Robert Collins
CSE486, Penn State

Lecture 31:
Object Recognition: SIFT Keys

Motivation

- Want to recognize a known objects from unknown viewpoints.

database of models

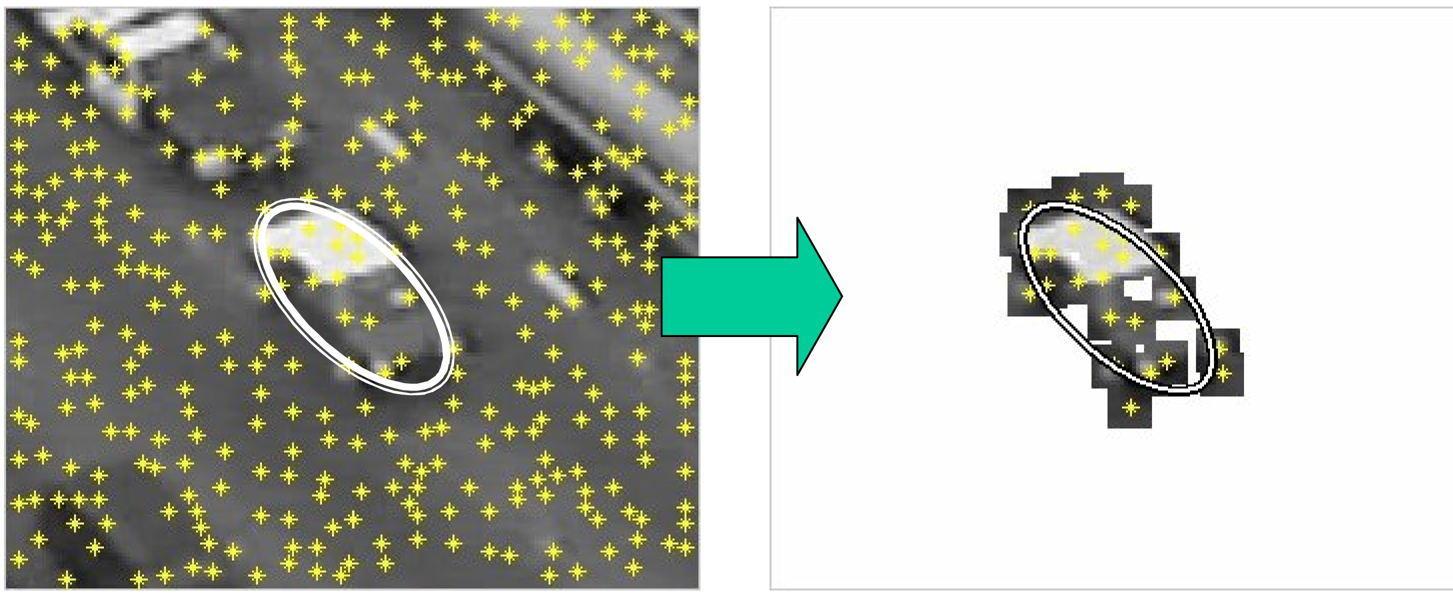
find them in an image

Local Feature based Approaches

- Represent appearance of object by little intensity/feature patches.
- Try to match patches from object to image
- Geometrically consistent matches tell you the location and pose of the object

Simple Example

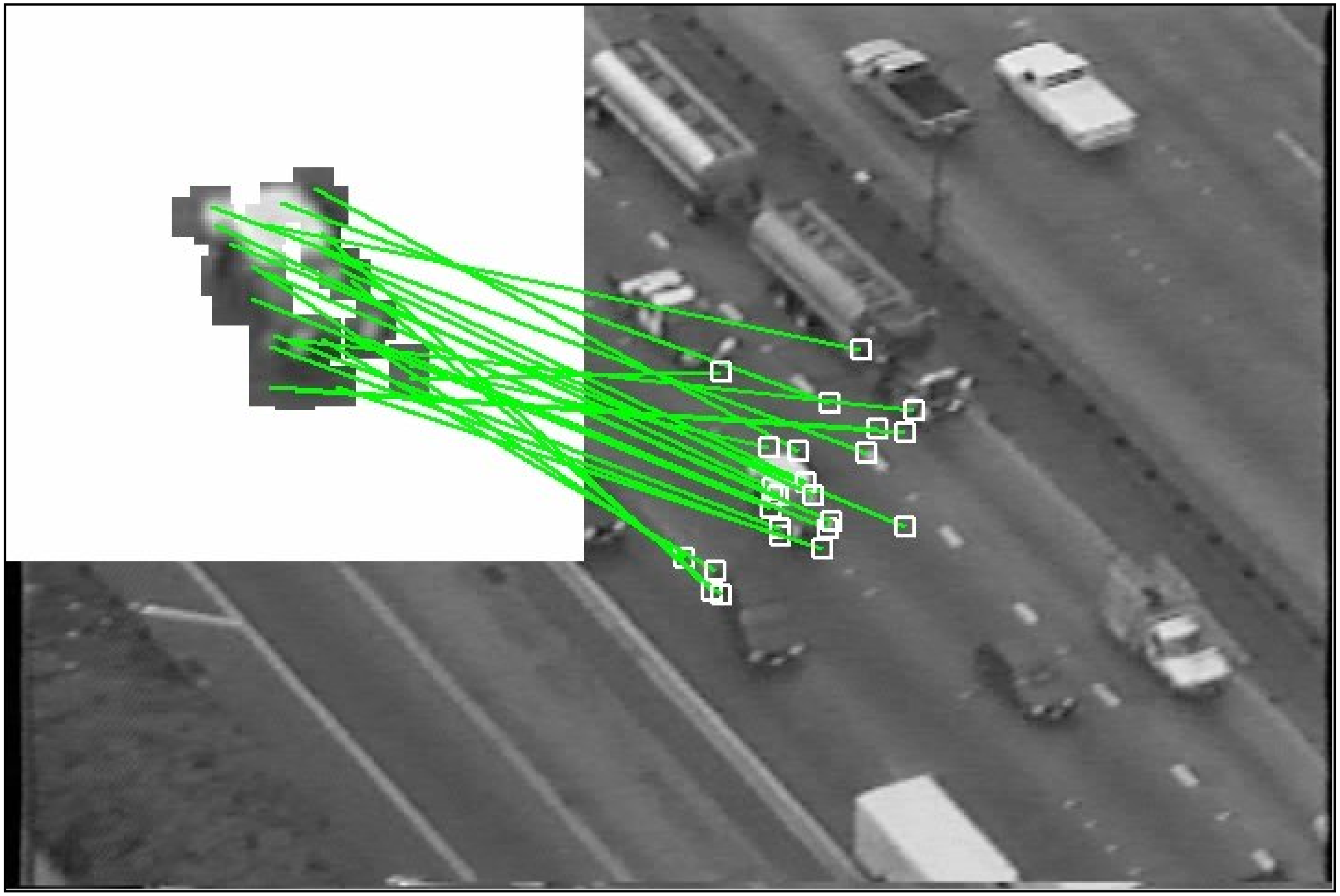
- Represent object by set of 11x11 intensity templates extracted around Harris corners.



harris corners **our object “model”**

Simple Example

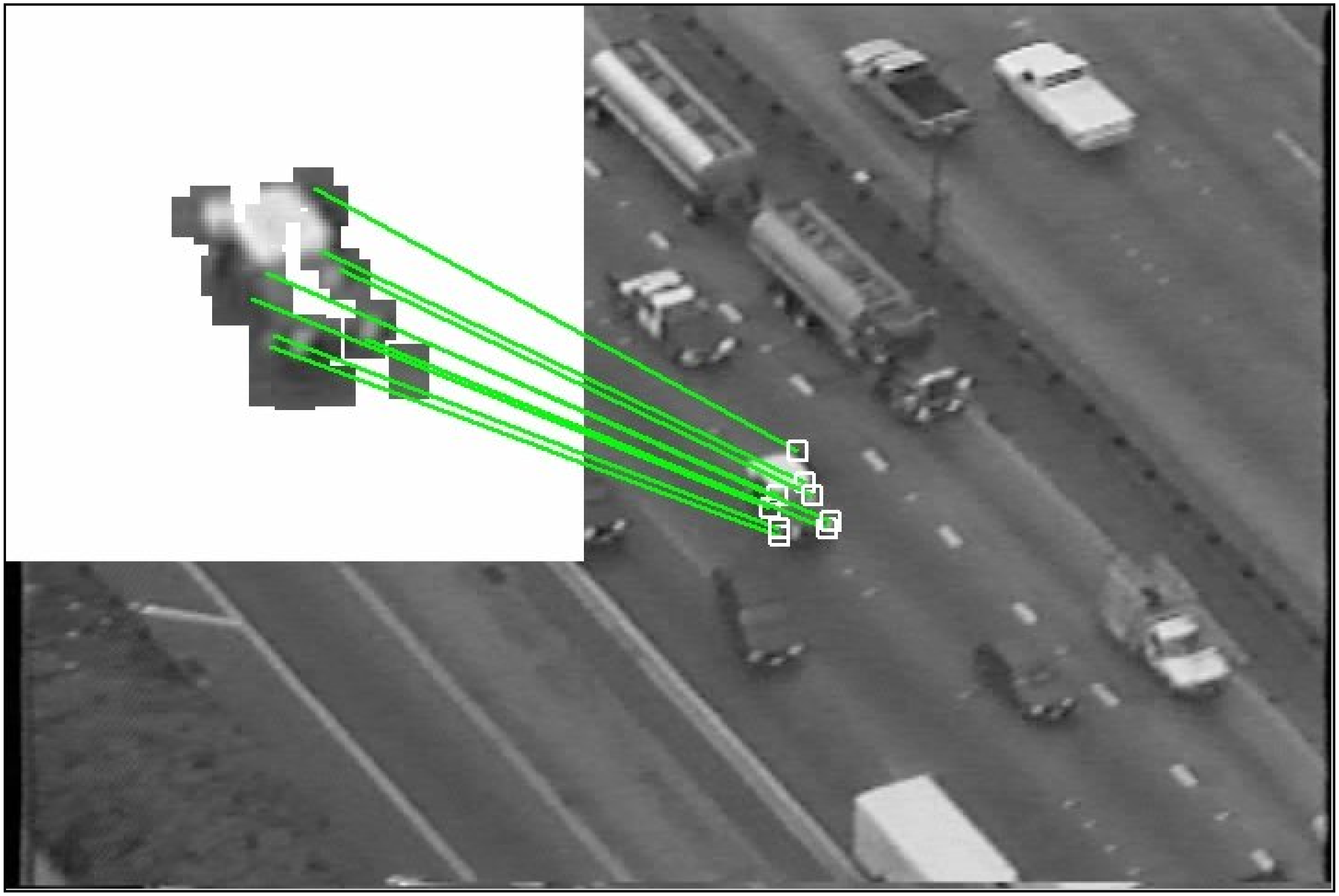
- Match patches to new image using NCC.



