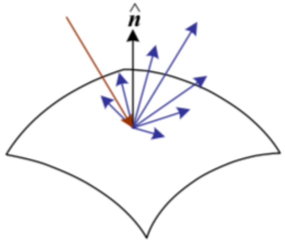
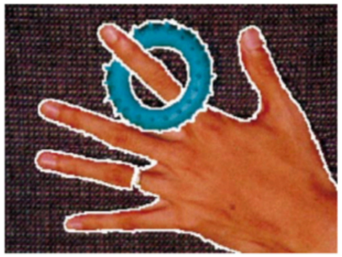


2. Image Formation

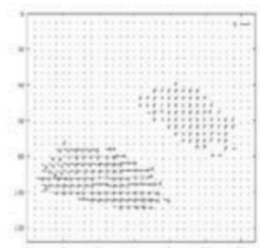
3. Image Processing

4. Features

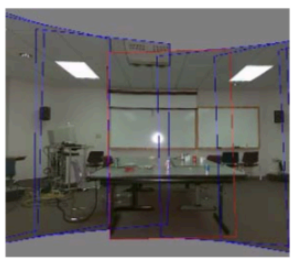


5. Segmentation

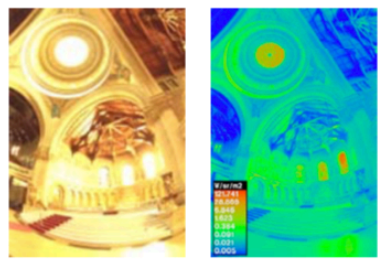
6-7. Structure from Motion



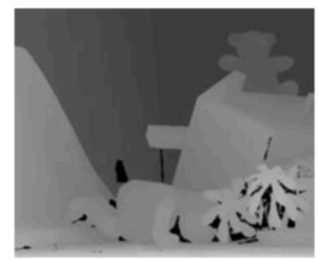
8. Motion



9. Stitching

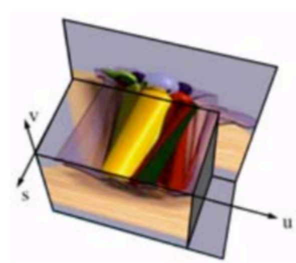


10. Computational Photography



11. Stereo

12. 3D Shape



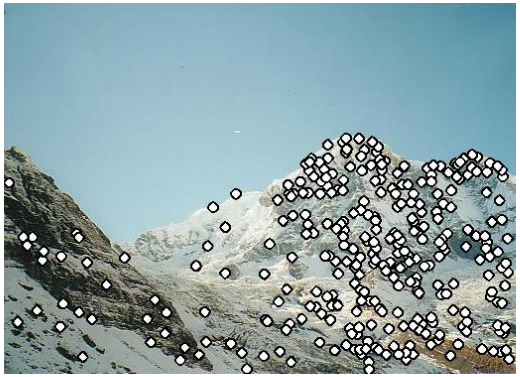
13. Image-based Rendering

14. Recognition

Detectors

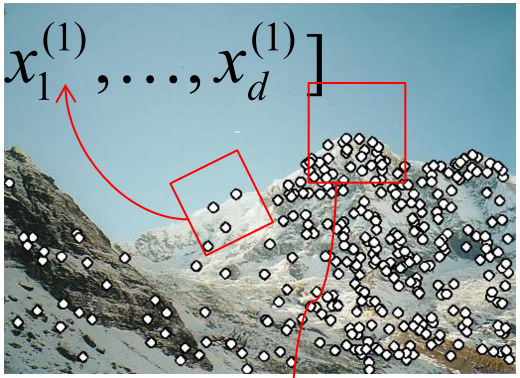
Local features: main components

1) **Detection: Identify the interest points**



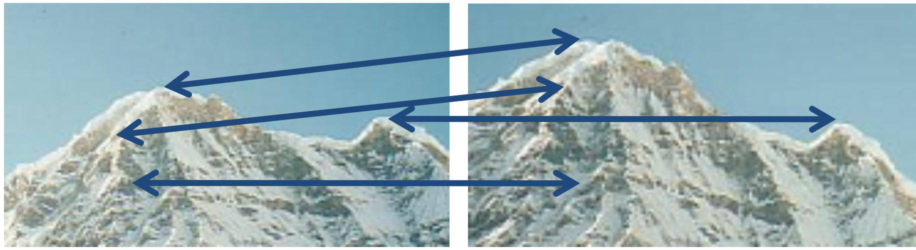
2) Description: Extract vector feature descriptor surrounding each interest point.

$$\mathbf{x}_1 = [x_1^{(1)}, \ldots, x_d^{(1)}]$$



$$\mathbf{x}_2 = [x_1^{(2)}, \ldots, x_d^{(2)}]$$

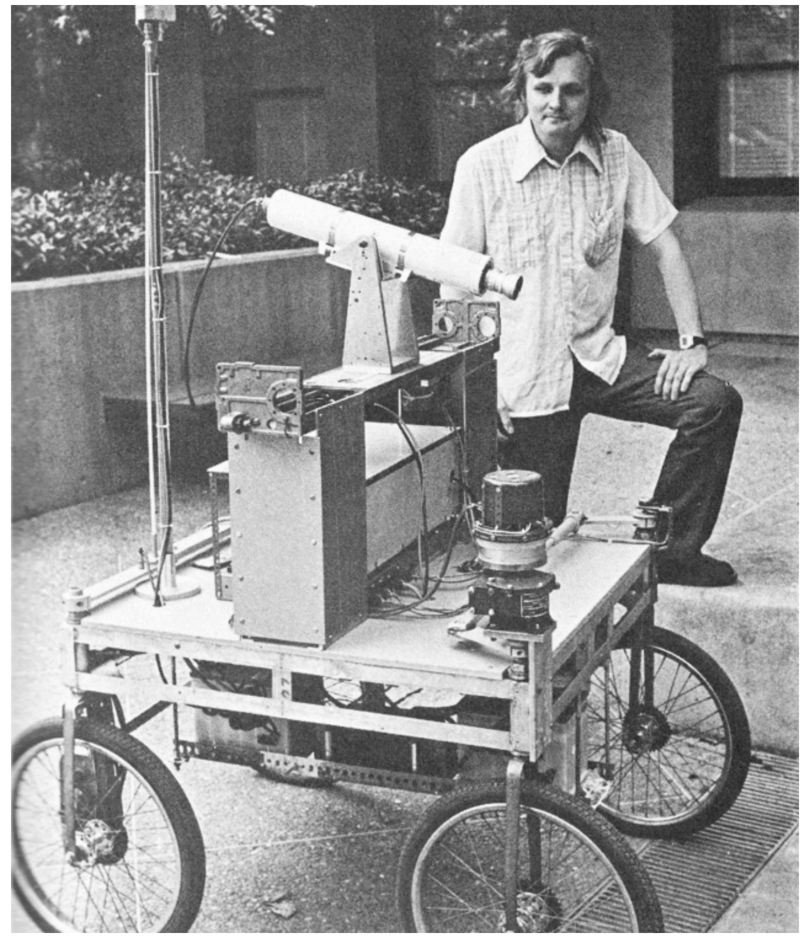
3) Matching: Determine correspondence between descriptors in two views



Kristen Grauman

History

- Moravec 1980
- Harris Corners 1988
- [Wolf & Platt 1993: FCN!]
- SIFT (Lowe) 2004
- FAST 2006 (learning!)
- SURF 2006
- ORB 2011



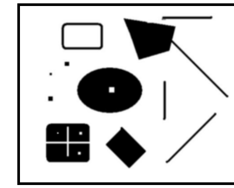
Harris corner detector

1) Compute *M* matrix for each image window to get their *cornerness* scores.
2) Find points whose surrounding window gave large corner response (*f*> threshold)
3) Take the points of local maxima, i.e., perform non-maximum suppression

C.Harris and M.Stephens. "A Combined Corner and Edge Detector." *Proceedings of the 4th Alvey Vision Conference*: pages 147—151, 1988.

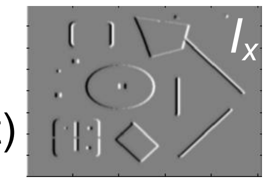
Harris Detector [Harris88]

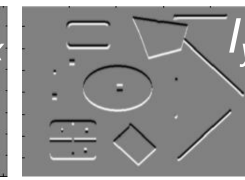
- Second moment matrix



$$\mu(\sigma_I,\sigma_D) = g(\sigma_I) * \begin{bmatrix} I_x^2(\sigma_D) & I_x I_y(\sigma_D) \\ I_x I_y(\sigma_D) & I_y^2(\sigma_D) \end{bmatrix}$$

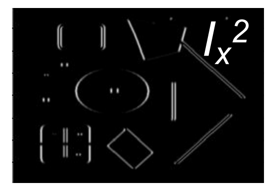
1. Image derivatives (optionally, blur first)

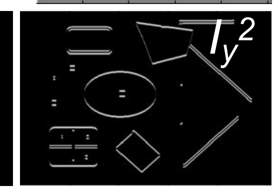


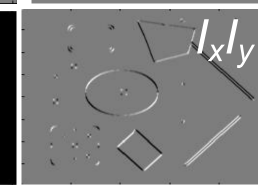


$$\det M = \lambda_1 \lambda_2$$
$$\operatorname{trace} M = \lambda_1 + \lambda_2$$

