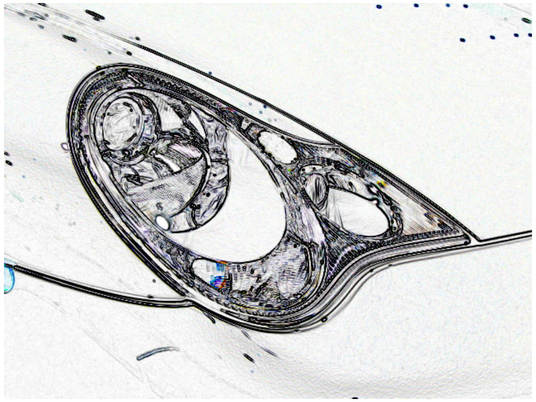
CSE578: Computer Vision

Spring 2017:

Feature Description and Matching



Anoop M. Namboodiri

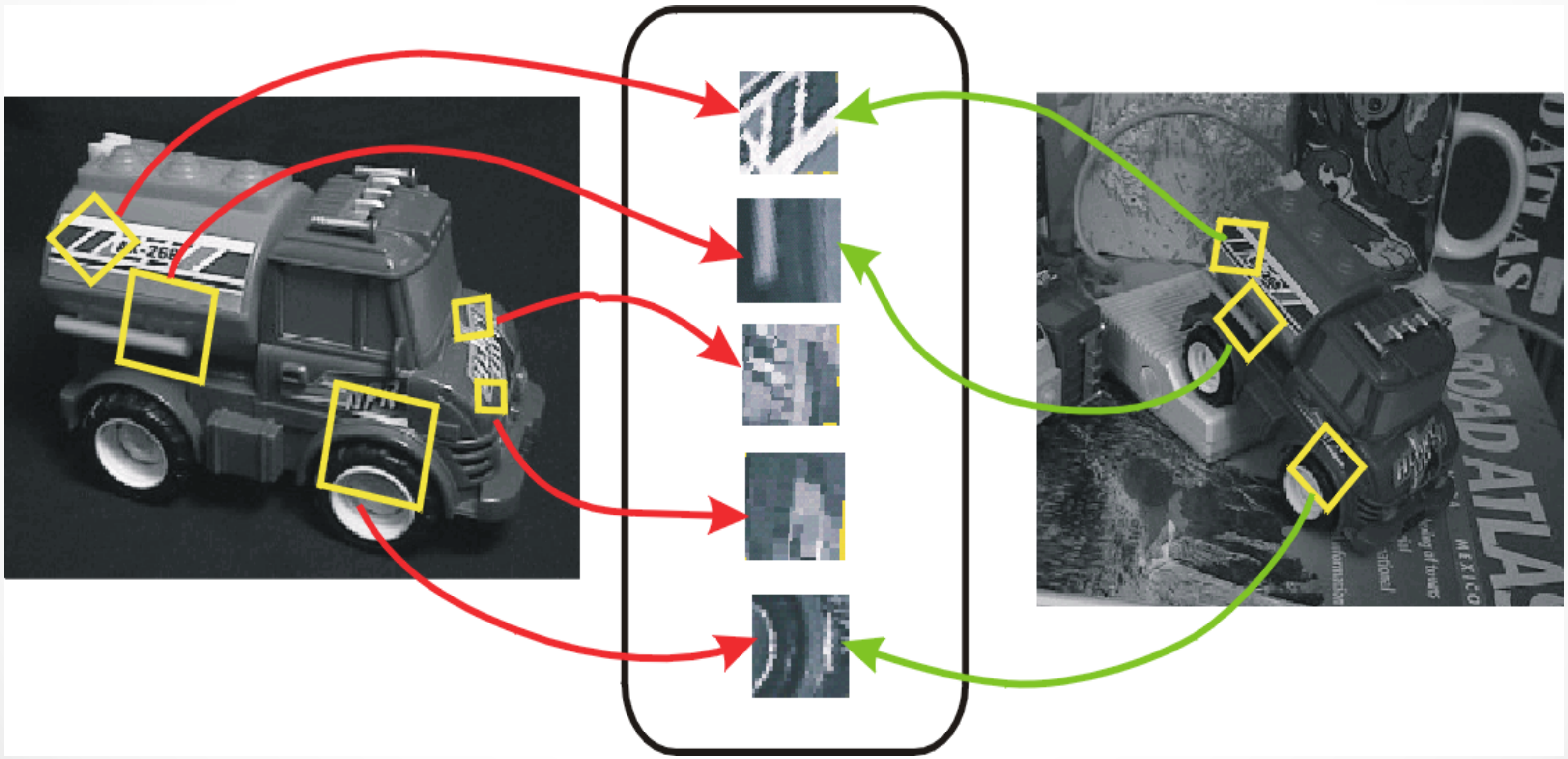
Center for Visual Information Technology

IIIT Hyderabad, INDIA

Invariant Local Features

Find features that are invariant to transformations

- geometric invariance: translation, rotation, scale
- photometric invariance: brightness, exposure, …



Feature Descriptors

Harris features (in red)

- The tops of the horns are detected in both images

Multi-Scale Oriented Patches (MOPS)