Robert Collins
CSE486, Penn State

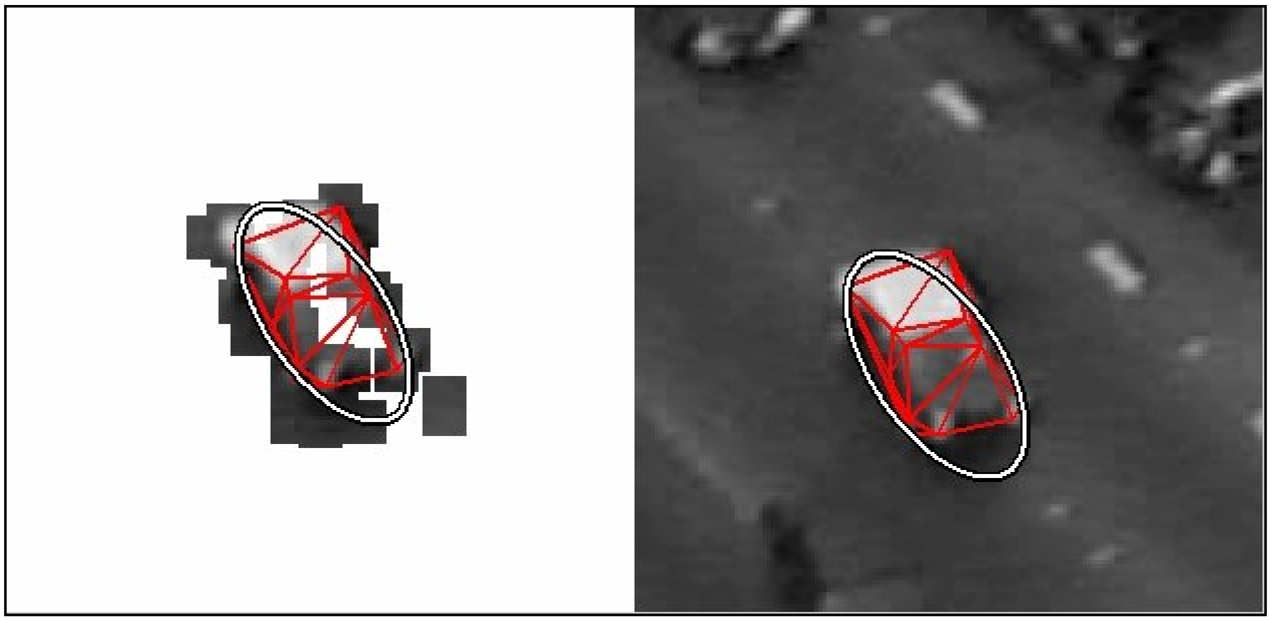
Simple Example

- Find matches consistent with affine transformation using RANSAC



Simple Example

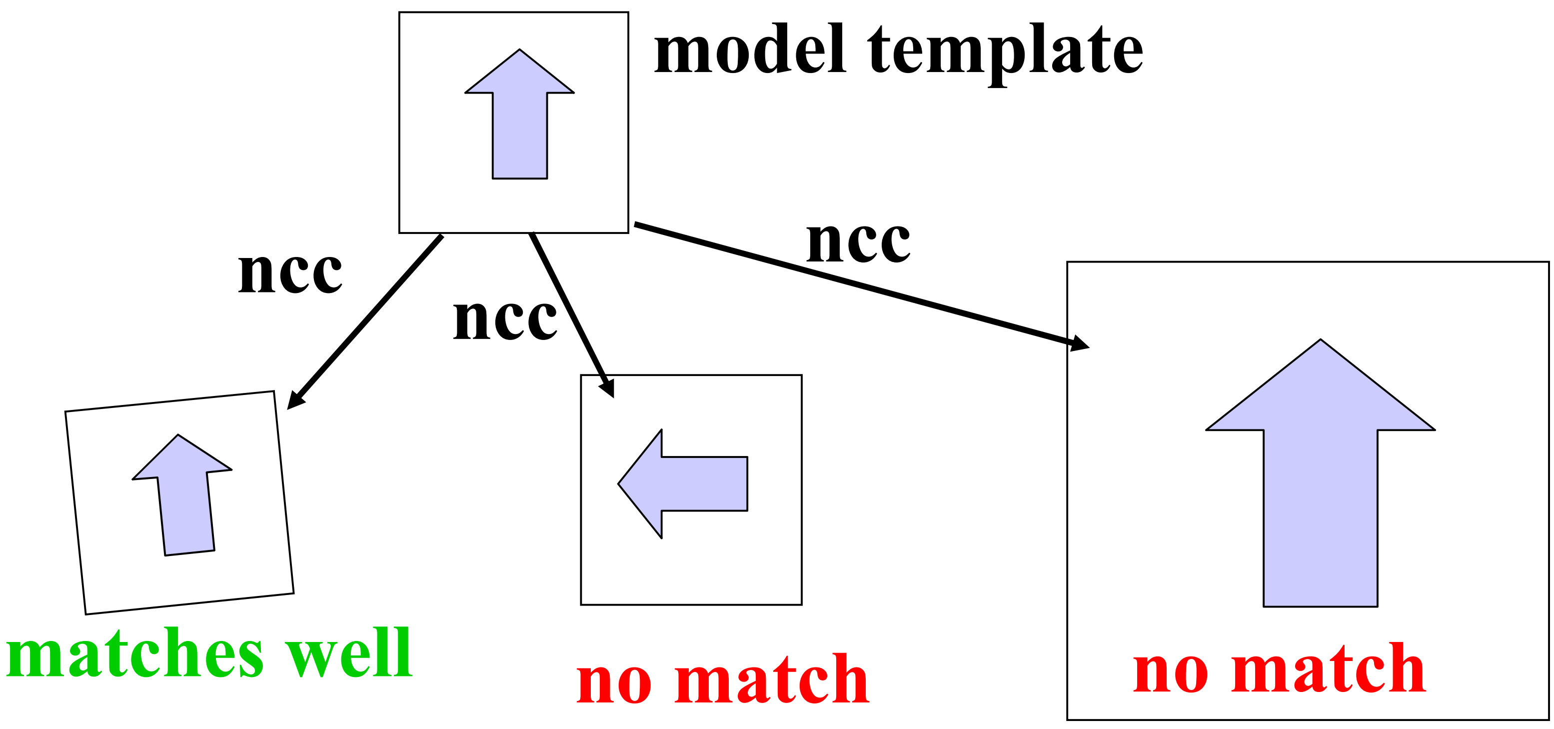
- Inlier matches let you solve for location and pose of object in the image.



Robert Collins
CSE486, Penn State

Problem with Simple Example

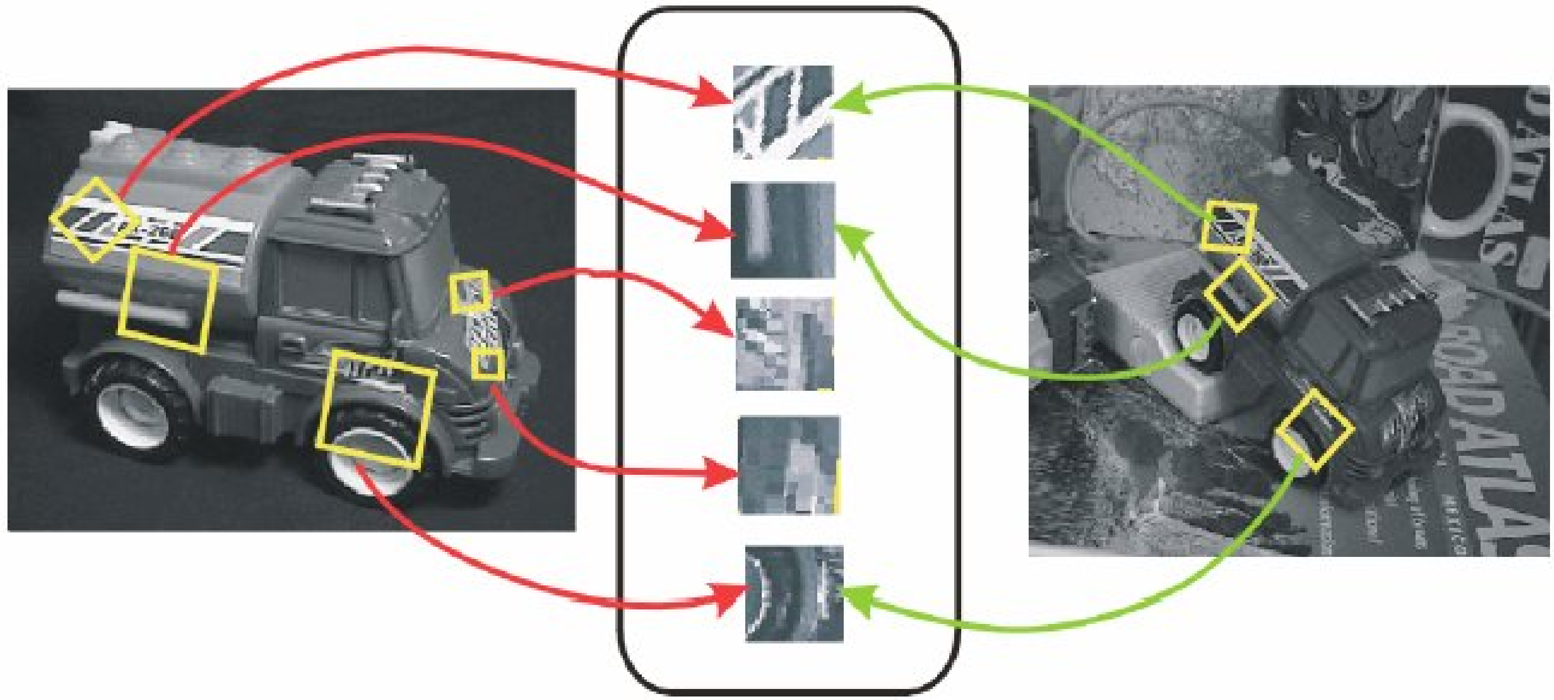
Using NCC to match intensity patches puts restrictions on the amount of overall rotation and scaling allowed between the model and the image appearance.



Robert Collins
CSE486, Penn State

More General : SIFT Keys

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



SIFT Features

David G. Lowe, **"Distinctive image features from scale-invariant keypoints,"** *International Journal of Computer Vision,* 60, 2 (2004), pp. 91-110.

Robert Collins
CSE486, Penn State

SIFT Keys: General Idea

- Reliably extract same image points regardless of new magnification and rotation of the image.
- Normalize image patches, extract feature vector
- Match feature vectors using correlation

SIFT Keys: General Idea

Want to detect/match same features regardless of

Translation : easy, almost every feature extraction and correlation matching algorithm in vision is translation invariant

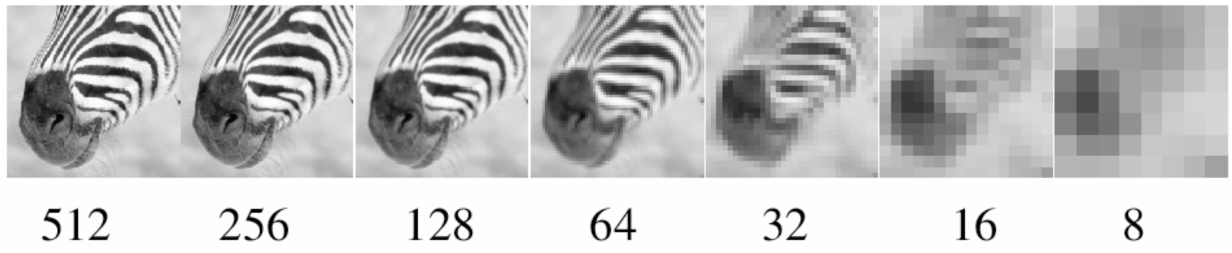
Rotation : harder. Guess a canonical orientation for each patch from local gradients

Scaling : hardest of all. Create a multi-scale representation of the image and appeal to scale space theory to determine correct scale at each point.

