Image Descriptors

- SIFT
- RANSAC
- Sparse descriptors
- Dense descriptors

Recall: Harris interest points

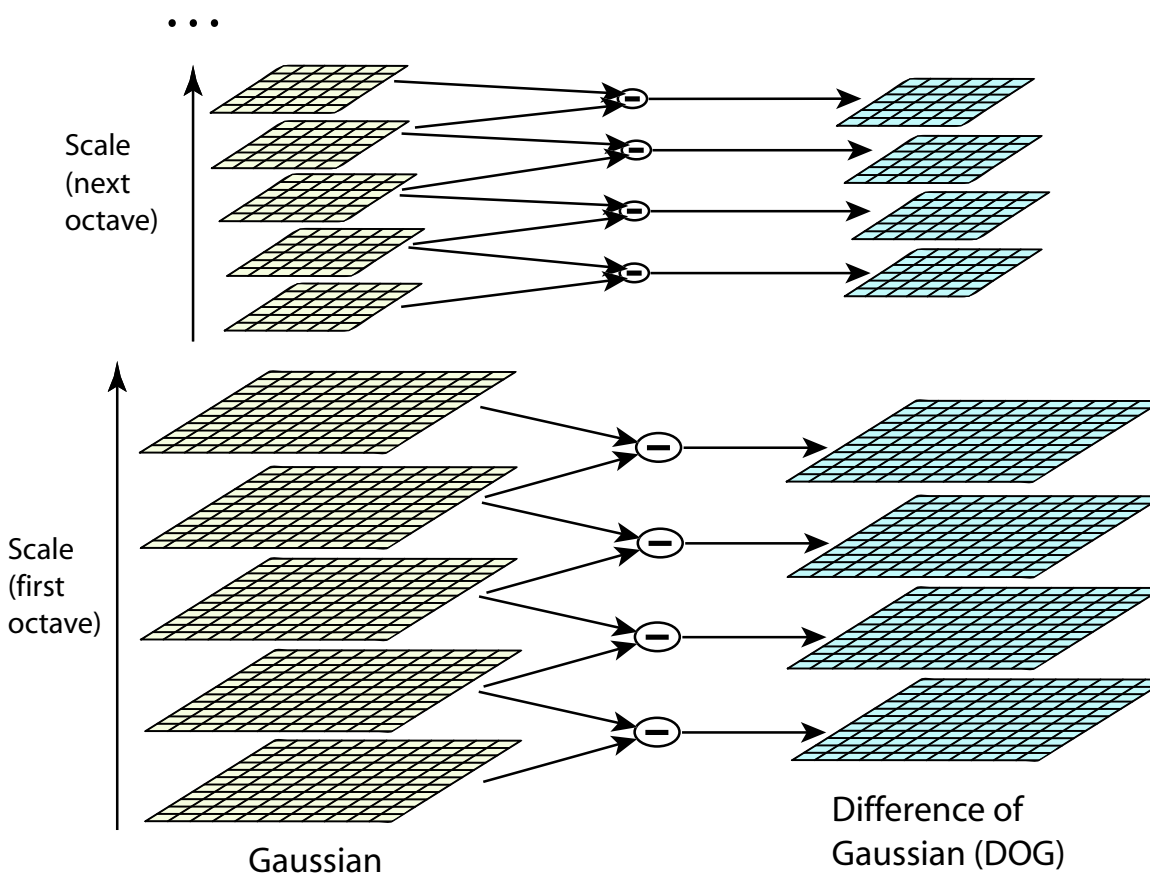
$$A(x, y, \sigma) = \sum_{x,y} \begin{bmatrix} \mathbf{I}_x(\sigma)^2 & \mathbf{I}_x\mathbf{I}_y(\sigma) \\ \mathbf{I}_y\mathbf{I}_x(\sigma) & \mathbf{I}_y^2(\sigma) \end{bmatrix}$$

second-moment matrix

$$\text{cornerness}(x, y, \sigma) = det(A(x, y, \sigma)) - \alpha Trace^2(A(x, y, \sigma))$$

Alternative: blob-based interest points

https://en.wikipedia.org/wiki/Scale-invariant_feature_transform



$$D(x, y, \sigma) \quad = \quad (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y)$$

$k = 2^{\frac{1}{s}}$ where s = # levels in an octave

Look for “blob detections” that are

locally maximal, high confidence, and localizeable

Local maxima of D(x,y,sigma)

D(x,y,sigma) > thresh

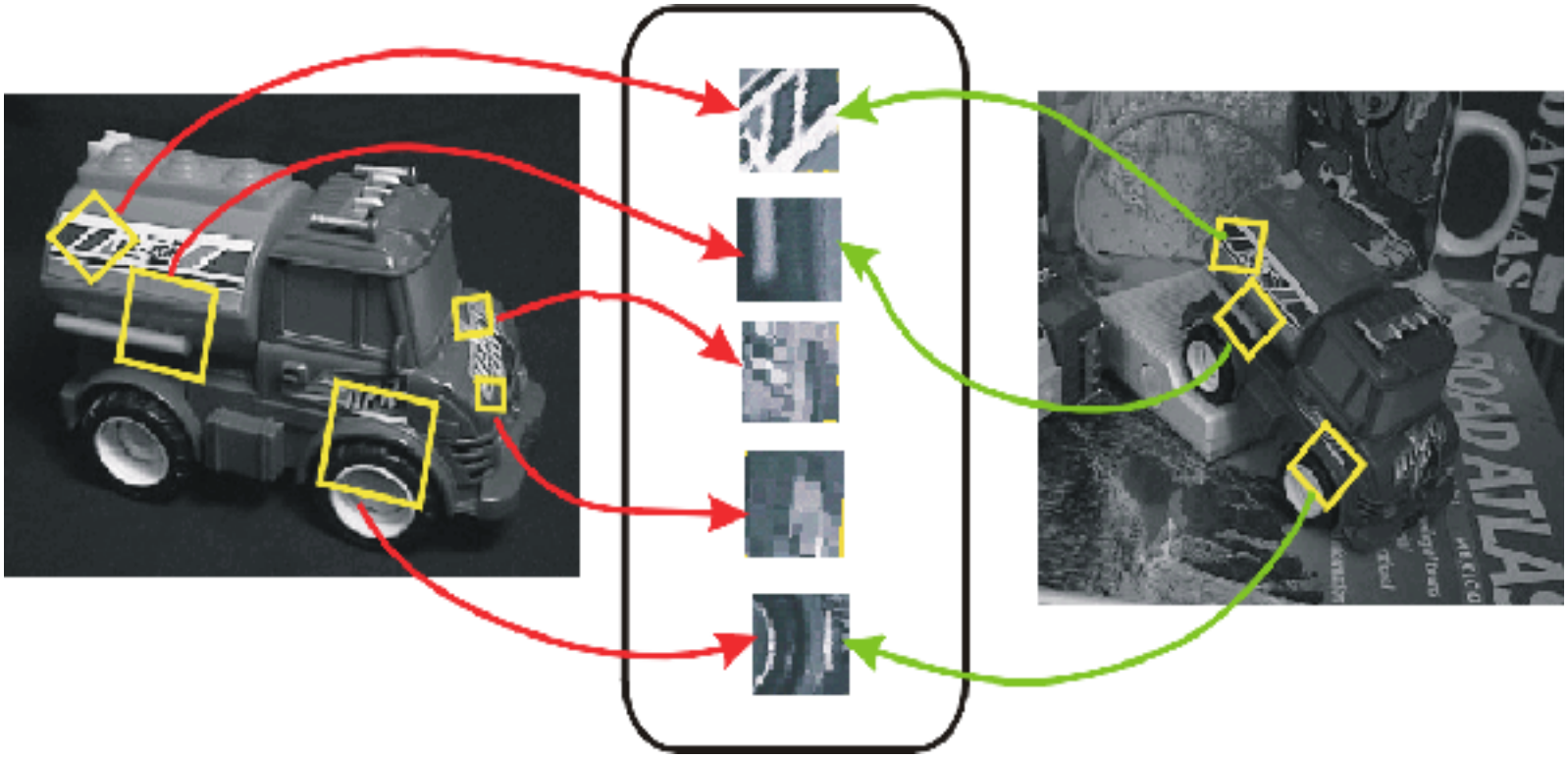
$$\begin{bmatrix} D_{xx} & D_{xy} \\ D_{yx} & D_{yy} \end{bmatrix}$$

min eigenvalue of Hessian > thresh

Added benefit of Hessian: use second-order taylor expansion to get “subpixel” accuracy

https://en.wikipedia.org/wiki/Scale-invariant_feature_transform

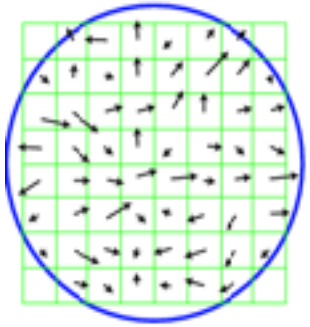
Coordinate frames

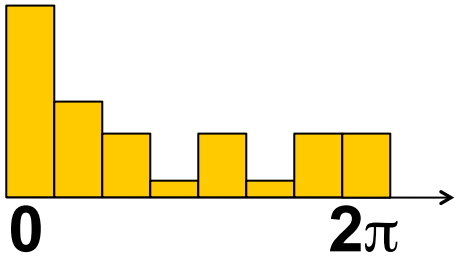


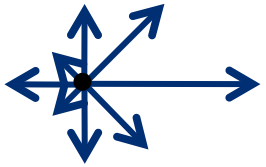
Represent each patch in a canonical scale and orientation (or general *affine* coordinate frame)

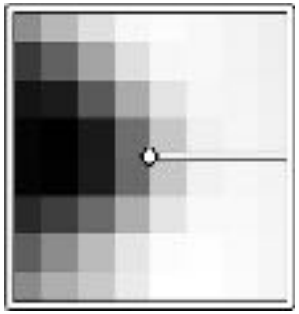
Find dominant orientation

Compute gradients for all pixels in patch. Histogram (bin) gradients by orientation



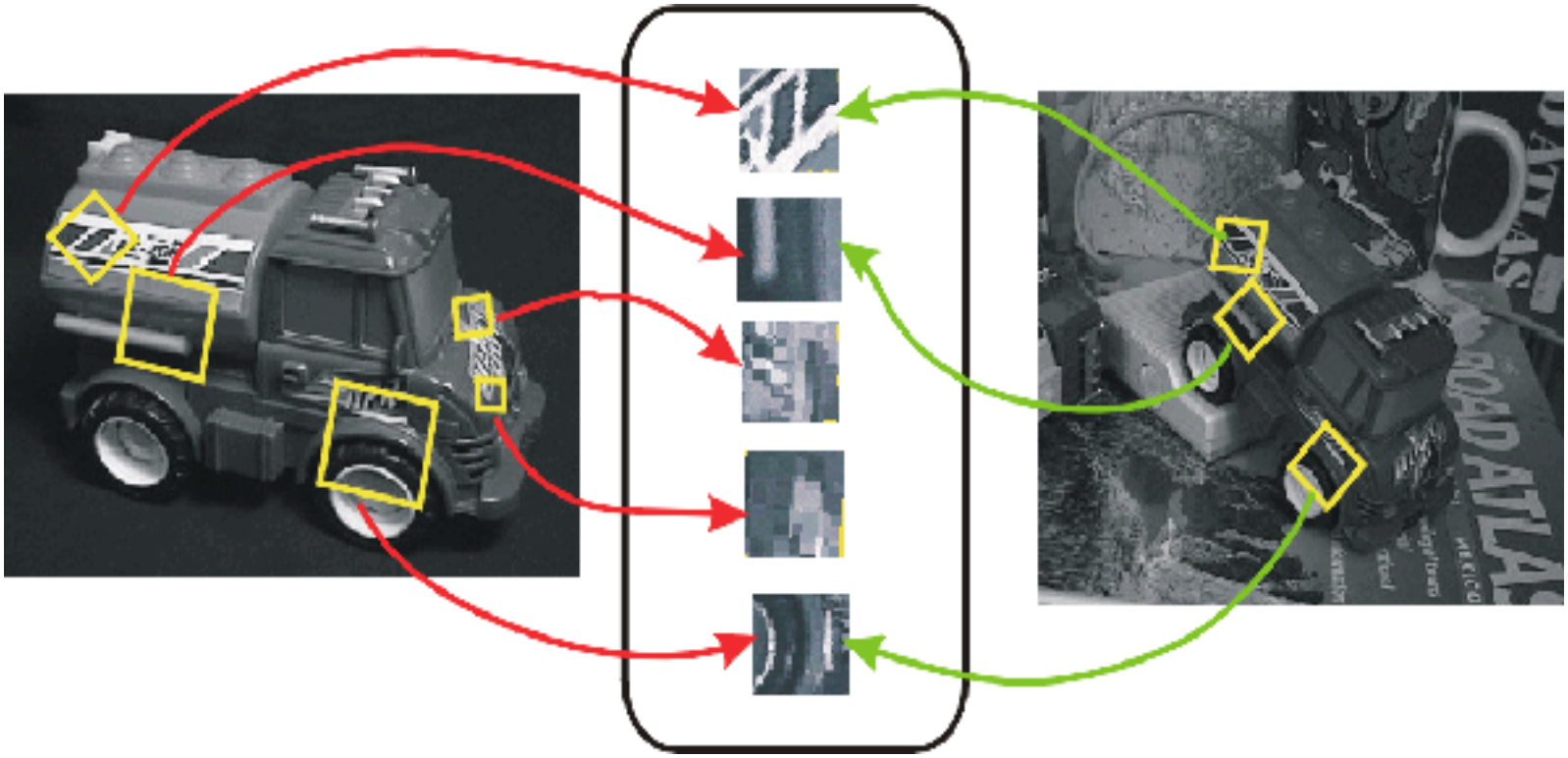






(I prefer this because you can look for multiple peaks)

