Features and Matching

CV @ NTHU

3 October 2019

Outline

Features and matching

Feature detectors

Feature representations

Histogram-based descriptors

Invariant-based descriptors

Evaluation

What is a feature

- Track a point on a set as a camera moves around during a shot
  - Automatically estimate the 3D path of a camera as it moves around a scene: *matchmoving*
- Not every point in the scene is a good choice for tracking

Features: regions of an image that can be reliably located in other images of the same scene

- Interest points, keypoints, tie points

Detecting, describing, matching, and tracking

Feature tracking

A subset of the more general problem of visual tracking

Differeces

- Visual trackers maintain a probabilistic representation of an object's state, e.g., Kalman filters
- Wide-baseline case: images are taken from cameras that are physically far apart, whereas visual tracking generally assumes the camera moves only slightly between images

Feature detection and description

Detector

- deciding which image regions are sufficiently distinctive

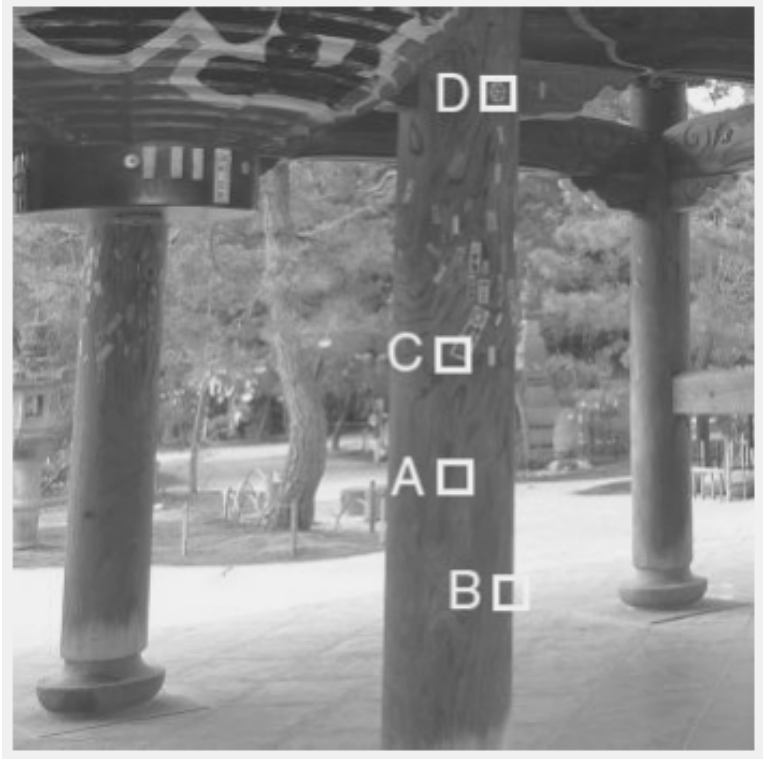
Descriptor

- deciding how to represent the image information inside each region for later matching

Properties of local features

A square block of pixels centered at a certain location in an image

- Repeatability
- Aperture problem
- Nearly-constant-intensity patch, edge, corner, blob



Harris corners

- H. Moravec.“ Obstacle avoidance and navigation in the real world by a seeing robot rover.” Technical Report CMU-RI-TR-3, Carnegie Mellon University, 1980.
- C. Harris and M. Stephens.“ A combined corner and edge detector.” In Alvey Vision Conference, 1988.

Cornerness

- Comparing a block of pixels to adjacent blocks in the horizontal, vertical, and diagonal directions. If the difference is high in all directions, the block is a good candidate to be a feature.

Harris corner detector

- Sum of squared differences obtained by a small shift of the block in the direction of vector (u, v)

$$E(u, v) = \sum_{(x,y)} w(x, y)(I(x + u, y + v) - I(x, y))^2 \qquad (1)$$

Harris corner detector

- Tylor expansion at $(0, 0)$

$$
\begin{aligned}
& E(u, v) && (2)\\
= & \sum_{(x,y)} w(x, y) \left(I(x, y) + u \frac{\partial I}{\partial x}(x, y) + v \frac{\partial I}{\partial y}(x, y) - I(x, y) \right)^2 && (3)\\
= & \sum_{(x,y)} w(x, y) \left(u \frac{\partial I}{\partial x}(x, y) + v \frac{\partial I}{\partial y}(x, y) \right)^2 && (4)\\
= & \begin{bmatrix} u \\ v \end{bmatrix}^T H \begin{bmatrix} u \\ v \end{bmatrix} && (5)
\end{aligned}
$$

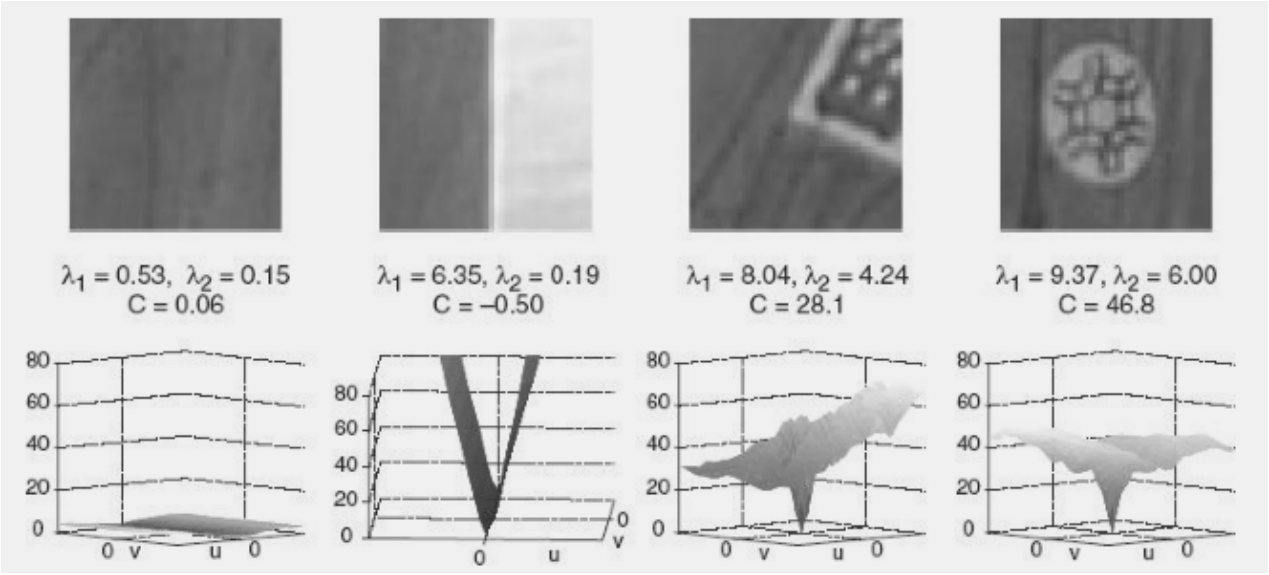
Harris matrix

H is a symmetric positive definite matrix defined by

$$\begin{bmatrix} \sum_{(x,y)} w(x,y) \left(\frac{\partial I}{\partial x}(x,y)\right)^2 & \sum_{(x,y)} w(x,y) \left(\frac{\partial I}{\partial x}(x,y)\frac{\partial I}{\partial y}(x,y)\right) \\ \sum_{(x,y)} w(x,y) \left(\frac{\partial I}{\partial x}(x,y)\frac{\partial I}{\partial y}(x,y)\right) & \sum_{(x,y)} w(x,y) \left(\frac{\partial I}{\partial y}(x,y)\right)^2 \end{bmatrix} \tag{6}$$

- The eigenvalues and eigenvectors of the Harris matrix H can be analyzed to assess the cornerness of a block.