Robert Collins
CSE486, Penn State

Recall: Scale Space

Basic idea: different scales are appropriate for describing different objects in the image, and we may not know the correct scale/size ahead of time.



Scale Selection

Scale Selection Principle (T. Lindeberg):

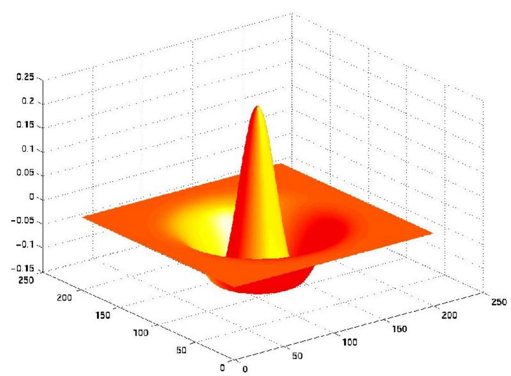
In the absence of other evidence, assume that a scale level, at which some (possibly non-linear) combination of normalized derivatives assumes a local maximum over scales, can be treated as reflecting a characteristic length of a corresponding structure in the data.

What are normalized derivatives?:

$$\sigma^{n+m} \frac{\partial^{(n+m)} f}{\partial x^n \partial x^m}$$

Example (using 2nd order derivatives):

$$\sigma^2 \nabla^2 f = \sigma^2 \underbrace{\left(\left(\frac{\partial^2 f}{\partial x^2} \right) + \left(\frac{\partial^2 f}{\partial y^2} \right) \right)}_{\text{"Laplacian" operator.}}$$



“Laplacian” operator.

Robert Collins
CSE486, Penn State

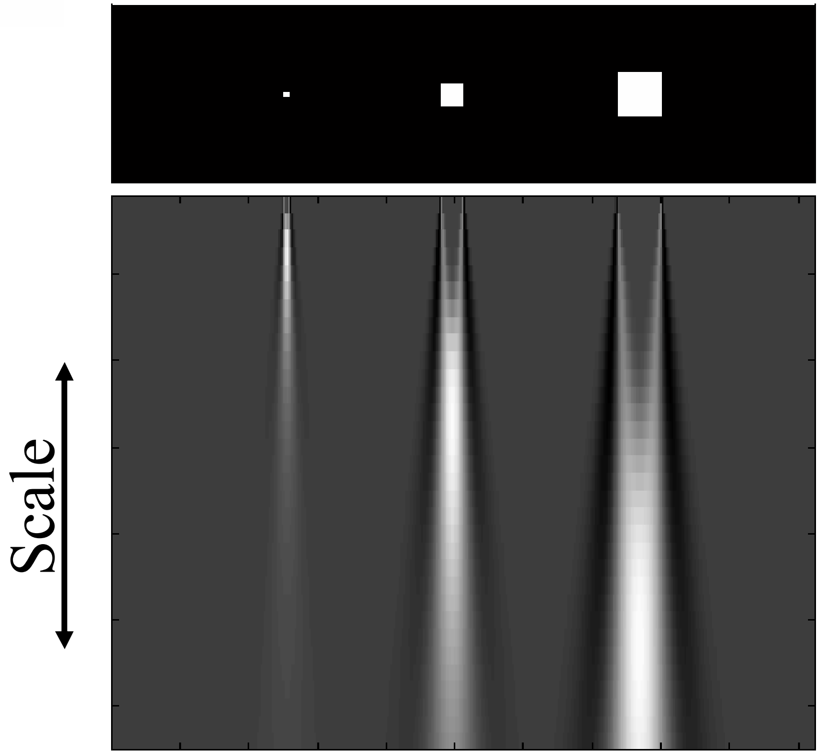
Local Scale Space Maxima

Lindeberg proposes that the natural scale for describing a feature is the scale at which a normalized derivative for detecting that feature achieves a local maximum both spatially and in scale.

$$\begin{cases} (\nabla(\mathcal{D}_{norm}L))(x_0;\ t_0) = 0, \\ (\partial_t(\mathcal{D}_{norm}L))(x_0;\ t_0) = 0. \end{cases}$$

DnormL is the DoG operator, in this case.

Example for blob detection



Robert Collins
CSE486, Penn State

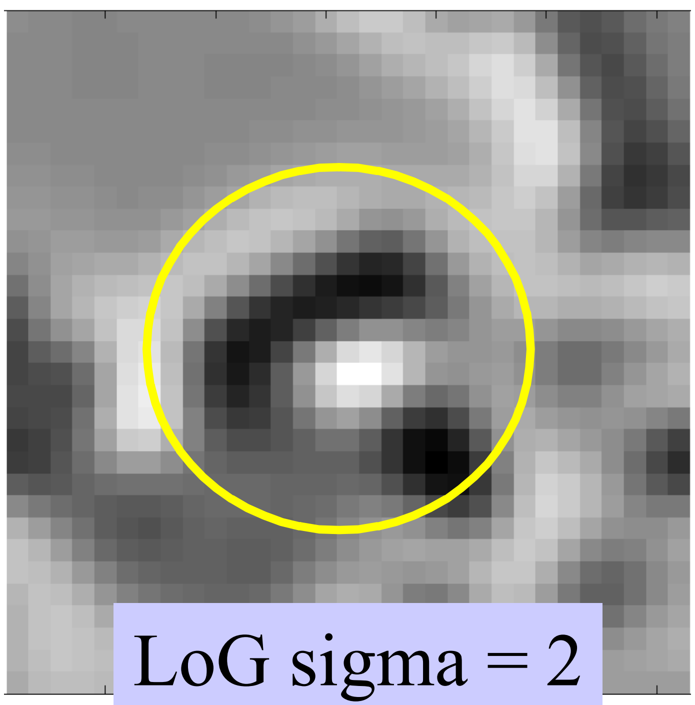
Recall: LoG Blob Finding

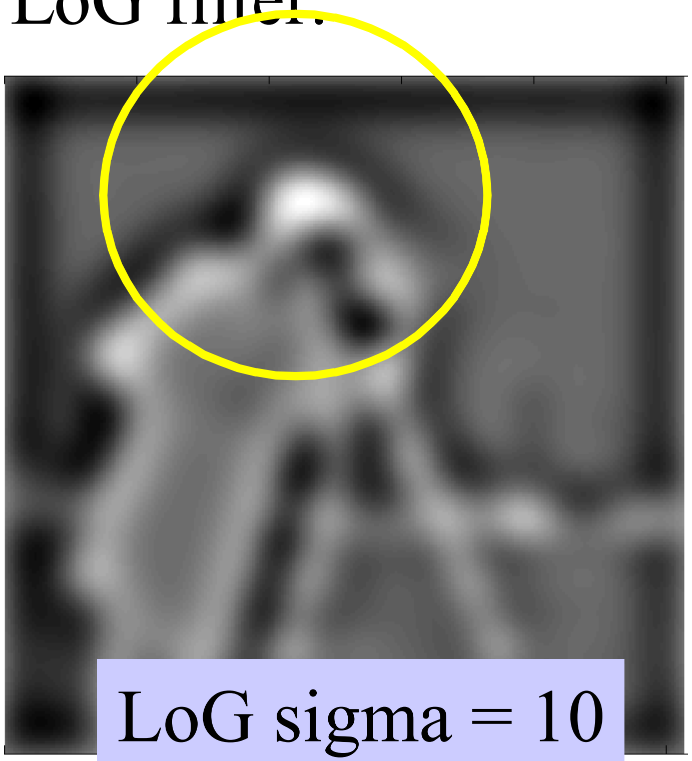
LoG filter extrema locates “blobs”

maxima = dark blobs on light background

minima = light blobs on dark background

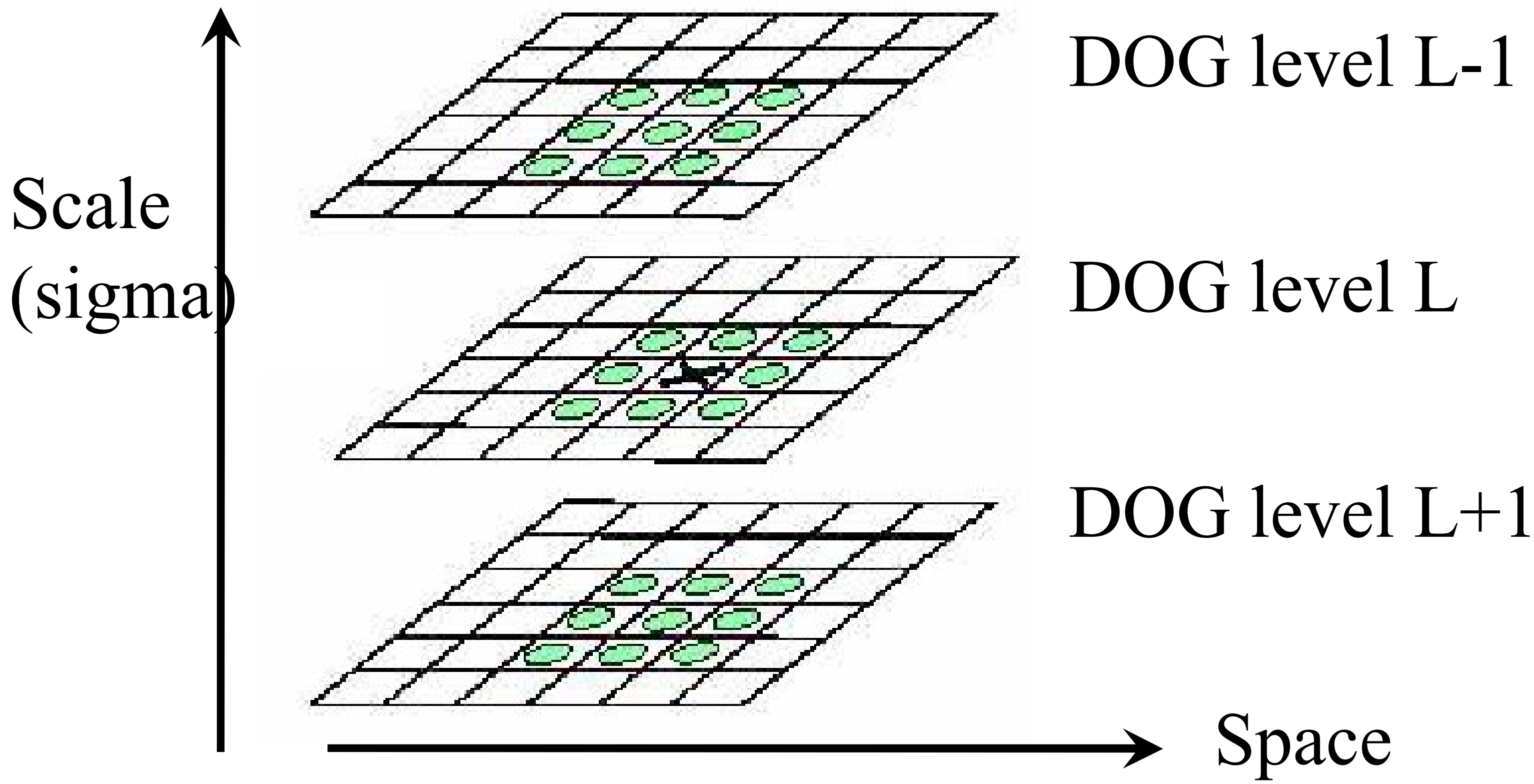
Scale of blob (size ; radius in pixels) is determined by the sigma parameter of the LoG filter.