Appearance descriptors



Represent each patch in a canonical scale and orientation (or general *affine* coordinate frame)

$$d(p_1, p_2) = \left\| \begin{bmatrix} \cdot \\ \cdot \end{bmatrix} - \begin{bmatrix} \cdot \\ \cdot \end{bmatrix} \right\|$$

Conflicting criteria

Discriminative: d(p1,p2) should be high when points p1 and p2 do not correspond

Repeatable: d(p1,p2) should be low when p1 and p2 correspond (but are different viewpoints, illumunations,… of same point)

Speed: Fast to compute p's

Storage: Low dimensional (easy to store p's)

Computing the SIFT Descriptor

Histograms of gradient directions over spatial regions

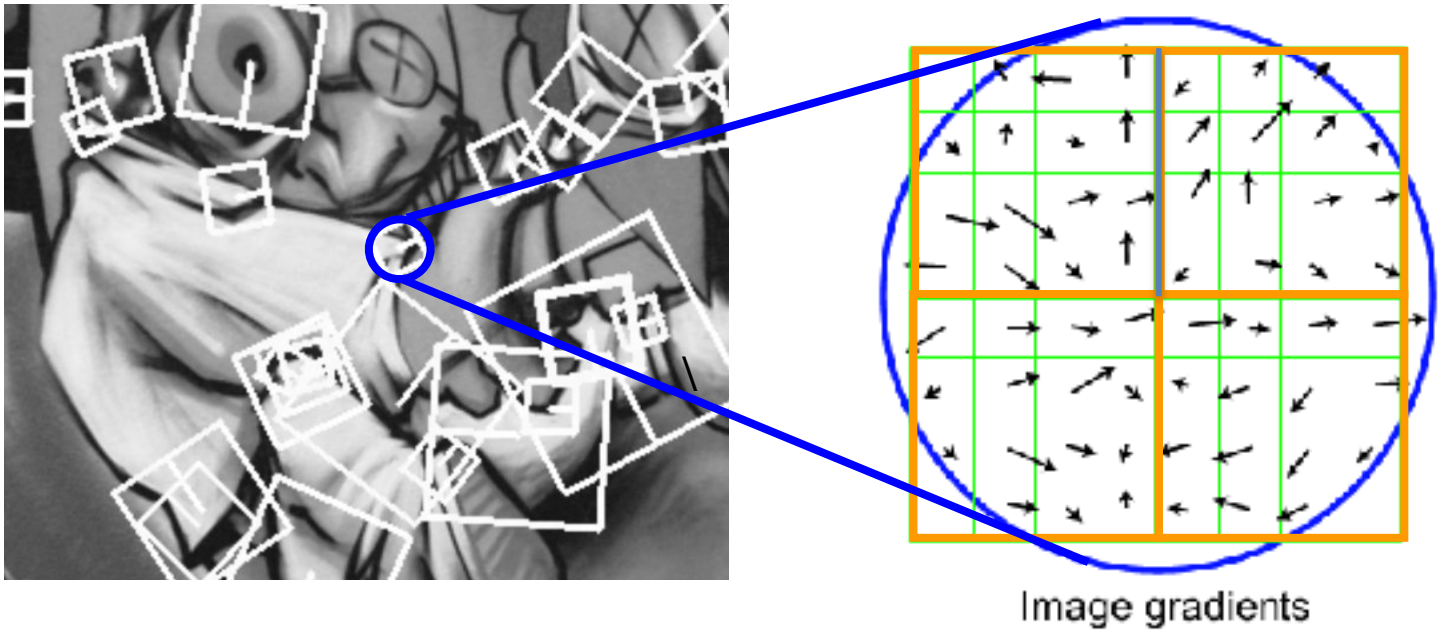
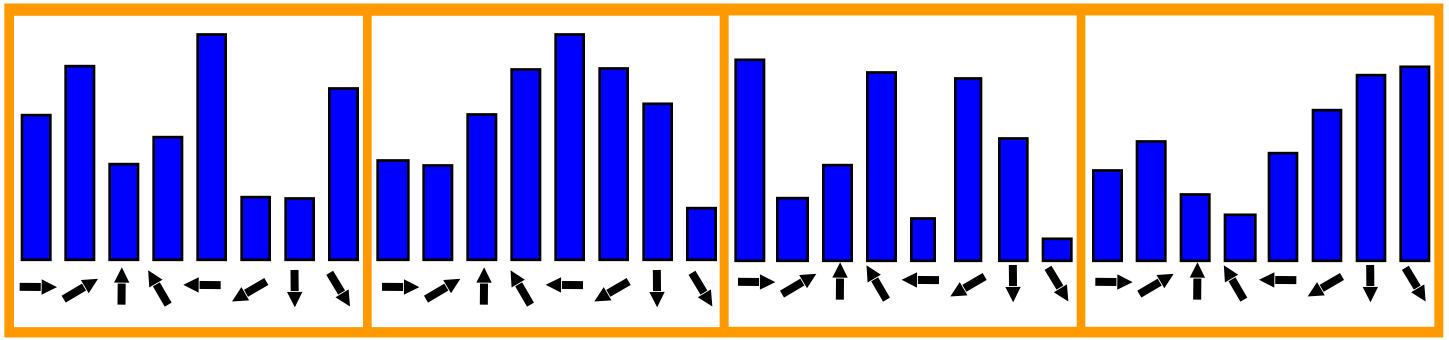


Image gradients

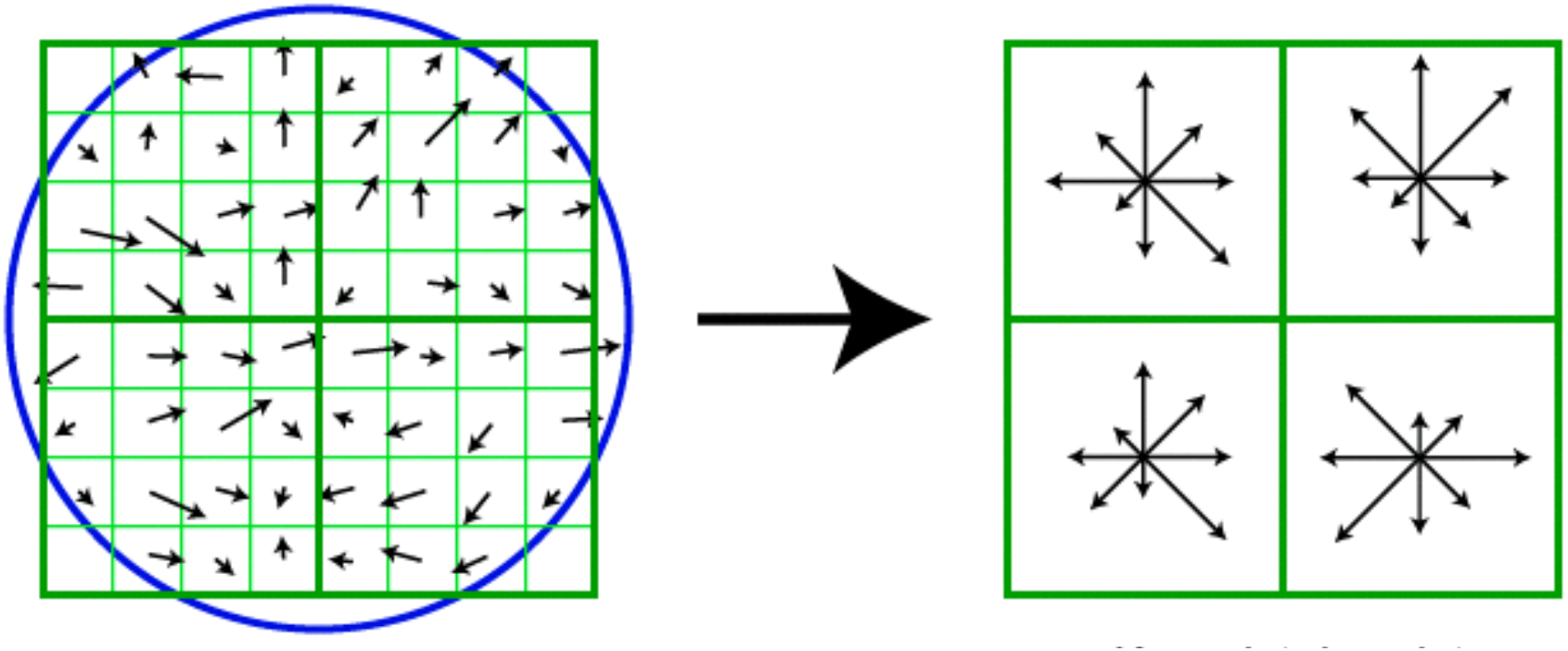


SIFT Descriptor - Details

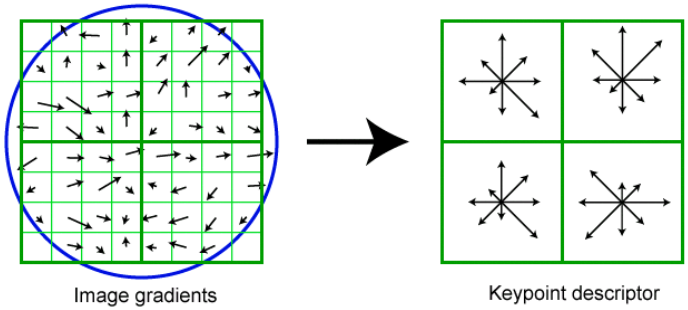
Assume that we bin “o” orientations over “kxk” cells, where each cell is “sxs” pixels

What’s final dimensionality of descriptor?

Common visualizations: o=8, k=2, s=4
Common implementations: o=8, k=4, s=4



Analogy with spatial pyramids and bag-of-words

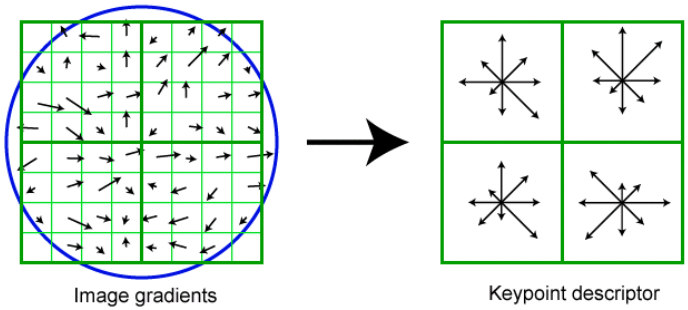


1. Use a pre-defined dictionairy of 8 oriented edges

2. Pixels that closely match their dictionairy element get a larger vote

“sparse” reconstructions rather than binary ones!