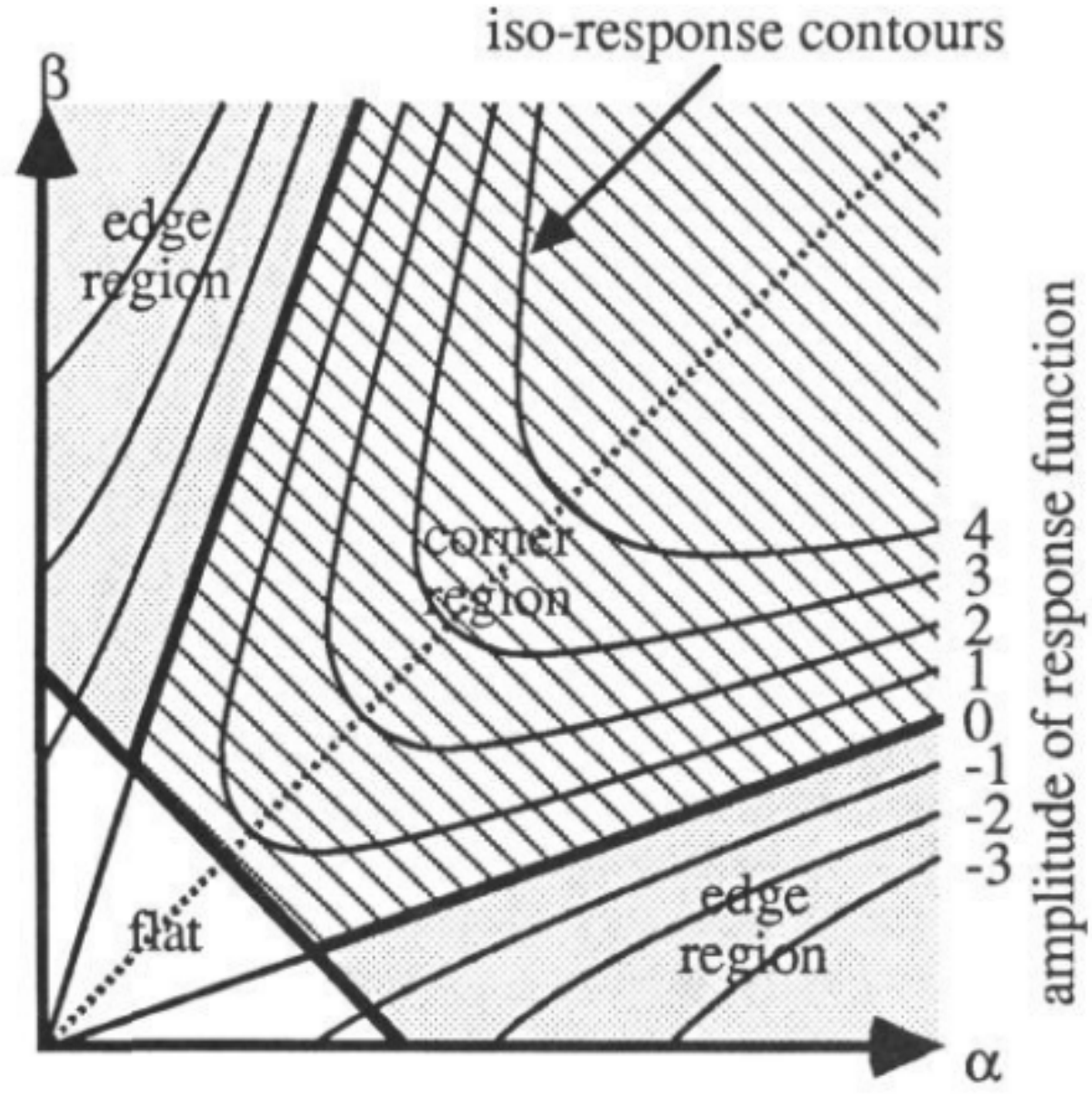
Cornerness



Let the eigenvalues and eigenvectors be (λ_1, λ_2) and (e_1, e_2), with $\lambda_1 \geq \lambda_2$.

- The block is nearly-constant-intensity: $\lambda_1 \approx \lambda_2 \approx 0$
- The block straddles a linear edge: The gradient will be perpendicular to the edge direction for pixels near the edge. $\lambda_1 > 0$ and $\lambda_2 \approx 0$ with e_2 along the edge.
- The block contains a corner or blob: $\lambda_1 > 0$ and $\lambda_2 > 0$.

Cornerness


iso-response contours
β
edge region
corner region
flat
edge region
α
4
3
2
1
0
-1
-2
-3
amplitude of response function

Avoiding explicitly computing the eigenvalues

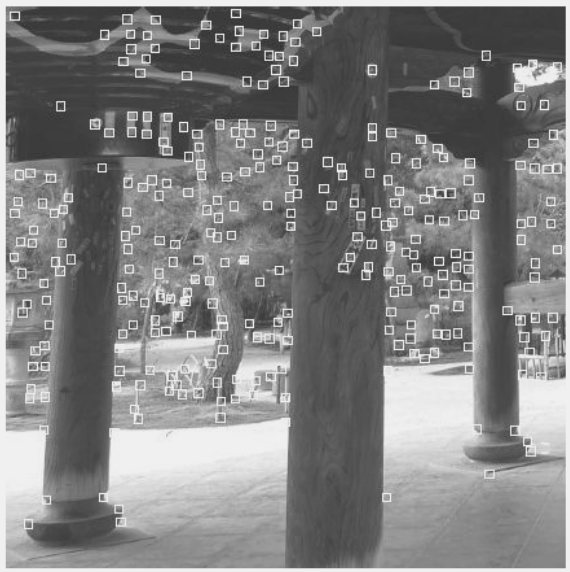
- Quality measure

$$C = \det(H) - \kappa \, \mathrm{trace}(H)^2 \tag{7}$$

To detect features in an image, we evaluate the quality measure at each block in the image, and select feature points where the quality measure is above a minimum threshold.

- ezcontour('x*y - 0.15*(x+y).^ 2',[0 10], [0 10]);

- Non-maximal suppression



Implementation details

Estimating the gradients:

$$\frac{\partial I}{\partial x}(x,y) = (I(x+1,y) - I(x-1,y))\ \cancel{/2} \tag{8}$$

An alternative:

$$\frac{\partial I}{\partial x}(x,y) = I(x,y) * \frac{\partial G(x,y,\sigma_D)}{\partial x} \tag{9}$$

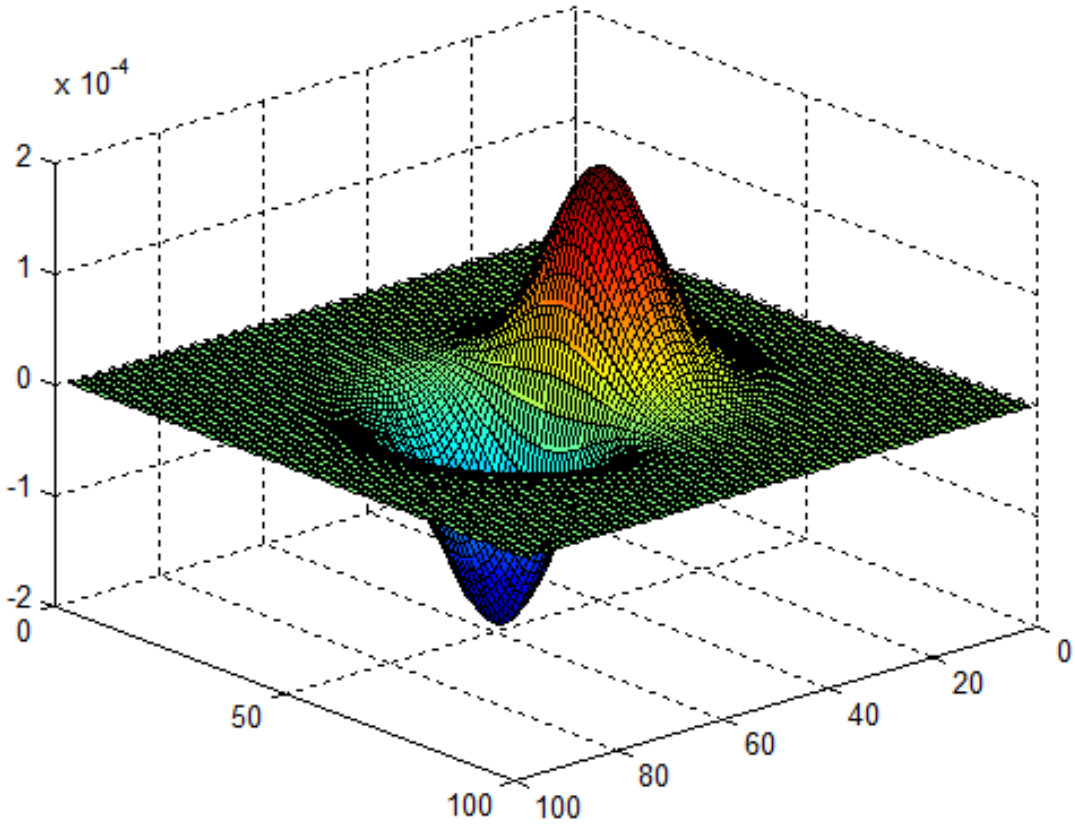
where $*$ indicates convolution, and

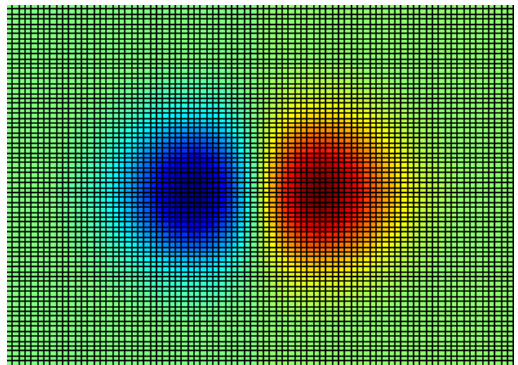
$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2}\exp(-\frac{1}{2\sigma^2}(x^2+y^2)) \tag{10}$$

