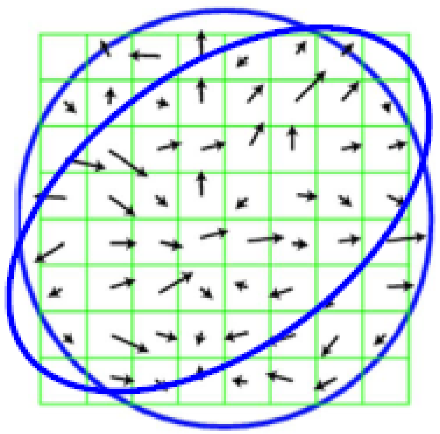


- Up to now, we had invariance to
  - Translation
  - Scale
  - Rotation
- Not sufficient to match regions under viewpoint changes
  - For this, we need also affine adaptation

Slide credit: Tinne Tuytelaars

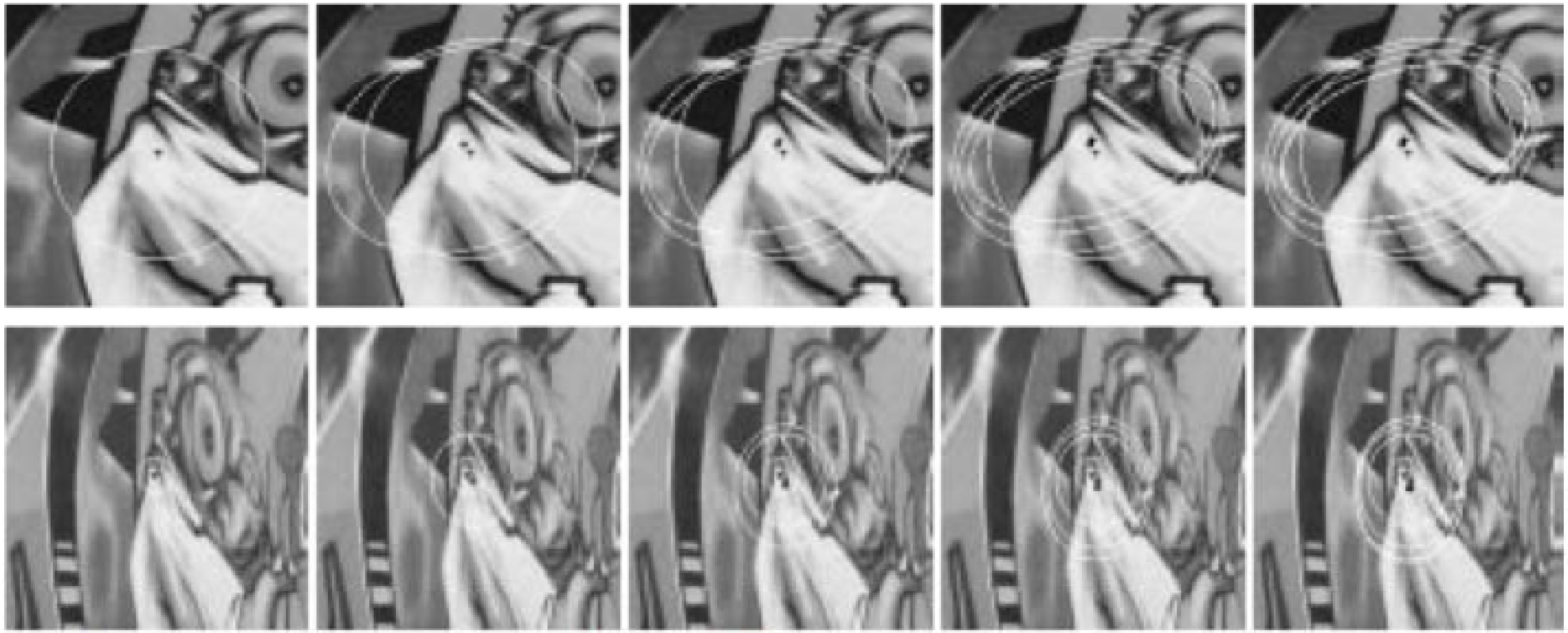
Affine Adaptation

- Problem:
  - Determine the characteristic shape of the region.
  - Assumption: shape can be described by “local affine frame”.
- Solution: iterative approach
  - Use a circular window to compute second moment matrix.
  - Compute eigenvectors to adapt the circle to an ellipse.
  - Recompute second moment matrix using new window and iterate…