Post-processing



1. Rescale 128-dim vector to have unit norm

$$x = \frac{x}{||x||}, \quad x \in R^{128}$$

"invariant to linear scalings of intensity"

2. Clip high values

$$x := \min(x, .2)$$

$$x := \frac{x}{||x||}$$

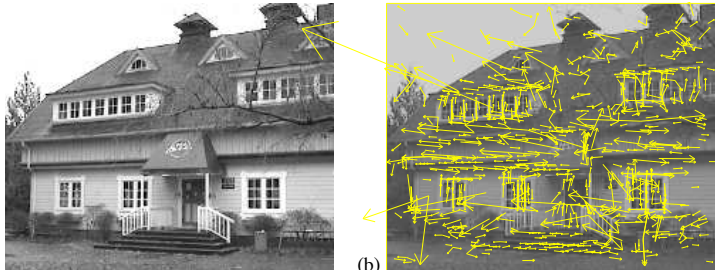
approximate binarization allows for for flat patches with small gradients to remain stable

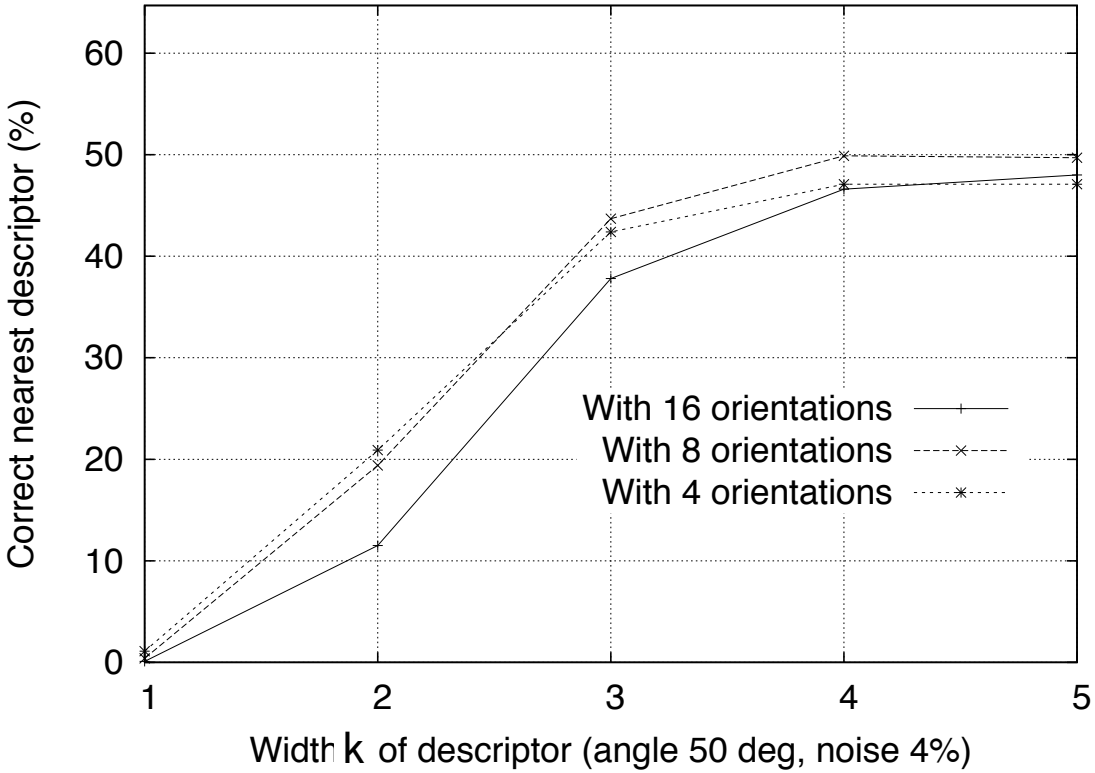
Evaluation

Historic problem in computer vision:
“wide-baseline matching”

SIFT

What made this work? Exhaustive evaluation of hyper-parameters on annotated dataset





SIFT

Where does scale, translation, and rotation invariance for the "Scale-Invariant Feature Transform" come from?

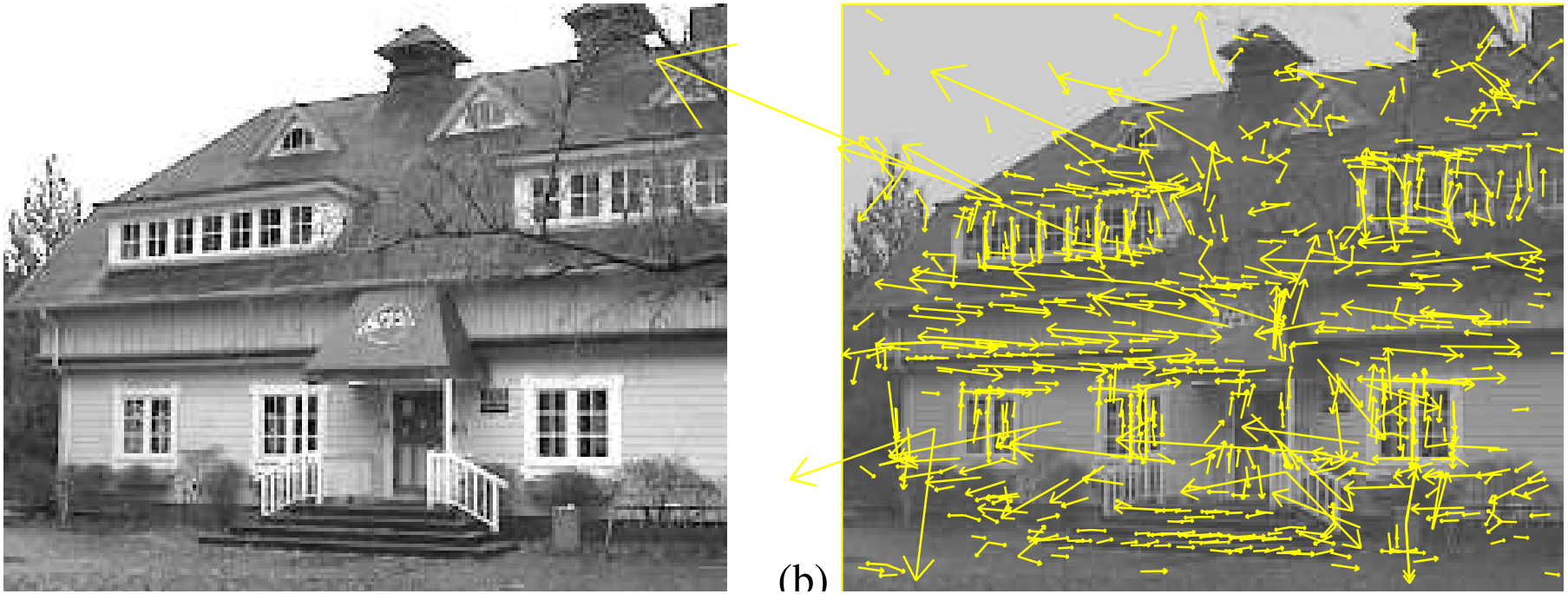
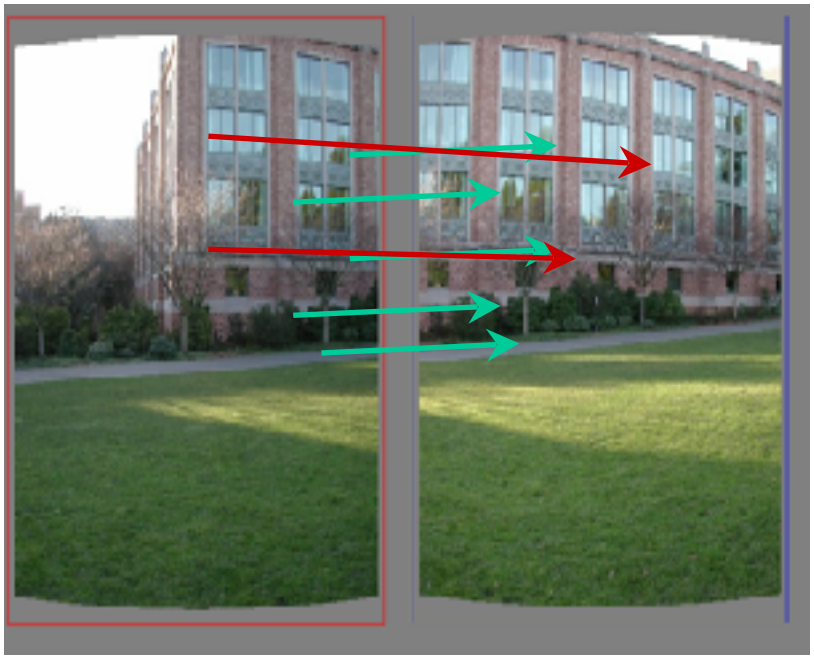


Image Descriptors

- SIFT
- RANSAC
- Sparse descriptors
- Dense descriptors

Feature matching

- Exhaustive search
  - for each feature in one image, look at *all* the other features in the other image(s)



How do we remove bad matches?

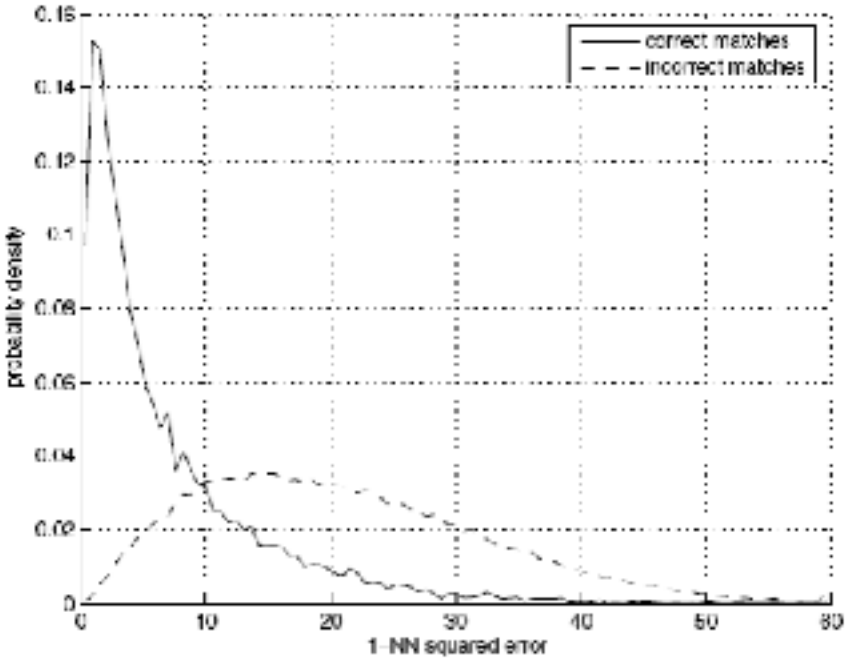
Repeated textures (like windows) are notoriously challenging!

Feature-space outlier rejection

Let's not match all features, but only these that have "similar enough" matches?

How can we do it?

- Distance(SIFT(patch1),SIFT(patch2)) < threshold
- Clean way to set threshold?

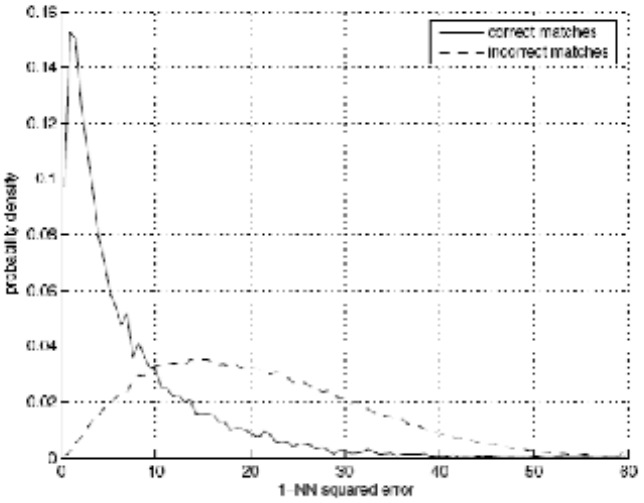


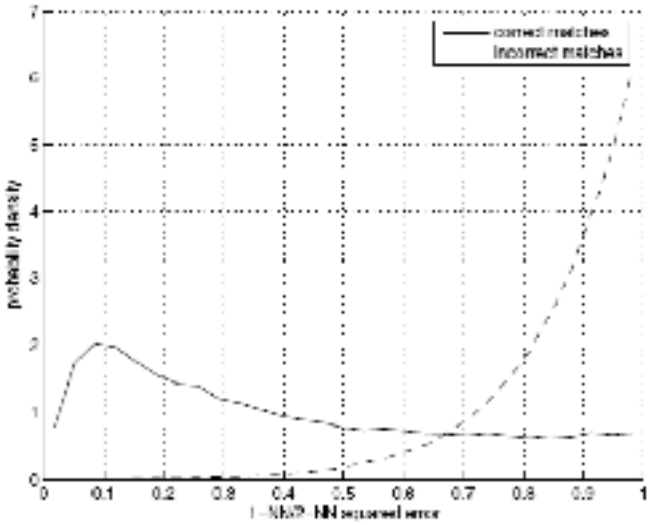
Rejecting unstable matches

"Look for patches that have a good match. If there's a good alternative match, don't trust it!"

