Object Recognition from Local Scale-Invariant Features (SIFT)

David G. Lowe

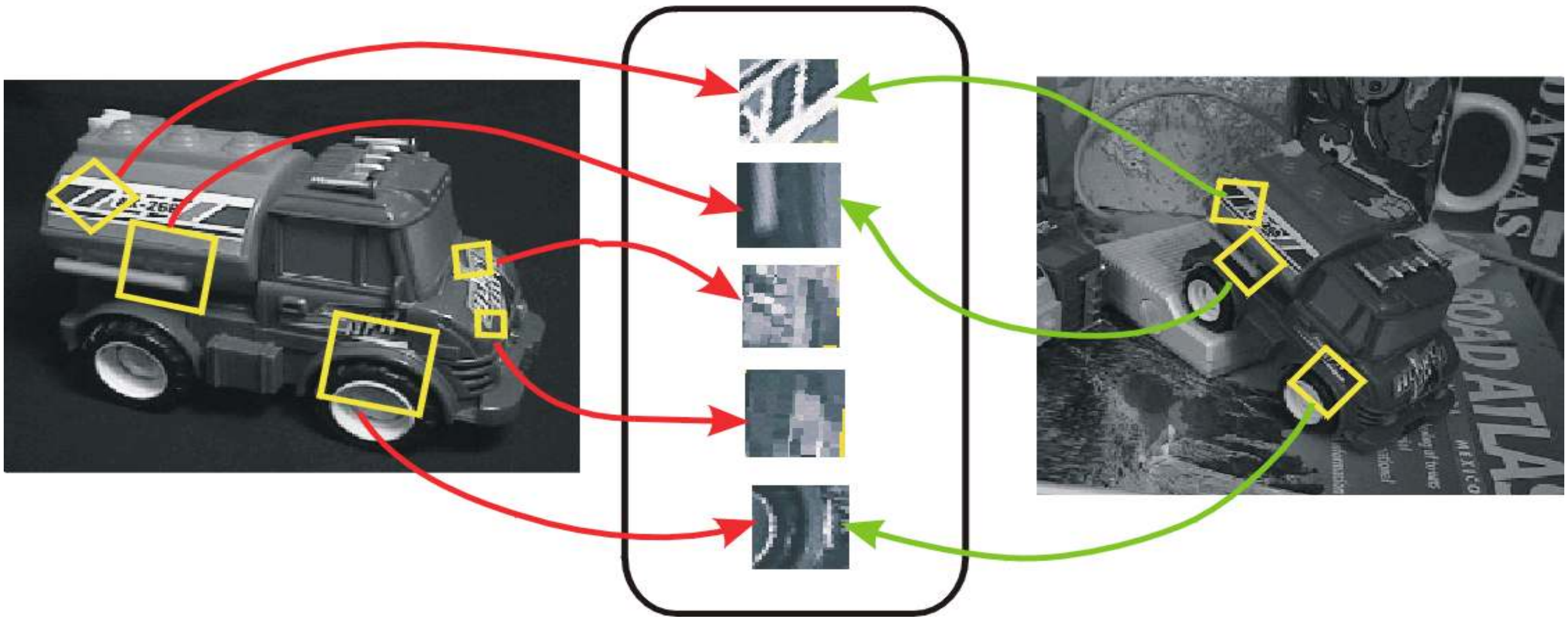
Presented by David Lee
3/20/2006

Introduction

- Well engineered local descriptor

Introduction

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

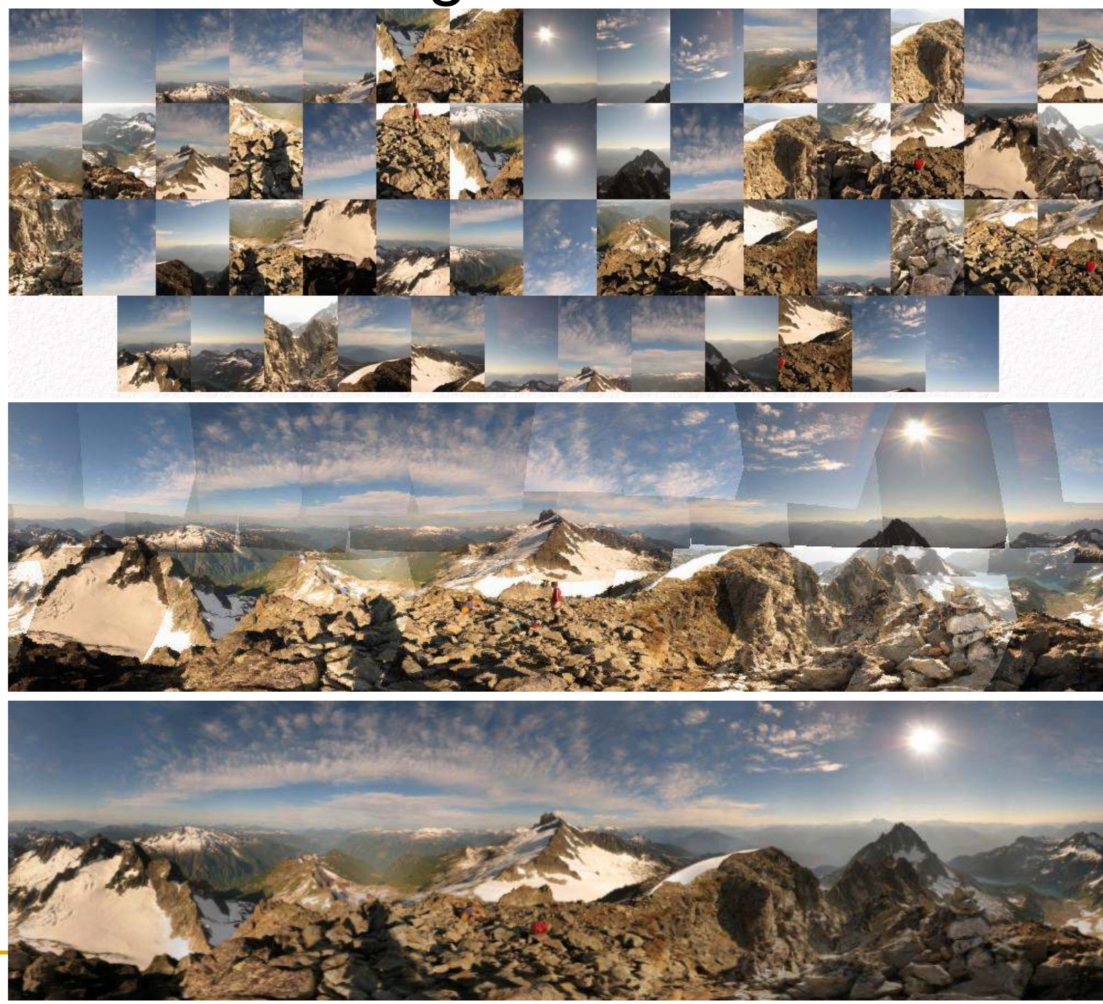


Introduction

- Initially proposed for correspondence matching
  - Proven to be the most effective in such cases according to a recent performance study by Mikolajczyk & Schmid (ICCV ’03)

Introduction

- Automatic Mosaicing



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

Introduction

- Now being used for general object class recognition (e.g. 2005 Pascal challenge)