Slide adapted from Svetlana Lazebnik

Iterative Affine Adaptation



1. Detect keypoints, e.g. multi-scale Harris
2. Automatically select the scales
3. Adapt affine shape based on second order moment matrix
4. Refine point location

K. Mikolajczyk and C. Schmid, Scale and affine invariant interest point detectors, IJCV 60(1):63-86, 2004.

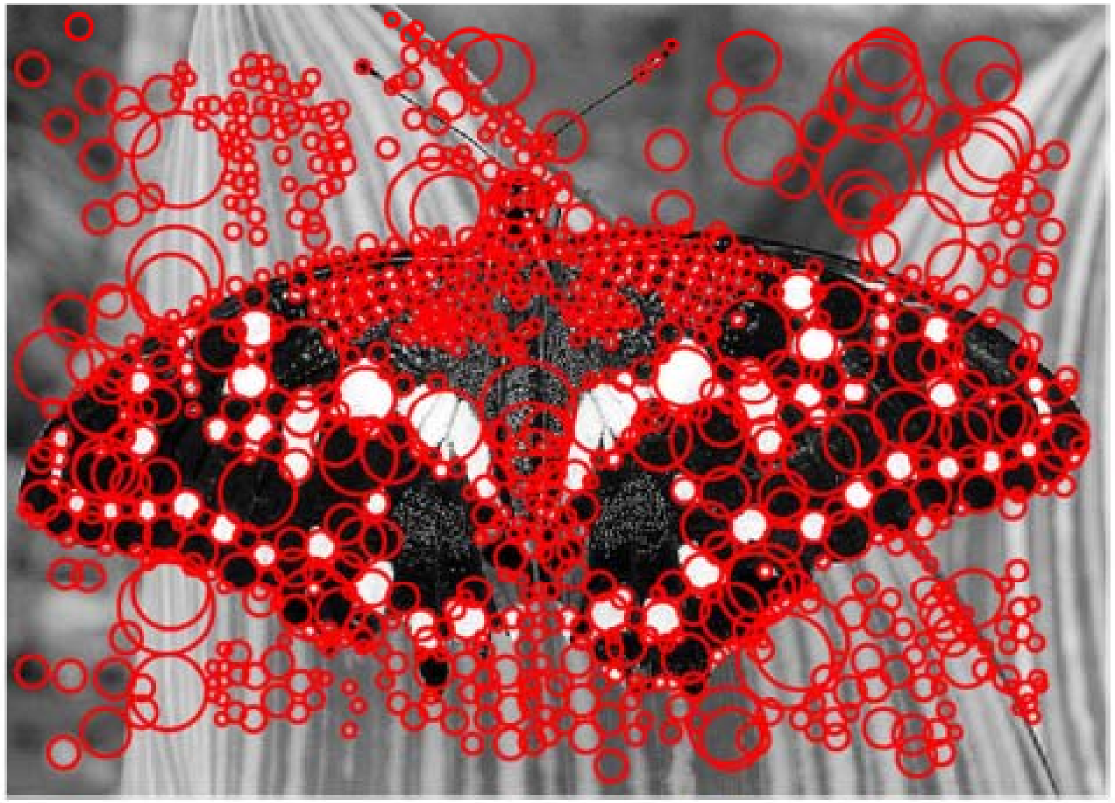
Slide credit: Tinne Tuytelaars

Affine Normalization/Deskewing

- Steps
  - Rotate the ellipse’s main axis to horizontal
  - Scale the x axis, such that it forms a circle