Intuition: Smoothing the image to remove high frequencies before taking the derivative.
Spatially weighted window:

$$w(x,y) = \frac{1}{2\pi\sigma_I^2}\exp(-\frac{1}{2\sigma_I^2}((x-x_0)^2+(y-y_0)^2)) \tag{11}$$

Gaussian derivatives





Good features to track

Investigating the properties of blocks of pixels that are good for tracking.

$$I(x+u, y+v, t+1) = I(x, y, t) \qquad \forall (x, y) \in \mathcal{W} \tag{12}$$

If we fix a block of pixels at time t and want to find its translation at time $t+1$, we minimize the cost function

$$F(u, v) = \sum_{(x,y)} w(x, y) I((x+u, y+v, t+1) - I(x, y, t))^2 \tag{13}$$

where $w(x, y)$ is 1 for pixels inside W and 0 otherwise.

Taylor series approximation

Setting the derivative equal to zero yields the linear system

$$H \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} \sum_{(x,y)} w(x,y)(\frac{\partial I}{\partial x}(x,y,t)\frac{\partial I}{\partial t}(x,y,t)) \\ \sum_{(x,y)} w(x,y)(\frac{\partial I}{\partial y}(x,y,t)\frac{\partial I}{\partial t}(x,y,t)) \end{bmatrix} \tag{14}$$

where H is

$$\begin{bmatrix} \sum_{(x,y)} w(x,y)(\frac{\partial I}{\partial x}(x,y,t))^2 & \sum_{(x,y)} w(x,y)(\frac{\partial I}{\partial x}(x,y,t)\frac{\partial I}{\partial y}(x,y,t)) \\ \sum_{(x,y)} w(x,y)(\frac{\partial I}{\partial x}(x,y,t)\frac{\partial I}{\partial y}(x,y,t)) & \sum_{(x,y)} w(x,y)(\frac{\partial I}{\partial y}(x,y,t))^2 \end{bmatrix} \tag{15}$$

KLT corners

$$\min(\lambda_1, \lambda_2) > \tau \tag{16}$$

KLT (Kanade-Lucas-Tomasi) tracker

Extended to affine transformation

Shi and Tomasi. "Good features to track." CVPR 1994.

$$I(ax + by + u, cx + dy + v, t + 1) = I(x, y, t) \quad \forall (x, y) \in \mathcal{W} \qquad (17)$$

Extended to local photometric changes

Jin et al., "Real-time feature tracking and outlier rejection with changes in illumination." ICCV 2001.

$$e \cdot I(ax + by + u, cx + dy + v, t + 1) + f = I(x, y, t) \quad \forall (x, y) \in \mathcal{W} \quad (18)$$

Harris-Laplace

Detecting features in scale space

Lindeberg. "Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention." IJCV, 1993.
Lindeberg. "Feature detection with automatic scale selection." IJCV, 1998.

Harris-Laplace

The key concept of scale space is the convolution of an image with a Gaussian function:

$$L(x, y, \sigma_D) = G(x, y, \sigma_D) * I(x, y) \tag{19}$$

where $\sigma_D \in \{\sigma_0, k\sigma_0, k^2\sigma_0, \ldots\}$.

$$H(x, y, \sigma_D, \sigma_I) = G(x, y, \sigma_I) * \begin{bmatrix} \left(\frac{\partial L}{\partial x}\right)^2 & \frac{\partial L}{\partial x}\frac{\partial L}{\partial y} \\ \frac{\partial L}{\partial x}\frac{\partial L}{\partial y} & \left(\frac{\partial L}{\partial y}\right)^2 \end{bmatrix} \tag{20}$$

where

$$\frac{\partial L}{\partial x} = \frac{\partial L(x, y, \sigma_D)}{\partial x} = \frac{\partial G(x, y, \sigma_D)}{\partial x} * I(x, y) \tag{21}$$

- Differentiation and convolution are commutative

Scale normalization and scale invariance

High-resolution $I(x, y)$ and low-resolution $I'(x', y')$ with $x = kx'$ and $y = ky'$.
If we consider a block centered at (x', y') with scale (σ'_D, σ'_I) in the low-resolution image, it will correspond to the block centered at (kx', ky') with scales $(k\sigma'_D, k\sigma'_I)$ in the high-resolution image.
Substituting everything and we have

$$H(x, y, k\sigma'_D, k\sigma'_I) = \frac{1}{k^2} H'(x', y', \sigma'_D, \sigma'_I) \tag{22}$$

where H and H' are the scale-dependent Harris matrices for high- and low-resolution images.

Scale-normalized Harris matrix

$$\hat{H}(x, y, k\sigma_D, k\sigma_I) = k^2 G(x, y, k\sigma_I) * \begin{bmatrix} \left(\frac{\partial L_k}{\partial x}\right)^2 & \frac{\partial L_k}{\partial x}\frac{\partial L_k}{\partial y} \\ \frac{\partial L_k}{\partial x}\frac{\partial L_k}{\partial y} & \left(\frac{\partial L_k}{\partial y}\right)^2 \end{bmatrix} \tag{23}$$

where

$$\frac{\partial L_k}{\partial x} = \frac{\partial L(x, y, k\sigma_D)}{\partial x} = \frac{\partial G(x, y, k\sigma_D)}{\partial x} * I(x, y) \tag{24}$$

Multi-scale Harris corners

Applying the Harris criterion

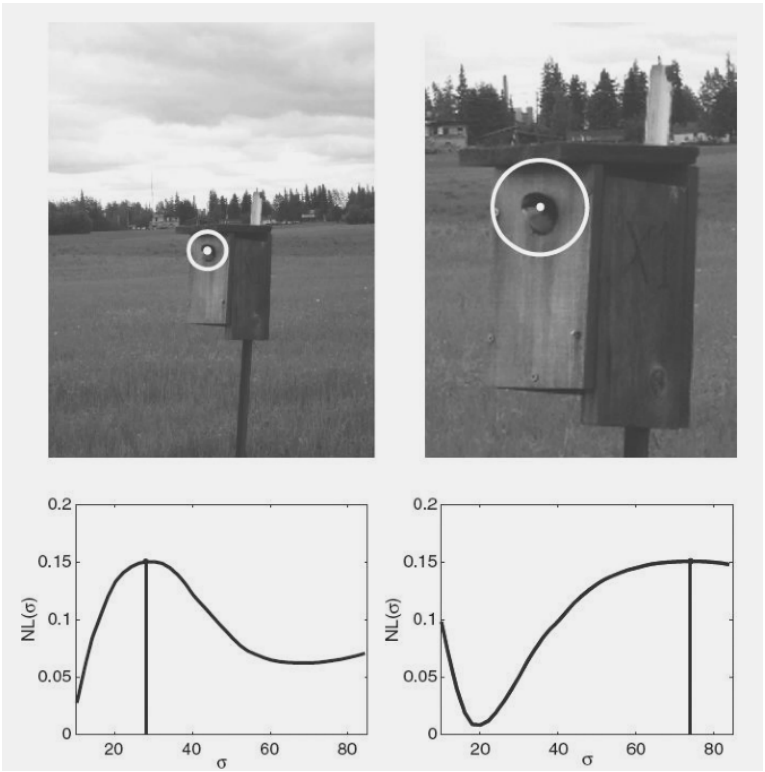
- Create the scale space of the image for a fixed set of scales $\sigma_D \in \{\sigma_0, k\sigma_0, k^2\sigma_0, \ldots\}$ with $\sigma_I = a\sigma_D$.
- For each scale, compute the scale-normalized Harris matrix and find all local maxima of the Harris criterion (quality measure) that are above a certain threshold.

This approach can detect multiple features centered at the same location (x, y) at different scales. However, we would often rather select a characteristic scale that defines the scale at which a given feature is most significant.