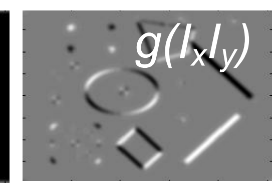
2. Square of derivatives







3. Gaussian filter $g(\sigma_I)$



4. Cornerness function – both eigenvalues are strong

$$har = \det[\mu(\sigma_I,\sigma_D)] - \alpha[\operatorname{trace}(\mu(\sigma_I,\sigma_D))^2] =$$
$$g(I_x^2)g(I_y^2) - [g(I_xI_y)]^2 - \alpha[g(I_x^2) + g(I_y^2)]^2$$

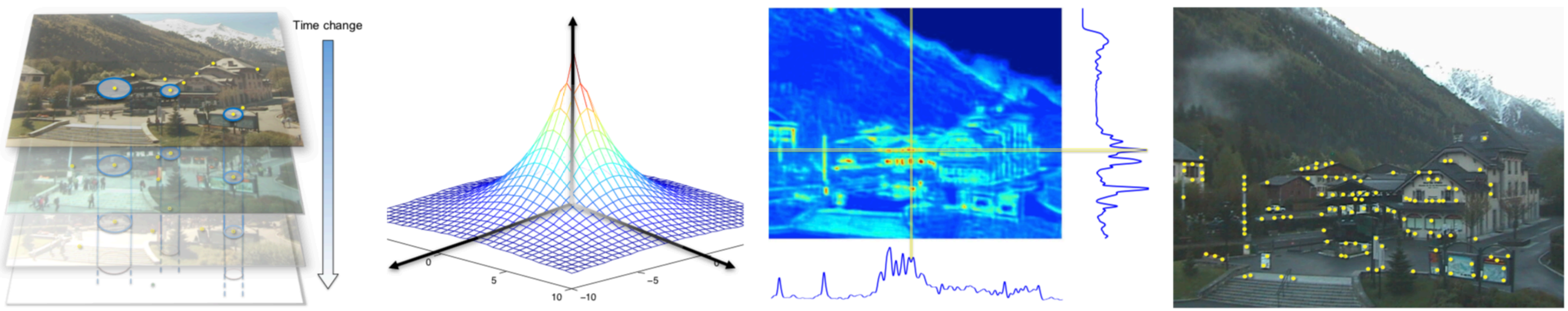
5. Non-maxima suppression

Deep Detectors

TILDE: A Temporally Invariant Learned DEtector

CVPR 2015

Yannick Verdie[1,*] Kwang Moo Yi[1,*] Pascal Fua[1] Vincent Lepetit[2]
[1]Computer Vision Laboratory, École Polytechnique Fédérale de Lausanne (EPFL)
[2]Institute for Computer Graphics and Vision, Graz University of Technology



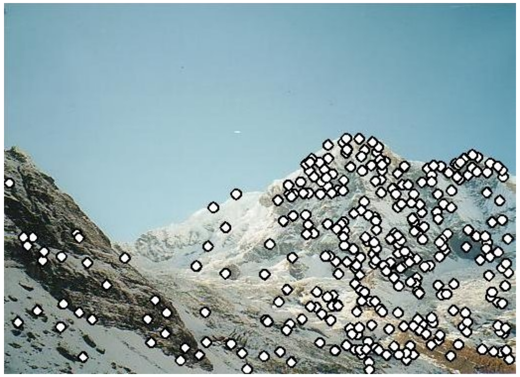
(a) Stack of training images (b) Desired response on positive samples (c) Regressor response for a new image (d) Keypoints detected in the new image

- Train on images from webcams: fixed view, different times
- Learn CNN-like regressor
- Loss = repeatability

Descriptors

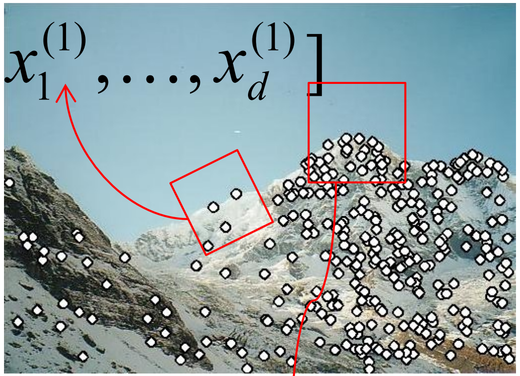
Local features: main components

1) Detection: Identify the interest points



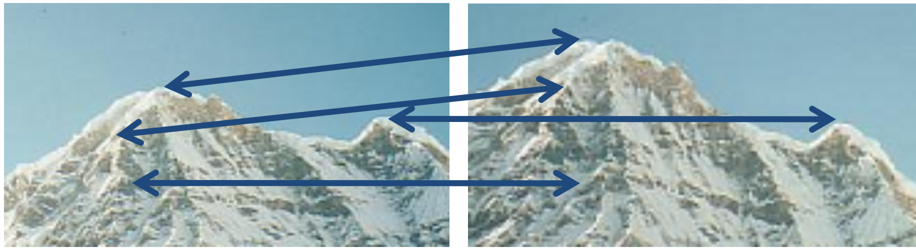
2) **Description: Extract vector feature descriptor surrounding each interest point.**

$$\mathbf{x}_1 = [x_1^{(1)}, \ldots, x_d^{(1)}]$$



$$\mathbf{x}_2 = [x_1^{(2)}, \ldots, x_d^{(2)}]$$

3) Matching: Determine correspondence between descriptors in two views



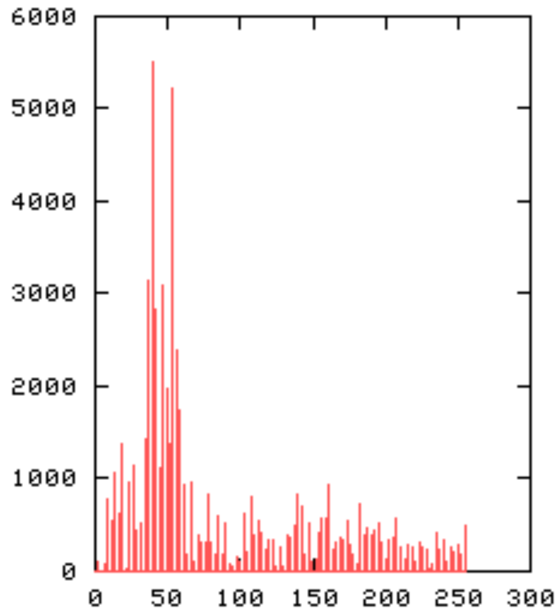
Kristen Grauman

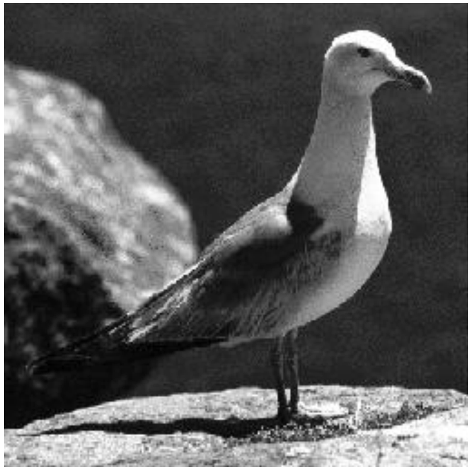
Image representations

- Templates
  - Intensity, gradients, etc.

- Histograms
  - Color, texture, SIFT descriptors, etc.

Image Representations: Histograms





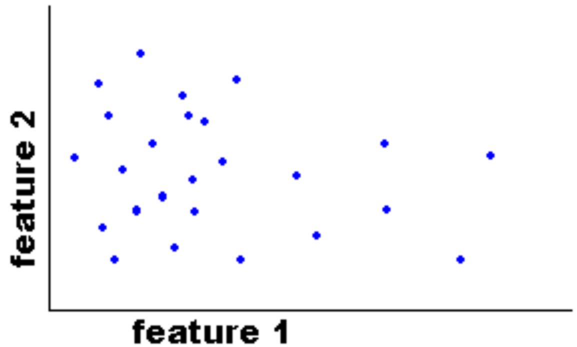
Global histogram

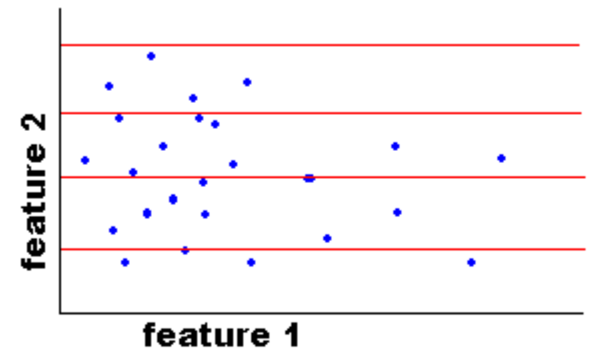
- Represent distribution of features
  - Color, texture, depth, …