Slide credit: Tinne Tuytelaars

Affine Adaptation Example



Scale-invariant regions (blobs)

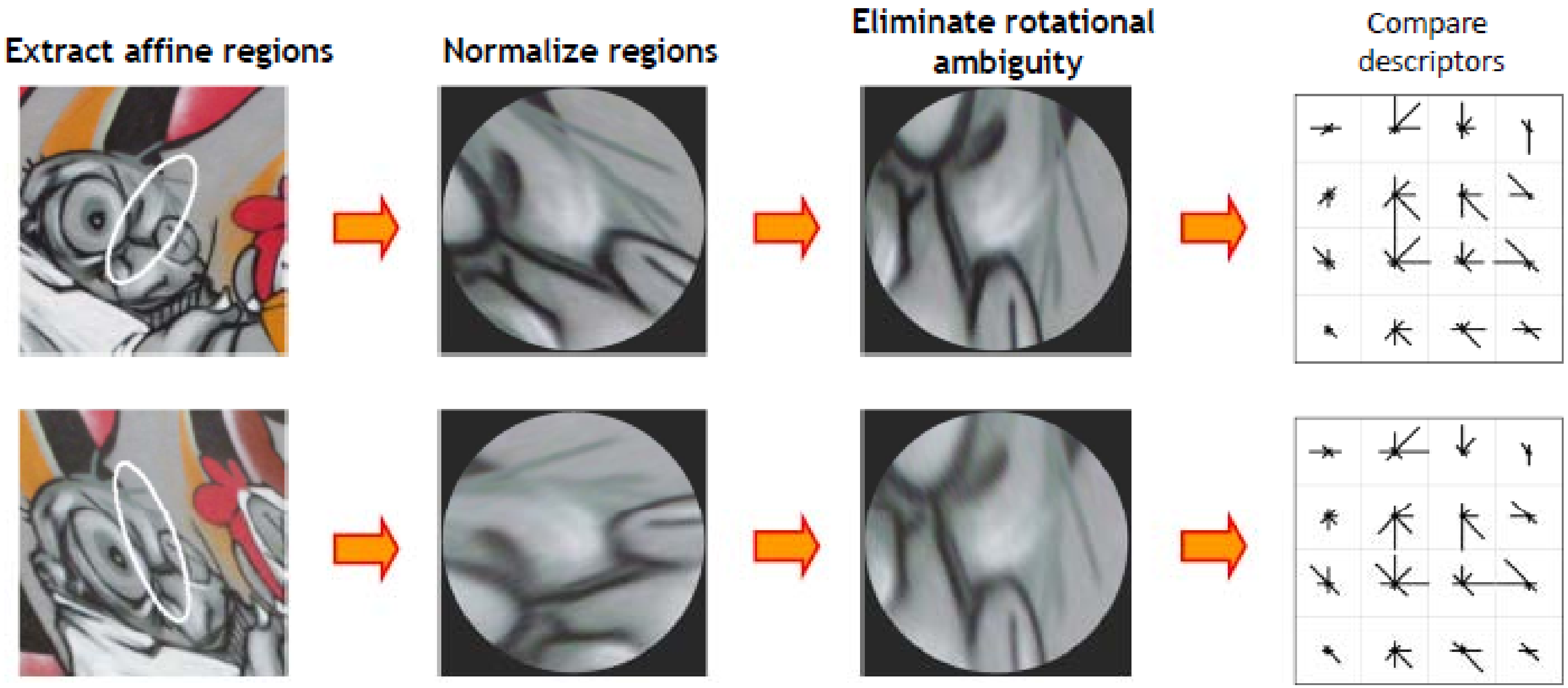
Slide credit: Svetlana Lazebnik

Affine Adaptation Example

Affine-adapted blobs

Slide credit: Svetlana Lazebnik