Robert Collins
CSE486, Penn State

Extrema in Space and Scale



Hint: when finding maxima or minima at level L, use DownSample or UpSample as necessary to make DOG images at level L-1 and L+1 the same size as L.

SIFT Keys: General Idea

Want to detect/match same features regardless of

Translation : easy, almost every feature extraction and correlation matching algorithm in vision is translation invariant

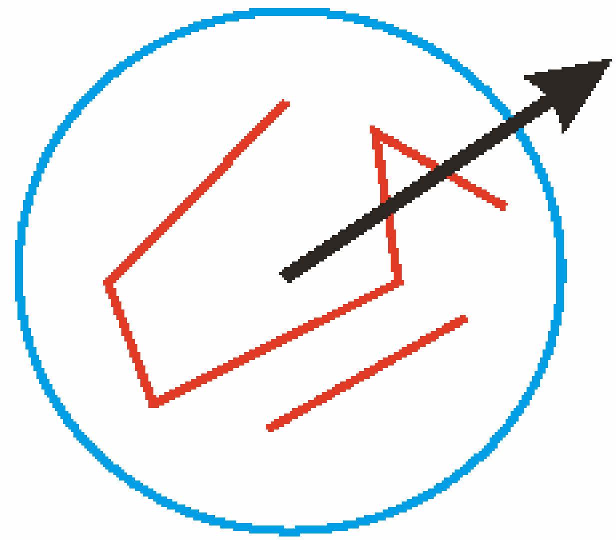
Rotation : harder. Guess a canonical orientation for each patch from local gradients

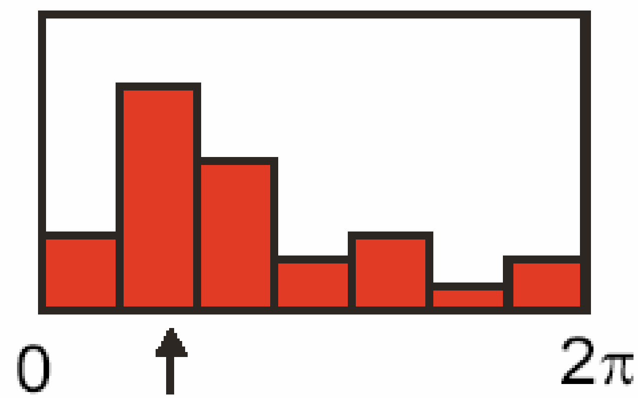
Scaling : hardest of all. Create a multi-scale representation of the image and appeal to scale space theory to determine correct scale at each point.

Sift Key Steps

Select canonical orientation

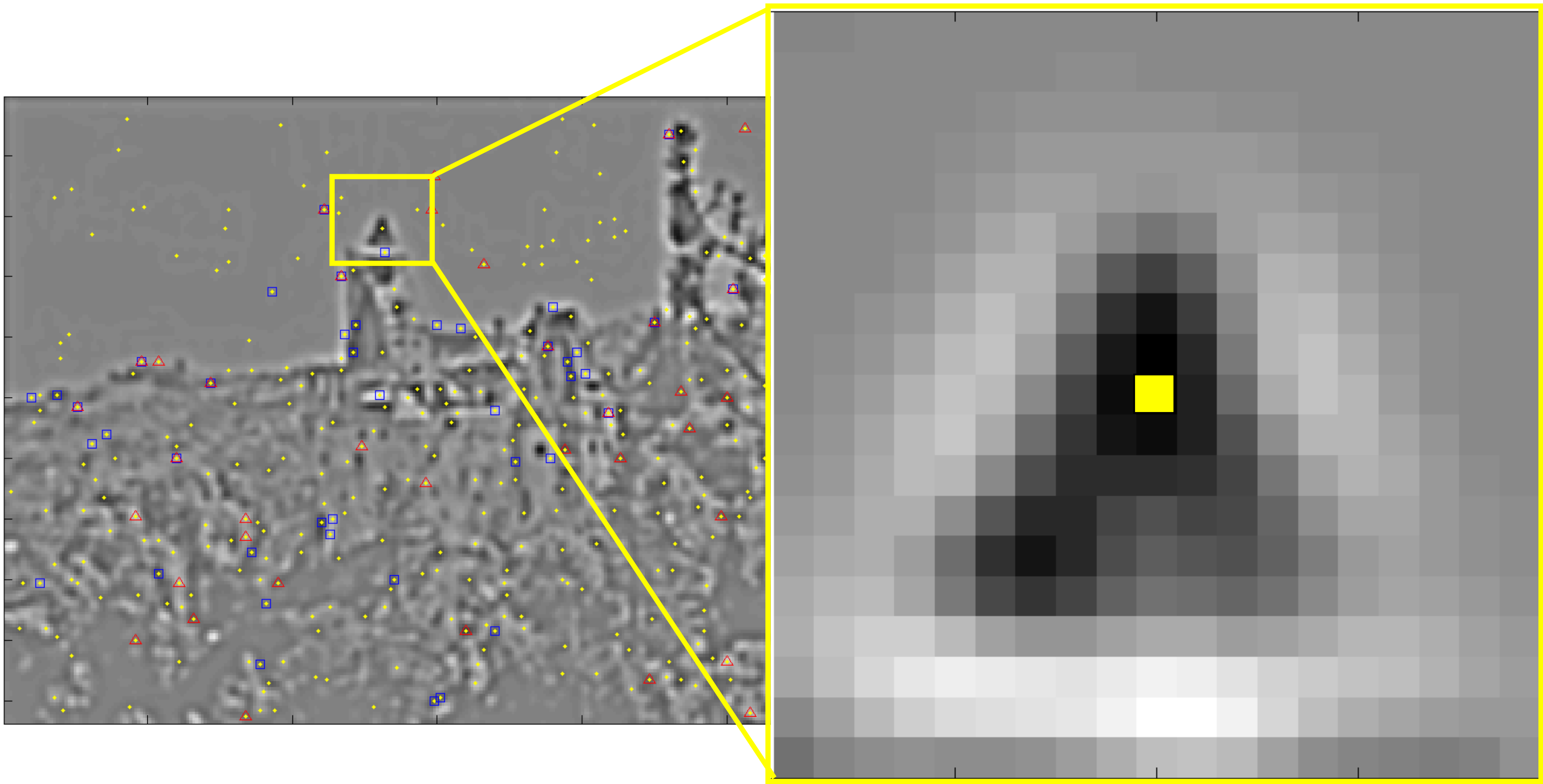
- Create histogram of local gradient directions computed at selected scale
- Assign canonical orientation at peak of smoothed histogram
- Each key specifies stable 2D coordinates (x, y, scale, orientation)





Robert Collins
CSE486, Penn State

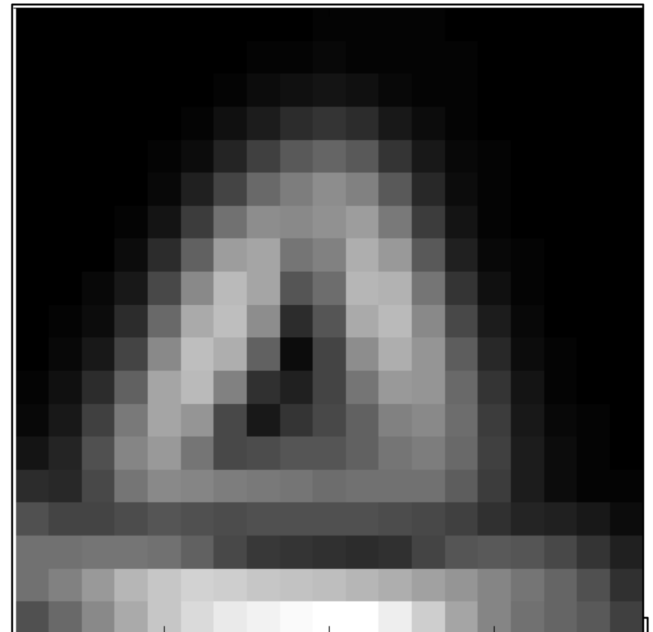
Example

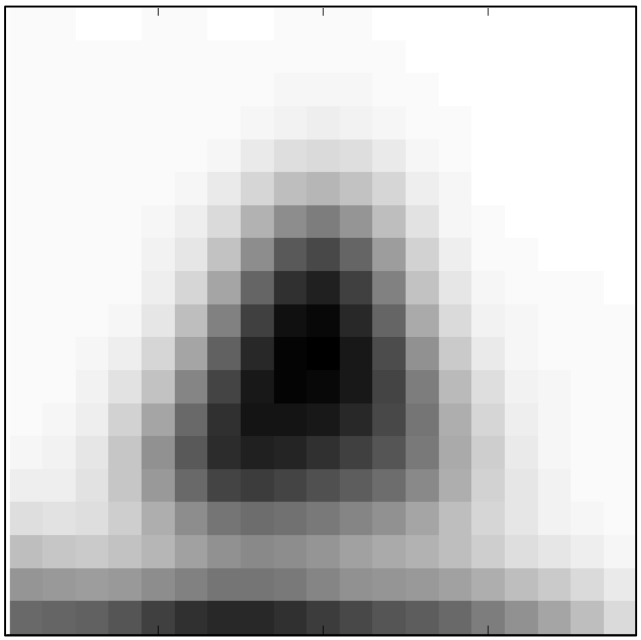


- Keypoint location = extrema location
- Keypoint scale is scale of the DOG image

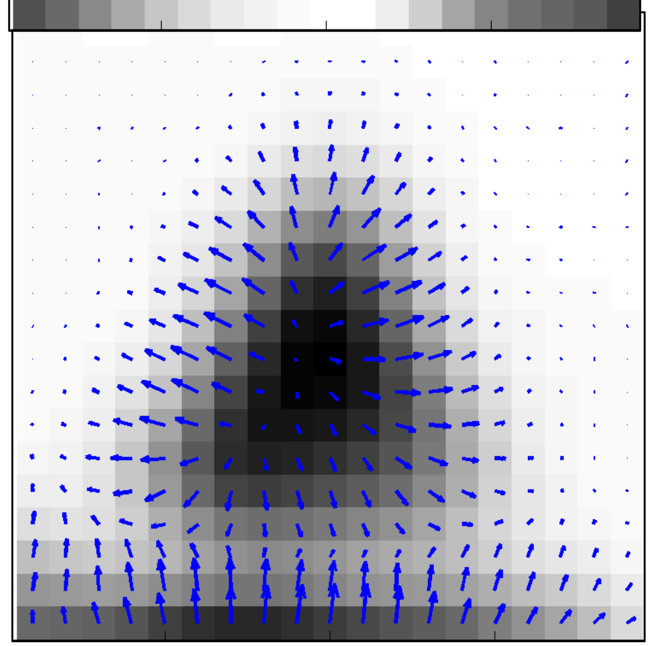
Example (continued)

gradient magnitude



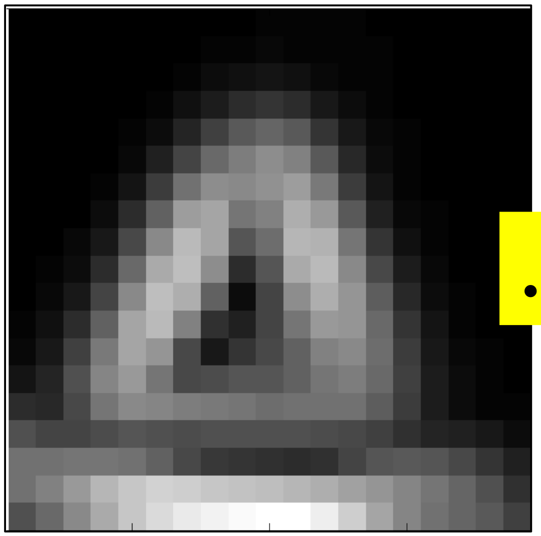


gaussian image (at closest scale, from pyramid)