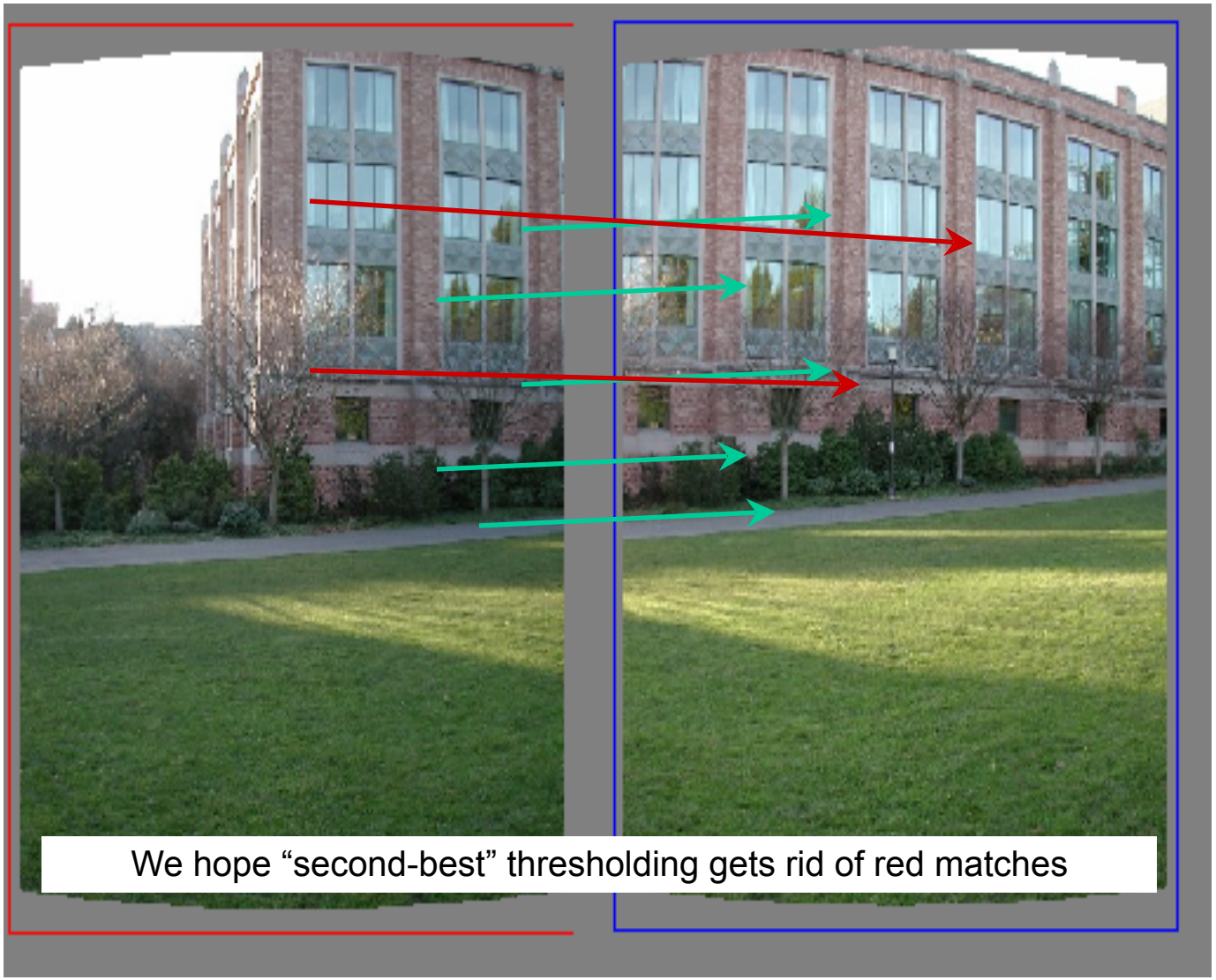
[Lowe, 1999]:

- 1-NN: SSD of the closest match
- 2-NN: SSD of the second-closest match
- Look at how much better 1-NN is than 2-NN, e.g. 1-NN/2-NN





Matching features

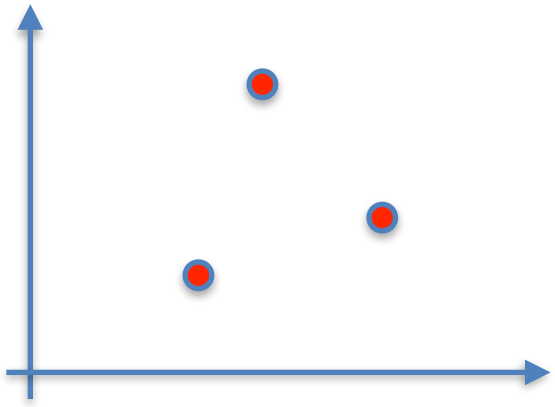


In general, we’ll have lots of bad matches.

RANSAC

https://en.wikipedia.org/wiki/Random_sample_consensus

Incredibly impactful procedure for fitting models under noise



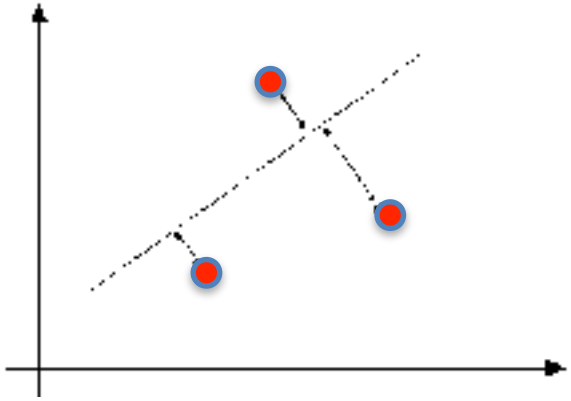
How would we fit a line to this data?

RANSAC

https://en.wikipedia.org/wiki/Random_sample_consensus

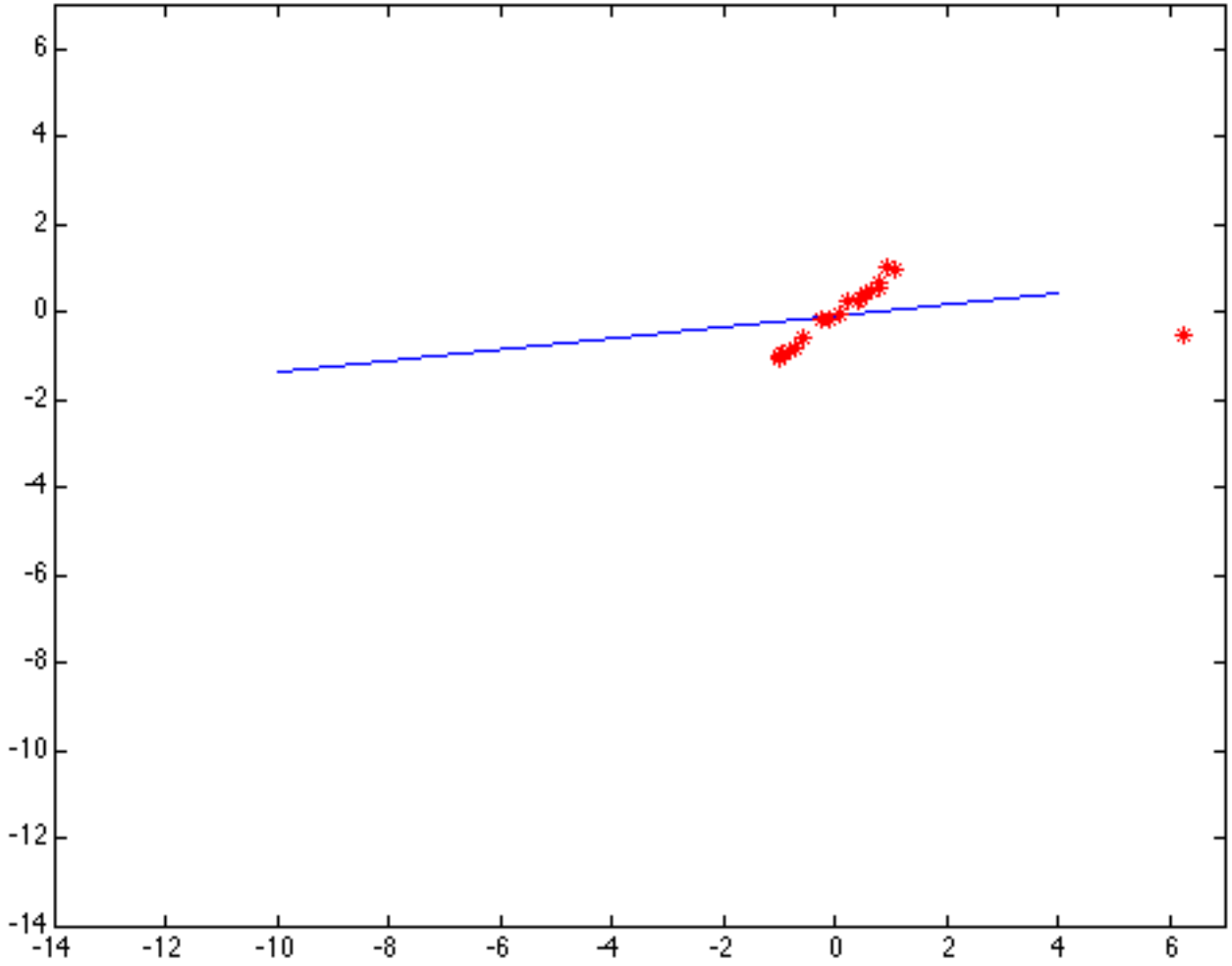
Formalize least squares vs robust statistics (ransac.pdf)

Least squares



Least squares: Robustness to noise

Least squares fit with an outlier:



Problem: squared error heavily penalizes outliers

Slide: S. Lazebnik

RANSAC for line fitting

Repeat ***N*** times:

- Draw ***s*** points uniformly at random
- Fit line to these ***s*** points
- Find inliers to this line among the remaining points (i.e., points whose distance from the line is less than ***t***)
- If there are ***d*** or more inliers, accept the line and refit using all inliers

Source: M. Pollefeys

Ransac parameters

https://en.wikipedia.org/wiki/RANSAC

p: prob that RANSAC returns a good model
w: fraction of inliers in data
n: number of points used to estimate model
k: number of random trials

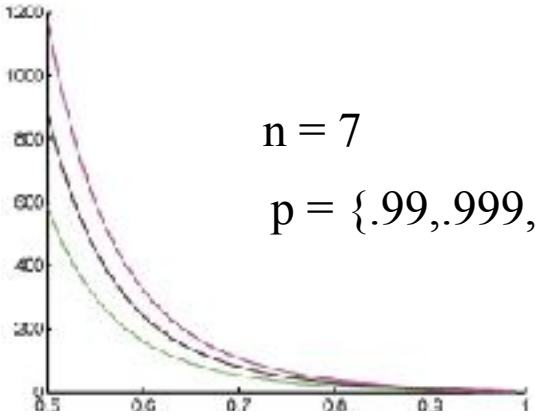
Goal: given a target ‘p’, compute number of needed trials ‘k’

$1- w^n$:prob that at least 1 out of n points is an outlier

$(1- w^n)^k$: prob that RANSAC fails across all k trials

$$(1-p) = (1- w^n)^k$$

$$k = \frac{\log(1-p)}{\log(1-w^n)}$$



n = 7

p = {.99,.999,.9999}

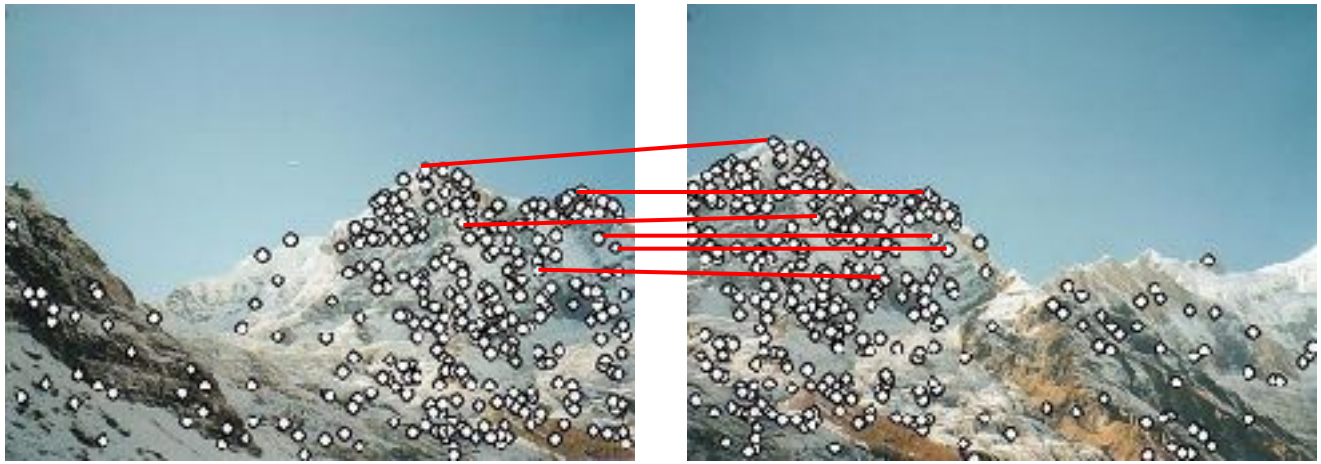
w

Example
from
Lowe2004

Panorama Stitching using SIFT

Image 1

Image 2



Match SIFT Interest Points

[Autopano]

Panorama Stitching using SIFT

Transform/Warp one or both images so that corresponding SIFT points in images are aligned.