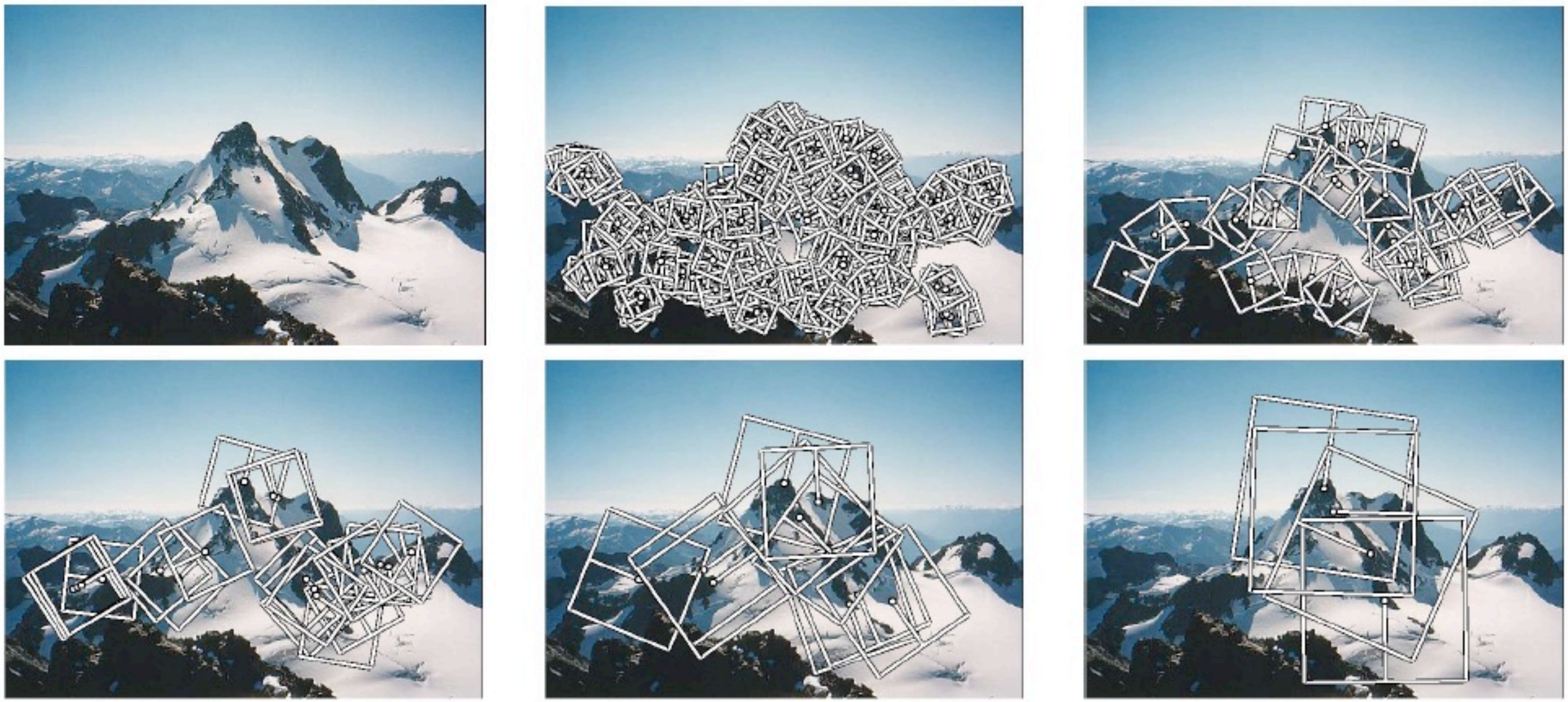
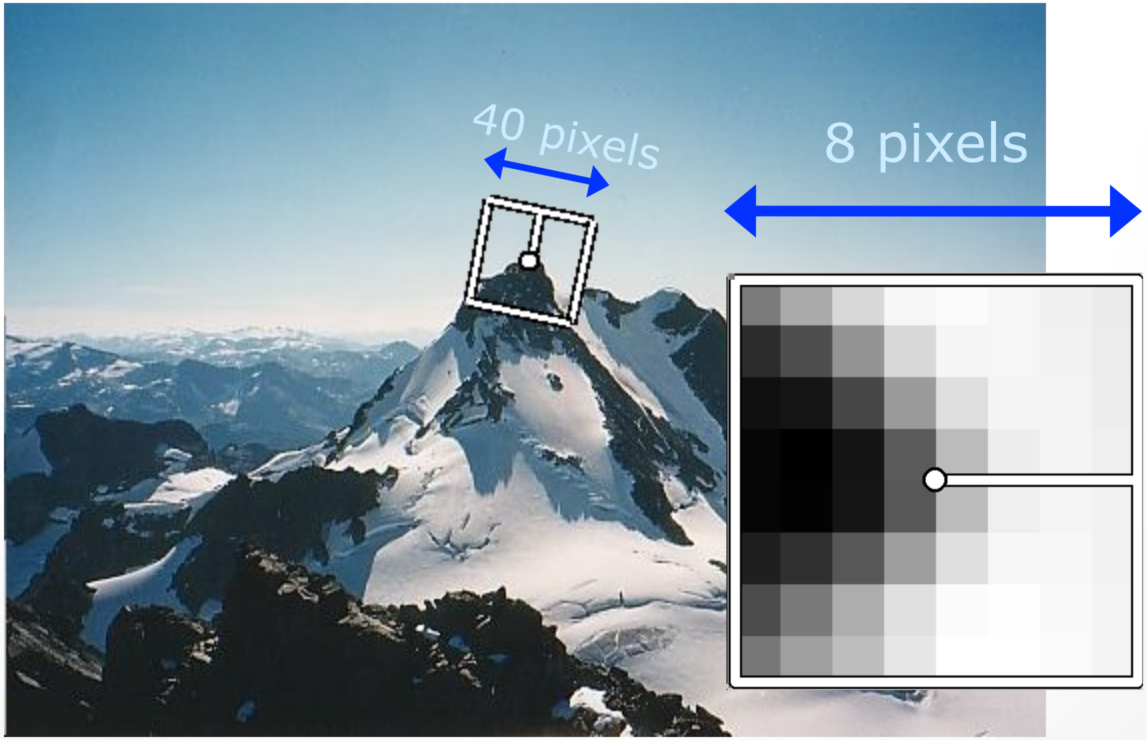


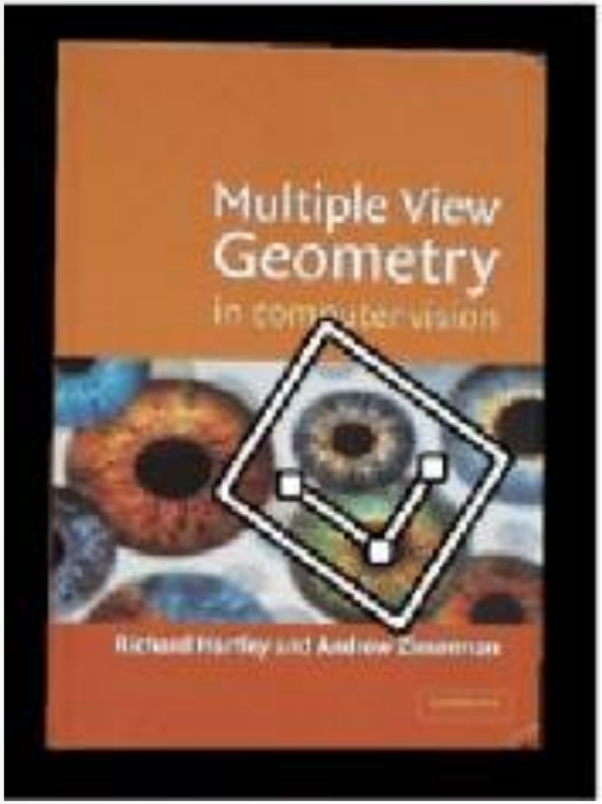
Figure 1. Multi-scale Oriented Patches (MOPS) extracted at five pyramid levels from one of the Matier images. The boxes show the feature orientation and the region from which the descriptor vector is sampled.

MOPS Descriptor Vector

- 8x8 oriented patch
  - Sampled at 5 x scale
- Bias/gain normalisation: $I' = (I - \mu)/\sigma$



Feature Description



Feature Description

Scale Invariant Feature Transform (SIFT)

Slides by Tom Duerig

Types of Invariances

- Illumination

Types of Invariances

- Illumination
- Scale

Types of Invariances

- Illumination
- Scale
- Rotation

Types of Invariances

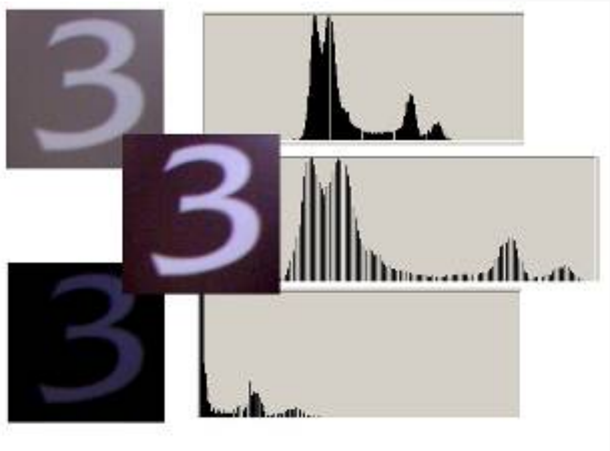
- Illumination
- Scale
- Rotation
- Affine

Types of Invariances

- Illumination
- Scale
- Rotation
- Affine
- Full Perspective

How to Achieve Illumination Invariance?