Summary: Affine-Inv. Feature Extraction

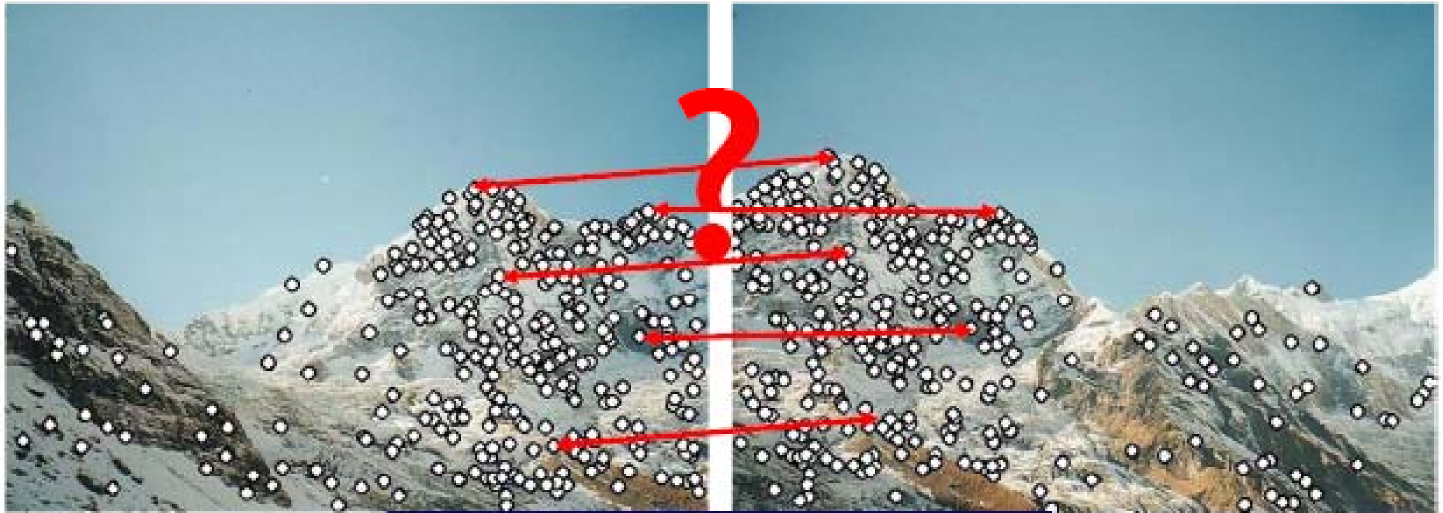


Slide credit: Svetlana Lazebnik

Local Descriptors

- We know how to detect points
- Next question:

How to *describe* them for matching?