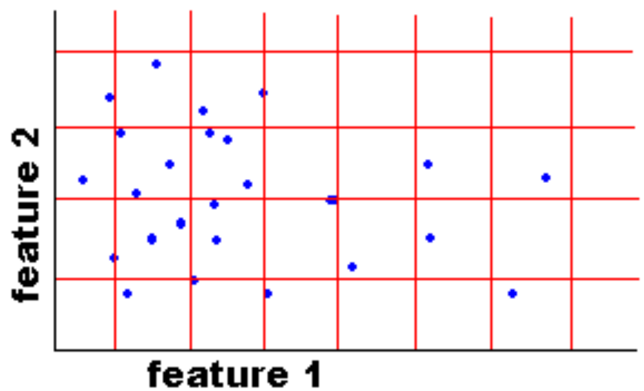
Images from Dave Kauchak

Image Representations: Histograms

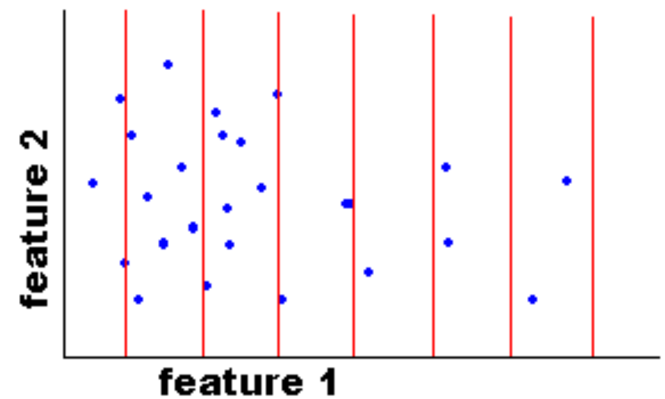
Histogram: Probability or count of data in each bin







- Joint histogram
  - Requires lots of data
  - Loss of resolution to avoid empty bins



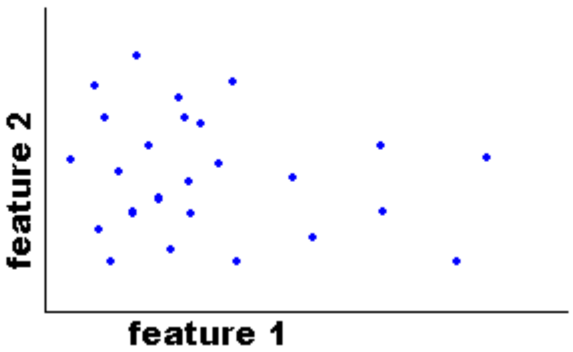
Marginal histogram

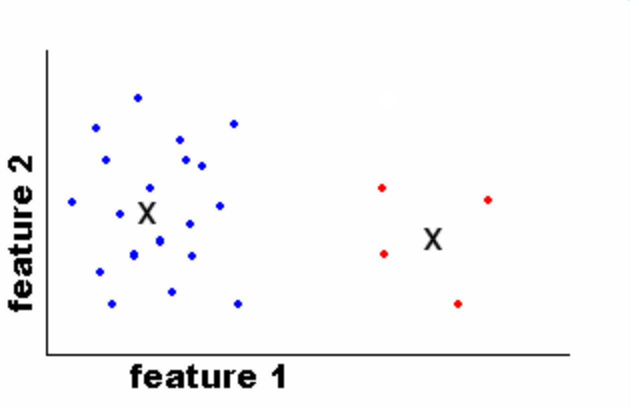
- Requires independent features
- More data/bin than joint histogram

Images from Dave Kauchak

Image Representations: Histograms

Clustering



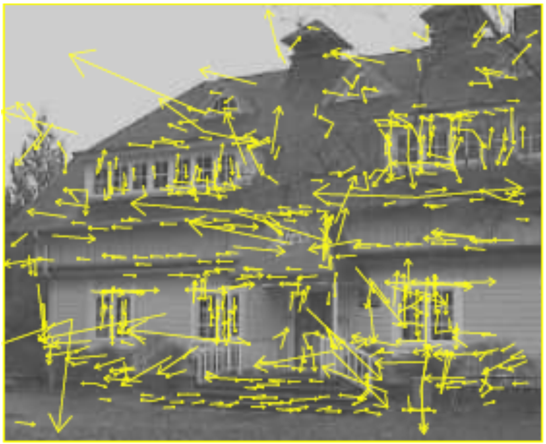


Use the same cluster centers for all images

Images from Dave Kauchak

What kind of things do we compute histograms of?

- Histograms of oriented gradients



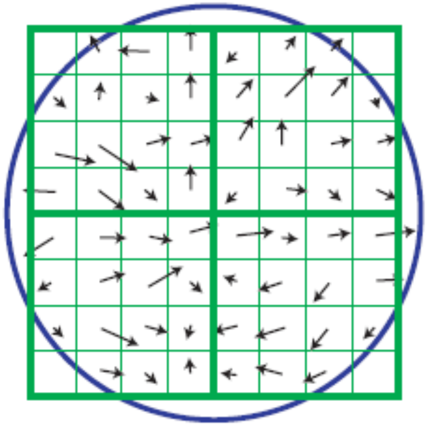
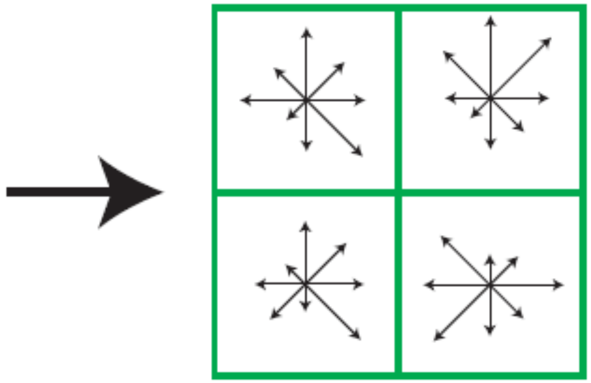


Image gradients



Keypoint descriptor

SIFT – Lowe IJCV 2004

SIFT vector formation

- Computed on rotated and scaled version of window according to computed orientation & scale
  - resample the window
- Based on gradients weighted by a Gaussian of variance half the window (for smooth falloff)

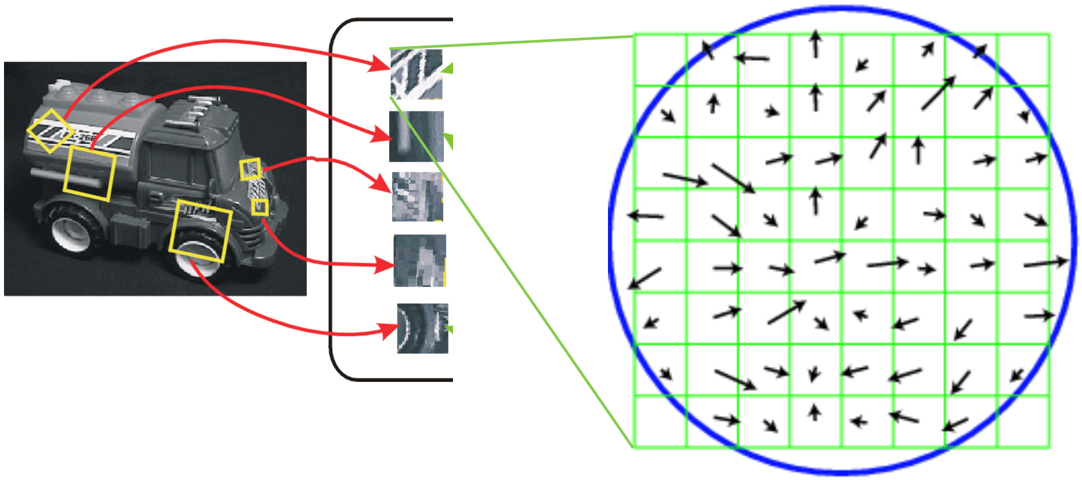
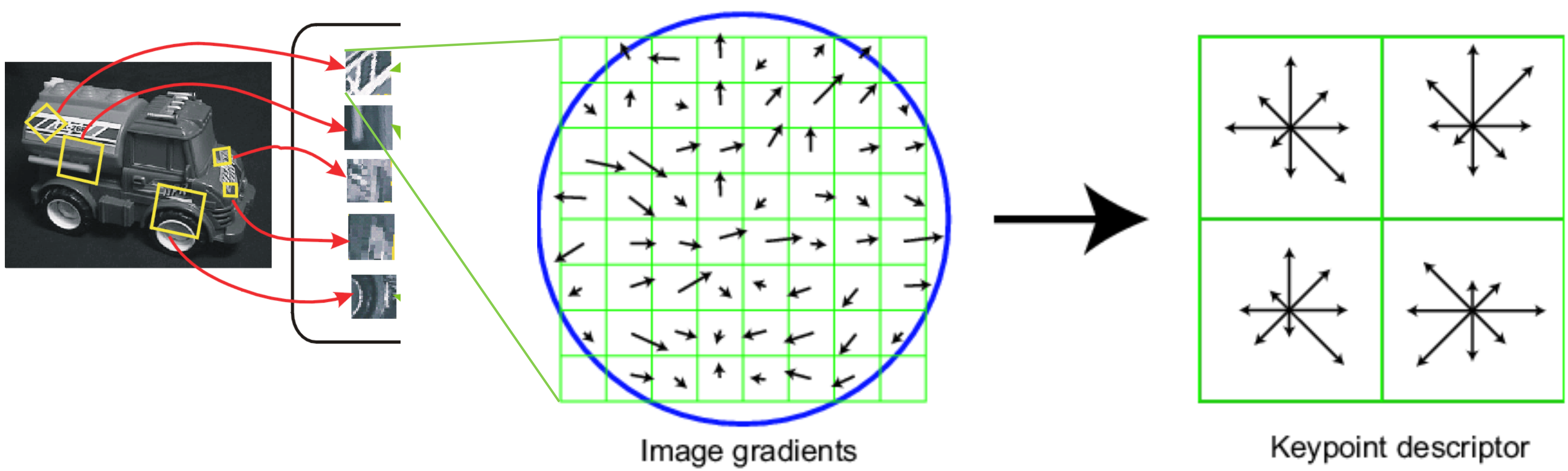