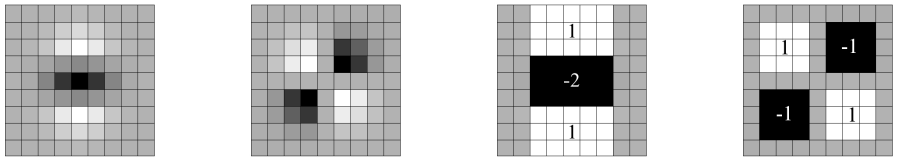
[Autopano]

Image Descriptors

- SIFT
- RANSAC
- (Other) sparse descriptors
- Dense descriptors

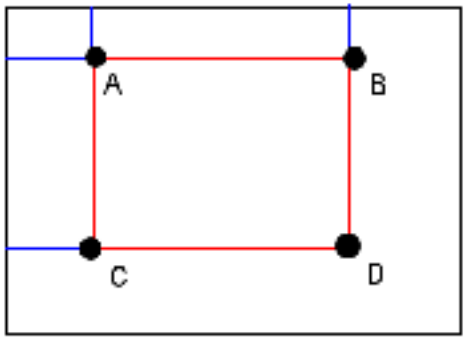
SURF (SIFT++)
(Speeded Up Robust Features)

- Replace Gaussians with box-filters



Fast approximation of first and second derivaties

Make use of integral images for (filter) size-indepdenant filtering



Sum = D - B - C + A

SURF

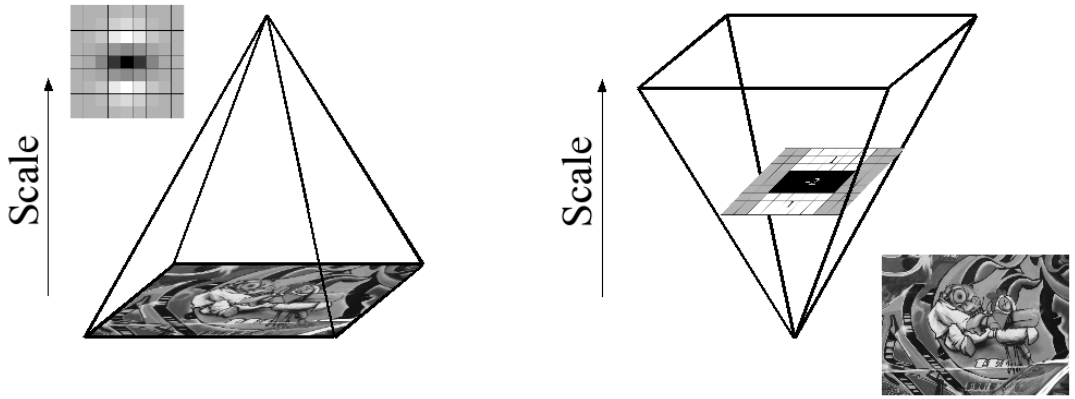


Fig. 4. Instead of iteratively reducing the image size (left), the use of integral images allows the up-scaling of the filter at constant cost (right).

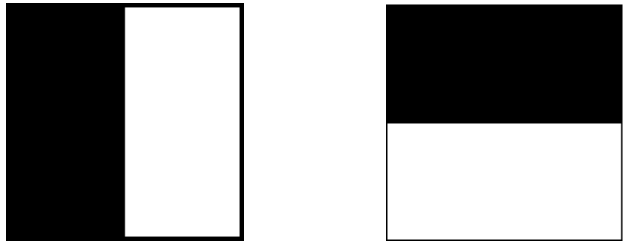
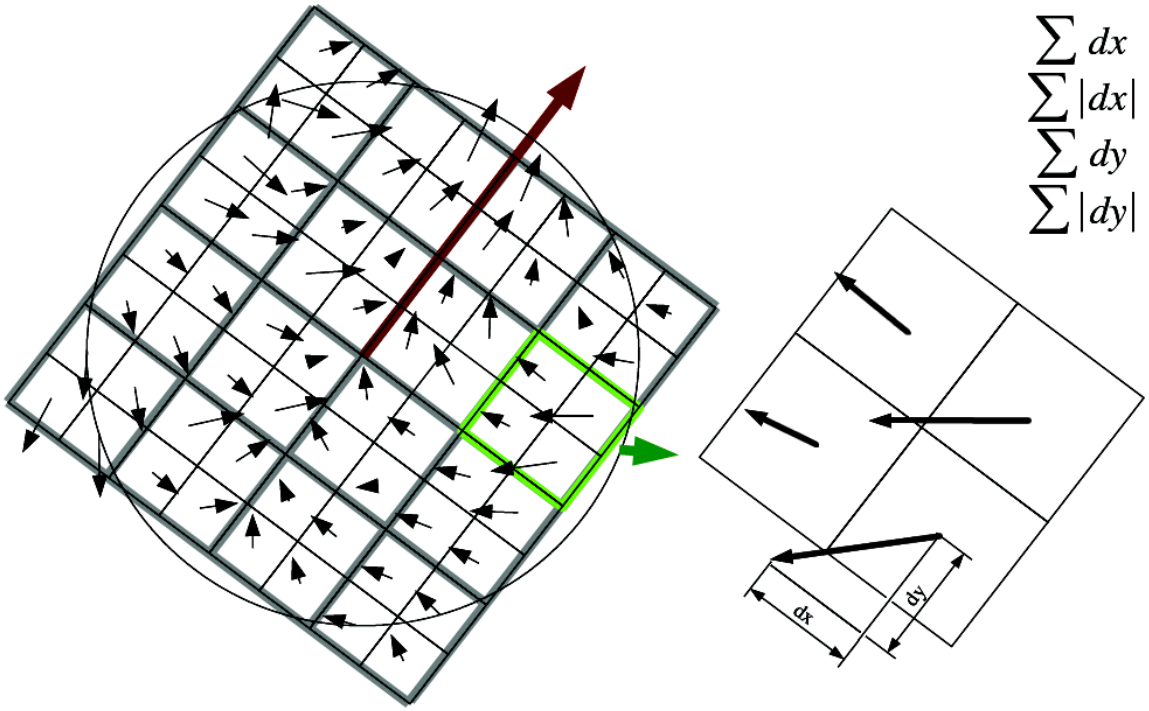


Fig. 9. Haar wavelet filters to compute the responses in x (left) and y direction (right). The dark parts have the weight -1 and the light parts $+1$.

Recuced histograms



Instead of recording histograms of 8 orientation inside each 4x4 block, record 4 marginal statistics

Final descriptor = 4x4x4 = 64 dim

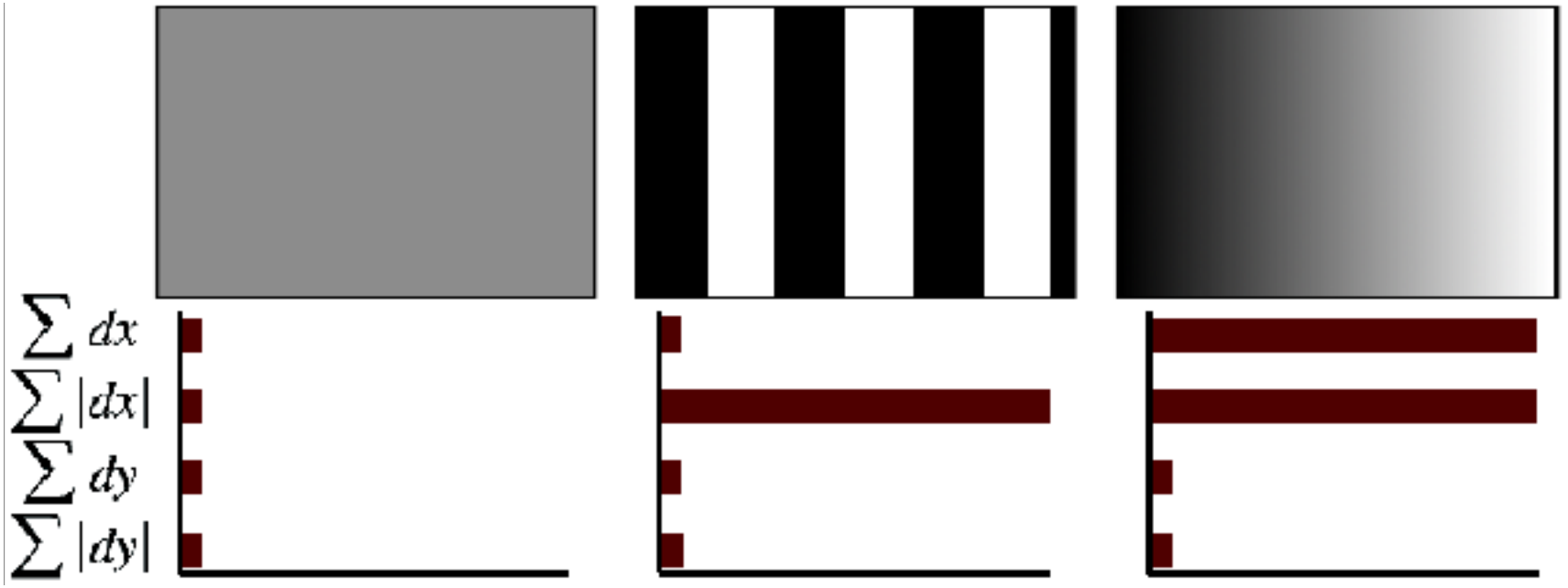

$\sum dx$
$\sum |dx|$
$\sum dy$
$\sum |dy|$

Image Descriptors

- SIFT
- RANSAC
- (Other) sparse descriptors (SIFT++, rank, region)
- Dense descriptors

Alternate family of approaches: rank-based representations

28	50	70
5	10	80
3	1	30

Order:
6,3,2,9,1,5,4,7,8

125	154	176
87	98	189
92	85	140

Order:
6,3,2,9,1,5,7,4,8

Posiiive: rank ordering is invariant to monotonic transformations of intensity (not just linear!)

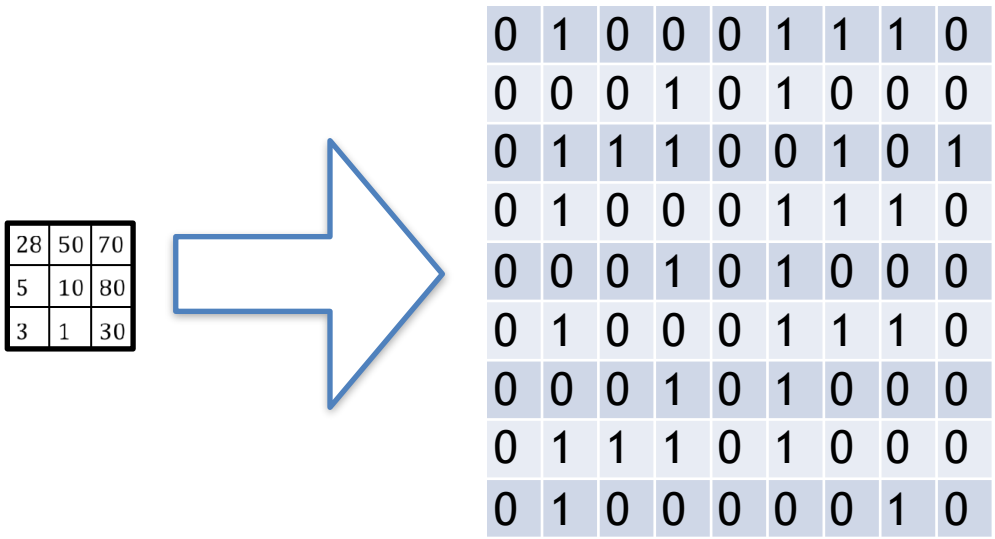
Negative: comparing two different ranks is expensive

Dinkar N. Bhat and Shree K. Nayar.
Ordinal Measures for Image Correspondence. PAMI Vol. 20, No. 4, April 1998

Alternate approach: binary patterns

Convert rank-order vector into a N^2 binary matrix of relative comparisons

Is pixel i > pixel j ?

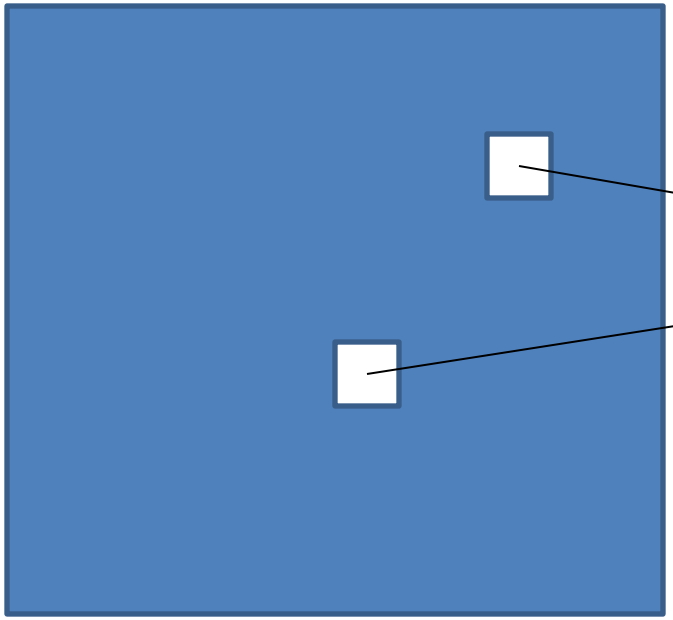


1. Create a descriptor that is a *random projection* of the vectorized matrix
2. Compare two descriptors by # of matching bits (Hamming distance with bitwise operations)

Ho, Tin (1995) "Random Decision Forests" Int'l Conf. Document Analysis and Recognition
Breiman, Leo (2001) "Random Forests" Machine Learning
Amit, Y. & Geman, D. (1997). Shape quantization and recognition with randomized trees. Neural Computation, 1997

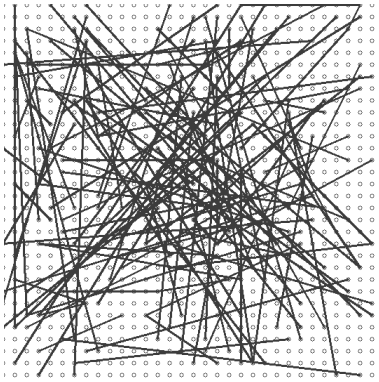
BRIEF:
Binary Robust Independent Elementary Features

- Randomly select pairs p and p' for comparison
- Design choice: How to sample p,p'?