Maximum of the normalized Laplacian

Find σ^* that maximizes

$$NL(x, y, \sigma) = \left| \sigma^2 \left(\frac{\partial^2 G(x, y, \sigma)}{\partial x^2} + \frac{\partial^2 G(x, y, \sigma)}{\partial y^2} \right) * I(x, y) \right| \qquad (25)$$

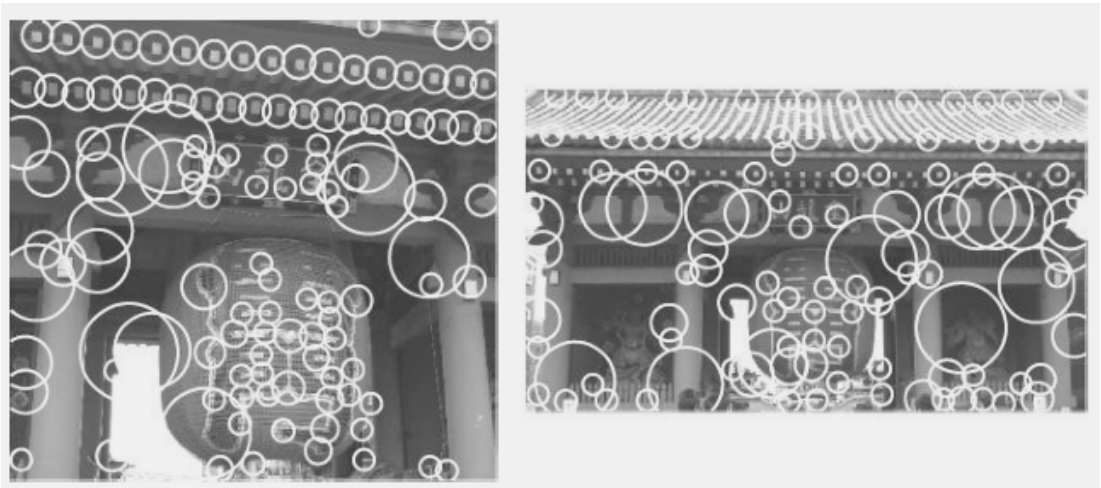


the ratio between the two characteristic scales is 2.64 similar to the actual scale factor of 2.59 relating the images

Harris-Laplace features

Mikolajczyk and Schmid. "Indexing based on scale invariant interest points." ICCV 2001.
For each detected feature (say at scale $k^n\sigma_0$), retain it only if its normalized Laplacian is above a certain threshold, and it forms a local maximum in the scale dimension, that is

$$NL(x, y, k^n\sigma_0) > NL(x, y, k^{n-1}\sigma_0) \text{ and } NL(x, y, k^n\sigma_0) > NL(x, y, k^{n+1}\sigma_0) \quad (26)$$



Scale-normalized Hessian matrix of second derivatives

Lindeberg. "Feature detection with automatic scale selection." IJCV, 1998.

$$\hat{S}(x, y, \sigma_D) = \sigma_D^2 \begin{bmatrix} \frac{\partial^2 L(x,y,\sigma_D)}{\partial x^2} & \frac{\partial^2 L(x,y,\sigma_D)}{\partial x \partial y} \\ \frac{\partial^2 L(x,y,\sigma_D)}{\partial x \partial y} & \frac{\partial^2 L(x,y,\sigma_D)}{\partial y^2} \end{bmatrix} \tag{27}$$

$L(x, y, \sigma_D)$ is the Gaussian-filtered image at the specified scale.

$$\text{trace}\hat{S}(x, y, \sigma_D) = \sigma_D^2 \left(\frac{\partial^2 L(x, y, \sigma_D)}{\partial x^2} + \frac{\partial^2 L(x, y, \sigma_D)}{\partial y^2} \right) \tag{28}$$

$$\det\hat{S}(x, y, \sigma_D) = \sigma_D^4 \left(\frac{\partial^2 L(x, y, \sigma_D)}{\partial x^2} \frac{\partial^2 L(x, y, \sigma_D)}{\partial y^2} - \left(\frac{\partial^2 L(x, y, \sigma_D)}{\partial x \partial y} \right)^2 \right) \tag{29}$$

The absolute value of the Hessian's trace is the same as the normalized Laplacian.

Laplacian-of-Gaussian and Hessian-Laplace

trace: Laplacian-of-Gaussian (LoG)

- compute the trace at every (x, y, σ_D) and find points where this function of three parameters is locally maximal
- same function for both the spatial and scale dimensions

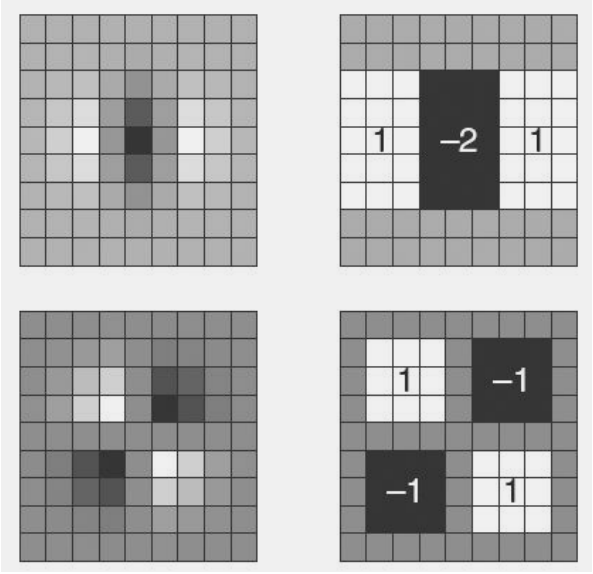
determinant: Hessian-Laplacian

- determinant for detection and trace for scale selection
- require that the trace and determinant are simultaneously maximized: features are scale-covariant

SURF

Bay et al. "SURF: Speeded up robust features." ECCV 2006.

- The discrete Gaussian filter used in the computation of the scale-normalized Hessian could be approximated by box filters.
- Integral images
- Fast Hessian detector: Using box filters in an approximation of the Hessian's determinant



Difference-of-Gaussians

Lowe. "Distinctive images features from scale-invariant keypoints." IJCV, 2004. SIFT (Scale Invariant Feature Transform)

DoG detector

Laplacian-of-Gaussian could be approximated by Difference-of-Gaussians

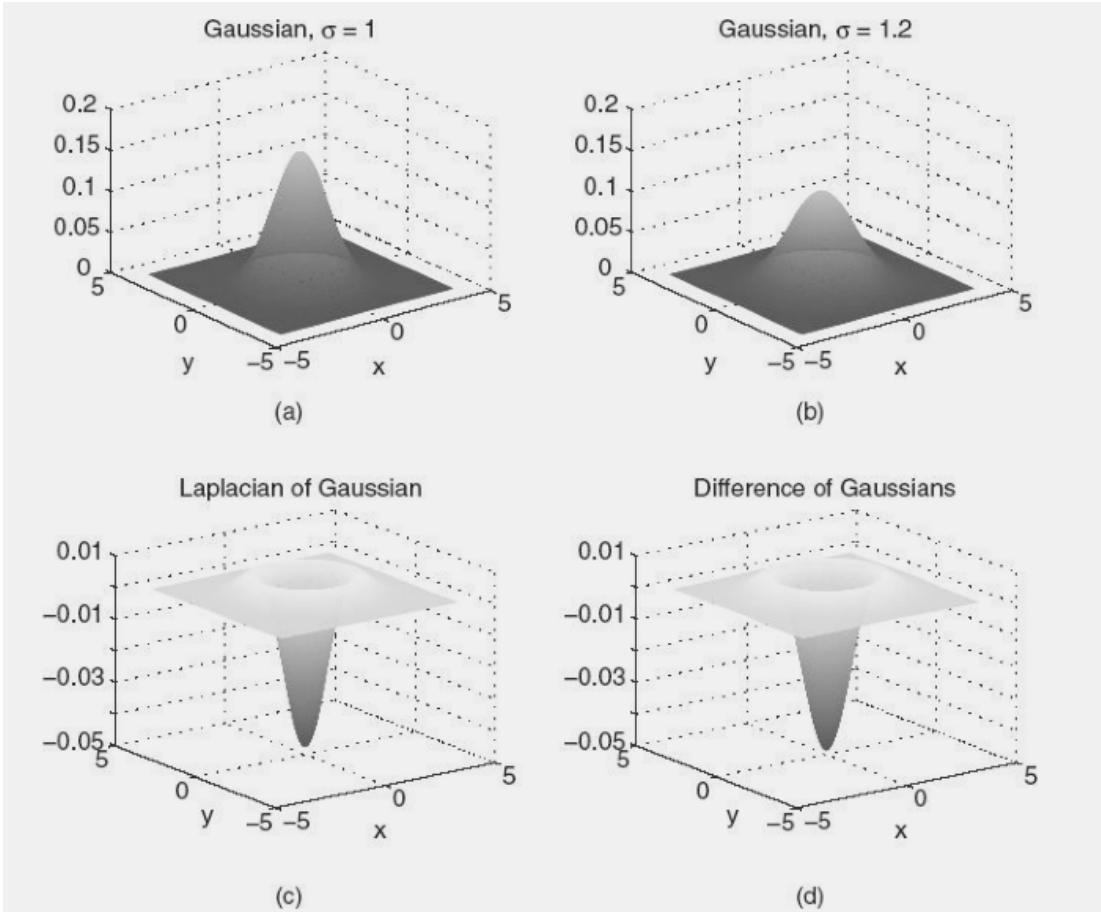
$$\frac{\partial G}{\partial \sigma} = \sigma \nabla^2 G \tag{30}$$