- Histogram of gradients
  - Human detection, Dalal & Triggs CVPR ’05

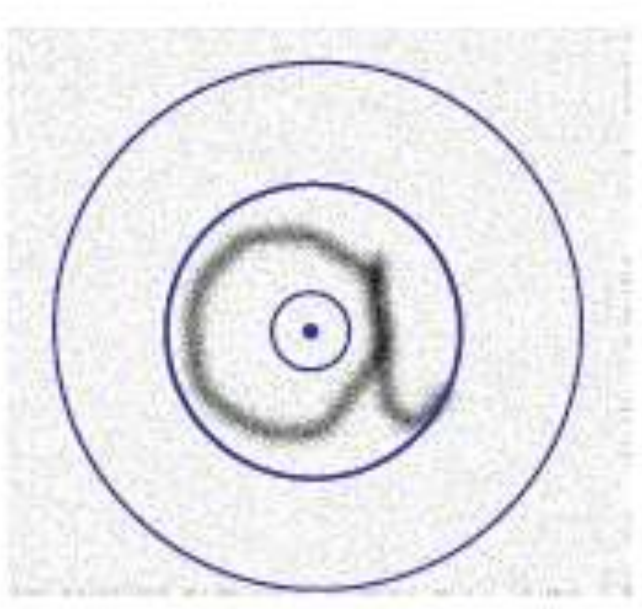
Introduction

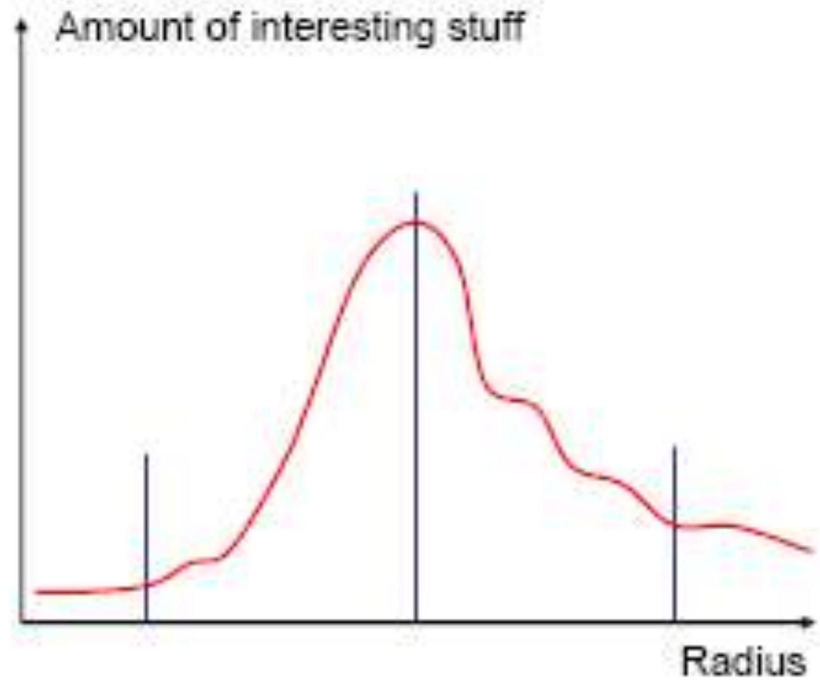
- SIFT in one sentence
  - Histogram of gradients @ Harris-corner-like

- Extract features
  - Find keypoints
    - Scale, Location
    - Orientation
  - Create signature
- Match features

Finding Keypoints – Scale, Location

- How do we choose scale?