- The easy way (normalized)
- Difference based metrics (random tree, Haar, and sift)



How to achieve scale invariance

- Pyramids
  - Divide width and height by 2
  - Take average of 4 pixels for each pixel (or Gaussian blur)
  - Repeat until image is tiny
  - Run filter over each size image and hope its robust
- Scale Space (DOG method)

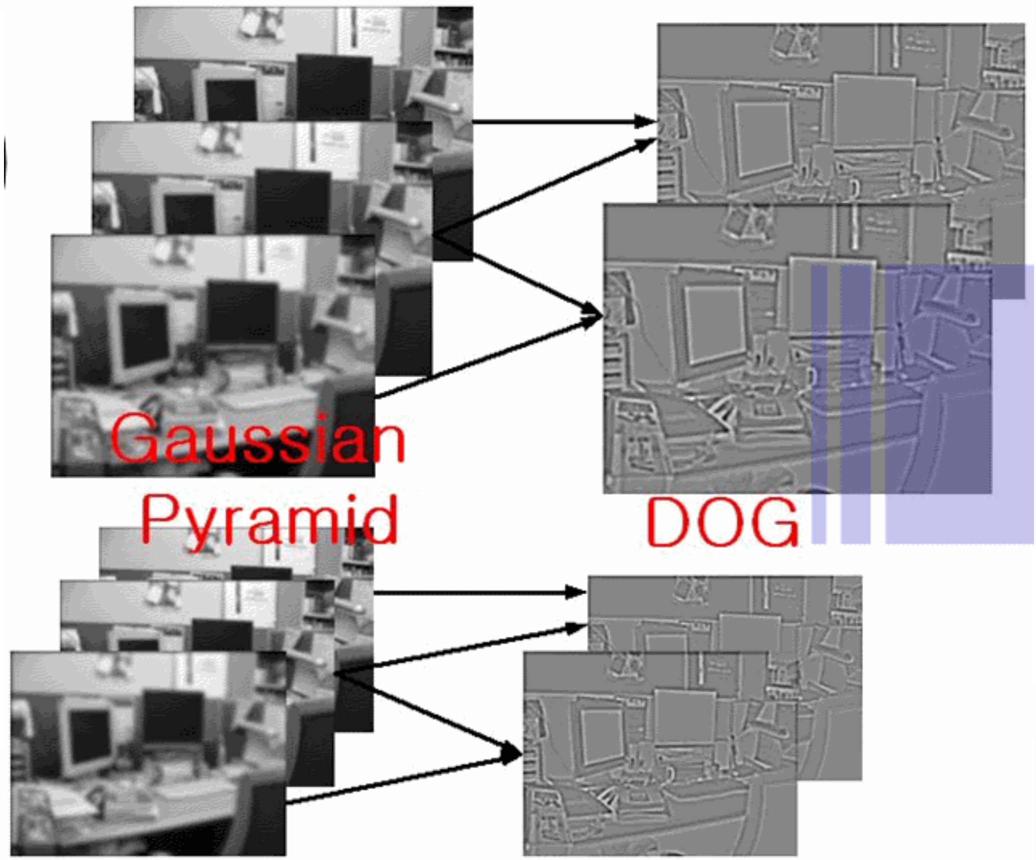
Pyramids


novelties
and
ice cream
novelties
and
ice cream
novelties
and
ice cream

How to achieve scale invariance

- Pyramids
- Scale Space (DOG method)
  - Pyramid but fill gaps with blurred images
  - Like having a nice linear scaling without the expense
  - Take features from differences of these images
  - If the feature is repeatably present in between Difference of Gaussians it is Scale Invariant and we should keep it.

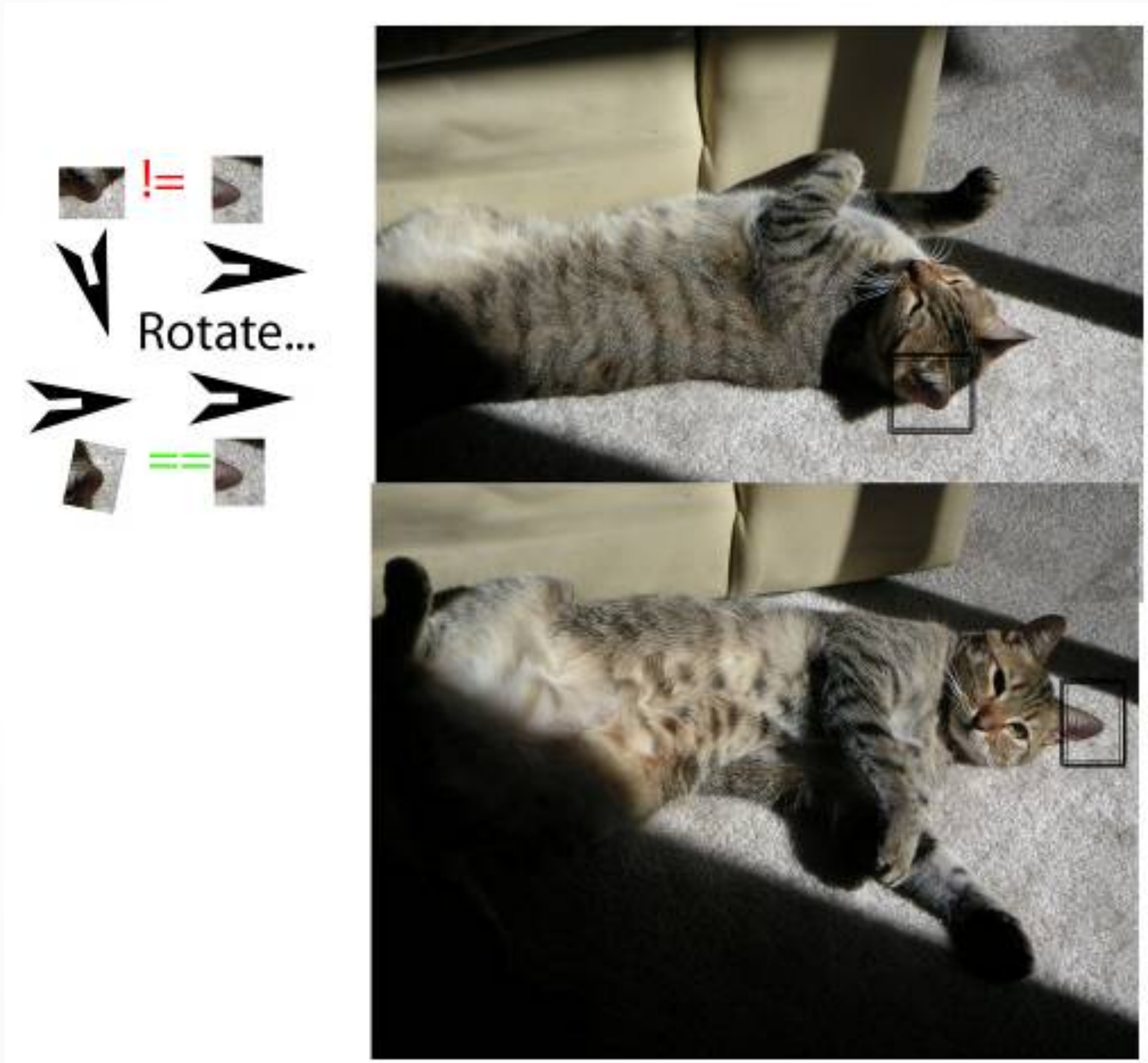
Differences Of Gaussians



Rotation Invariance

- Rotate all features to go the same way in a determined manner
- Take histogram of Gradient directions (36 in paper for 1 every 10 degrees)
- Rotate to most dominant (maybe second if its good enough, sub-Bin accuracy)