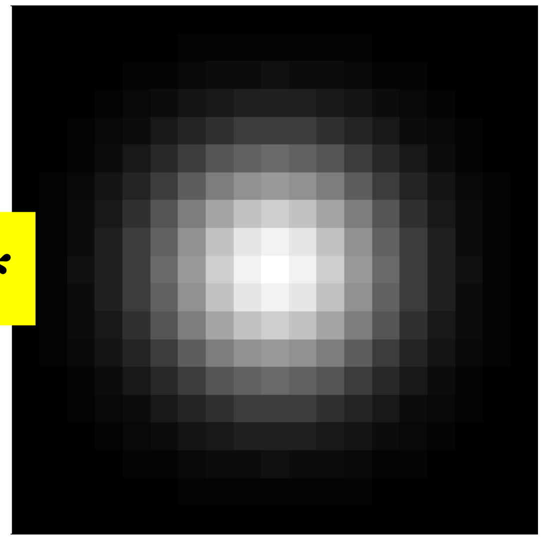
gradient orientation

Example (continued)



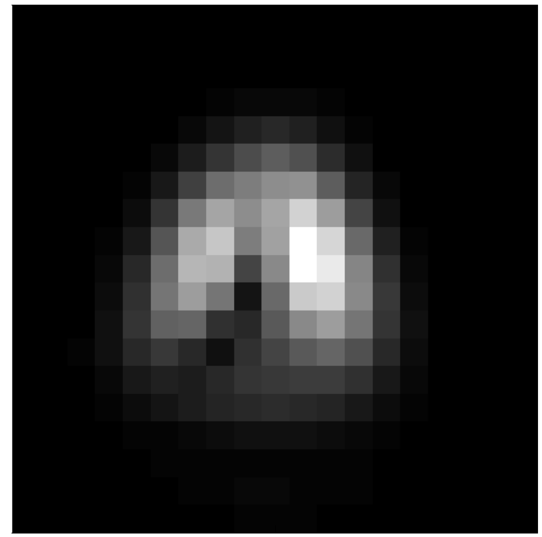
gradient magnitude

.*



weighted by 2D gaussian kernel

=

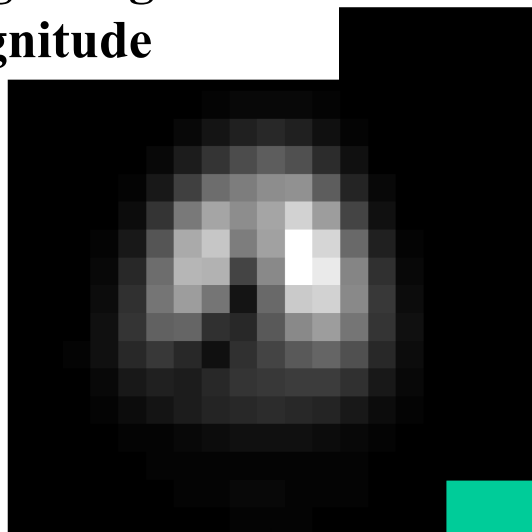


weighted gradient magnitude

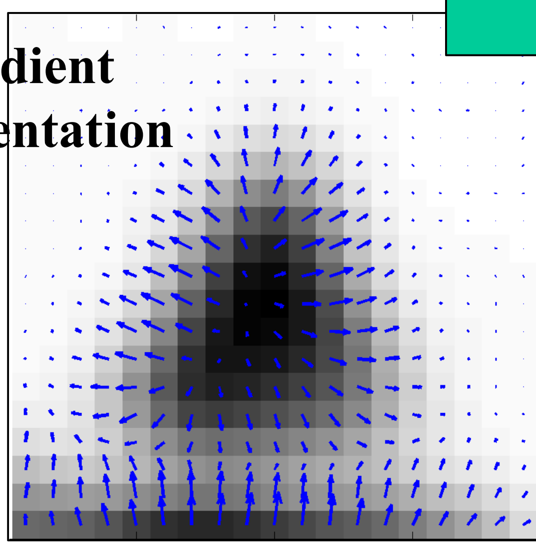
Robert Collins
CSE486, Penn State

Example (continued)

weighted gradient magnitude

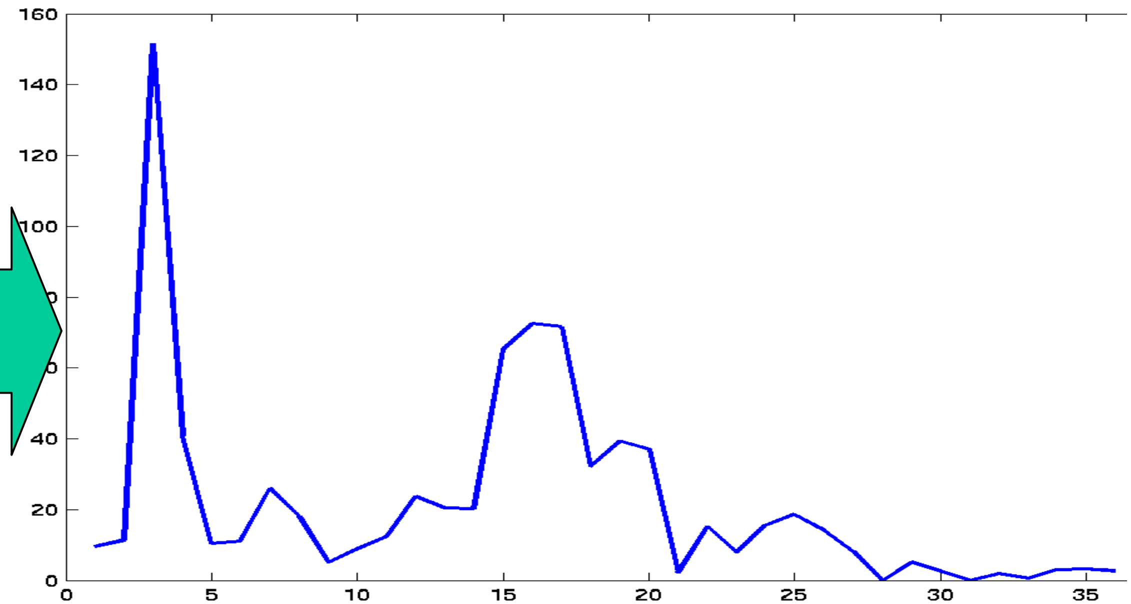


gradient orientation



weighted orientation histogram.

Each bucket contains sum of weighted gradient magnitudes corresponding to angles that fall within that bucket.



36 buckets

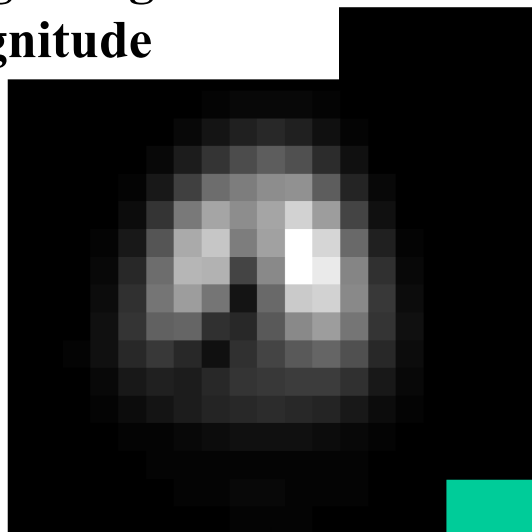
10 degree range of angles in each bucket, i.e.

- 0 <=ang<10 : bucket 1
- 10<=ang<20 : bucket 2
- 20<=ang<30 : bucket 3 …

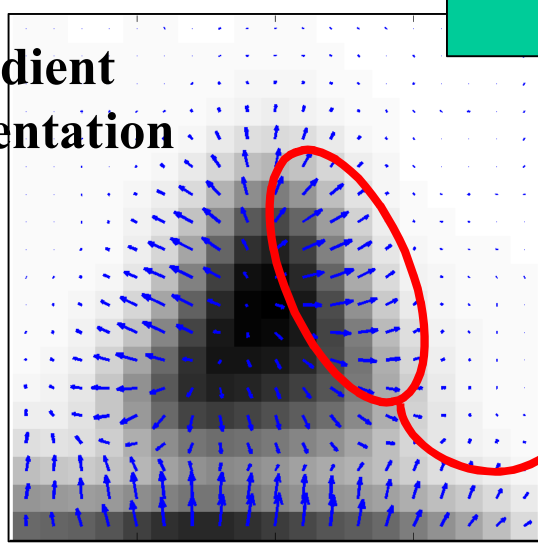
Robert Collins
CSE486, Penn State

Example (continued)

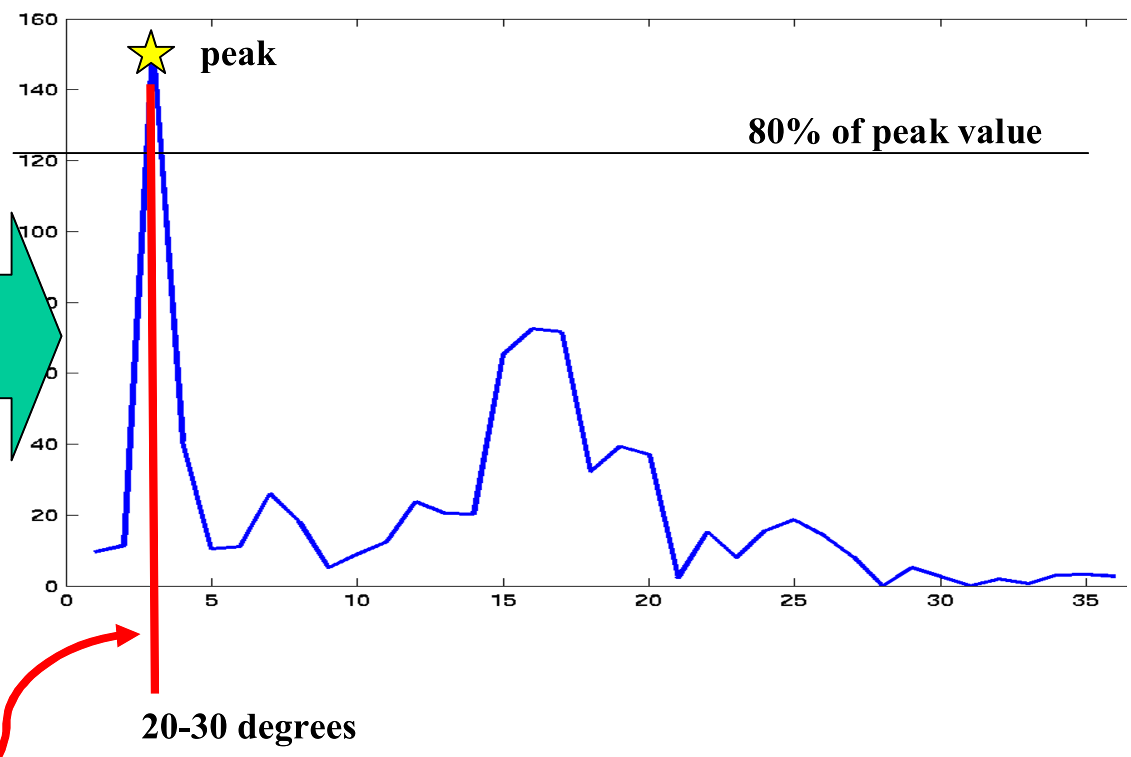
weighted gradient magnitude



gradient orientation



weighted orientation histogram.