$$\frac{\partial G}{\partial \sigma} \approx \frac{G(x, y, k\sigma) - G(x, y, \sigma)}{k\sigma - \sigma} \tag{31}$$

$$(k-1)\sigma^2 \nabla^2 G \approx G(x, y, k\sigma) - G(x, y, \sigma) \tag{32}$$

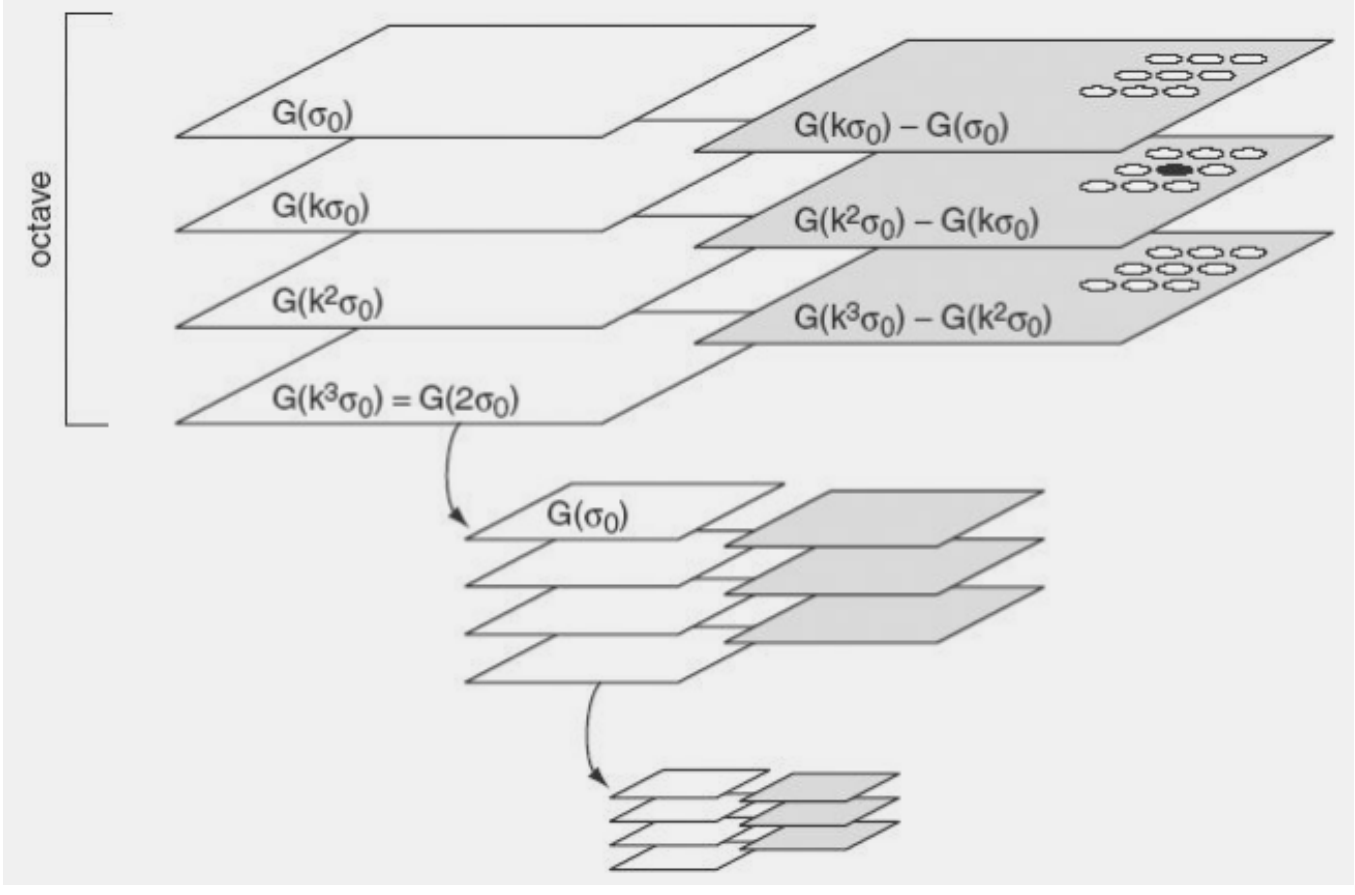
DoG detector

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (33)$$



Octave of a Gaussian scale space

$k = 2^{1/S}$, $S = 3$, $\sigma_0 = 1.6$



Additional refinements of DoG detector

- Local extrema with repect to both space and scale: larger or smaller than all twenty-six neighbors
- Fitting a quadratic function to the twenty-seven values of $D(x, y, \sigma)$ at and around the detected point (x_i, y_i, σ_i).

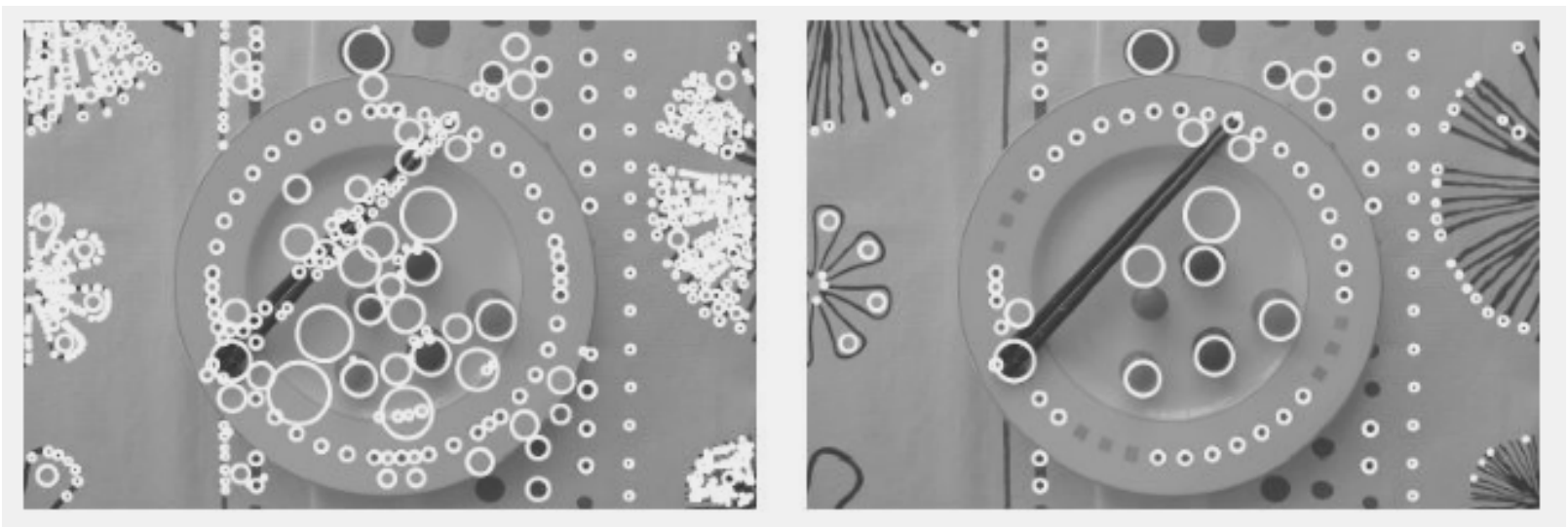
$$Q(x, y, \sigma) = D(x_i, y_i, \sigma_i) + g^T \begin{bmatrix} x \\ y \\ \sigma \end{bmatrix} + \frac{1}{2} \begin{bmatrix} x \\ y \\ \sigma \end{bmatrix}^T \Gamma \begin{bmatrix} x \\ y \\ \sigma \end{bmatrix} \tag{34}$$

where g is the gradient and Γ is the Hessian of D with respect to (x, y, σ), evaluated at (x_i, y_i, σ_i) using finite-difference approximations. The updated location and scale (sub-pixel accuracy)

$$\begin{bmatrix} \hat{x}_i \\ \hat{y}_i \\ \hat{\sigma}_i \end{bmatrix} = -\Gamma^{-1} g|_{(x_i, y_i, \sigma_i)} \tag{35}$$