Point descriptor should be:
1. Invariant
2. Distinctive

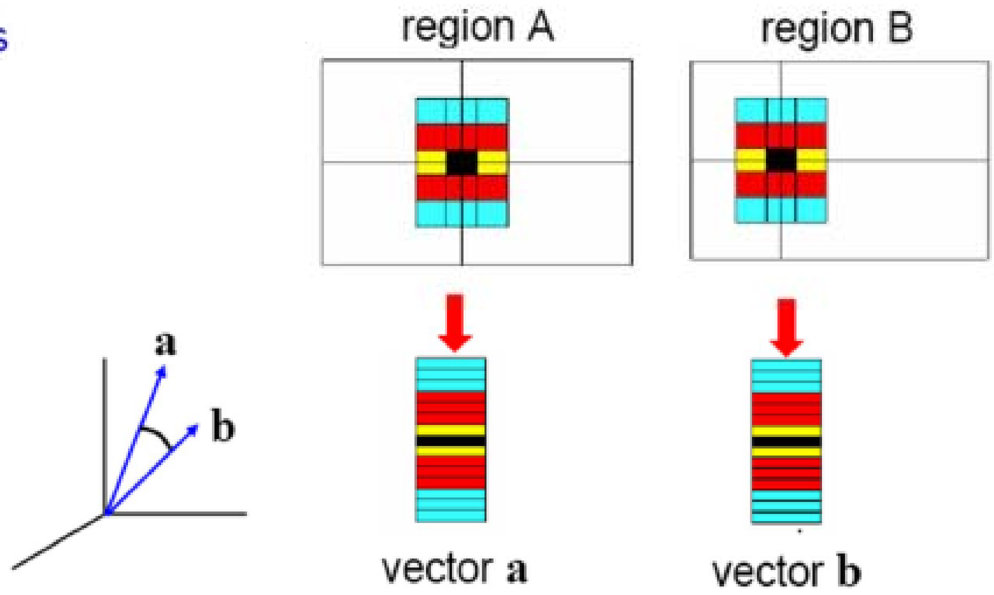
Slide credit: Kristen Grauman

Local Descriptors

- Simplest descriptor: list of intensities within a patch.
- What is this going to be invariant to?

Write regions as vectors

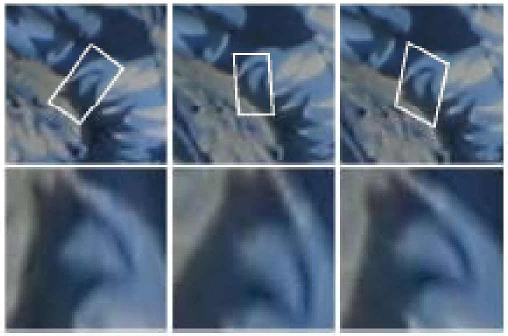
$$\mathtt{A} \rightarrow \mathbf{a}, \ \ \mathtt{B} \rightarrow \mathbf{b}$$



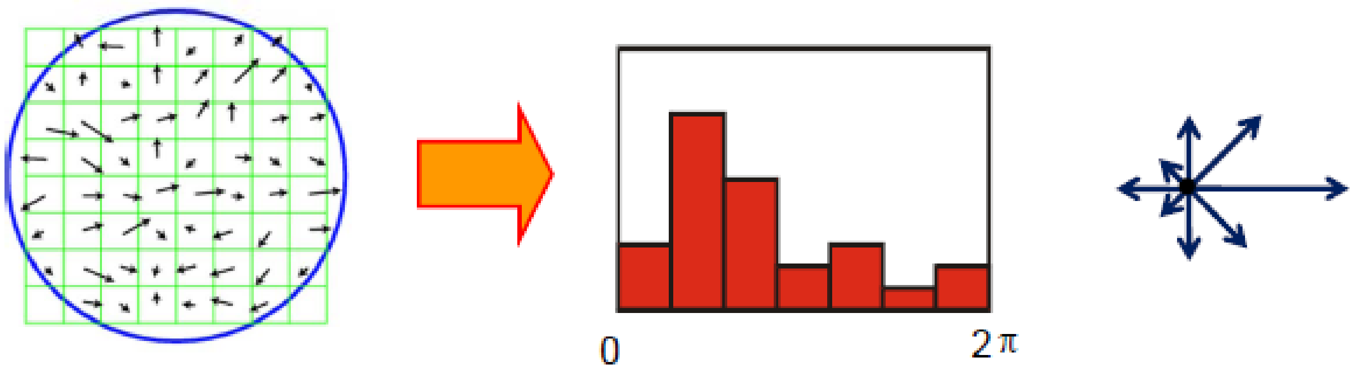
Slide credit: Kristen Grauman

Feature Descriptors

- Disadvantage of patches as descriptors:
  - Small shifts can affect matching score a lot