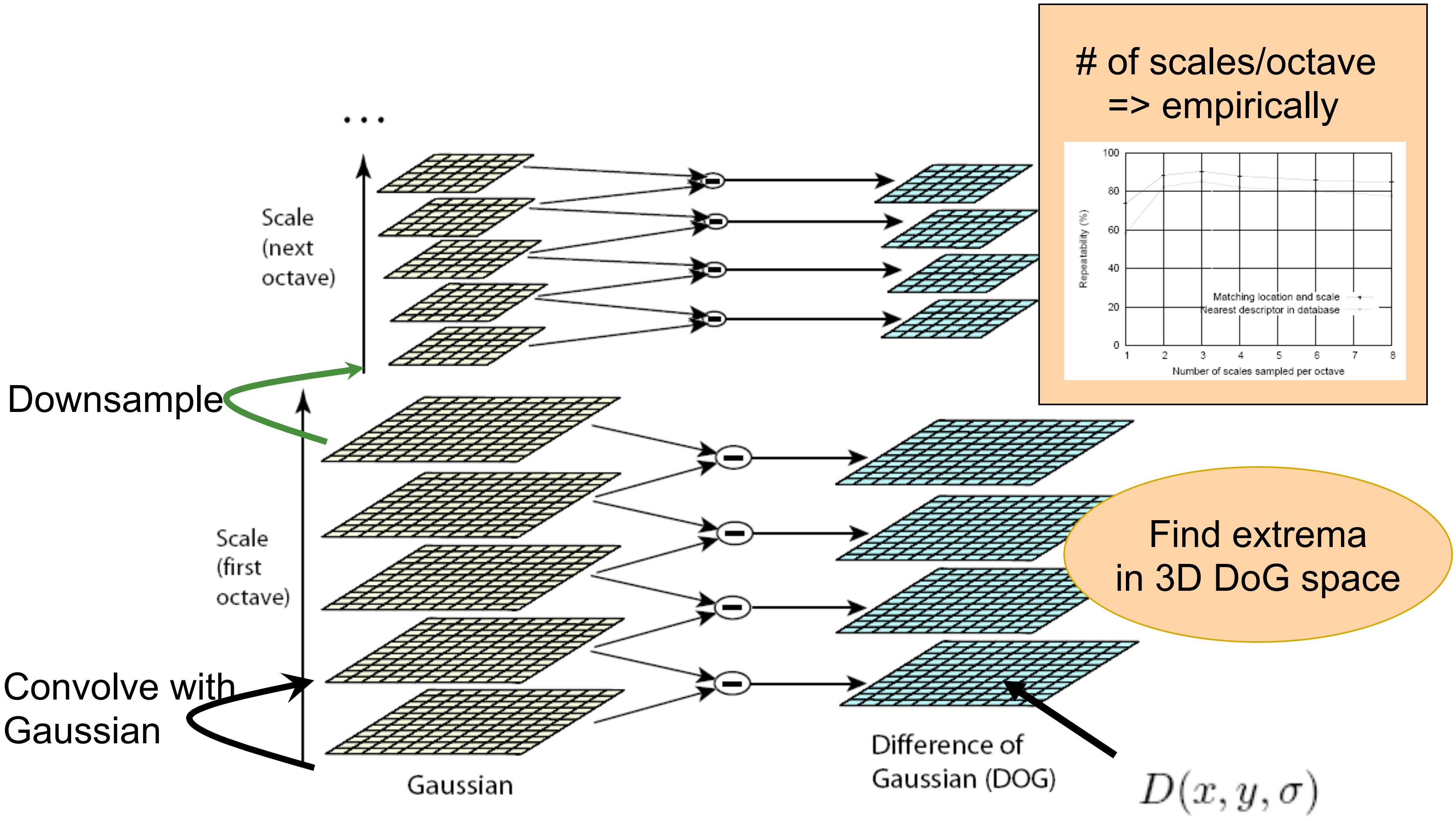


Finding Keypoints – Scale, Location

- Scale selection principle (T. Lindeberg '94)
  - In the absence of other evidence, assume that a scale level, at which (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.

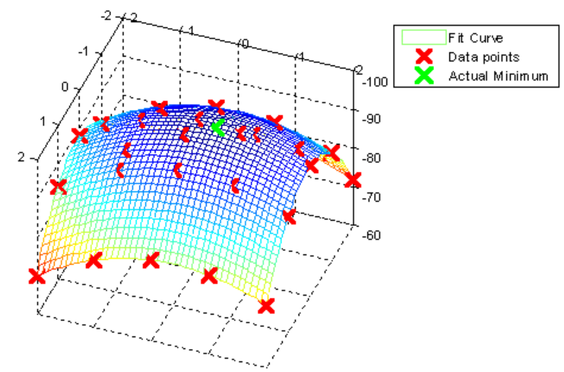
➔ Maxima/minima of Difference of Gaussian

Finding Keypoints – Scale, Location



Finding Keypoints – Scale, Location

- Sub-pixel Localization
  - Fit Trivariate quadratic to find sub-pixel extrema



- Eliminating edges
  - Similar to Harris corner detector

$$\mathbf{H} = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \qquad \frac{\mathrm{Tr}(\mathbf{H})^2}{\mathrm{Det}(\mathbf{H})} < \frac{(r+1)^2}{r}$$