Image gradients

SIFT vector formation

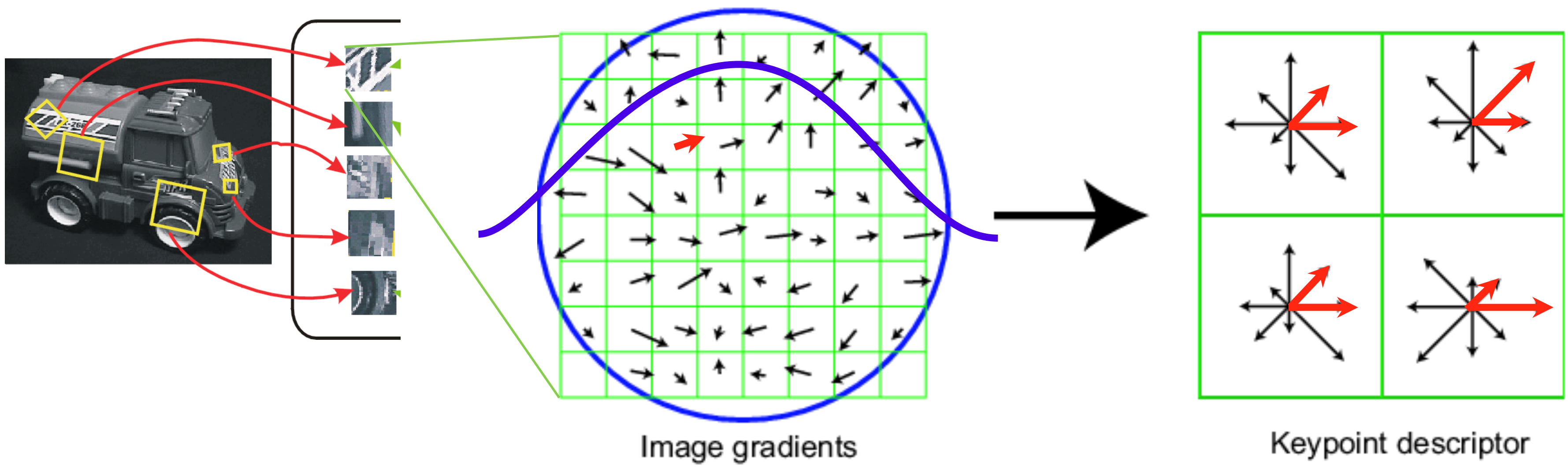
- 4x4 array of gradient orientation histogram weighted by magnitude
- 8 orientations x 4x4 array = 128 dimensions
- Motivation: some sensitivity to spatial layout, but not too much.



showing only 2x2 here but is 4x4

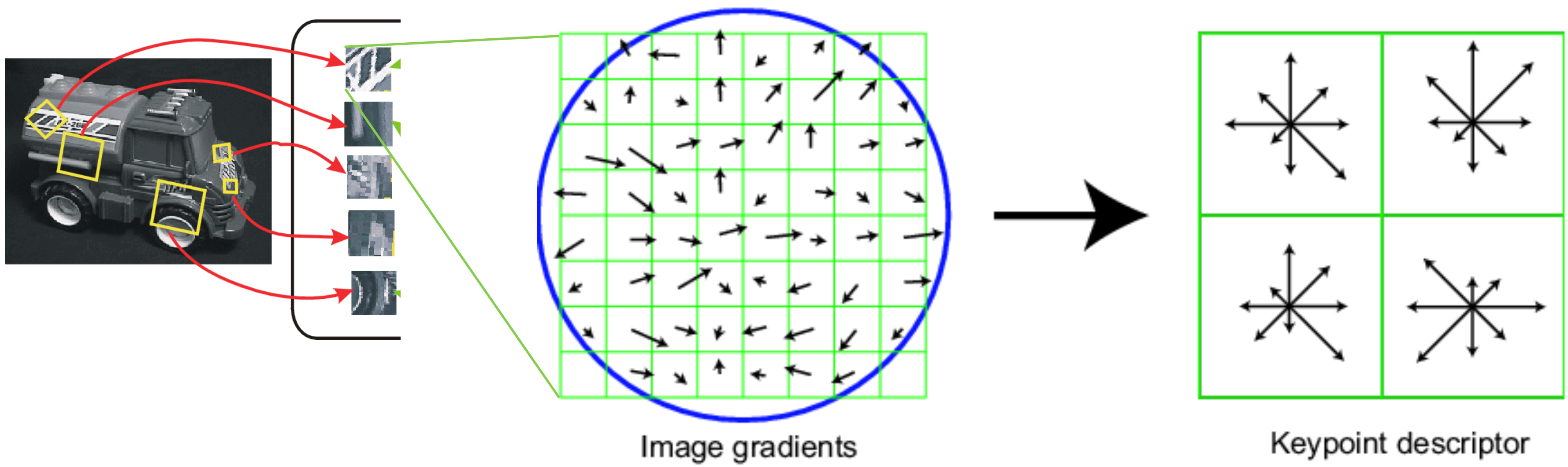
Ensure smoothness

- Gaussian weight
- Interpolation
  - a given gradient contributes to 8 bins: 4 in space times 2 in orientation

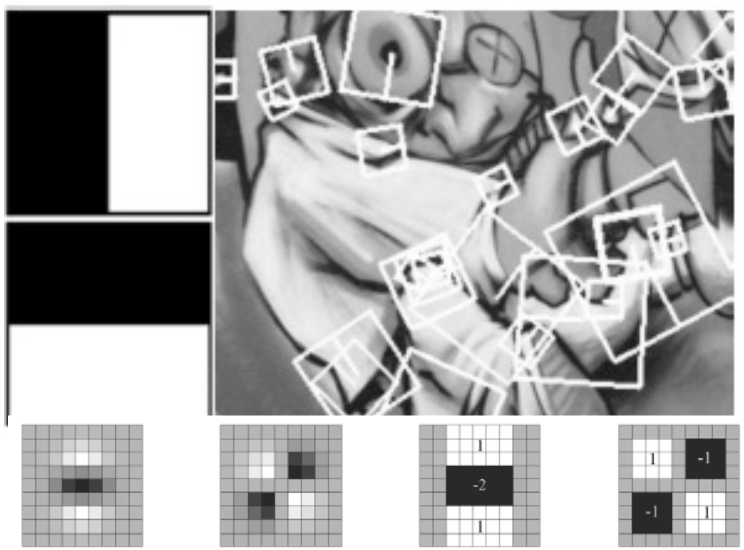


Reduce effect of illumination

- 128-dim vector normalized to 1
- Threshold gradient magnitudes to avoid excessive influence of high gradients
  - after normalization, clamp gradients >0.2
  - renormalize



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**

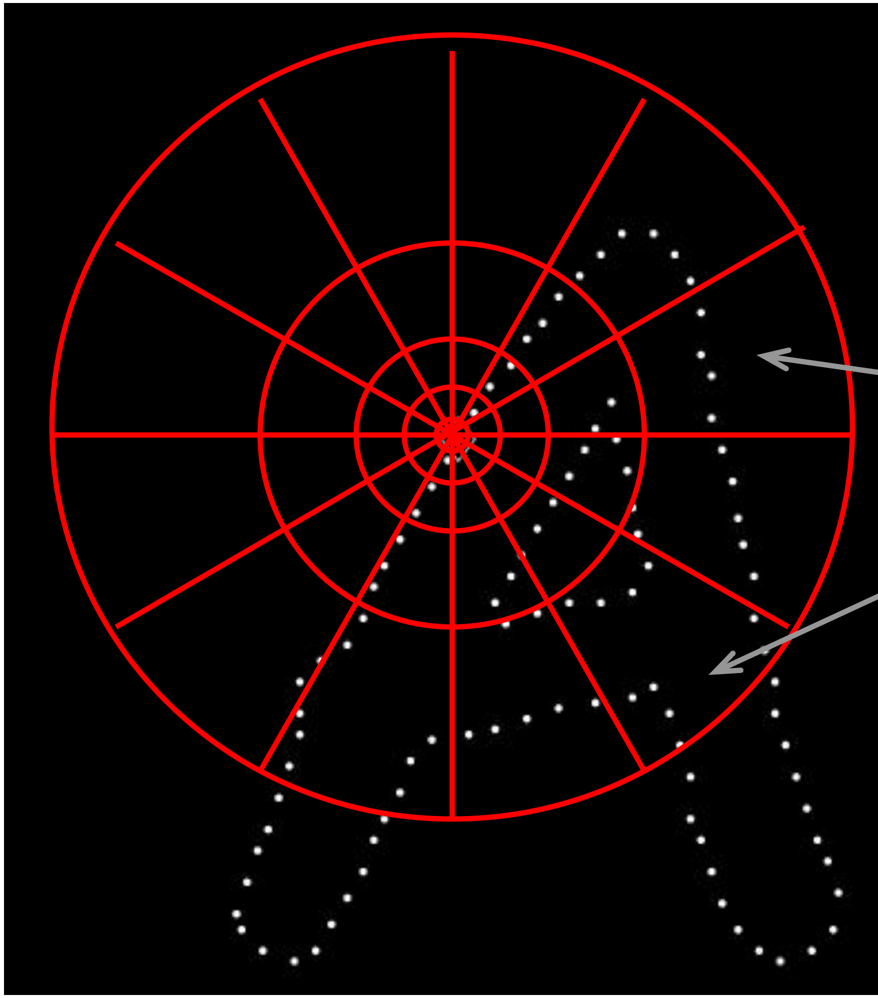
GPU implementation available

- **Feature extraction @ 200Hz**
  (detector + descriptor, 640×480 img)
- http://www.vision.ee.ethz.ch/~surf