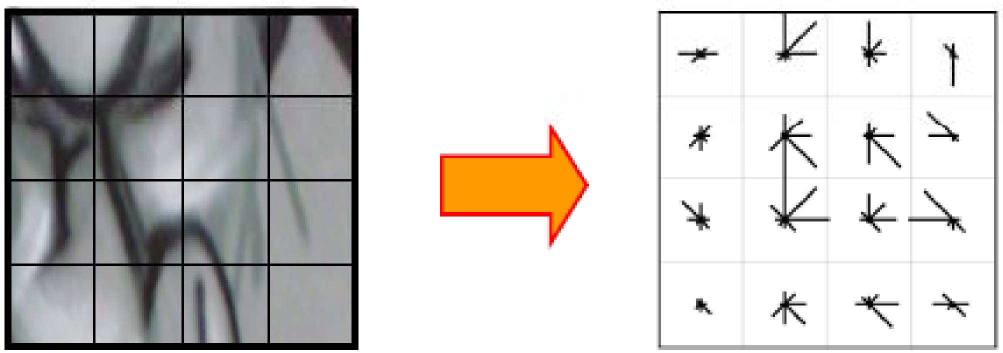
- Solution: histograms



Slide credit: Svetlana Lazebnik

Feature Descriptors: SIFT

- Scale Invariant Feature Transform
- Descriptor computation:
  - Divide patch into 4x4 sub-patches: 16 cells
  - Compute histogram of gradient orientations (8 reference angles) for all pixels inside each sub-patch
  - Resulting descriptor: 4x4x8 = 128 dimensions



David G. Lowe. "Distinctive image features from scale-invariant keypoints." *IJCV* 60 (2), pp. 91-110, 2004.

Slide credit: Svetlana Lazebnik

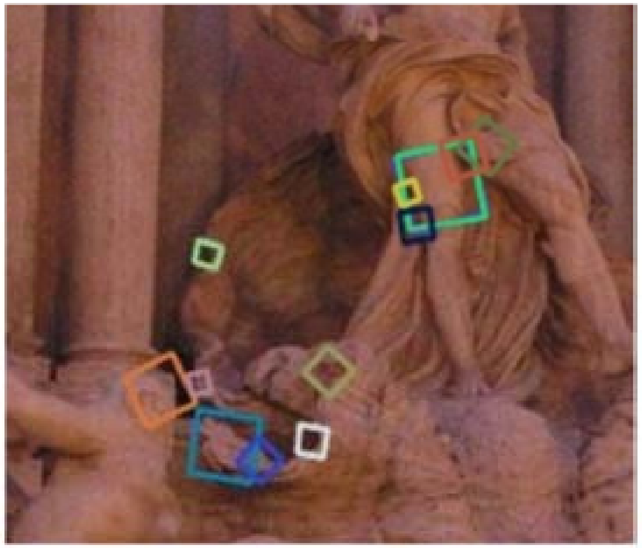
Overview: SIFT

- Extraordinarily robust matching technique
  - Can handle changes in viewpoint up to ~60 deg. out-of-plane rotation
  - Can handle significant changes in illumination
    - Sometimes even day vs. night (below)
  - Fast and efficient—can run in real time
  - Lots of code available
    - http://people.csail.mit.edu/albert/ladypack/wiki/index.php/Known_implementations_of_SIFT