Additional refinements of DoG detector

- Reject features that have poor contrast or correspond to edges rather than blobs.
  - poor contrast: small Q
  - edge-like: eigenvalues of the 2×2 spatial Hessian, one eigenvalue is much larger than the other, using trace and determinant



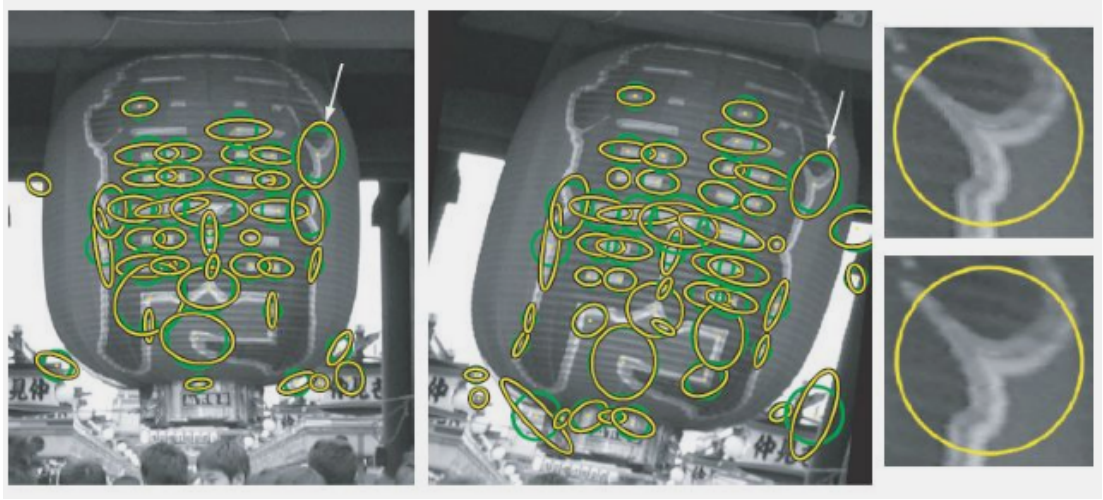
Affine-invariant regions

- Scale-invariant: Hessian-Laplace, LoG, DoG

Wide baseline

Suboptimal matches

Hessian-affine (yellow elipses)



Affine-invariant regions

Covariant: $E(I)$ is an elliptical region produced by a detector in an image I, and T is an affine transformation

$$T(E(I)) = E(T(I)) \tag{36}$$

[Lindeberg and Garding], [Baumberg], [Schaffalitzky and Zisserman], [Mikolajczyk and Schmid]

Affine adaptation:

1. Detect the feature point position and its characteristic scale
2. Compute the local second-moment matrix H at the given scale (the scale-normalized Harris matrix). Scale H so it has unit determinant.
3. Compute the Cholesky factorization $H = CC^T$, where C is a lower-triangular matrix with non-negative diagonal elements. C is sometimes called the matrix square root of H.
4. Warp the image structure around the feature point using the linear transformation C. $I_{\text{new}}(x_{\text{new}}) = I_{\text{old}}(Cx_{\text{old}})$
5. Compute the local second-moment matrix H for the new image and scale H so it has unit determinant.
6. If H is sufficiently close to the identity (its eigenvalues are nearly equal), stop. Otherwise, go to Step 3.

Harris-affine and Hessian-affine features

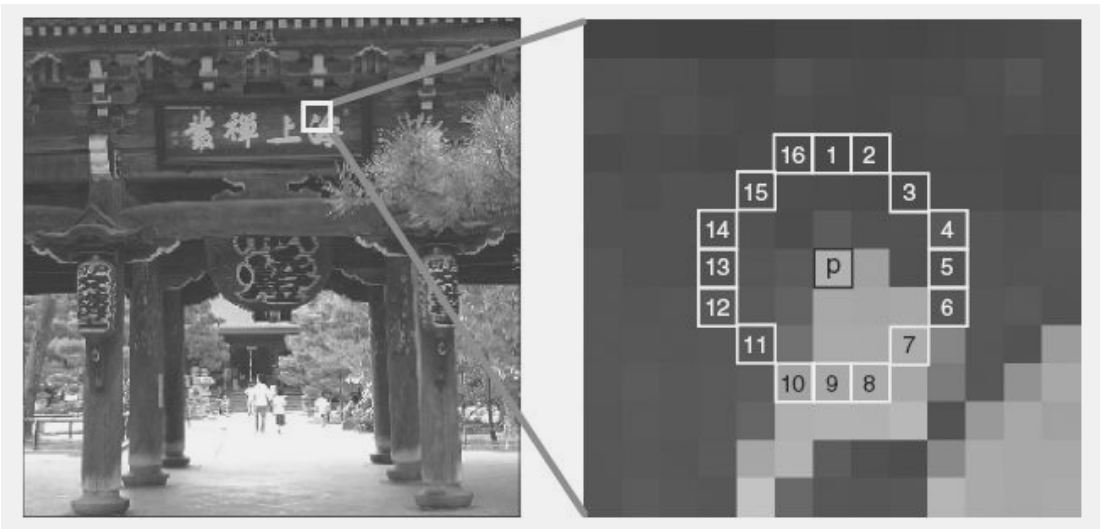
Mikolajczyk and Schmid.
Simultaneously detect feature point locations and corresponding affine-invariant regions using an iterative algorithm
Re-estimate the location and characteristic scale of the feature point.

FAST Corners

Rosten and Drummond. "Fusing points and lines for high performance tracking." ICCV 2005.

FAST: Features from Accelerated Segment Test

- A candidate pixel p is compared to a discretized circle of pixels around it.
- If all the pixels on a contiguous arc of n pixels around the circle are significantly darker or lighter than the candidate pixel, it is detected as a feature.
- $n = 12$ FAST-12: 1, 5, 9, 13 must pass the test



Using machine learning

- Create a database of a large number of pixel patches labeled as corners or not-corners
- Learn the structure of a decision tree based on the intensities of the sixteen surrounding pixels that was able to correctly classify all the patches in this training set.
- On the average, fewer than three intensity comparisons need to be made to determine if a candidate pixel is a FAST corner

Maximally Stable Extremal Regions (MSER)

Metas et al. "Robust wide-baseline stereo from maximally stable extremal regions." IVC, 2004.

- An extremal region is defined as a connected subset of pixels Ω such that for all $p \in \Omega$ and q adjacent to but not in Ω, the image intensities all satisfy either $I(p) < I(q)$ or $I(p) > I(q)$.
- Thresholding the image at a given pixel value
- Finding connected components as extremal regions
- Choosing extremal regions that are stable: as the threshold is varied, the connected component changes little.

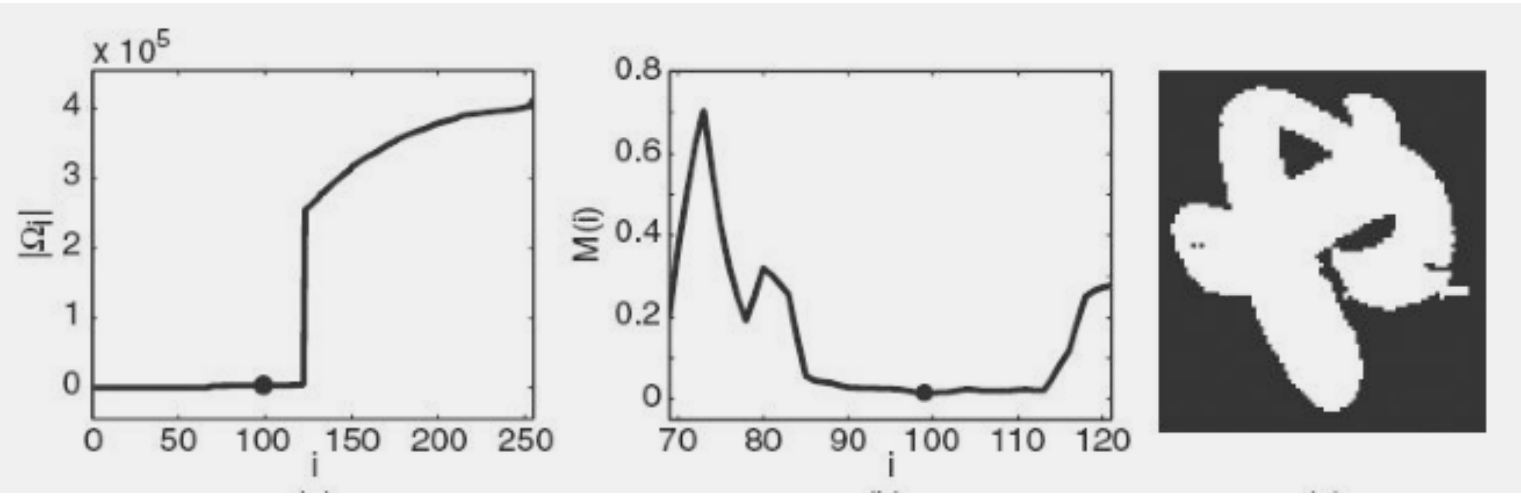
$$M(i) = \frac{|\Omega_{i+1} - \Omega_{i-1}|}{|\Omega_i|} \tag{37}$$

where $\{\Omega_i\}$ is a nested sequence of corresponding extremal regions obtained by thresholding the image at intensity i, and $|\Omega_i|$ is the area of Ω_i.