Rotation Invariance

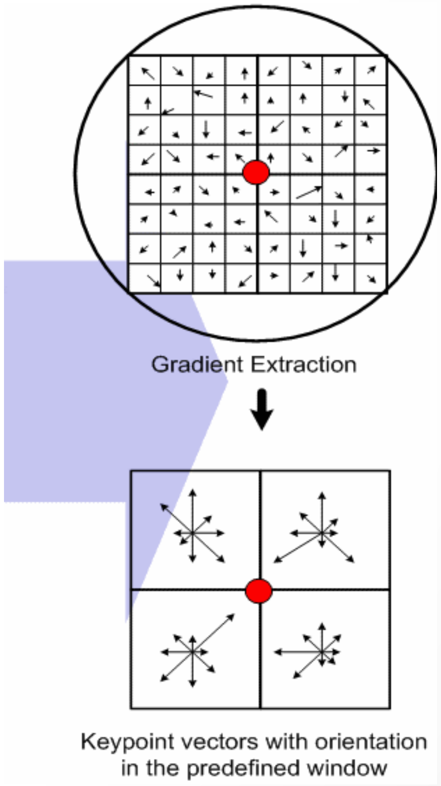


Affine Invariance

- Easy way: Warp your training and hope
- Fancy way: design your feature itself to be robust against affine transformations (SIFT method)

Actual SIFT features

- Remember the gradient histograms we used for rotation invariance?
- Same theory, except keep N^2 histograms (4 shown, 16 used)
- Note, use weighted contributions to avoid edge nastiness



SIFT algorithm overview

- Get tons of points from maxima+minima of DOGS
- Threshold on simple contrast (low contrast is generally less reliable than high for feature points)
- Threshold based on principal curvatures (technical term is linyness)

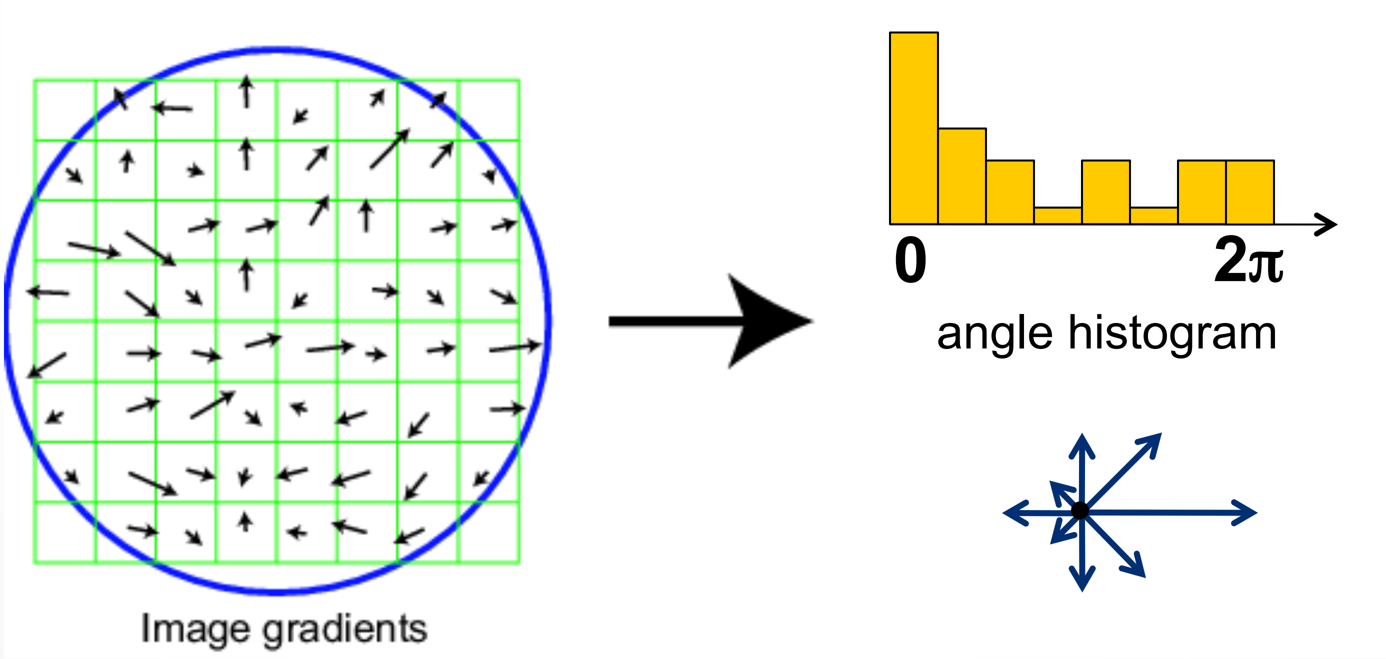
SIFT algorithm overview

- Gradient, histogram, Rotate
- Take old gradient, histogram regions using Gaussian weighting
- Hand off beautiful robust feature to someone who cares (i.e. object recognizer, navigation software (SFM), or stereo matching algorithm)

Scale Invariant Feature Transform

Basic idea:

- Take 16x16 square window around detected feature
- Compute edge orientation (angle of the gradient - 90°) for each pixel
- Throw out weak edges (threshold gradient magnitude)
- Create histogram of surviving edge orientations