Slide credit: Steve Seitz

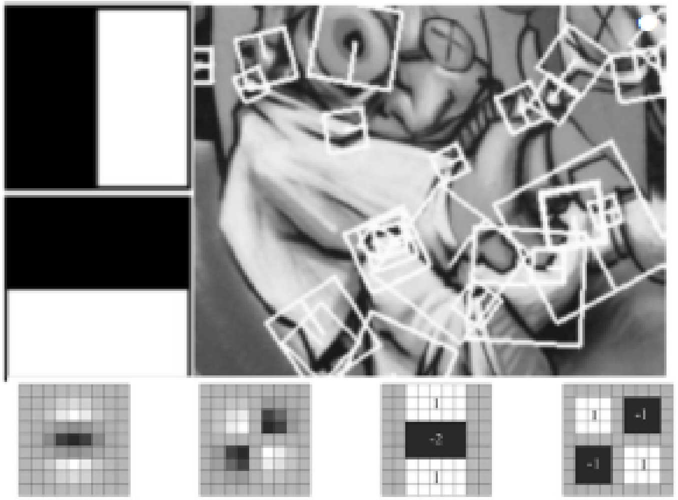
Working with SIFT Descriptors

- One image yields:
  - n 128-dimensional descriptors: each one is a histogram of the gradient orientations within a patch
    - [n x 128 matrix]
  - n scale parameters specifying the size of each patch
    - [n x 1 vector]
  - n orientation parameters specifying the angle of the patch
    - [n x 1 vector]
  - n 2D points giving positions of the patches
    - [n x 2 matrix]



Slide credit: Steve Seitz

Local Descriptors: SURF



- **Fast approximation of SIFT idea**
  - Efficient computation by 2D box filters & integral images
    ⇒ 6 times faster than SIFT
  - Equivalent quality for object identification
  - http://www.vision.ee.ethz.ch/~surf

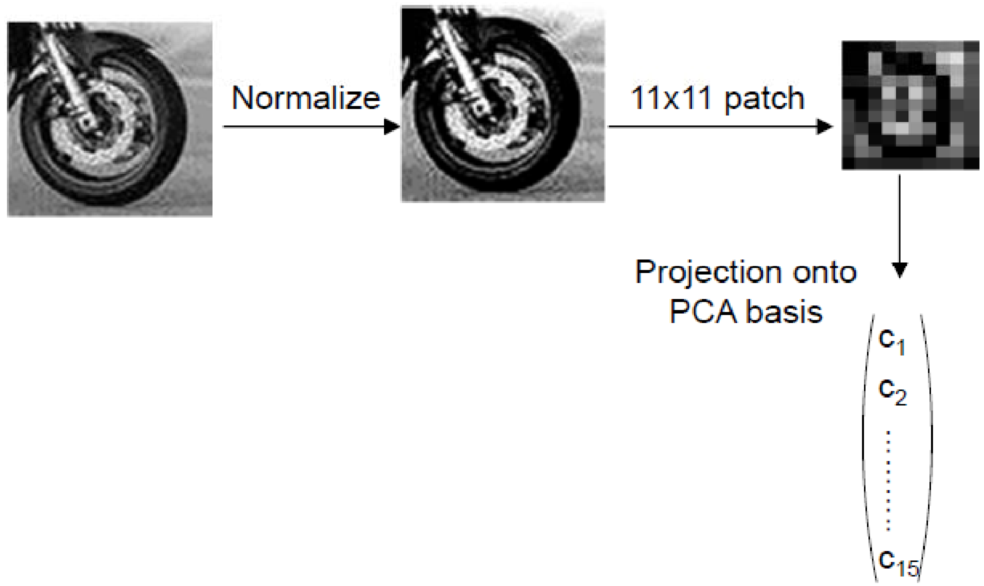
http://www.vision.ee.ethz.ch/~surf

- **GPU implementation available**
  - Feature extraction @ 100Hz (detector + descriptor, 640×480 img)
  - http://homes.esat.kuleuven.be/~ncorneli/gpusurf/

[Bay, ECCV'06], [Cornelis, CVGPU'08]

Other Descriptors

- Gray-scale intensity



- GIST
- Many others

Applications of Local Invariant Features

- Wide baseline stereo
- Motion tracking
- Panoramas
- Mobile robot navigation
- 3D reconstruction
- Recognition
  - Specific objects
  - Textures
  - Categories
- ...