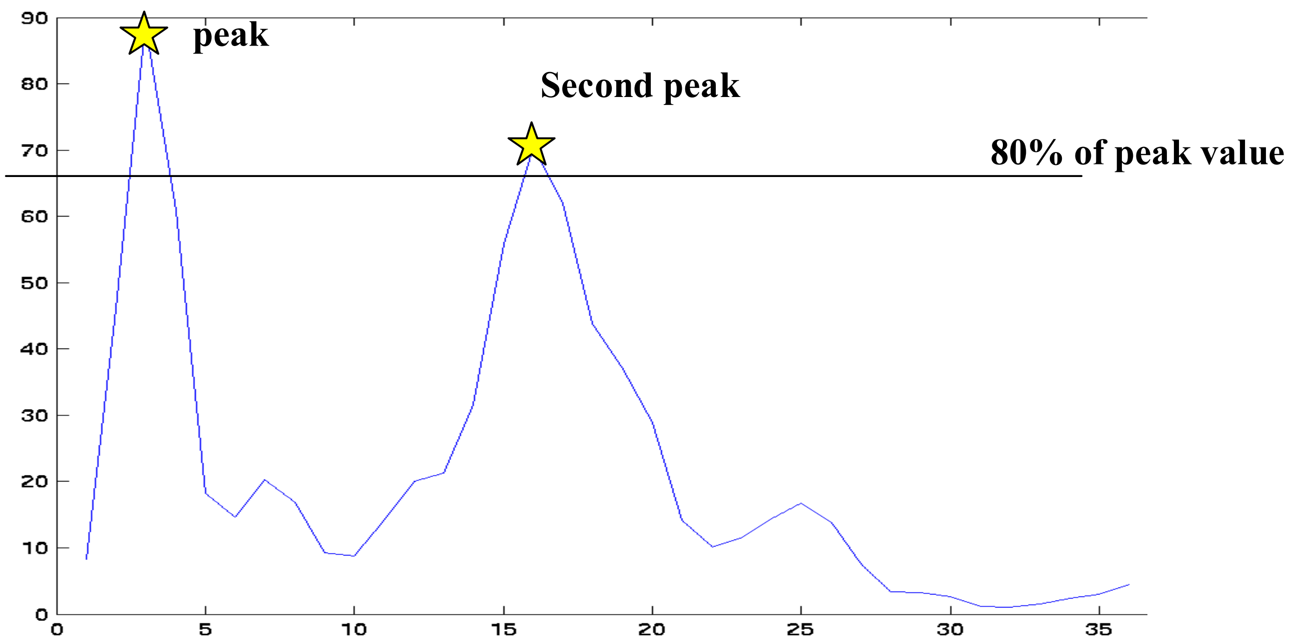


Orientation of keypoint is approximately 25 degrees

Robert Collins
CSE486, Penn State

Example (continued)

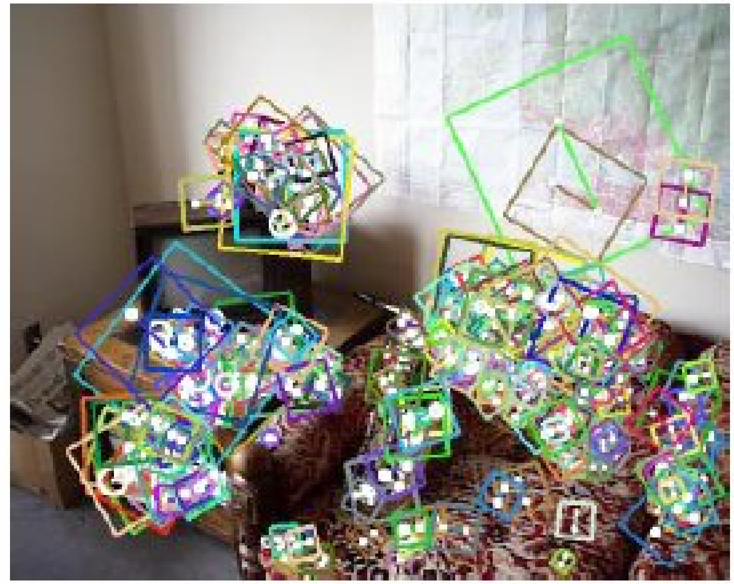
There may be multiple orientations.



In this case, generate duplicate keypoints, one with orientation at 25 degrees, one at 155 degrees.

Design decision: you may want to limit number of possible multiple peaks to two.

Example of KeyPoint Detection



Each keypoint has a center point (location), an orientation (rotation) and a radius (scale).

At this point, we could try to correlate patches (after first normalizing to a canonical orientation and scale).

SIFT Vector

to make things more insensitive to changes in lighting or small changes in geometry, Lowe constructs feature vector from image gradients.