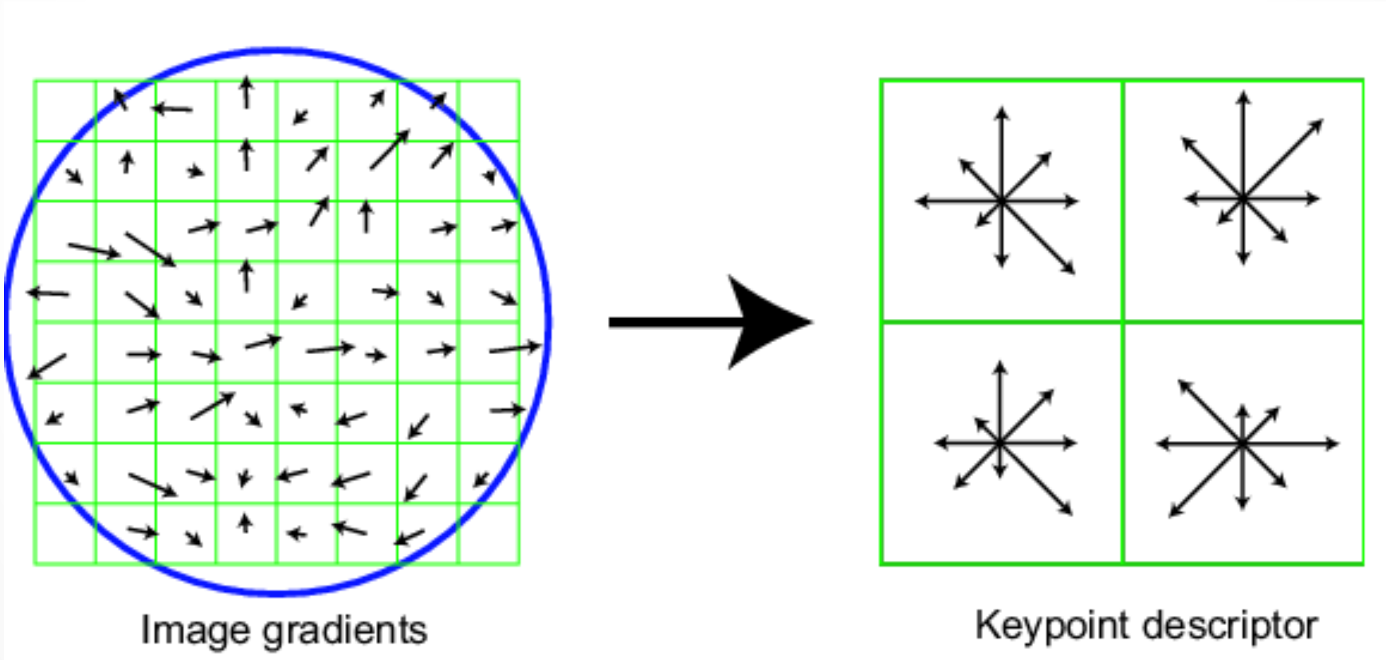


Adapted from slide by David Lowe

SIFT descriptor

Full version

- Divide the 16x16 window into a 4x4 grid of cells (2x2 case shown below)
- Compute an orientation histogram for each cell
- 16 cells * 8 orientations = 128 dimensional descriptor



Adapted from slide by David Lowe

Actual SIFT stage output

Keypoint detection

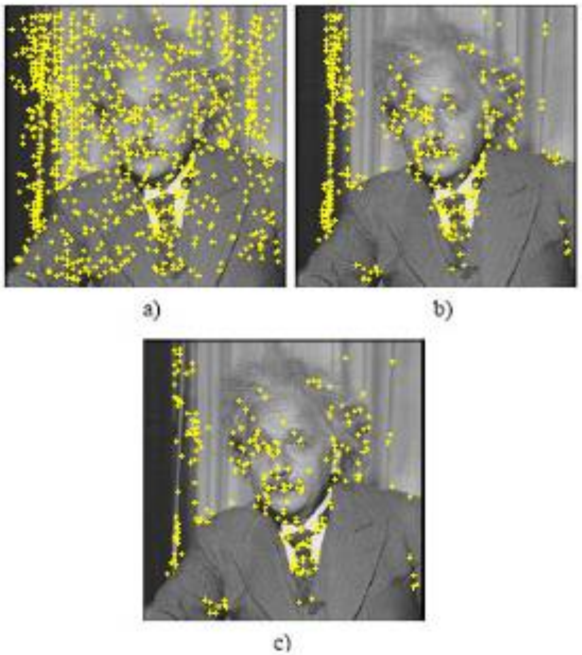


Figure 5: a) Maxima of DoG across scales. b) Remaining keypoints after removal of low contrast points. C) Remaining keypoints after removal of edge responses (bottom).

Final keypoints with selected orientation and scale

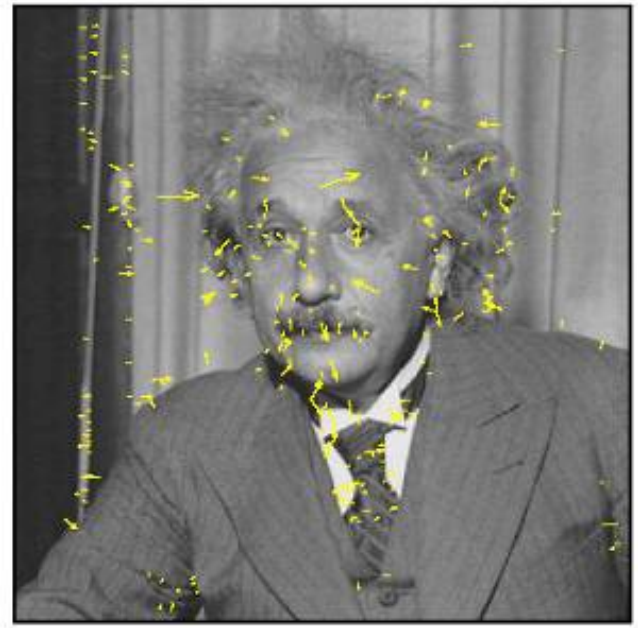
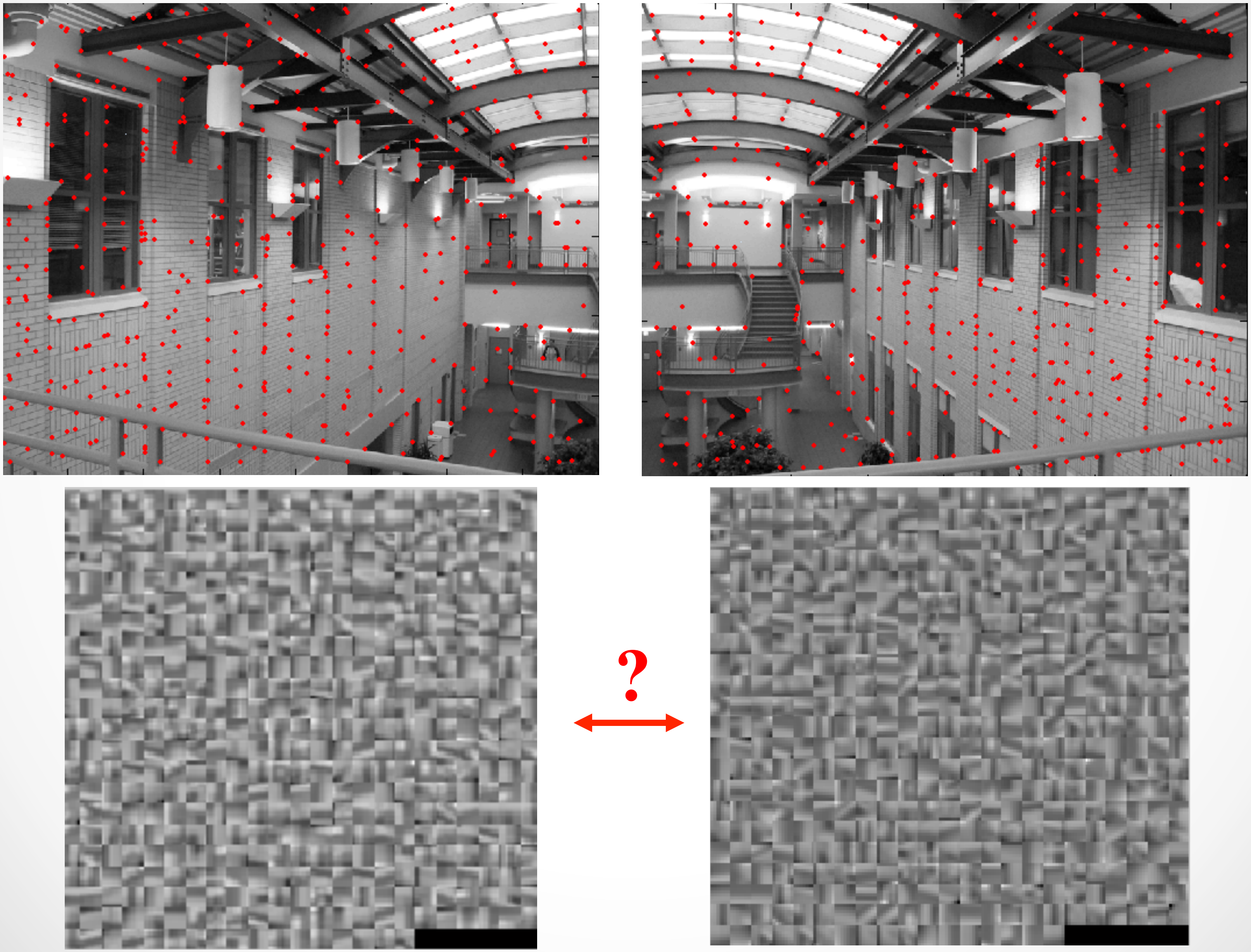


Figure 6: Extracted keypoints, arrows indicate scale and orientation.