Slide credit: Kristen Grauman

Wide-Baseline Stereo

Image from T. Tuytelaars ECCV 2006 tutorial

Automatic Mosaicing

[Brown & Lowe, ICCV'03]

Panorama Stitching

(a) Matier data set (7 images)

iPhone version available

(b) Matier final stitch

[Brown, Szeliski, and Winder, 2005]

http://www.cs.ubc.ca/~mbrown/autostitch/autostitch.html

Recognition of Specific Objects, Scenes

Schmid and Mohr 1997

Sivic and Zisserman, 2003

Rothganger et al. 2003

Lowe 2002

Slide credit: Kristen Grauman

Value of Local Features

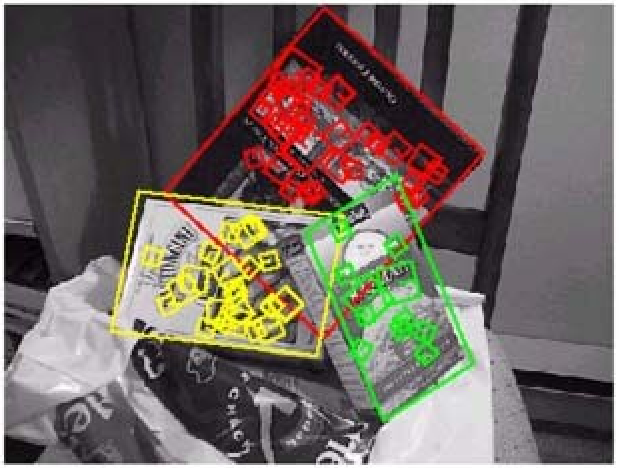
- Advantages
  - Critical to find distinctive and repeatable local regions for multi-view matching.
  - Complexity reduction via selection of distinctive points.
  - Describe images, objects, parts without requiring segmentation; robustness to clutter & occlusion.
  - Robustness: similar descriptors in spite of moderate view changes, noise, blur, etc.

- How can we use local features for such applications?

  matching and recognition

Alignment Problem

- **Fit different images into one canonical image**



Alignment Problem

- Many different approaches exist

- Simple fitting procedure in the linear least square sense
  - Approximates viewpoint changes for roughly planar objects and roughly orthographic cameras
  - Can be used to initialize fitting for more complex models

- We do not discuss this issue here
  - Will be discussed in a computer vision course

Time for a Demo...

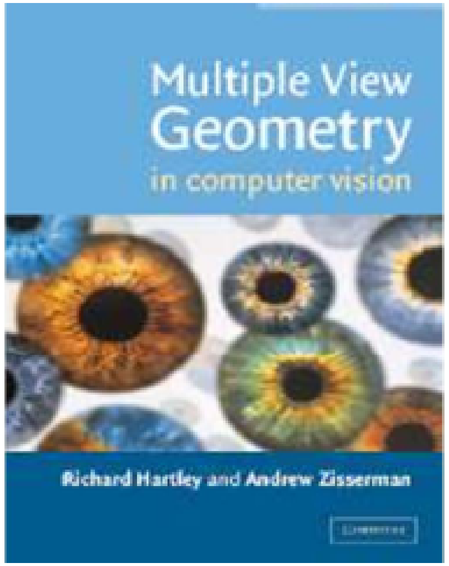
Automatic panorama stitching

Matthew Brown: http://cvlab.epfl.ch/~brown/autostitch/autostitch.html

References and Further Reading

•More details on the alignment problem can be found in:

- R. Hartley, A. Zisserman
  Multiple View Geometry in Computer Vision
  2nd Ed., Cambridge Univ. Press, 2004



- Details about the DoG detector and the SIFT descriptor can be found in
  - D. Lowe, Distinctive image features from scale-invariant keypoints, *IJCV* 60(2), pp. 91-110, 2004

- Try the available local feature detectors and descriptors
  - http://www.robots.ox.ac.uk/~vgg/research/affine/detectors.html#binaries

What we have learned today

- Local descriptor
  - SIFT
  - An assortment of other descriptors
  - Applications

Next Time…

- Object recognition
- Bag-of-Words (BoW) models

Homework for Every Class

- Go over the next lecture slides
- Come up with one question on what we have discussed today
  - 0 for no questions
  - 2 for typical questions
  - 3 for questions with thoughts
  - 4 for questions that surprised me