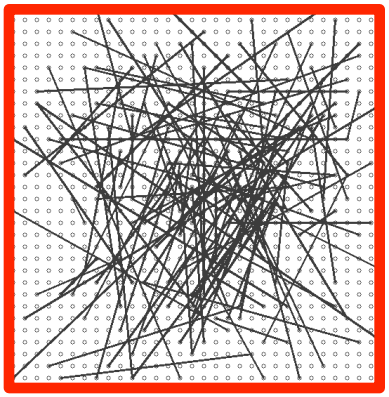


$$b = \begin{cases} 1 \; if \; I(p) > I(p') \\ 0 \; otherwise \end{cases}$$

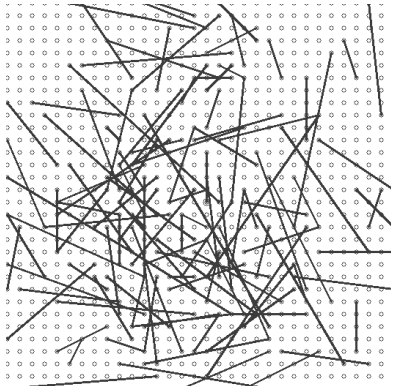
Sampling strategies



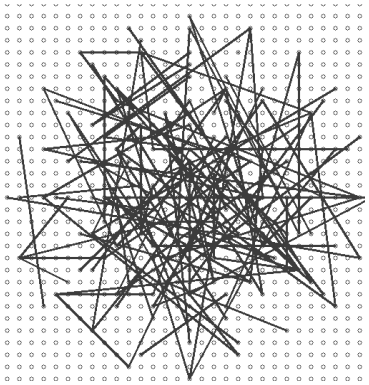
p1,p2 ~ uniform(x,y)



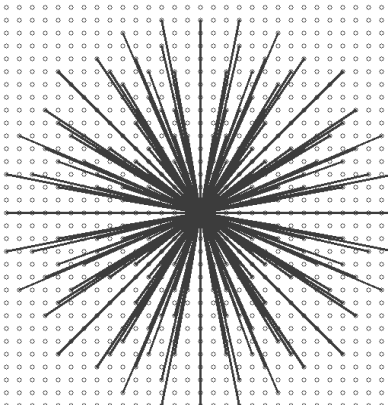
p1,p2 ~ N(0,sigma)



p1 ~ N(0,sigma)
p2 ~N(p1,sigma$_2$)



p1,p2 ~ uniform(r,theta)

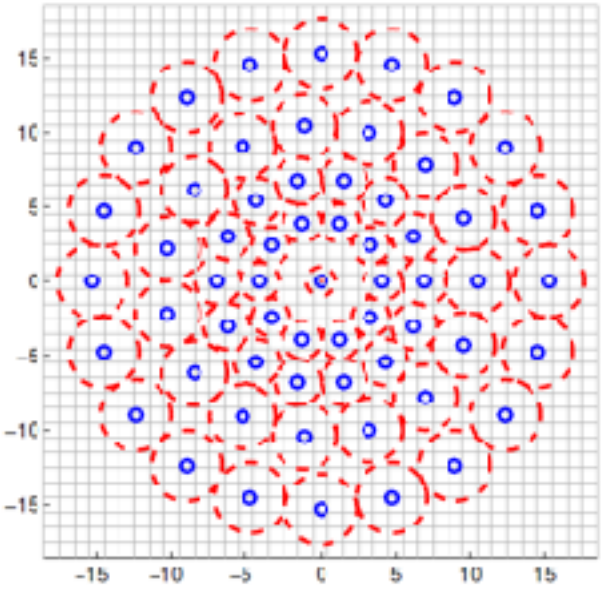


p1 = 0
p2 = grid(r,theta)

BRISK: Binary Robust Invariant Scalable Keypoints

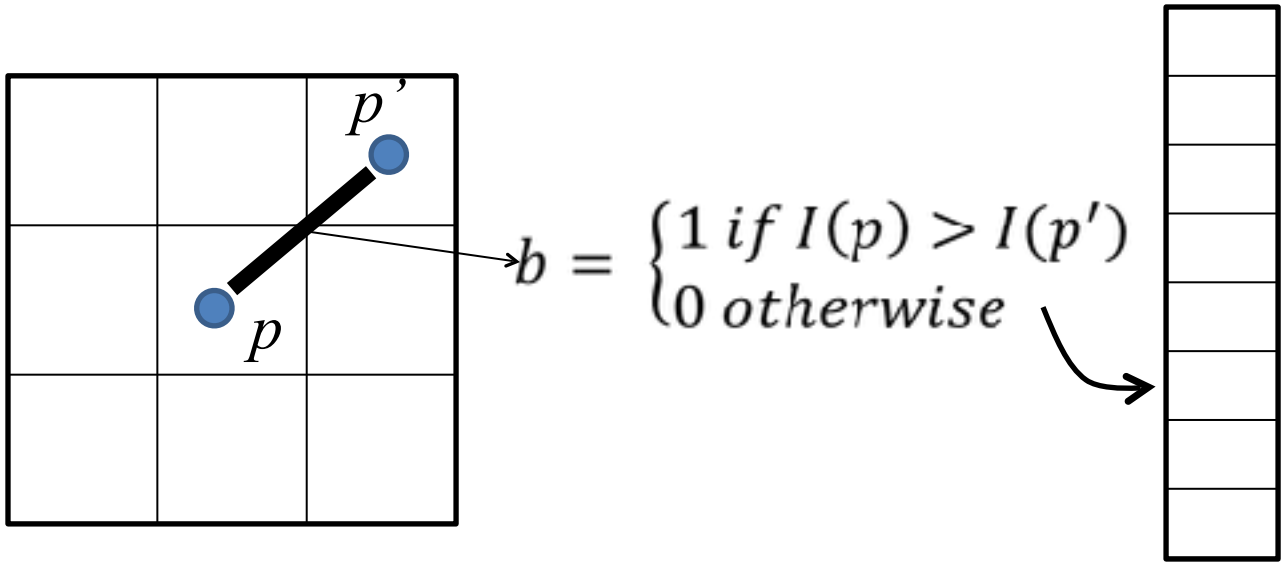
- Many variations on the BRIEF-like descriptors
- Example BRISK:
  - 60 sample points regularly spaced in log-polar space
  - Intensity smoothed by Gaussian proportional to distance to interest point
  - Binary code computed over the 512 pairs of closely points such that:

$$||p_1 - p_2|| < 9.75\sigma$$



LBP: Local Binary Patterns

- Special case: always compare with center pixel
- Originally used for texture classification
- Now standard descriptor for vision tasks
- Simplest form: record binary comparions of central pixel with its (8) neighbors
  - Represent 3x3 patch with 8-bit binary vector

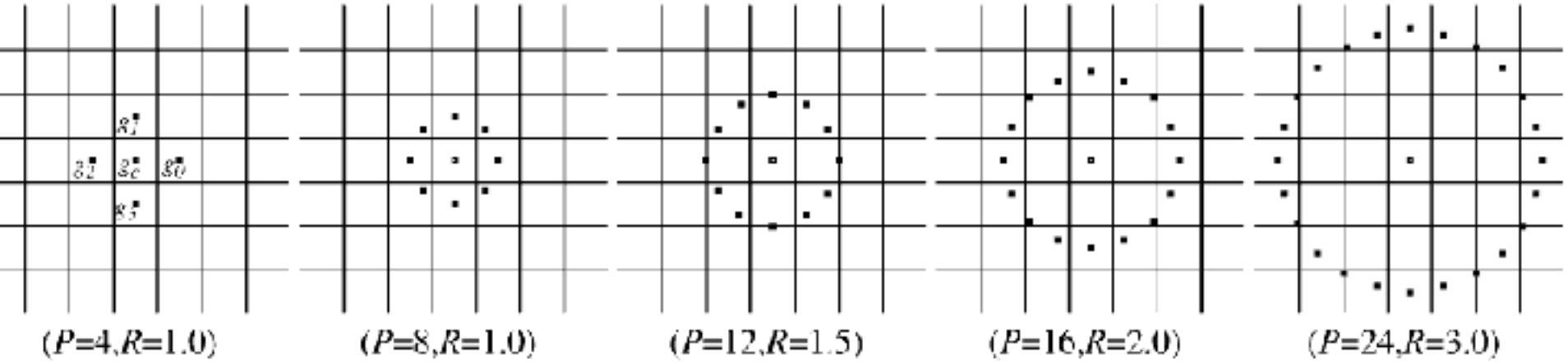


Multiscale LBPs

$LBP_{P,R}$

P = Number of sample points

R = Radius



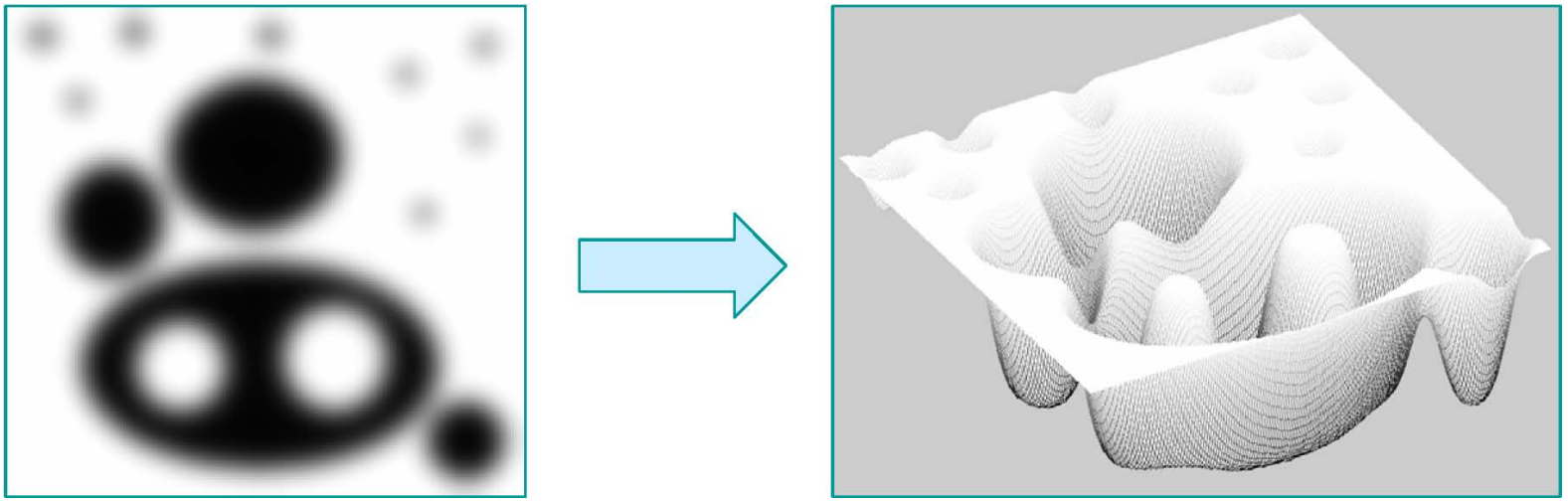
(P=4,R=1.0) (P=8,R=1.0) (P=12,R=1.5) (P=16,R=2.0) (P=24,R=3.0)

Image Descriptors

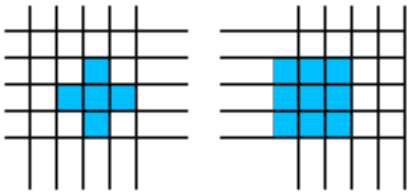
- SIFT
- RANSAC
- (Other) sparse descriptors (SIFT++, rank, region)
- Dense descriptors