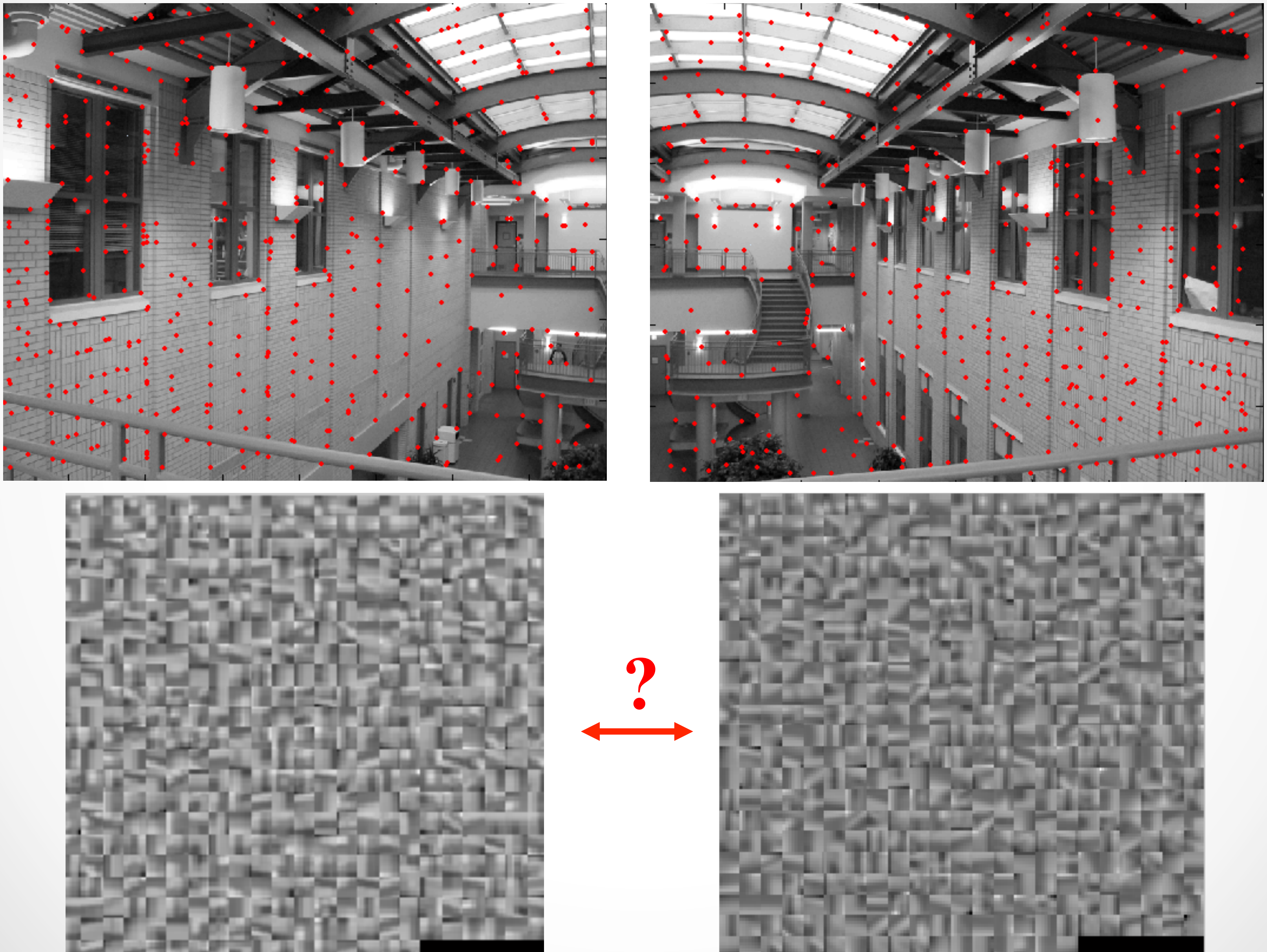
Feature matching



Feature matching

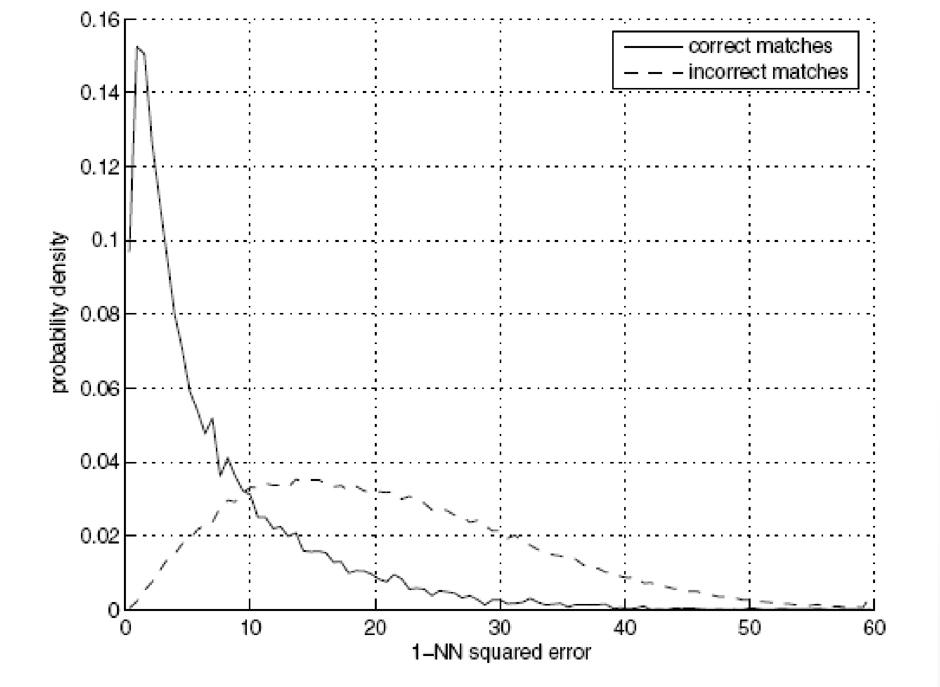
- Exhaustive search
  - for each feature in one image, look at *all* the other features in the other image(s)
- Hashing
  - compute a short descriptor from each feature vector, or hash longer descriptors (randomly)
- Nearest neighbor techniques
  - *kd*-trees and their variants

What about outliers?