MSER

Thresholds 20, 125, and 200

$M(i)$ is minimized at inensity level 99



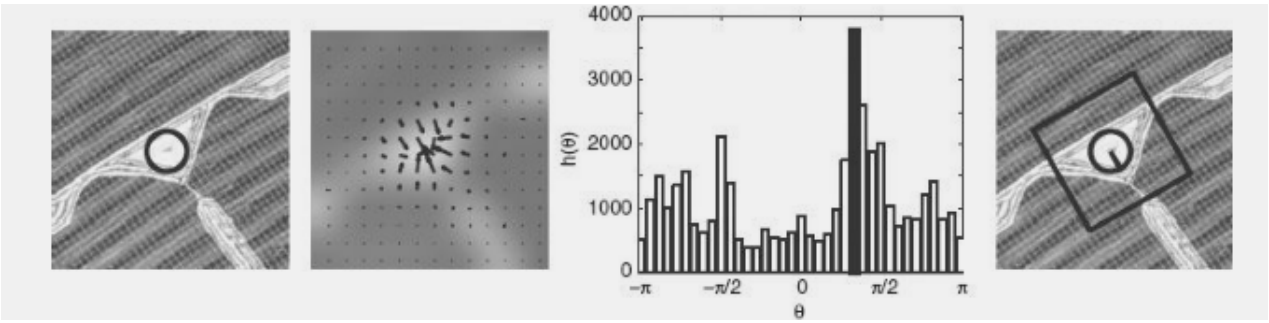
Feature descriptors

- To enable high-quality feature matching
- Features arising from the same 3D location in different views of the same scene result in very similar descriptor vectors
- $D(f) \approx D(Tf)$
- feature detection to be covariant to geometric transformations, feature description to be invariant to geometric transformations
- matching two descriptors from different images to form correspondences

Support regions

- Scale-invariant features: Harris-Laplace, Hessian-Laplace, DoG, detected at a characteristic scale
- Affine-covariant regions: MSER, Hessian-Affine
- Dominant gradient orientation of a patch: in SIFT, the orientation is estimated by a histogram of pixel gradient orientations over the support region of the scale-covariant circle.

The histogram $h(\theta)$ derived from
$\theta(x, y) = M(x, y)G(x - x_0, y - y_0, 1.5\sigma)$



Compensating for affine illumination changes

- Compute the mean μ and standard deviation s of the pixels in the support region, and rescale the intensities of the patch by

$$I_{\text{new}} = \frac{I_{\text{old}} - \mu}{s} \tag{38}$$

Matching criteria

Correspondences

- Given two sets of feature descriptors $\mathcal{A} = \{a^1, \ldots, a^{N_1}\}$ and $\mathcal{B} = \{b^1, \ldots, b^{N_2}\}$ in two images, generate a list of feature correspondences

$\{(c_a^j, c_b^j), j = 1, \ldots, N_3\}$, which says that $a^{c_a^j} \in \mathcal{A}$ and $b^{c_b^j} \in \mathcal{B}$ are a match.

Goal: to design a matching criterion that produces a high-quality set of correspondences

- few false matches and few missed matches

Matching criteria

Euclidean distance

$$D_{\text{euc}}(a, b) = \|a - b\|_2 = \left(\sum_{i=1}^{n} (a_i - b_i)^2 \right)^{1/2} \tag{39}$$

Sum-of-squared-differences distance (SSD)

$$D_{\text{ssd}}(a, b) = \|a - b\|_2^2 = D_{\text{euc}}(a, b)^2 \tag{40}$$

Mahalanobis distance

$$D_{\text{mahal}}(a, b) = \left((a - b)^T \Sigma^{-1} (a - b) \right)^{1/2} \tag{41}$$

where Σ is the covariance matrix

Matching criteria

Nearest neighbor

Use the method of nearest neighbors to find the match to a descriptor $a \in \mathcal{A}$

$$b^* = \arg\min_{b \in \mathcal{B}} D(a, b) \tag{42}$$

Reducing ambiguities

SIFT: Nearest neighbor distance ratio: accept (a, b^*) as a match if $D(a, b^*)/D(a, b^{**})$ is below a threshold (e.g. 0.8), where b^{**} is the descriptor with the second closestdistance to a.

Matching criteria

Normalized cross-correlation (NCC)

$$NCC(a,b) = \sum_{i=1}^{n} \frac{1}{s_a s_b}(a_i - \mu_a)(b_i - \mu_b) \tag{43}$$

where μ_a and s_a are the mean and standard deviation of the elements of a.

- often for matching raw blocks of intensities
- normalized for affine illumination changes, dot product between two blocks
- computed very efficiently using FFT

Other criteria

Epipolar geometry constraint

RootSIFT

- L1 normalize SIFT
- Square root each element
- Hellinger kernel or Bhattacharyya coefficient

$$H(a, b) = \sum_{i=1}^{n} \sqrt{a_i b_i} \tag{44}$$

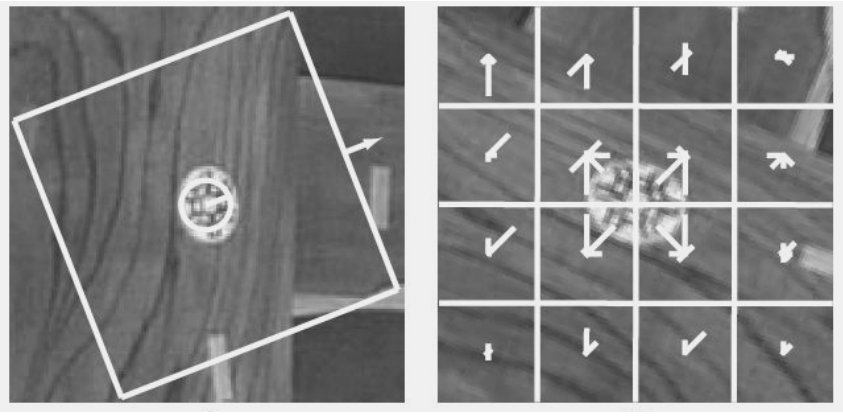
where $\sum_{i=1}^{n} a_i = 1$.
Arandjelovic and Zisserman. "Three things everyone should know to improve object retrieval." CVPR 2012.