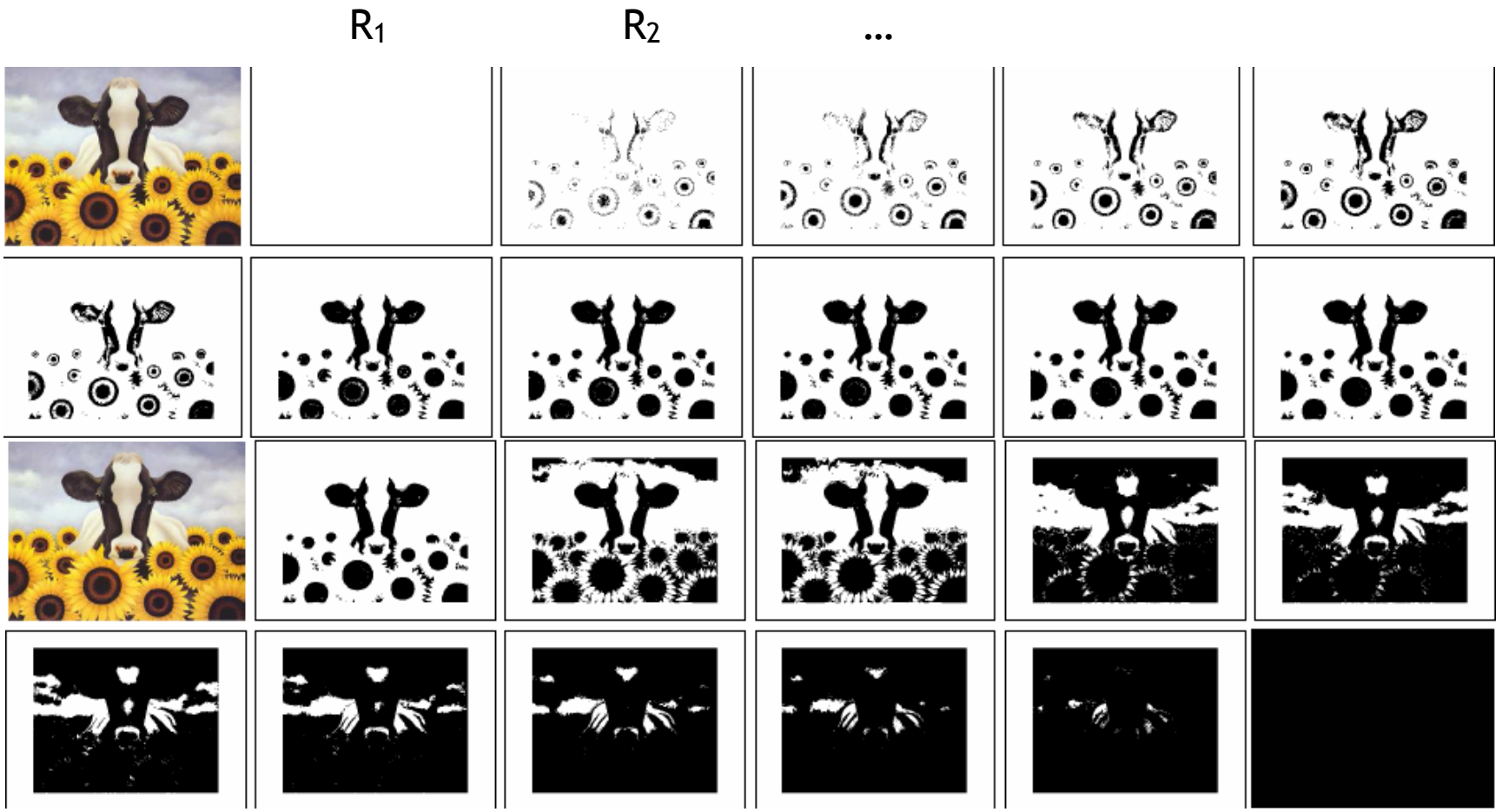
Regions instead of points

Maximally Stable Extremal Regions (MSER)



Extremal region R_i : Set of connected pixels such that the intensity values inside R_i are greater (or lower) than those at the boundary of R_i

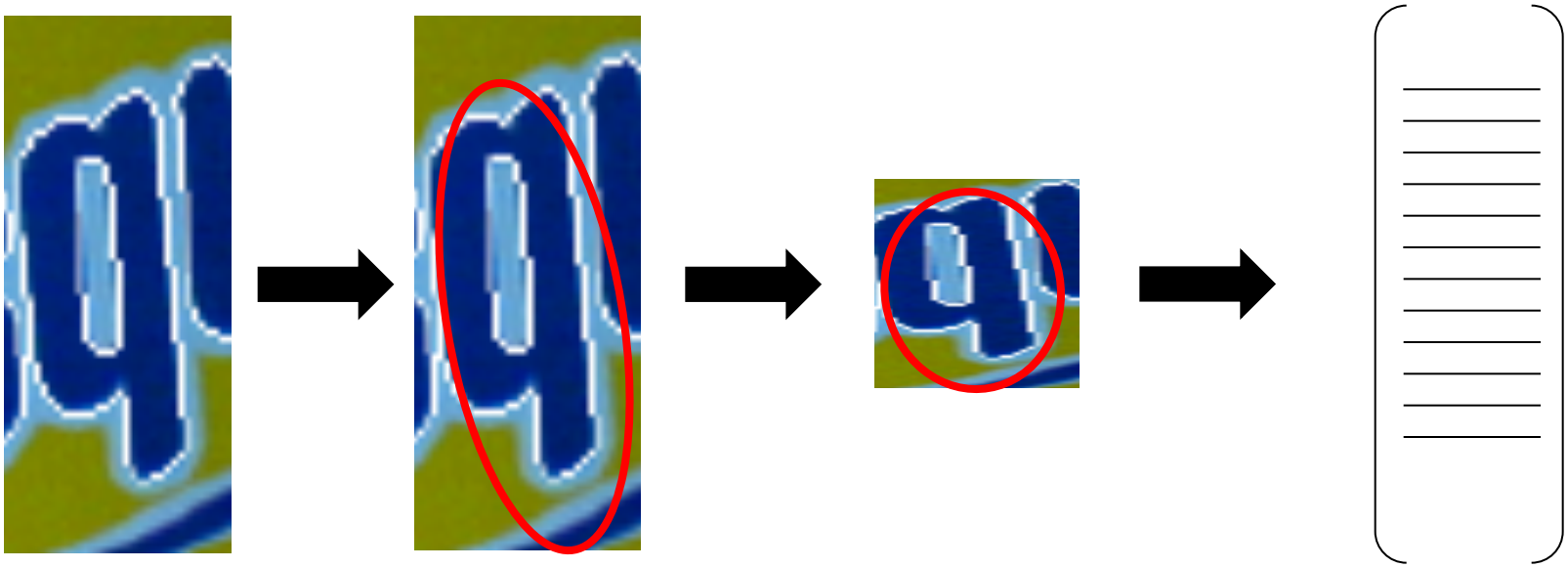




Stable region: connected region of pixels with stable size for different thresholds

$$(|R_{i+1}| - |R_i|)/ |R_i|$$

From MSER to SIFT



Find extremal region

Fit ellipse

Eigenthings of:

Rotate + scale to canonical patch

Compute 128-dim SIFT vector

Examples from image matching

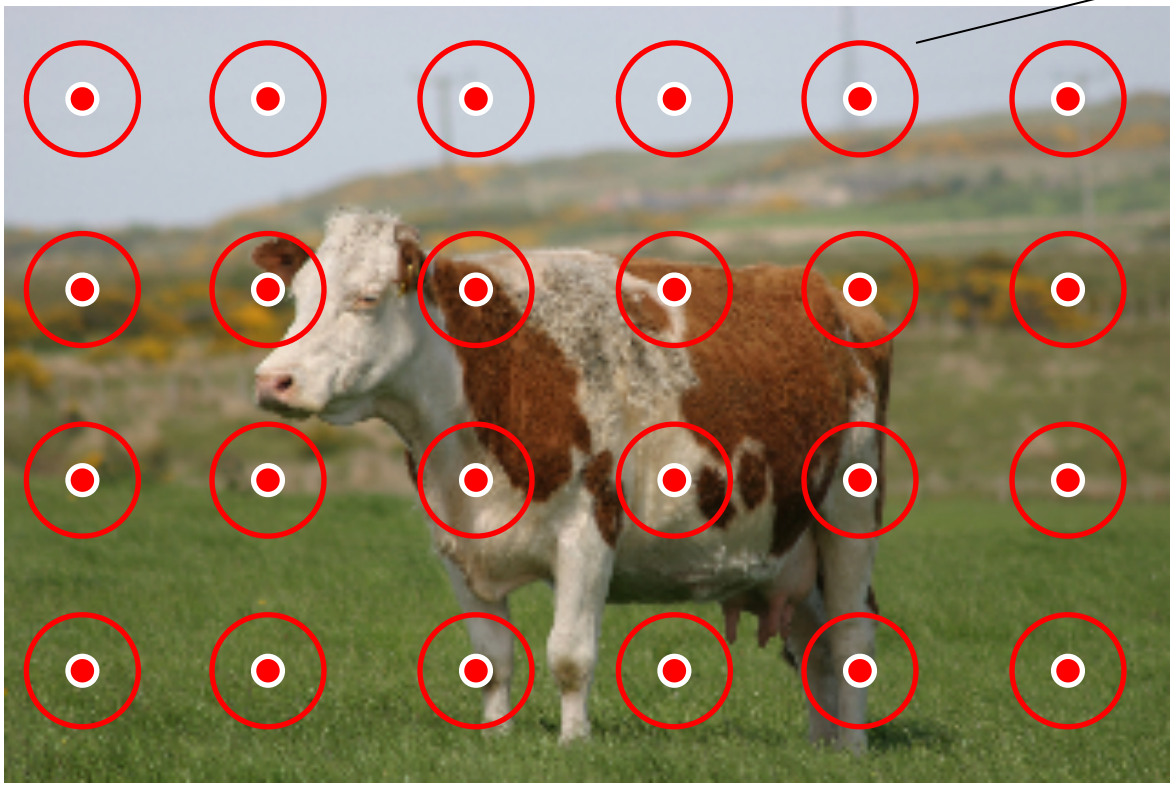
Example from Matas et al. BMCV 2002

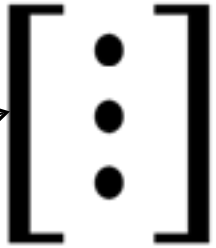
Image Descriptors

- SIFT
- RANSAC
- (Other) sparse descriptors (SIFT++, rank, region)
- Dense descriptors

Dense sampling

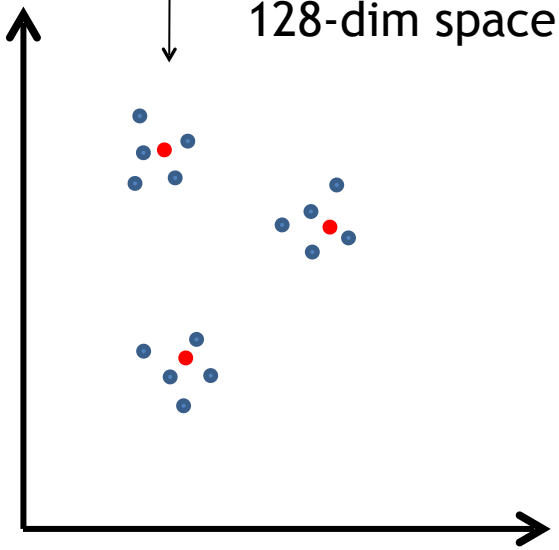
- So far: Descriptors of patches centered at sparse interest points
- But we can use the descriptors at any point
- Common case:
  - Regularly sampled grid of points
  - Dense SIFT (or LBP, or…)