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

K. Grauman, B. Leibe

Local Descriptors: Shape Context



Count the number of points inside each bin, e.g.:

Count = 4

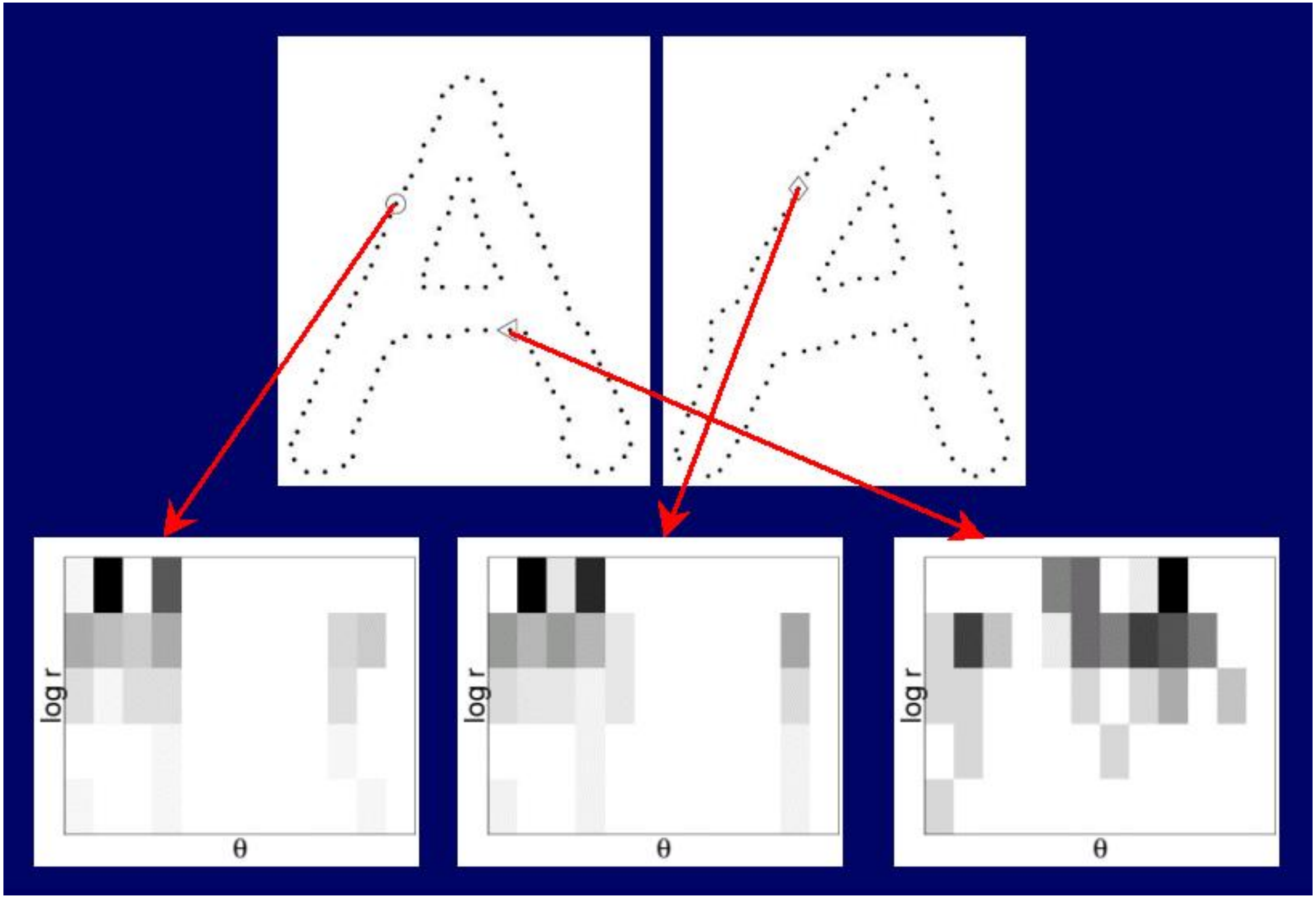
⋮

Count = 10

Log-polar binning: more precision for nearby points, more flexibility for farther points.

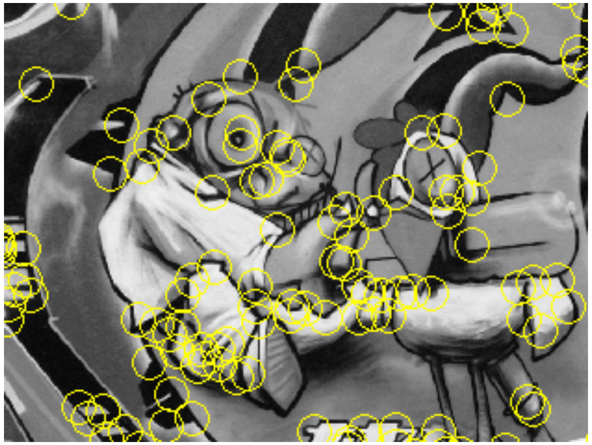
Belongie & Malik, ICCV 2001

Shape Context Descriptor

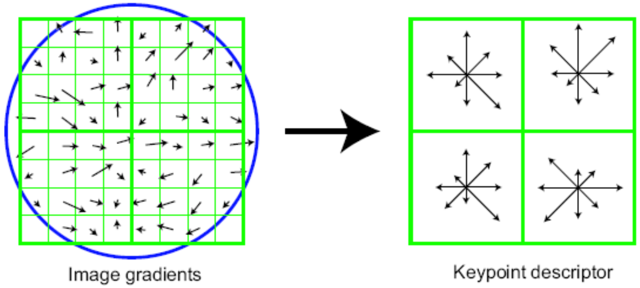


Things to remember

- Keypoint detection: repeatable and distinctive
  - Corners, blobs, stable regions
  - Harris, DoG



- Descriptors: robust and selective
  - spatial histograms of orientation
  - SIFT