Finding Keypoints – Scale, Location

- Key issue: **Stability (Repeatability)**
- Alternatives
  - Multi-scale Harris corner detector
  - Harris-Laplacian
  - Kadir & Brady Saliency Detector
  - …
  - Uni
  - Ra

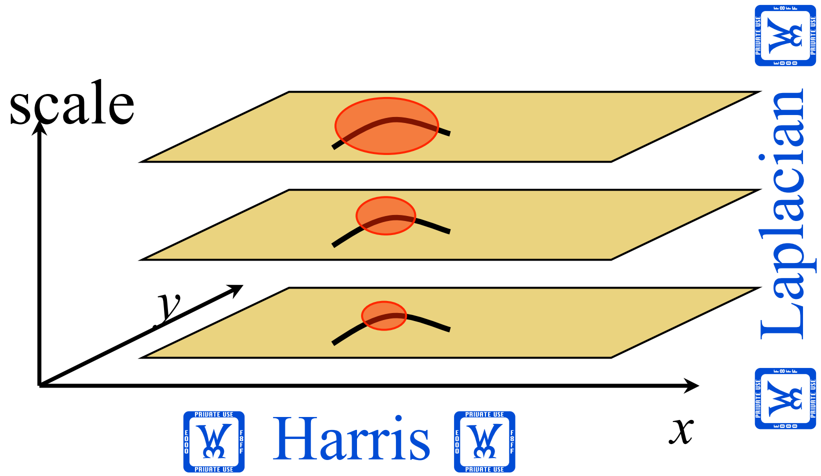
Recall Fei-fei's pLSA paper

Descriptor	Grid	Random	Saliency [4]	DoG [7]
11×11 Pixel	64.0%	47.5%	45.5%	N/A
128-dim Sift	65.2%	60.7%	53.1%	52.5%

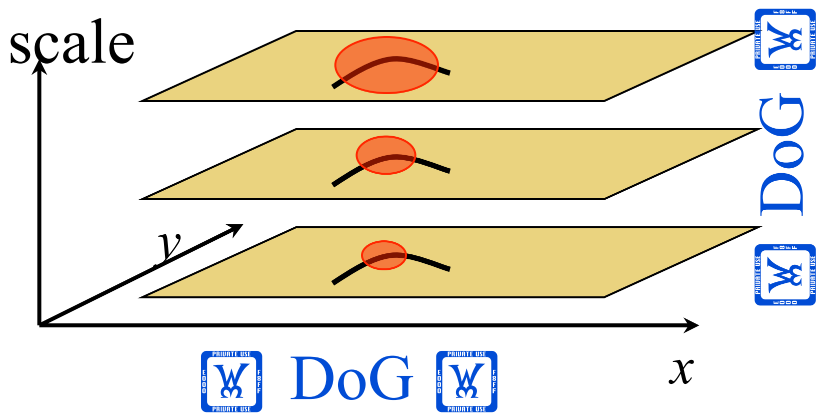
** Important Note ** Their application was scene classification
NOT correspondence matching