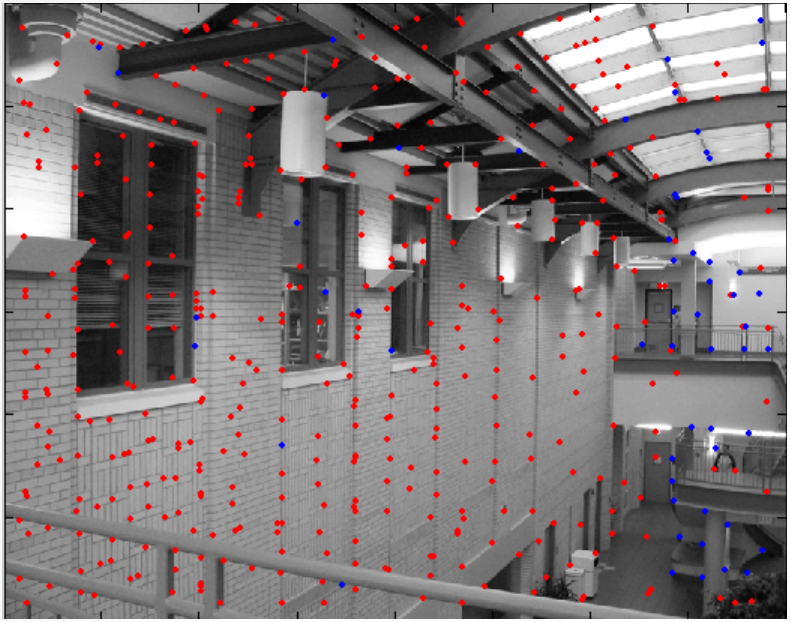


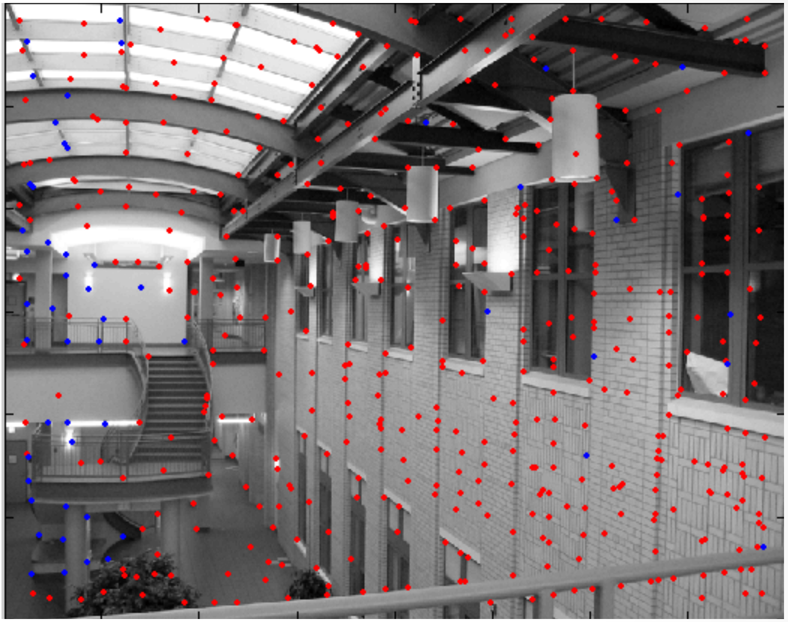
Feature-space outlier rejection

- Let us not match all features, but only these that have “similar enough” matches?
- How can we do it?
  - SSD(patch1,patch2) < threshold
  - How to set threshold?



Feature-space outliner rejection





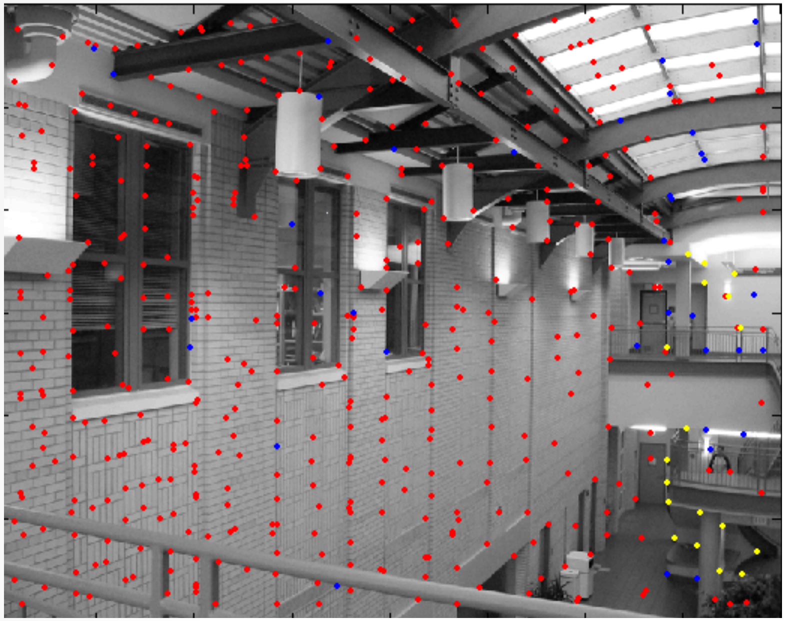
- Can we now compute H from the blue points?
  - No! Still too many outliers…
  - What can we do?

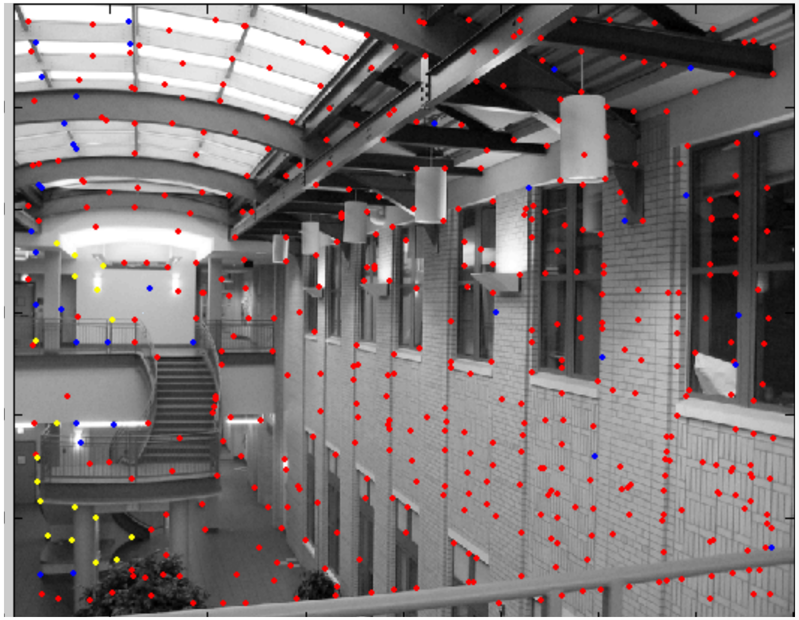
RANSAC for estimating homography

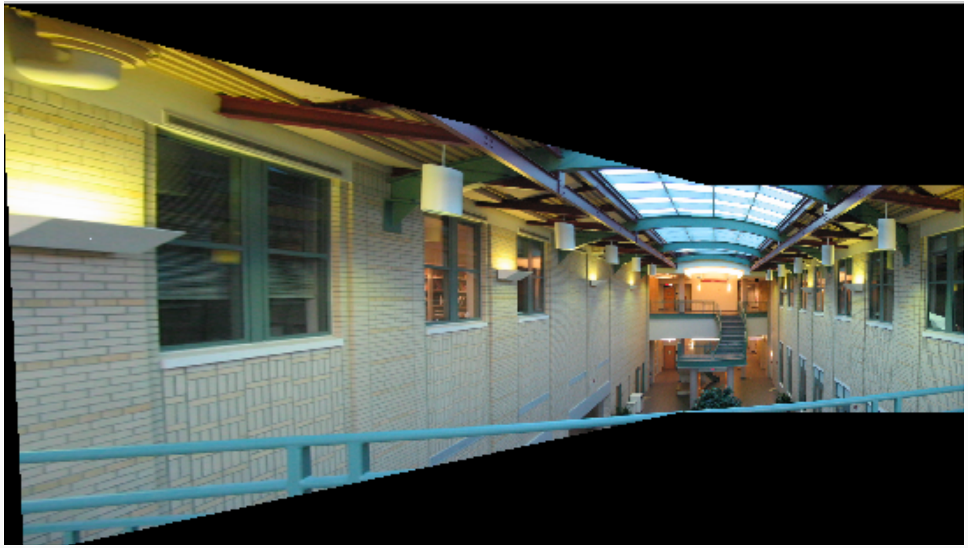
- RANSAC loop:

1. Select four feature pairs (at random)
2. Compute homography H (exact)
3. Compute *inliers* where $\mathrm{SSD}(p_i', \boldsymbol{H}\, p_i) < \varepsilon$
4. Keep largest set of inliers
5. Re-compute least-squares H estimate on all of the inliers

RANSAC







Mosaicing Example

 . . .  . . .  . . . 

Properties of SIFT

Extraordinarily robust matching technique

- Can handle changes in viewpoint
  - Up to about 60 degree 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

Stereo Matching

- Constrained by the Fundamental Matrix
- What is the nature of disparity/depth?

MSER (a one minute survey)

- Maximally Stable Extremal Regions
- Go through thresholds, grab regions which stay nearly the same through a wide range of thresholds (connected components then bounding ellipses)
- Keep those regions descriptors as features
- As regions are illumination based warp ellipse to circle for affine invariance

Actual MSER output

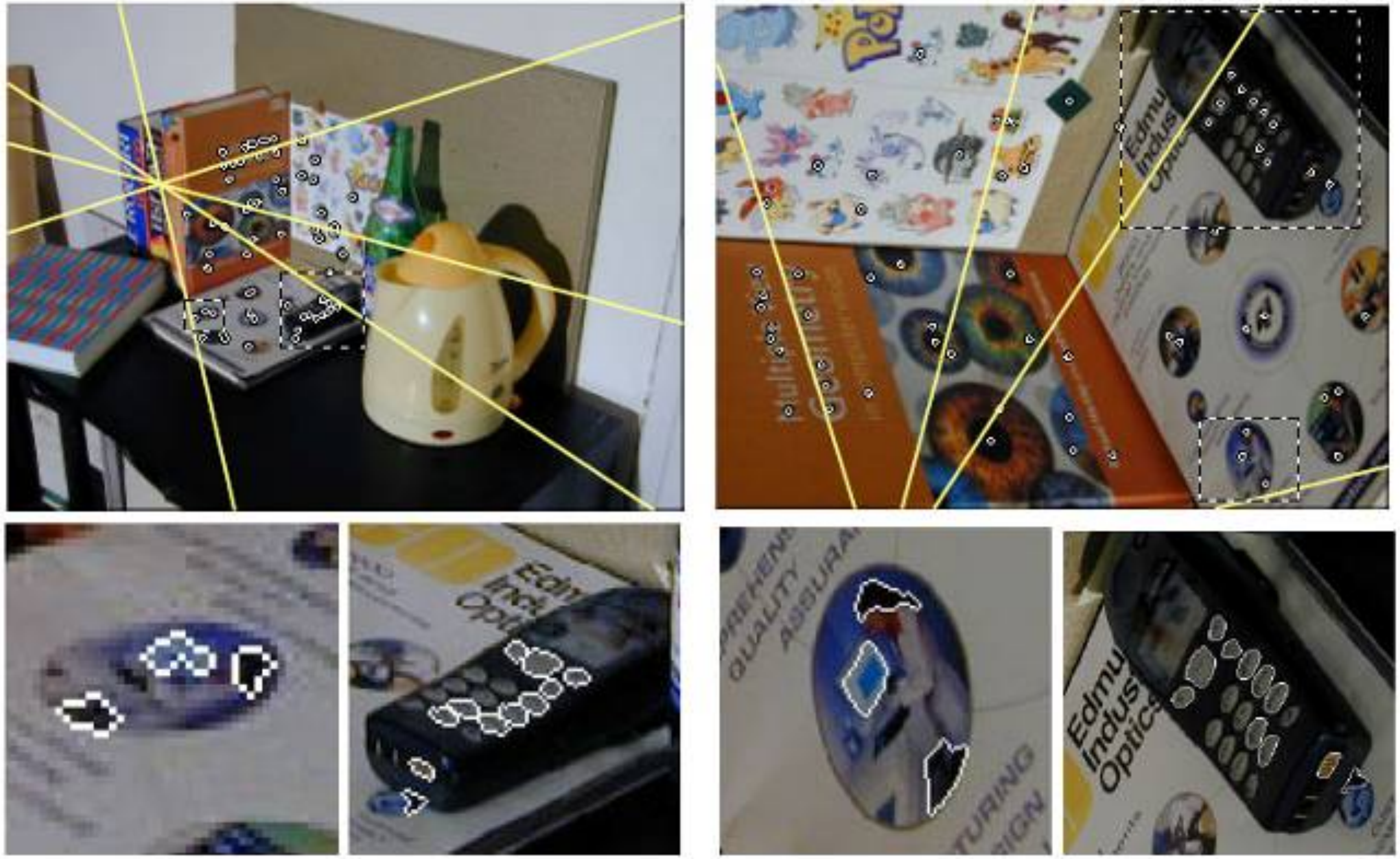


Figure 1: BOOKSHELF: Estimated epipolar geometry on indoor scene with significant scale change. In the cutouts the change in the resolution of detected DRs is clearly visible.

Initial Image

Threshold 1

Threshold 2

Threshold 3

Ears and Square
(note: not actual algorithm output)

How to use these features?

- Distance could be L2 norm on histograms
- Match by (nearest neighbor distance)/(2nd nearest neighbor distance) ratio
- Object recognize with Hough of pose of points (ie, these three should be in line on object, gee… they are that’s the object all right)

But where does the magic end?

- SIFT is relatively expensive (computationally) and copyrighted (the other type of expensive)
- MSER does not work well with images with any motion blur
- Interesting alternatives:
  - GLOH (Gradient Location and Orientation Histogram)
    - larger initial descriptor + PCA
  - SURF (Speeded Up Robust Features)
    - possibly faster AND more robust?

Credits + References

- MSER image from J. Matas, O. Chum, M. Urban, T. Pajdla "Robust Wide Baseline Stereo from Maximally Stable Extremal Regions", BMVC, 2002
- SIFT keypoint images from F. Estrada & A Jepson & D. Fleet 's SIFT tutorial, 2004
- GLOH "A performance evaluation of local descriptors", Krystian Mikolajczyk and Cordelia Schmid, IEEE tran. On Pattern Analysis and Machine Intelligence, pp 1615—1630, 2005
- SURF "SURF: Speeded Up Robust Features", Herbert Bay, Tinne Tuytelaars and Luc Van Gool Proceedings of the 9th European Conference on Computer Vision 2006