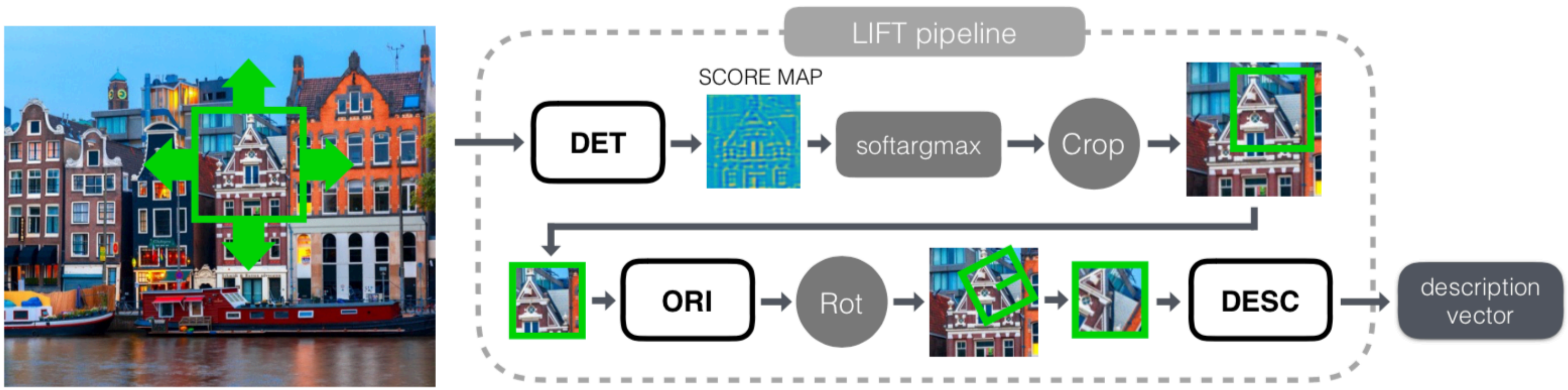
Deep Descriptors

LIFT: Learned Invariant Feature Transform

ECCV 2016

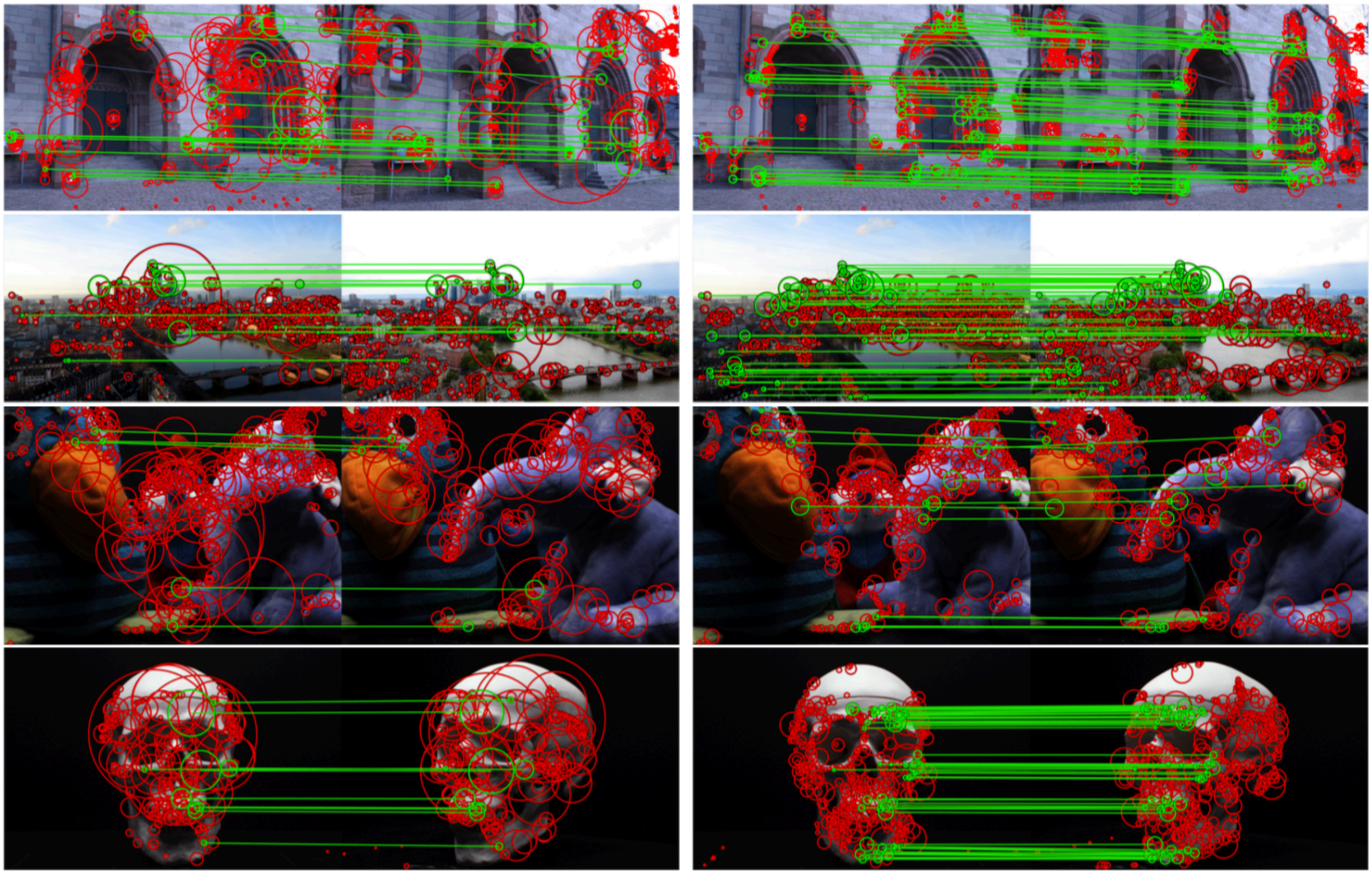
Kwang Moo Yi[*,1], Eduard Trulls[*,1], Vincent Lepetit[2], Pascal Fua[1]

[1]Computer Vision Laboratory, Ecole Polytechnique Fédérale de Lausanne (EPFL)
[2]Institute for Computer Graphics and Vision, Graz University of Technology



- Three networks: detection, orientation, description
- detection+orientation -> STN -> descriptor
- Trained separately :-(

SIFT vs. LIFT



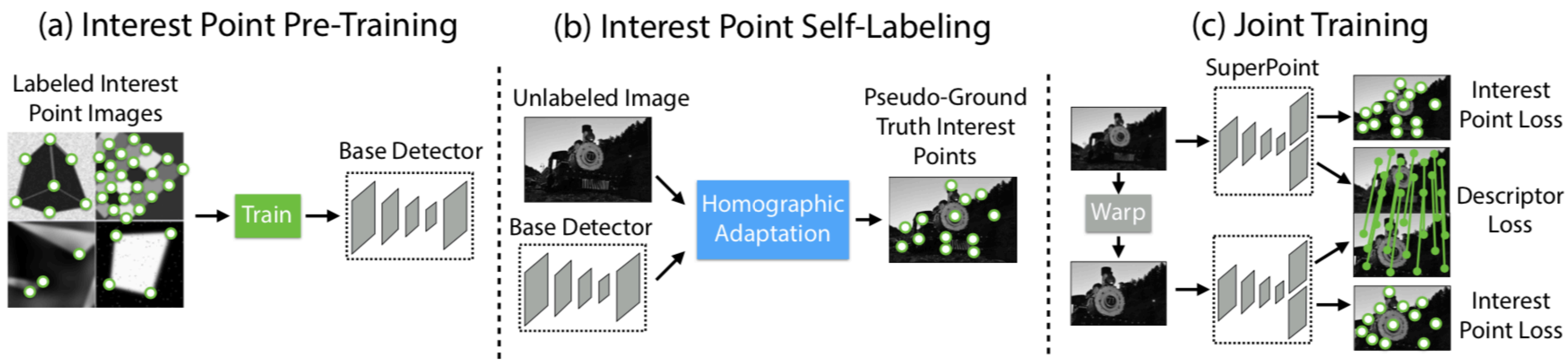
SuperPoint: Self-Supervised Interest Point Detection and Description

2018 CVPR Workshop

Daniel DeTone
Magic Leap
Sunnyvale, CA

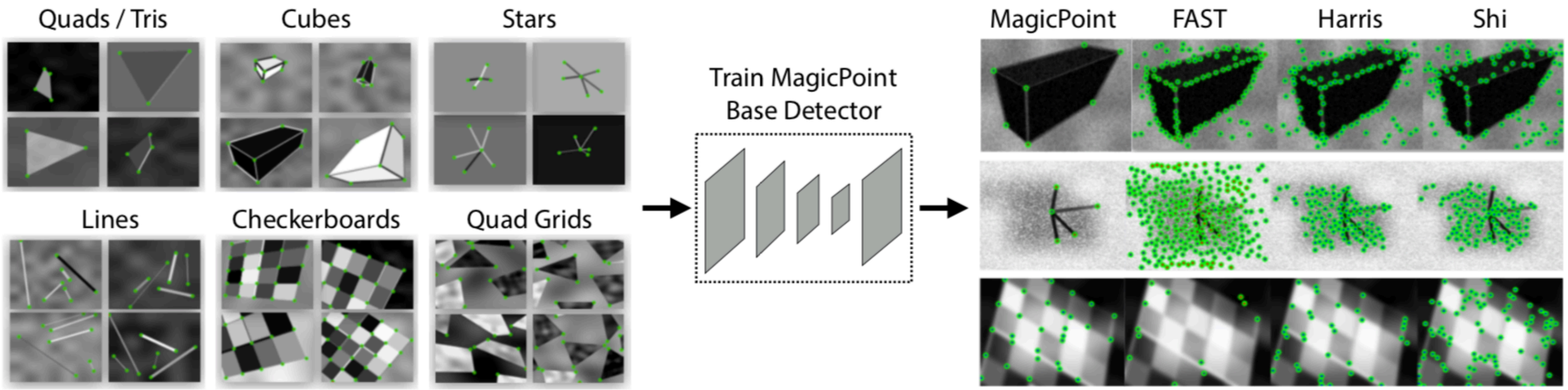
Tomasz Malisiewicz
Magic Leap
Sunnyvale, CA