Finding Keypoints – Scale, Location

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



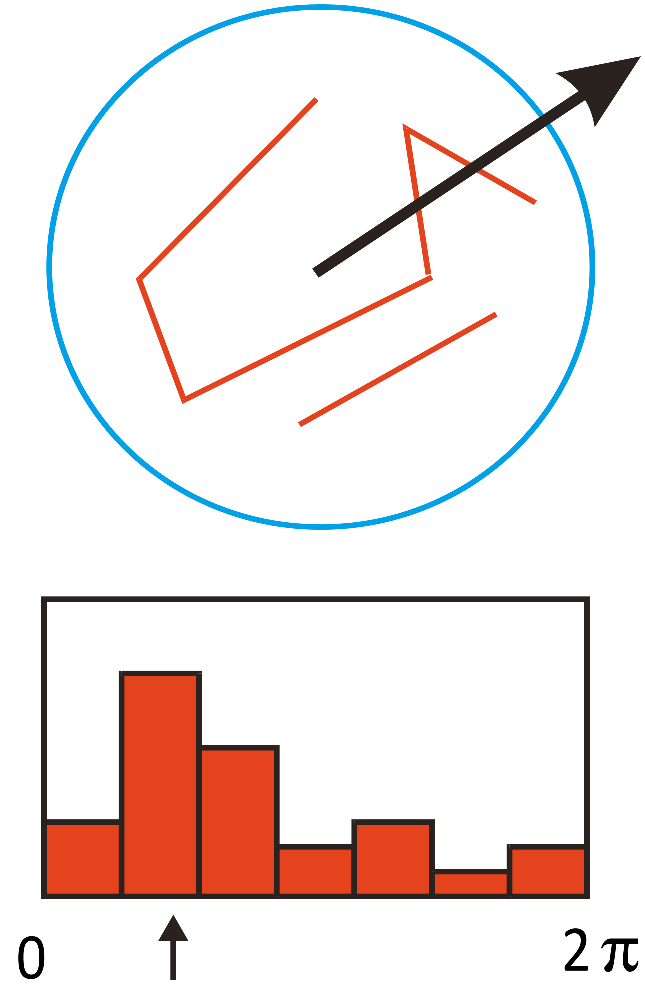
- SIFT[2]
  Find local maximum of:
  – Difference of Gaussians in space and scale



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

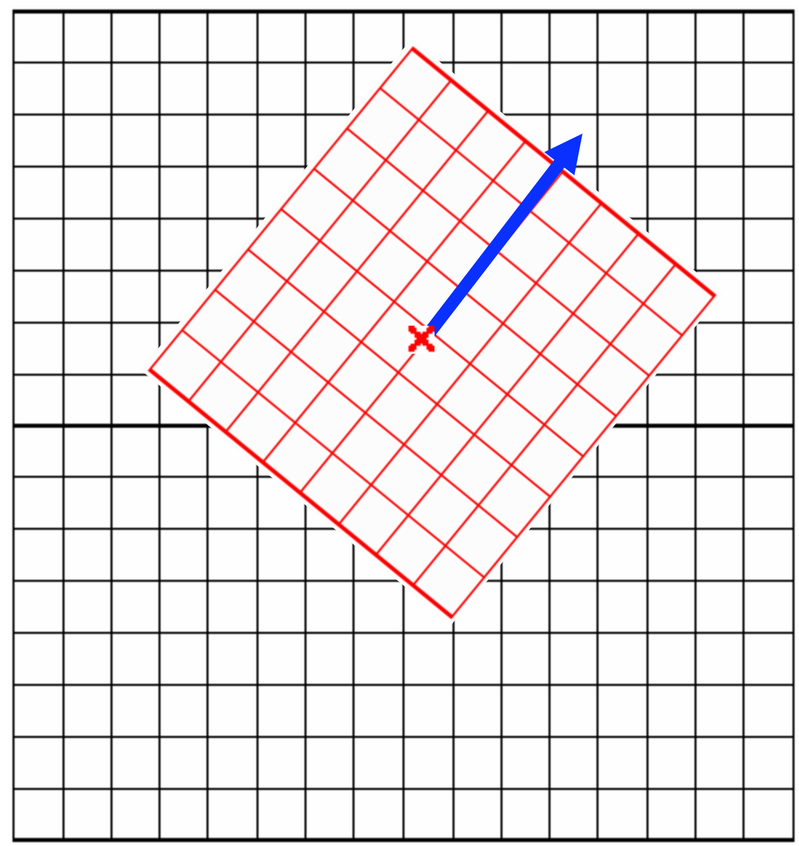
Finding Keypoints – Orientation

- 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)



Finding Keypoints – Orientation

- Assign dominant orientation as the orientation of the keypoint



Finding Keypoints

- So far, we found…
  - where interesting things are happening
  - and its orientation

- With the hope of
  - Same keypoints being found, even under some scale, rotation, illumination variation.