128-dim SIFT feature

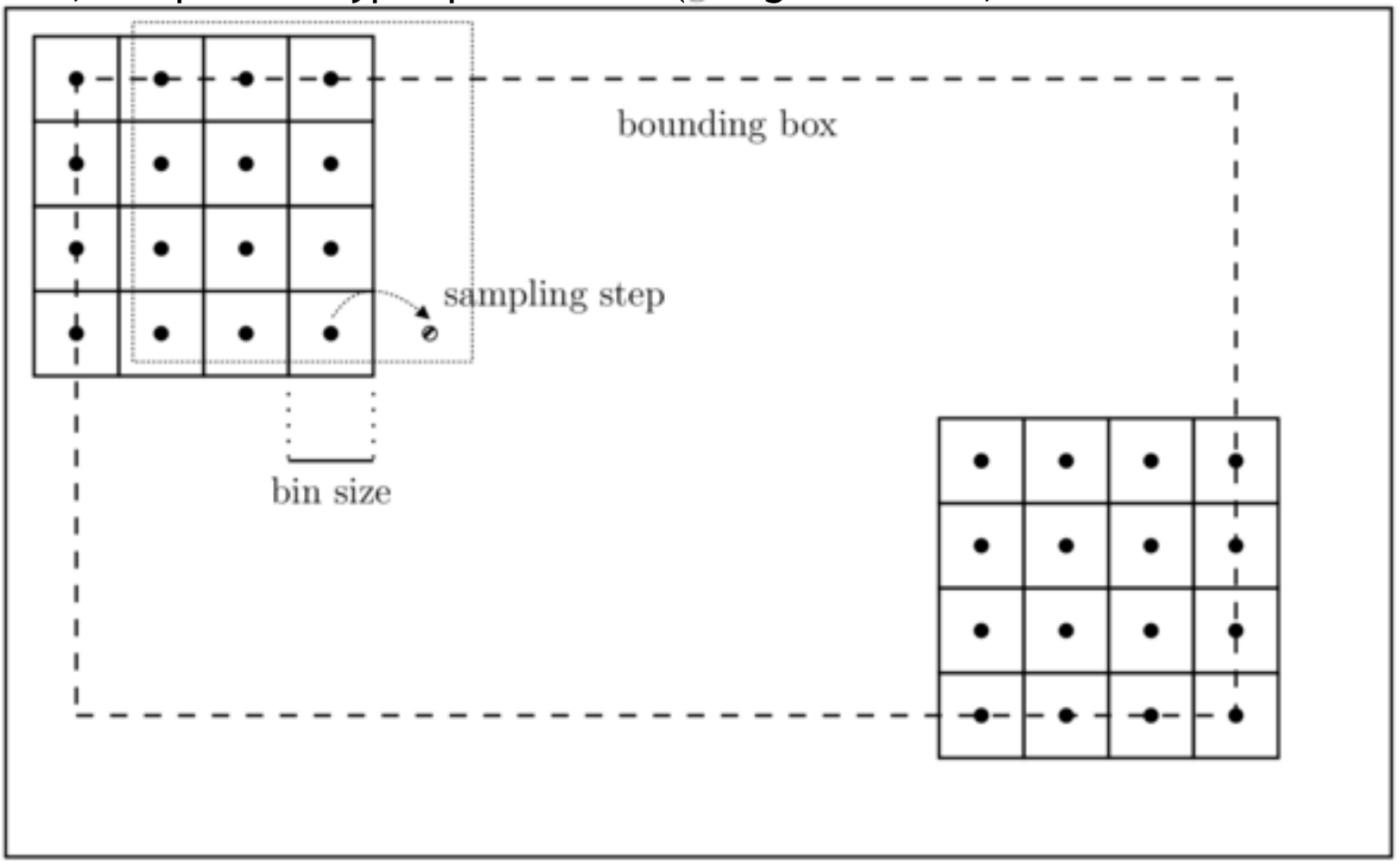
Visual words from clusters in 128-dim space



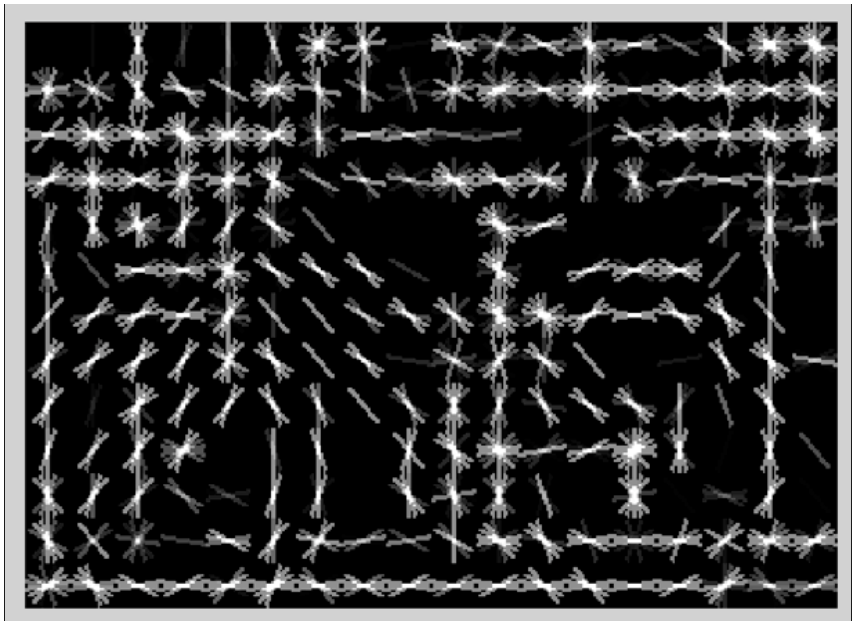
HOG

Compute SIFT descriptors on a grid equal to size of individual “cell”

In practice, re-optimize hyper-parameters (2x2 grid of cells, with each cell of 8x8 pixels)

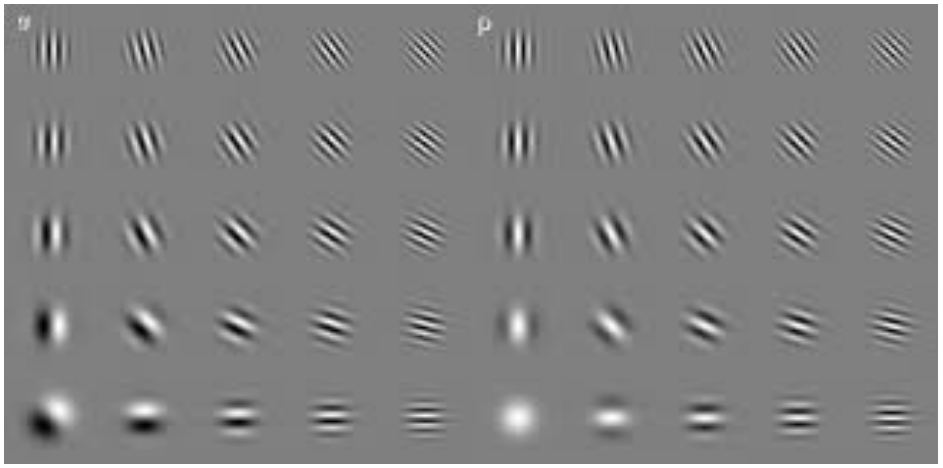


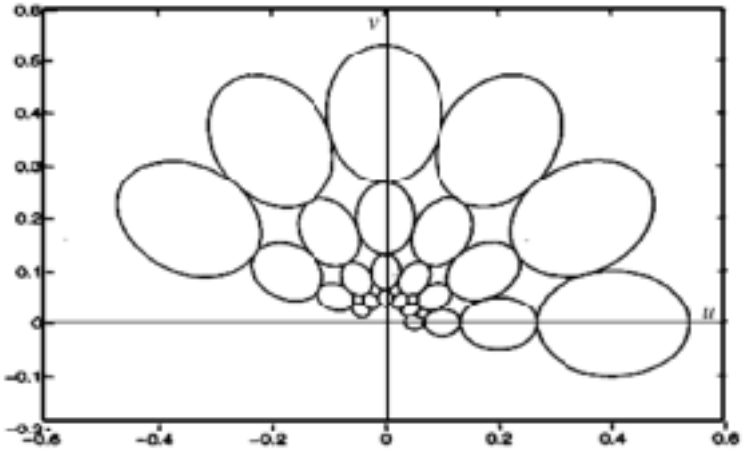
Common visualization



Alternative global desciptor: Gist

Oliva and Torralba, 2001





1.Compute frequency energy (magnitude) at each spatial (x,y) location with gabor filters

2. Average energy over 4x4 spatial grids

8	orientations
4	scales
x 16	spatial bins
512	dimensions

Common visualization

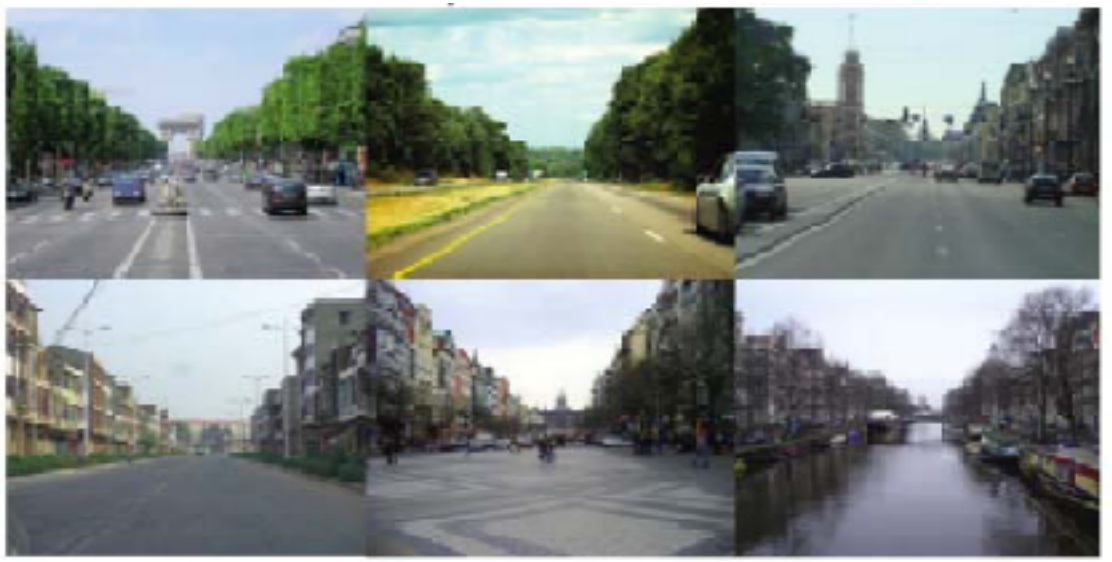
Global features (I) ~ global features (I')

Oliva & Torralba (2001)

Example: Image matching

Query:

Matches:



- **BRIEF, (D-BRIEF, binboost, LBGM, …):**
  - Michael Calonder, Vincent Lepetit, Christoph Strecha, and Pascal Fua . BRIEF: Binary Robust Independent Elementary Features, Proc. ECCV 2010.
  - http://cvlabwww.epfl.ch/~lepetit/
- **BRISK:**
  - Stefan Leutenegger, Margarita Chli and Roland Y. Siegwart. BRISK: Binary Robust Invariant Scalable Keypoints. Proc. ICCV 2011.
- **GIST:**
  - Aude Oliva, Antonio Torralba. Modeling the shape of the scene: a holistic representation of the spatial envelope. International Journal of Computer Vision, Vol. 42(3): 145-175, 2001.
  - http://people.csail.mit.edu/torralba/code/spatialenvelope/

- K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, T. Kadir and L. Van Gool, A Comparison of Affine Region Detectors; International Journal of Computer Vision (IJCV), Volume 65, Number 1, 2005.
- Major resource for all the features (Vision Lab Features Library (VLFeat)):

http://www.vlfeat.org/api/index.html

- **SIFT:**
  - David G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. IJCV (International Journal of Computer Vision), 2004.
  - http://www.cs.ubc.ca/~lowe/keypoints/
- **MSER:**
  - P-E. Forssen, P-E. and D. Lowe, Shape Descriptors for Maximally Stable Extremal Regions International Conference on Computer Vision (ICCV), 2007.
- **SURF:**
  - Herbert Bay, Andreas Ess, Tinne Tuytelaars, Luc Van Gool, SURF: Speeded Up Robust Features, Computer Vision and Image Understanding (CVIU), Vol. 110, No. 3, 2008.
  - http://www.vision.ee.ethz.ch/~surf/
  - **DAISY:** Engin Tola, Vincent Lepetit, Pascal Fua, DAISY: An Efficient Dense Descriptor Applied to Wide-Baseline Stereo, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 2010.
- **LBP:**
  - Timo Ojala, Matti Pietikainen, and Topi Maenpa. Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol. 24, No. 7, July 2002
  - http://www.cse.oulu.fi/CMV/Research/LBP

References

[Autopano] Software to make panaromas using SIFT. http://user.cs.tu-berlin.de/~nowozin/autopano-sift/

[Brown and Lowe 2002] M. Brown and D. Lowe. "Invariant Features from Interest Point Groups". *BMVC*, 2002.

[Harris and Stephens 1988] C. Harris and M. Stephens. "A Combined Corner and Edge Detector". *4th Alvey Vision Conference*, 1988.

[Lowe 2004] D. Lowe."Distinctive Image Features from Scale-Invariant Keypoints". *IJCV*, 2004.

[Lindeberg 1994] T. Lindeberg. "Scale-Space Theory: A Basic Tool for Analysing Structures at Different Scales." *J. of Applied Statistics,* 1994.

[Matas 2002] J. Matas, O. Chum, M. Urban, and T. Pajdla. "Robust Wide Baseline Stereo from Maximally Stable Extremal Regions. *BMVC*, 2002.

[Mikolajczyk 2002] K. Mikolajczyk. "Detection of Local Features Invariant to Affine Transformations." *Ph.D. Thesis*, 2002.

References

[Mikolajczyk 2004] K. Mikolajczyk and C. Schmid. “Scale and Affine Invariant Interest Point Detectors.” *IJCV,* 2004*.*

[Mikolajczyk 2005] K. Mikolajczyk and C. Schmid. “A Performance Evaluation of Local Descriptors.” *PAMI,* 2005*.*

[SIFT] SIFT Binaries. http://www.cs.ubc.ca/~lowe/keypoints/

[Witkin 1983] A. Witkin. “Scale-Space Filtering”. *IJCAI*, 1983.