Andrew Rabinovich
Magic Leap
Sunnyvale, CA

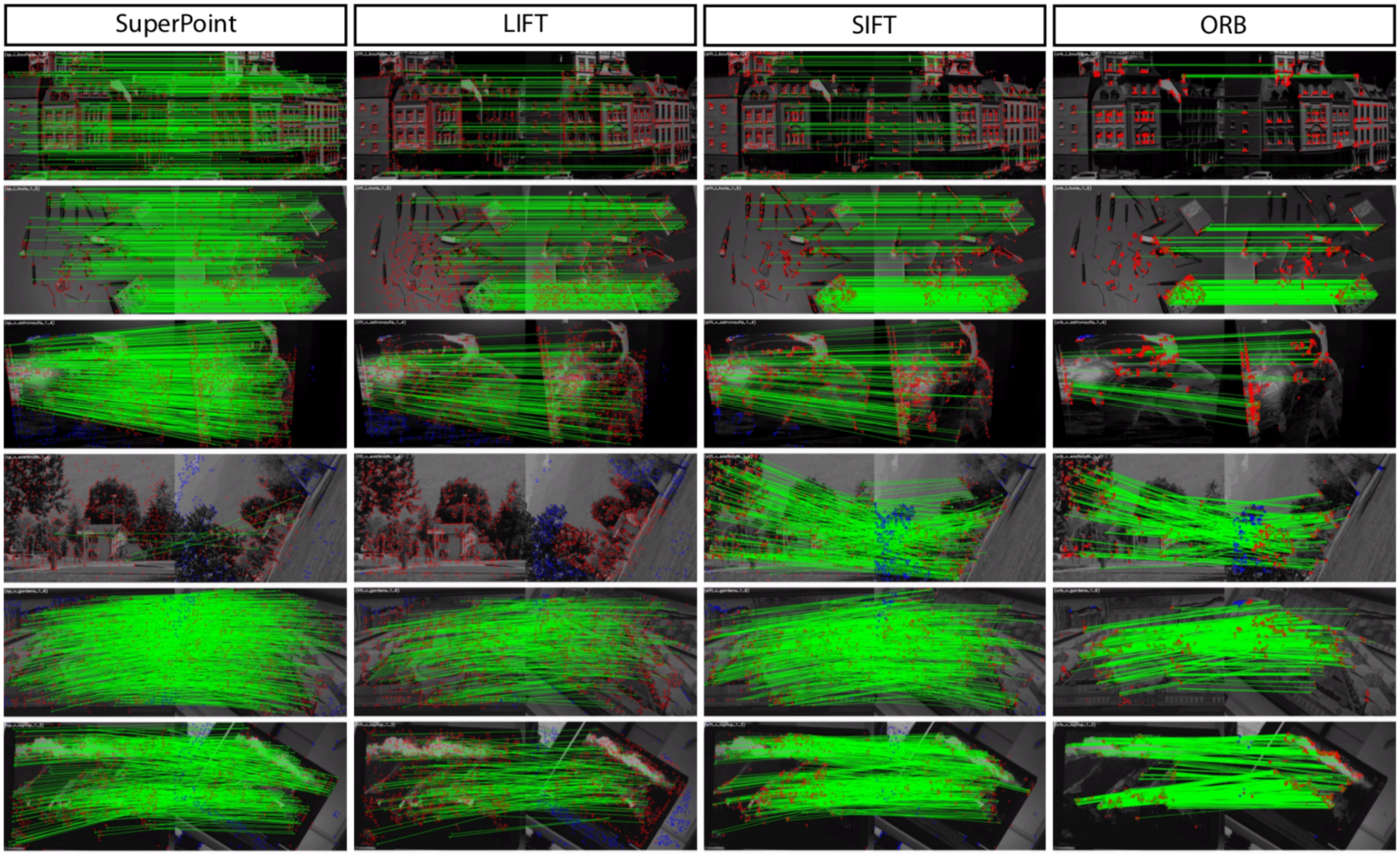


- Interest point = ill-defined -> self-supervised
- MagicPoint -> SuperPoint

MagicPoint


Quads / Tris
Cubes
Stars
Lines
Checkerboards
Quad Grids
Train MagicPoint
Base Detector
MagicPoint
FAST
Harris
Shi

SuperPoint Results

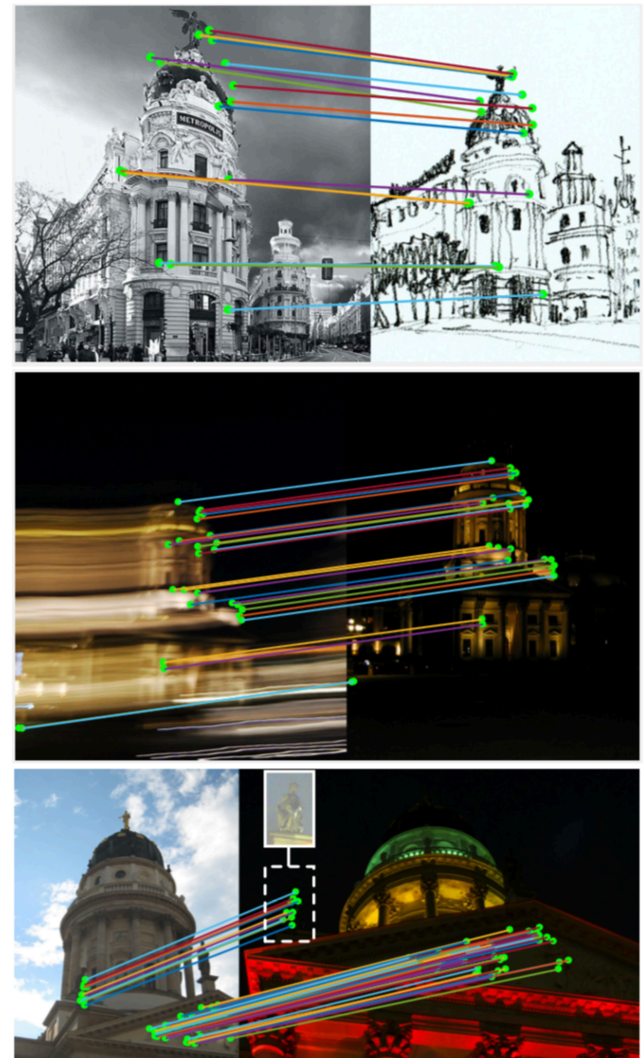


D2-Net: A Trainable CNN for *Joint Description and Detection* of Local Features

CVPR 2019

Mihai Dusmanu[1,2,3] Ignacio Rocco[1,2] Tomas Pajdla[4] Marc Pollefeys[3,5]
Josef Sivic[1,2,4] Akihiko Torii[6] Torsten Sattler[7]

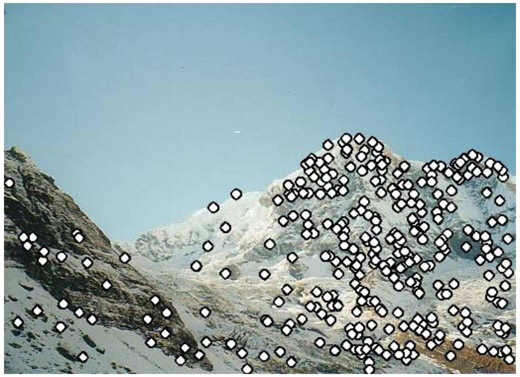
- Tensor viewed as descriptors and detector maps
- VGG16-based, loss encourages distinctiveness and repeatability
- Results beat the star of the art in day-night and indoor localization, but not in more traditional settings (Superpoint shines for HPatches, **GeoDesc** for SFM)



Matching

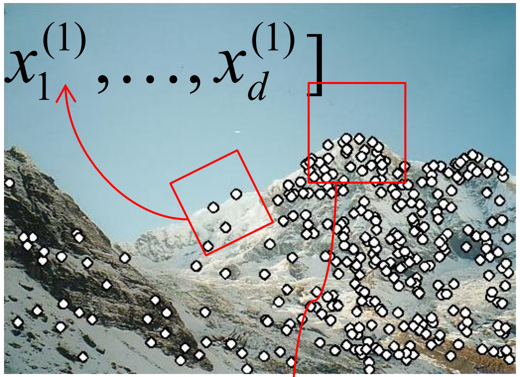
Local features: main components

1) **Detection:** Identify the interest points



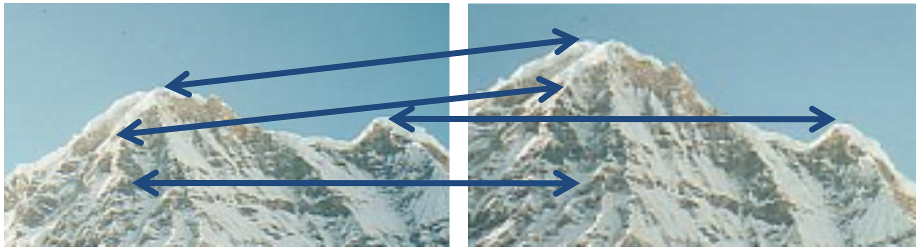
2) **Description:** Extract vector feature descriptor surrounding each interest point.

$$\mathbf{x}_1 = [x_1^{(1)}, \ldots, x_d^{(1)}]$$



$$\mathbf{x}_2 = [x_1^{(2)}, \ldots, x_d^{(2)}]$$

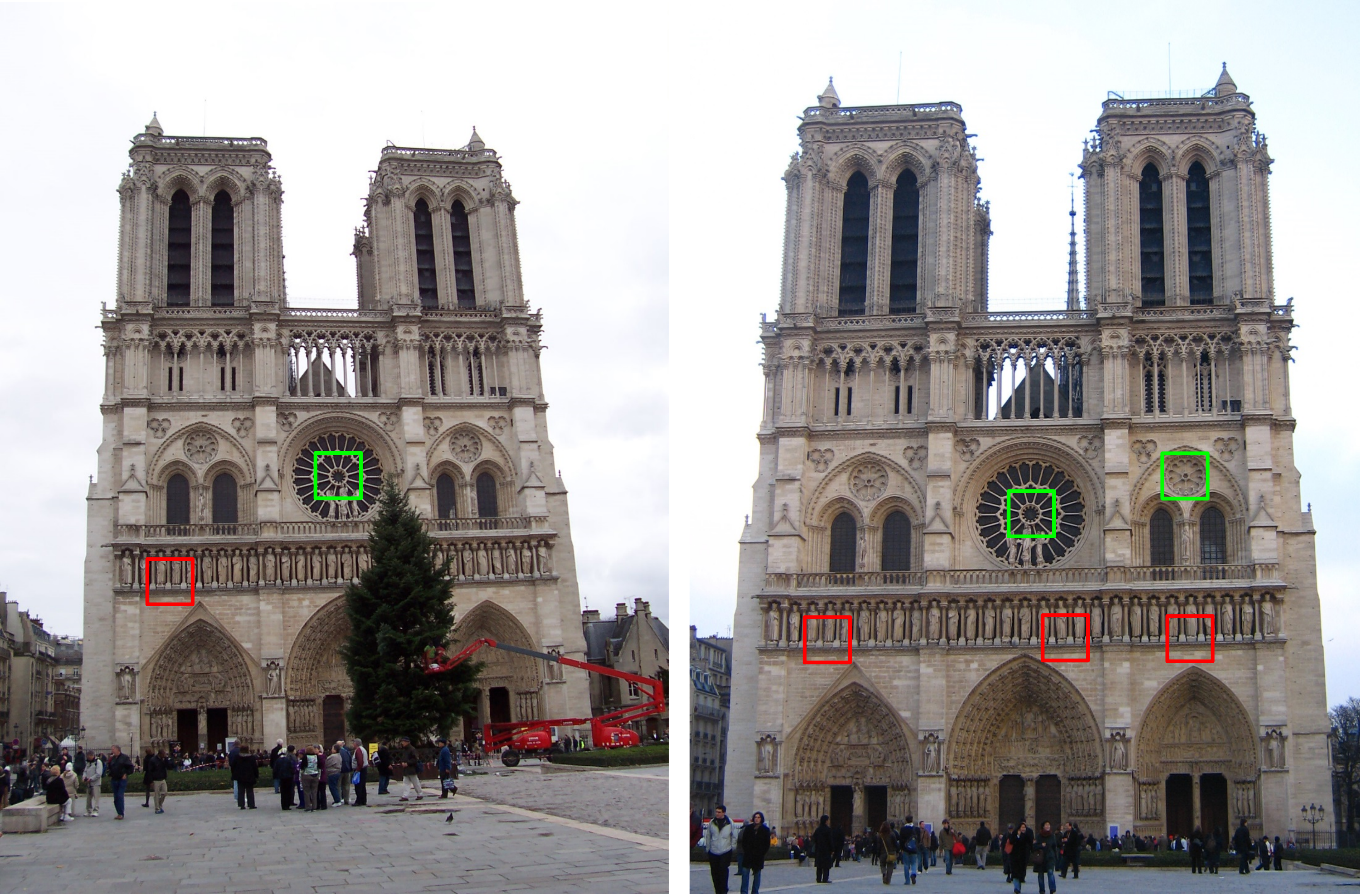
3) **Matching: Determine correspondence between descriptors in two views**