- Extract features
  - ☑ Find keypoints
    - ☑ Scale, Location
    - ☑ Orientation
  - ☐ Create signature
- Match features

Creating Signature

- 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

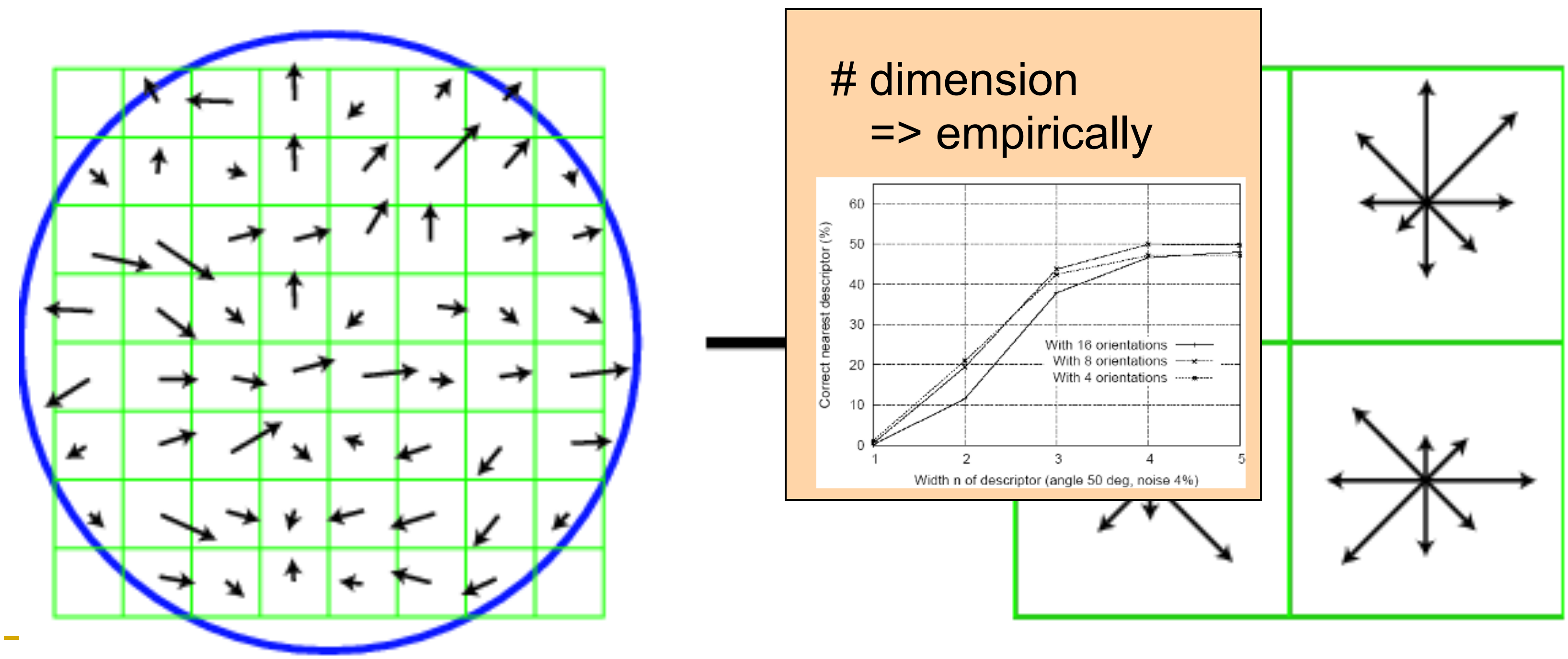
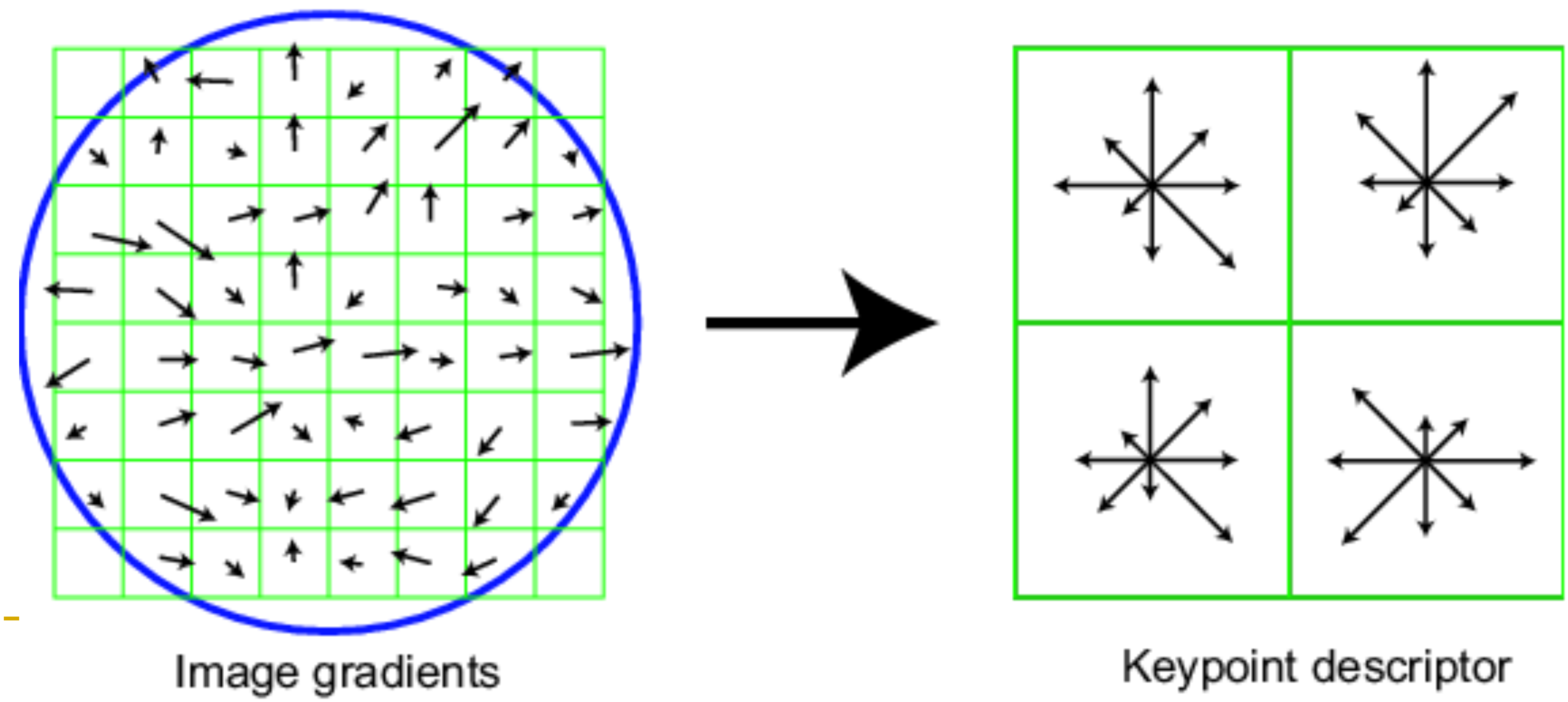


Image gradients

Keypoint descriptor

Creating Signature

- What kind of information does this capture?



Comparison with HOG (Dalal '05)

- Histogram of Oriented Gradients
- General object class recognition (Human)
  - Engineered for a different goal

- Uniform sampling
- Larger cell (6-8 pixels)
- Fine orientation binning
  - 9 bins/180^{O} vs. 8 bins/360^{O}

- Both are well engineered

Comparison with MOPS (Brown '05)

- Mul
  Orie
- Sim
  - Mu
  - No
  - Sr
- Goo ng

- Extract features
  - ☑ Find keypoints
    - ☑ Scale, Location
    - ☑ Orientation
  - ☑ Create signature

- Match features
  - Nearest neighbor, Hough voting, Least-square affine parameter fit

Conclusion

- A novel method for detecting interest points
- Histogram of Oriented Gradients are becoming more popular
- SIFT may not be optimal for general object classification