SIFT Vector

- 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

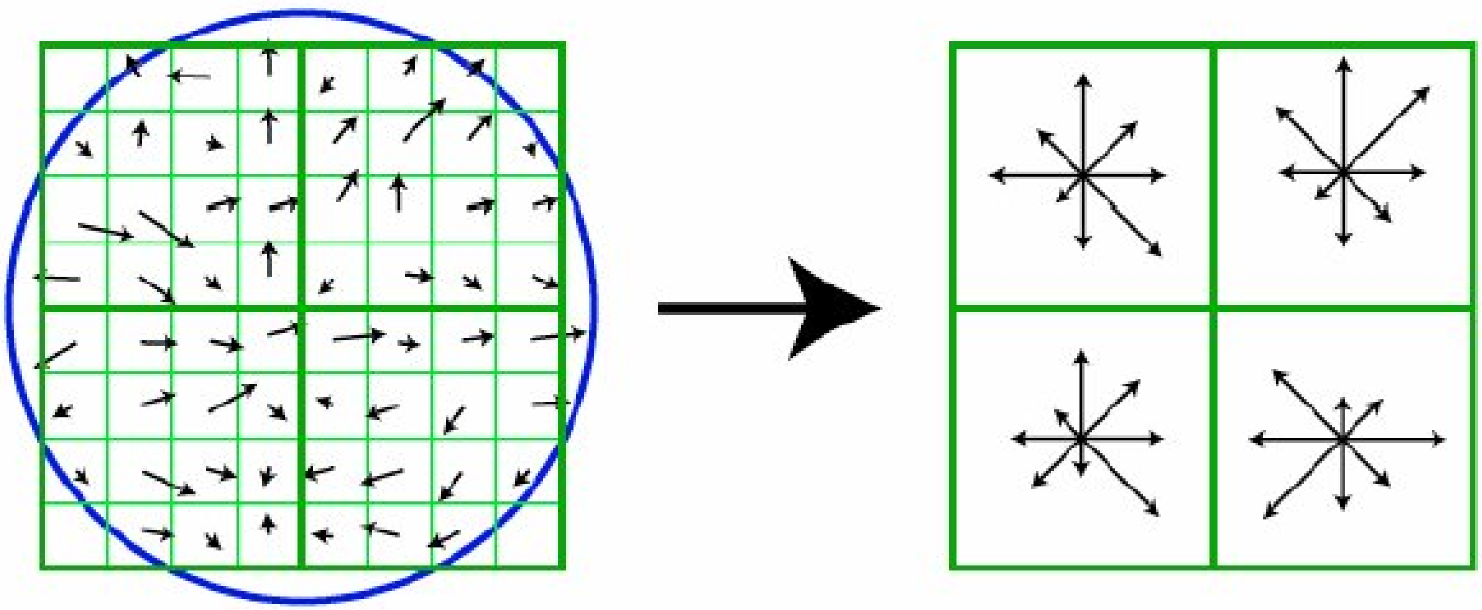


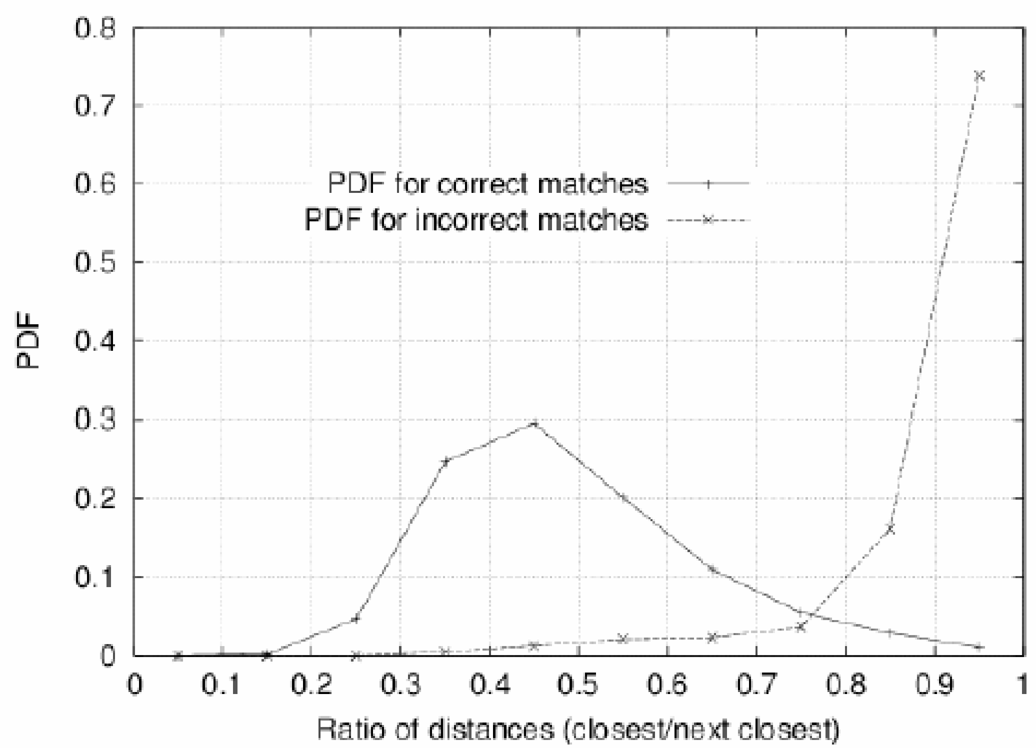
Image gradients

Keypoint descriptor

Robert Collins
CSE486, Penn State

Sift Key Matching

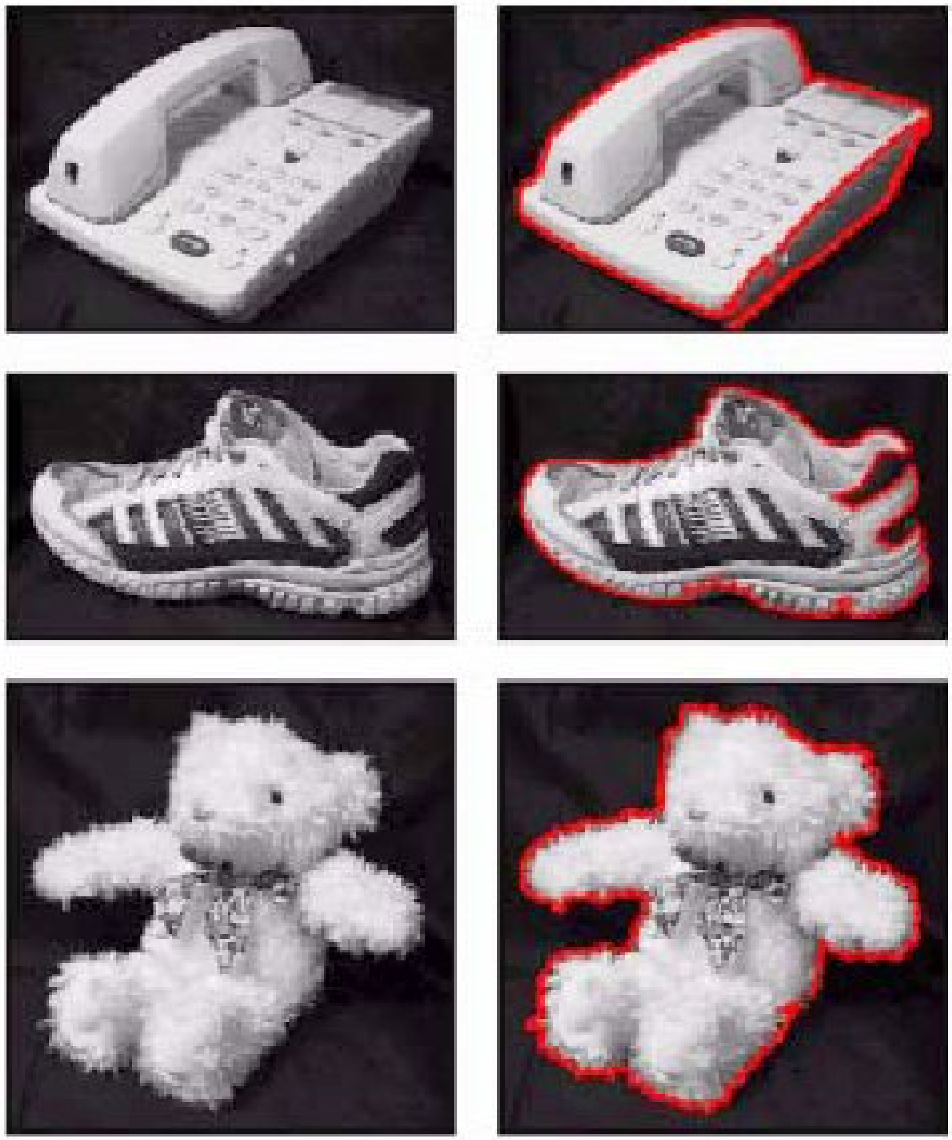
- Hypotheses are generated by **approximate nearest neighbor** matching of each feature to vectors in the database
- Compare distance of **nearest** neighbor to **second** nearest neighbor (from different object)
- Threshold of 0.8 provides excellent separation



Model Verification

1. Examine all clusters with at least 3 features
2. Perform least-squares affine fit to model.
3. Discard outliers and perform top-down check for additional features.
4. Evaluate probability that match is correct
   - Use Bayesian model, with probability that features would arise by chance if object was *not* present (Lowe, CVPR 01)

Application: Object Recognition



Compute SIFT keys of models and store in a database

Application: Object Recognition

database of models

find them in an image

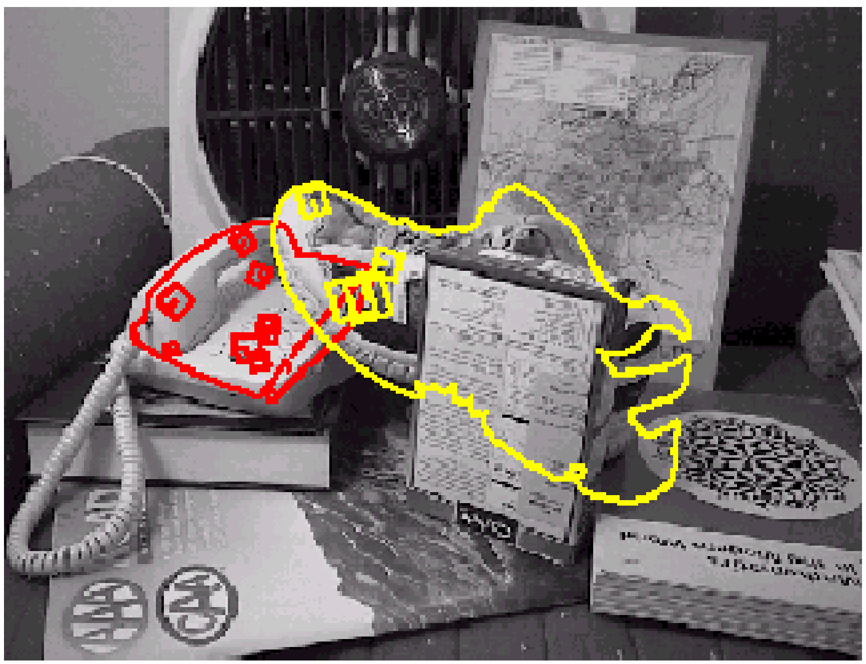
Robert Collins
CSE486, Penn State

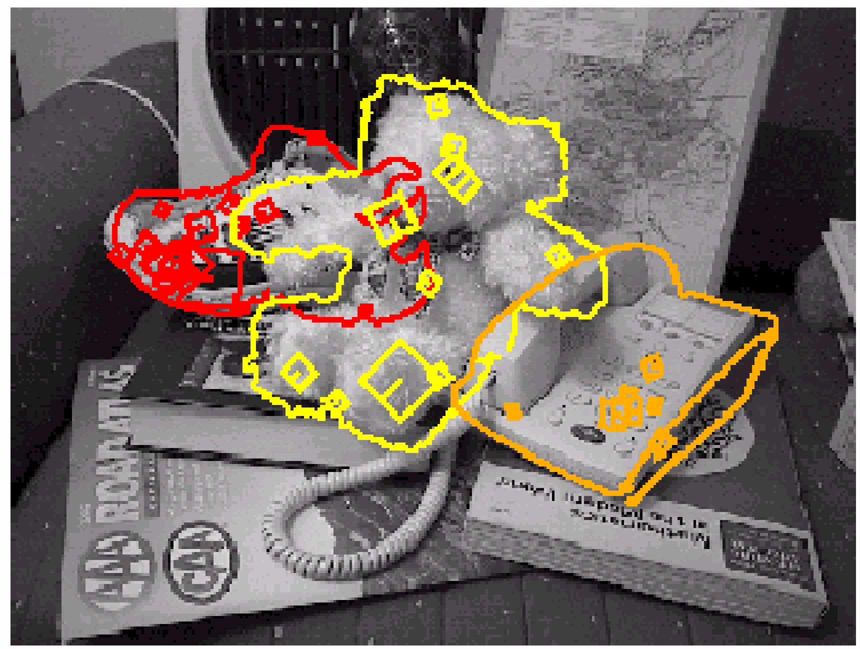
Application: Object Recognition

For sets of 3 SIFT key matches, compute affine transformation and perform model verification.

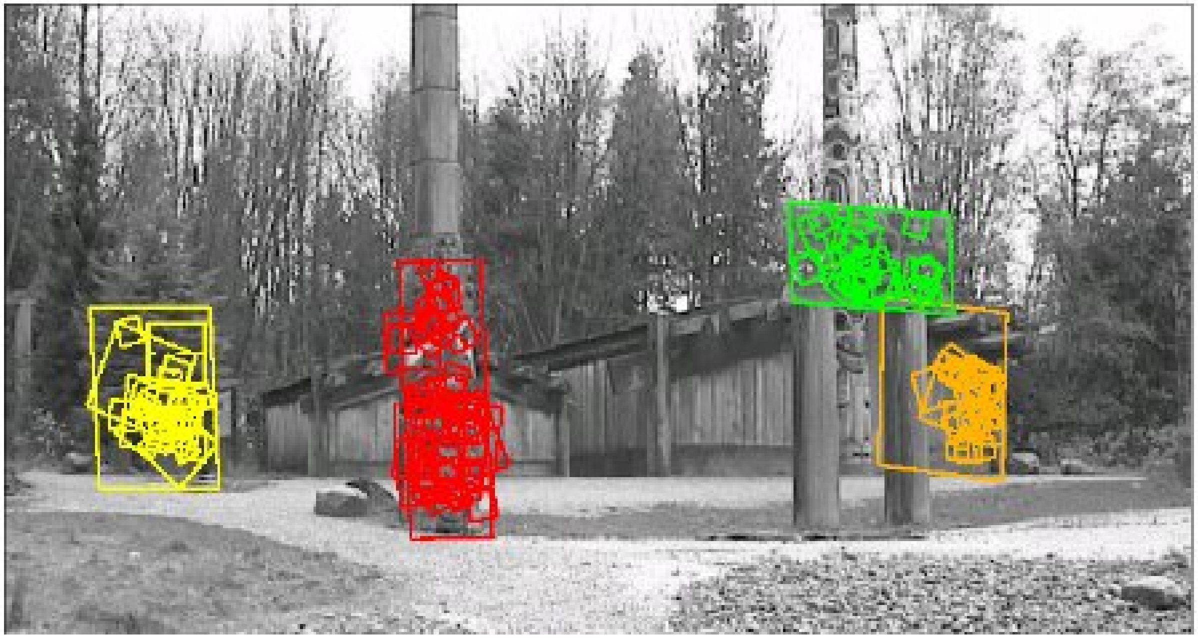
Application: Object Recognition

Note: since these are local, parts-based descriptors, they perform well even when some parts are missing (i.e. under occlusion).