Kristen Grauman

Matching

- Simplest approach: Pick the nearest neighbor. Threshold on absolute distance
- Problem: Lots of self similarity in many photos



Distance: 0.34, 0.30, 0.40

Distance: 0.61

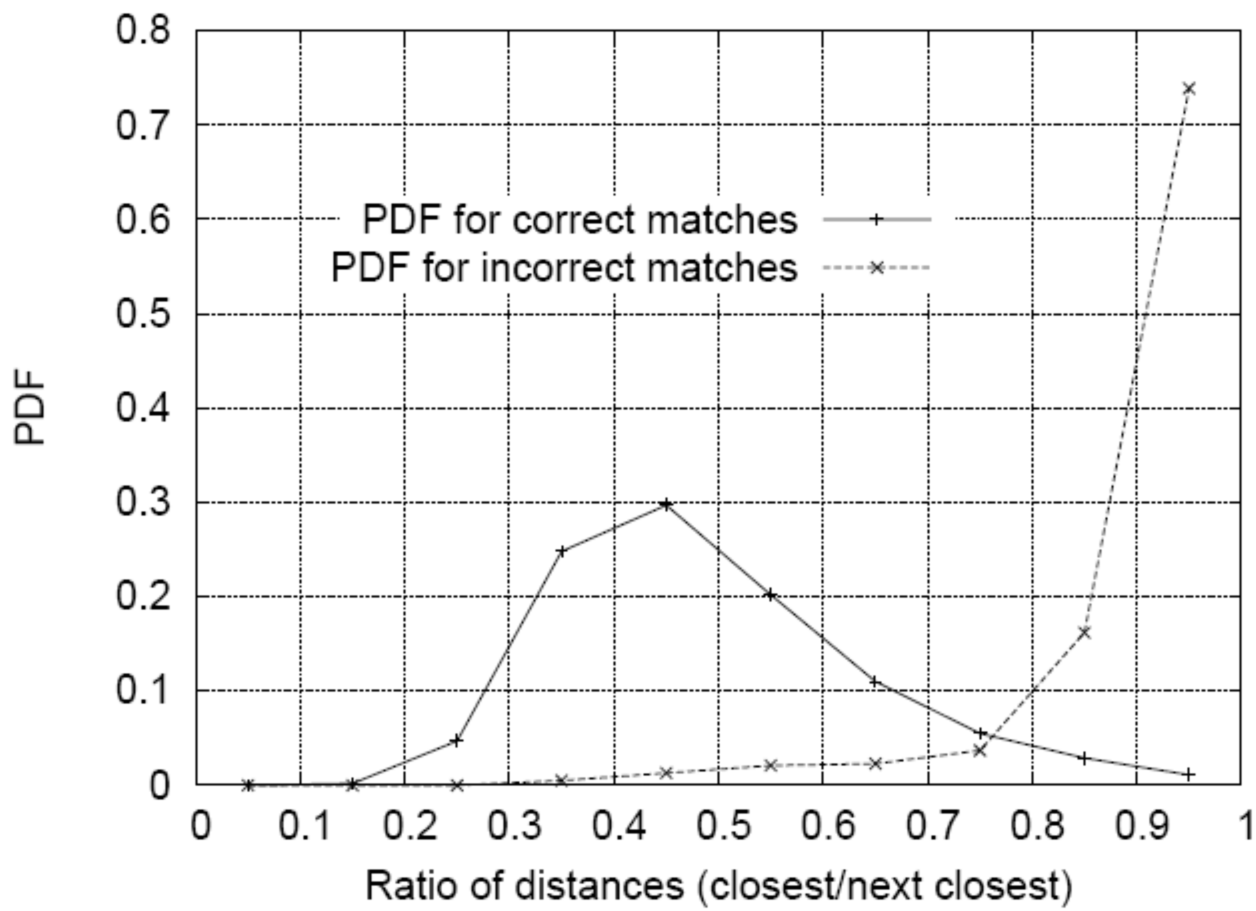
Distance: 1.22

Nearest Neighbor Distance Ratio

- $\frac{NN1}{NN2}$ where NN1 is the distance to the first nearest neighbor and NN2 is the distance to the second nearest neighbor.
- Sorting by this ratio puts matches in order of confidence.

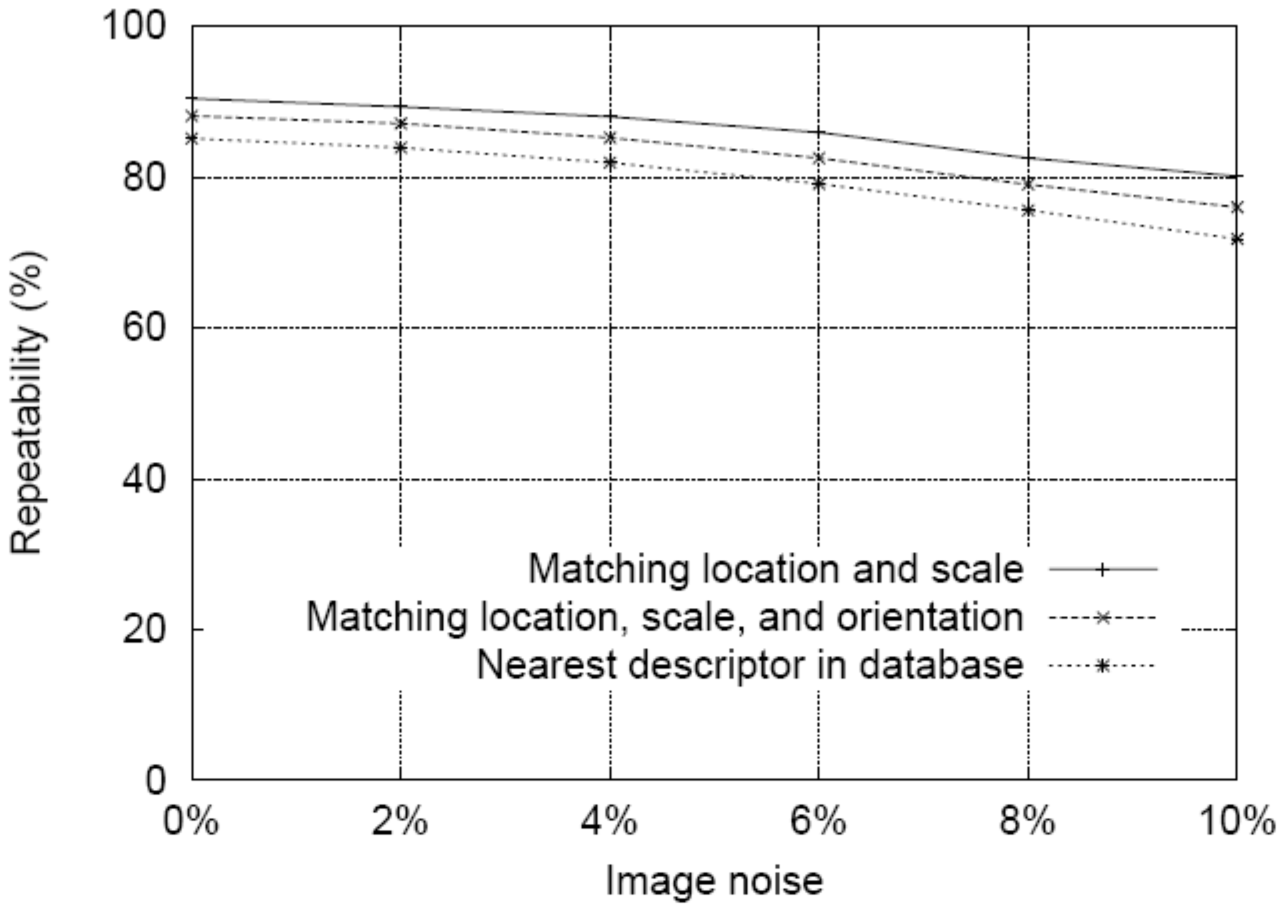
Matching Local Features

- Nearest neighbor (Euclidean distance)
- Threshold ratio of nearest to 2nd nearest descriptor



Lowe IJCV 2004

SIFT Repeatability



Lowe IJCV 2004

SIFT Repeatability