Application: Landmark Recognition

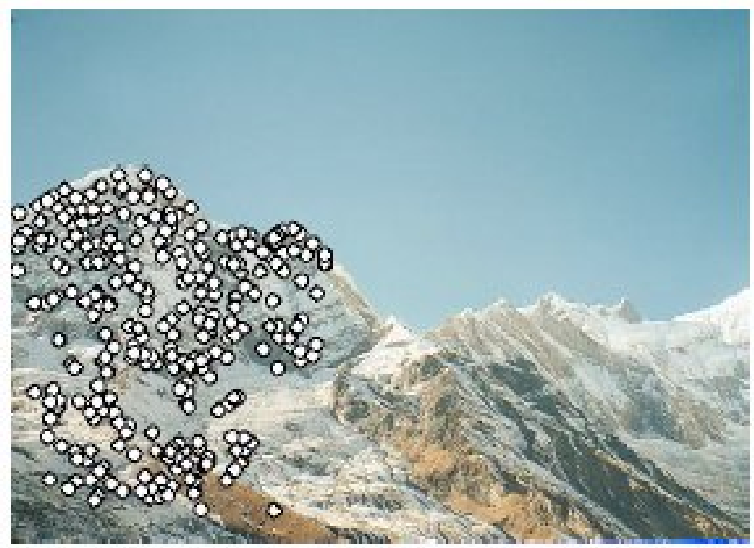


Application: Generating Panoramas

(a) Image 1

(b) Image 2

(c) SIFT matches 1



(d) SIFT matches 2

Brown and Lowe, ICCV03

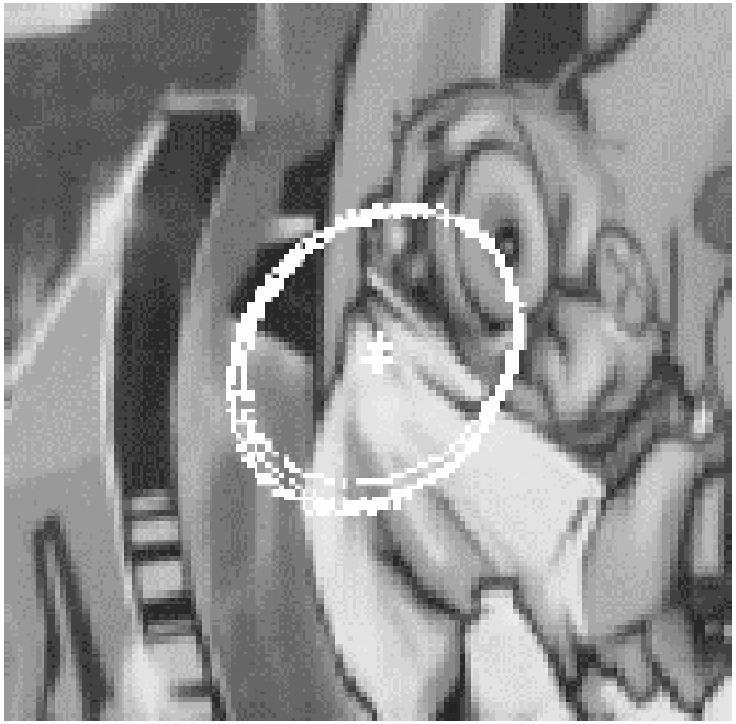
Application: Generating Panoramas

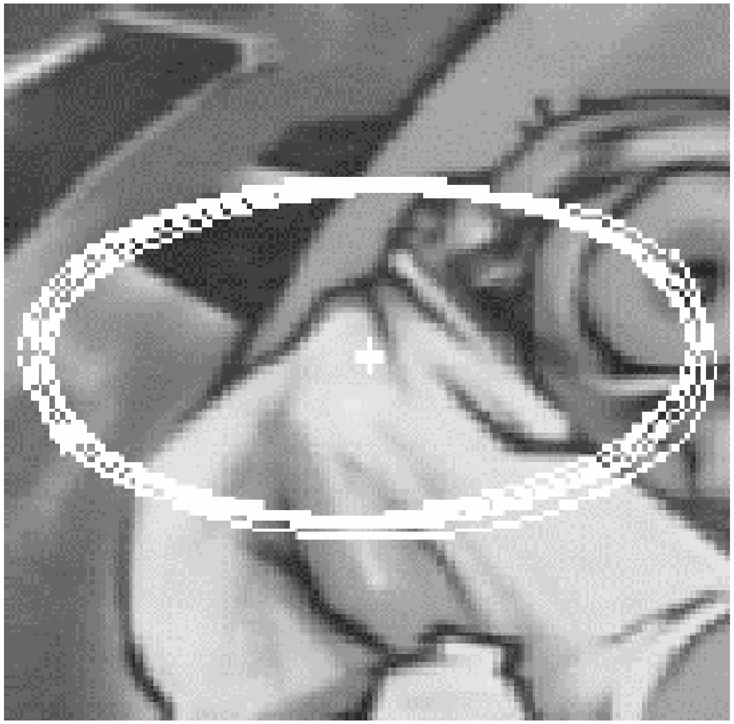
(e) Images aligned according to a homography

Brown and Lowe, ICCV03

State of the Art

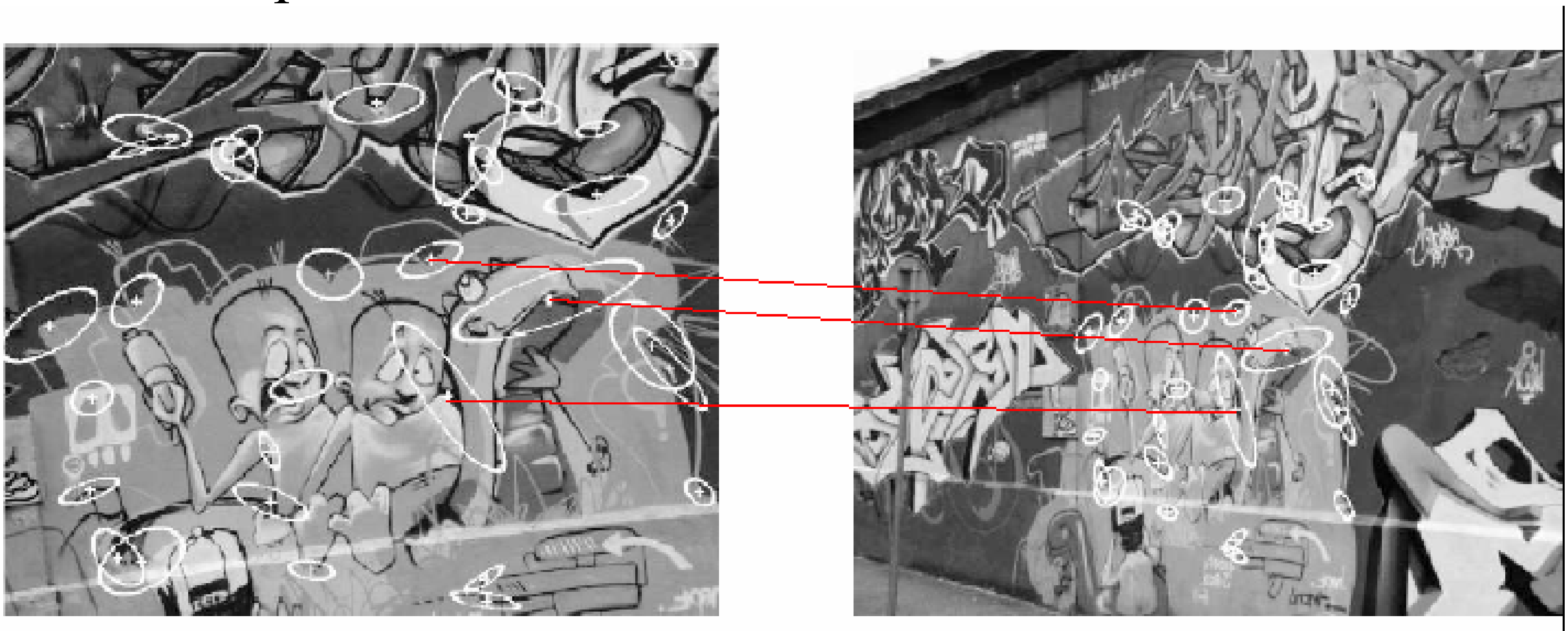
Fully affine invariant local feature descriptors.



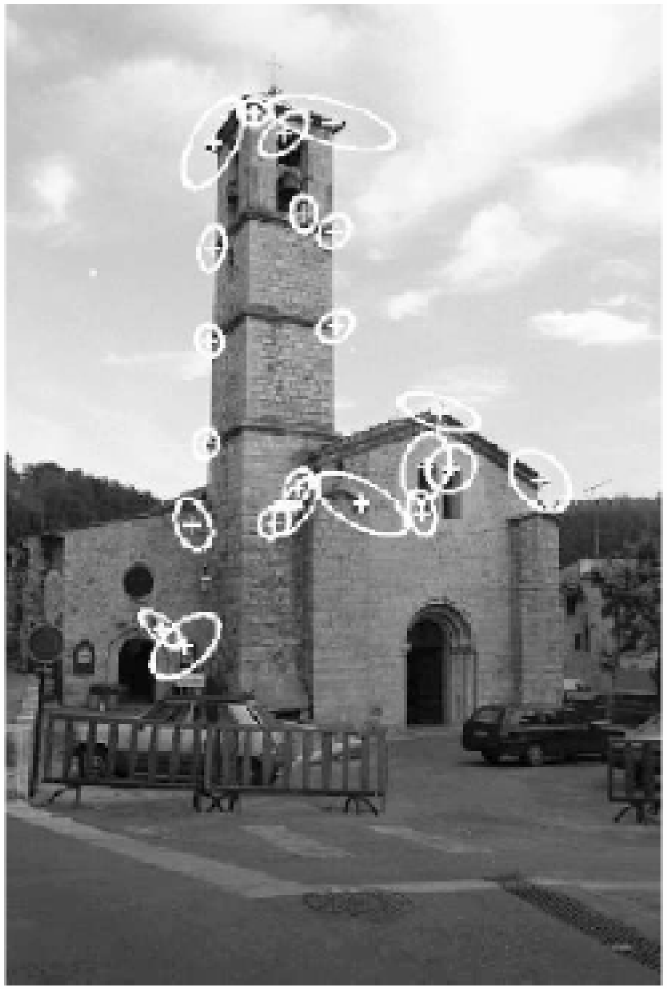


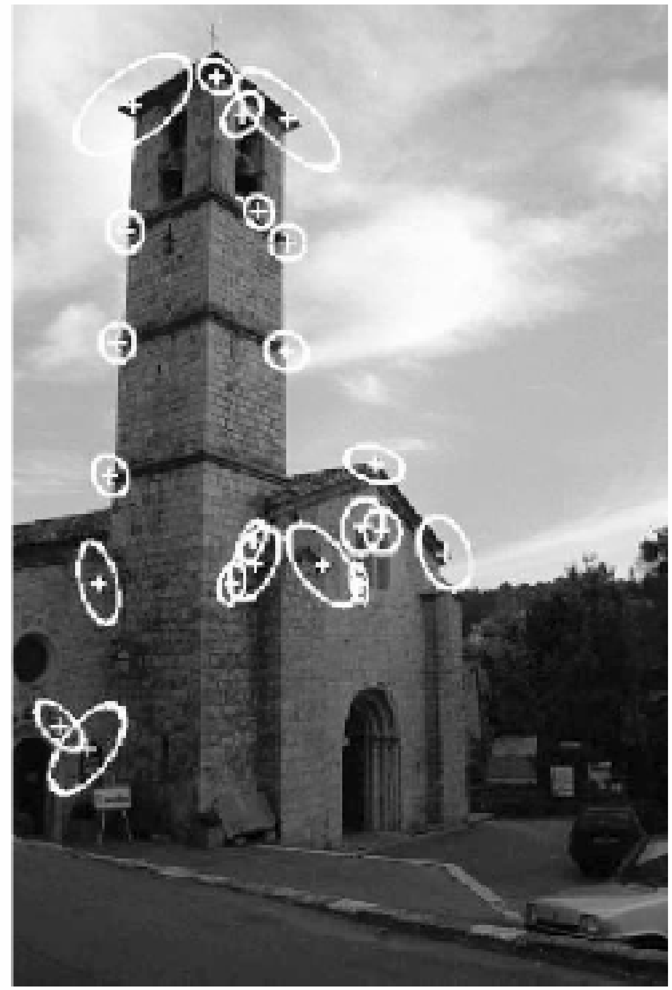
State of the Art

Affine invariant descriptors can handle larger changes in viewpoint.



State of the Art





For More Information

SIFT keys

David G. Lowe, **"Distinctive image features from scale-invariant keypoints,"** *International Journal of Computer Vision,* 60, 2 (2004), pp. 91-110.

Affine local-feature methods

- K. Mikolajczk and C. Schmid. A performance evaluation of local descriptors. In IEEE Conference on Computer Vision and Pattern Recognition, June 2003.
- K.Mikolajczyk and C.Schmid. An affine invariant interest point detector. In European Conference on Computer Vision, vol. 1, 128--142, 2002.