Histogram-based descriptors

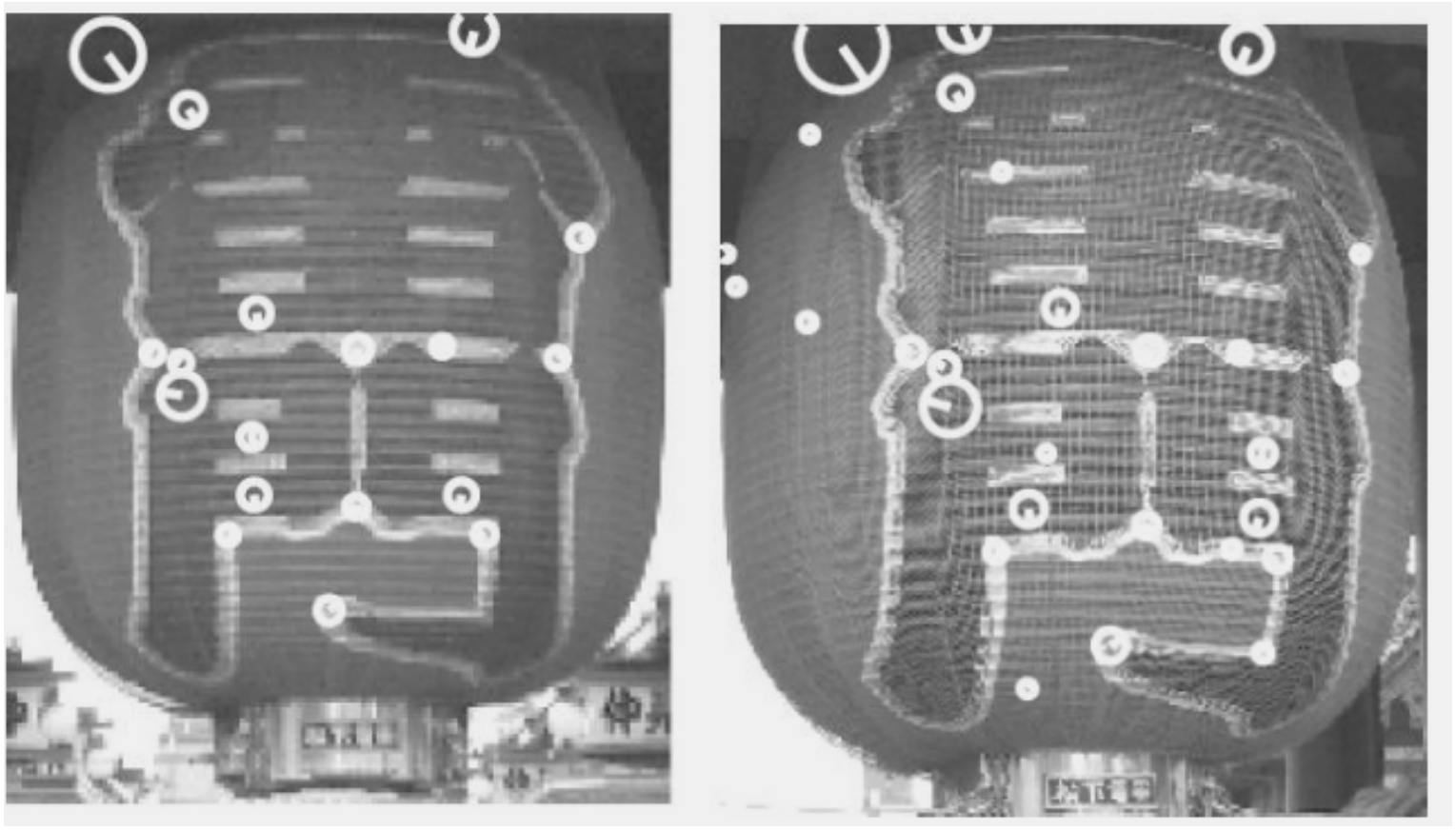
- SIFT
- GLOH
- Shape contexts
- Spin images

SIFT descriptor

- Estimating the location and scale σ
- Deciding the dominant orientation
- Support region: oriented square at the feature location location, side $= 6\sigma$
- Spatially Gaussian weighted gradients
- 4×4 grid of cells, eight gradient orientations, \$128\$-dimensional vector
- the gradient at each pixel contributes to multiple cells and multiple histogram bin based on trilinear interpolation.
- normalizeded to unit length, zeroing out any extremely large values ($>$0.2), and renormalizing to unit length.



Correspondences botained by SIFT

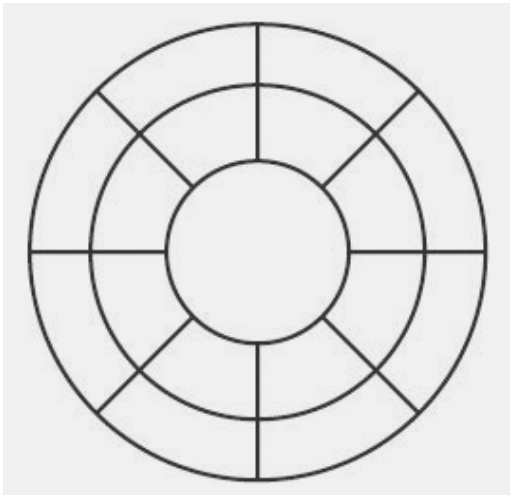


GLOH

Gradient Location and Orientation histogram

- log-polar grid
- sixteen angles
- $17 \times 16 = 272$ dimensions.
- dimensionality reduced to 128 using PCA

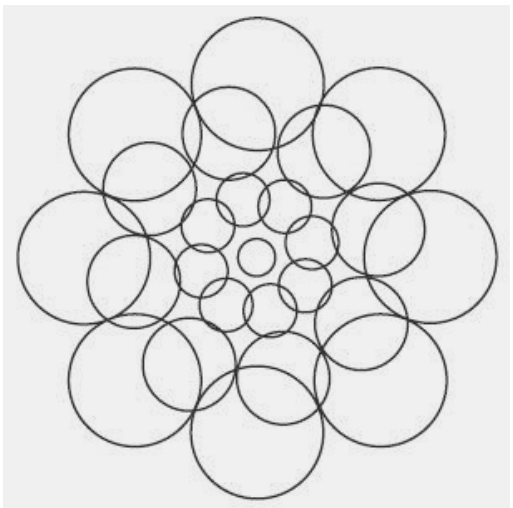
Mikolajczyk and Schmid. "A performance evaluation of local descriptors." PAMI, 2005.



DAISY

- Soft support regions

Winder and Brown. “Learning local image descriptors.” CVPR 2007.
Toal, Lepetit, and Fua. “DAISY: An efficient dense descriptor applied to wide-baseline stereo.” PAMI, 2010.



Shape contexts

- log-polar location grid
- histogram of edge points, weighted by gradient of edge point

Belongie, Malik, Puzicha. "Shape matching and object recognition using shape contexts." PAMI, 2002.

Spin images

- Rotation-invariant
- histogram of quantized intensities for each of several rings around the feature location

Johnson and Hebert. "Using spin images for efficient object recognition in cluttered 3D scenes." PAMI, 2002.
Lazebnik, Schmid, and Ponce. "A sparse texture representation using local affine regions." PAMI, 2005.

Invariant-based descriptors

- bypassing the estimation of rotation
- invariant function of the patch pixels with respect to a class of geometric transformations, e.g., rotation or affine

Differential invariants

- combinations of increasingly higher-order derivatives of the Gaussian-smoothed image $L(x, y)$
- sum of squared gradient magnitude

$$\sum_{(x,y)} \left(\frac{\partial L(x,y)}{\partial x} \right)^2 + \left(\frac{\partial L(x,y)}{\partial y} \right)^2 \tag{45}$$

- sum of Laplacians

$$\sum_{(x,y)} \frac{\partial^2 L(x,y)}{\partial x^2} + \frac{\partial^2 L(x,y)}{\partial y^2} \tag{46}$$

- issues: accuracy in estimation of higher-order derivatives, noise

Schmid and Mohr. "Local grayvalue invariants for image retrieval." PAMI, 1997.

Moment invariants

- using both image intensities and spatial coordinates
- the (m,n)-th moment of a function defined over a region is the average value of $x^m y^n f(x, y)$
- $(0, 0)$ moment?
- $(0, 1)$ and $(1, 0)$ moments?
- Schaffalitzky and Zisserman: a bank of complex filters whose magnitude responses are invariant to rotation, similar to derivatives of a Gaussian

$$K_{mn}(x, y) = (x + iy)^m (x - iy)^n G(x, y) \tag{47}$$

Flusser. "On the independence of rotation moment invariants." Pattern Recognition, 2000.
Van Gool, Moons, and Ungureanu. "Affine photometric invariants for planar intensity patterns." ECCV 1996.
Schaffalitzky and Zisserman. "Multi-view matching for unordered image sets, or 'How do I organize my holiday snaps?'." ECCV 2002.

Other approaches

Steerable filters

- sum of responses to a small number of basis filter at canonical orientations.

Freeman and Adelson. "The design and use of steerable filters." PAMI, 1991.

- Filter bank of Gaussian derivatives with respect to the angel

Mikolajczyk and Schmid. "Indexing based on scale invariant interest points." ICCV 2001.

SURF (Speeded-up robust featuers)

- 64-dimensional descriptor

Bay, Tuytelaars, Van Gool. "SURF: Speeded up robust features." ECCV 2006.

Other approaches

PCA-SIFT

- PCA performed on the raw gradients of a scale- and rotation-normalized patch.
- collect a large number of DoG keypoints and construct 41×41 patches;
- the x and y gradients at the interior pixels: $39 \times 39 \times 2 = 3042$

Ke and Sukthankar. "PCA-SIFT: a more distinctive representation for local image descriptors." CVPR 2004.

Evaluating detectors and descriptors

Repeatability score

$$RS = \frac{\text{repeated detections}}{\min(N_1, N_2)} \tag{48}$$

Matching score

$$MS = \frac{\text{correct nearest-neighbor matches}}{\min(N_1, N_2)} \tag{49}$$

- precision and recall

$$\text{precision} = \frac{\text{correct matches}}{\text{total matches}} \tag{50}$$

$$\text{recall} = \frac{\text{correct matches}}{\text{true correspondences}} \tag{51}$$

Recent detectors and descriptors

BRIEF, BRISK, FREAK

ORB (Oriented FAST and Rotated BRIEF)

SIFER/D-SIFER (Scale-Invariant Feature detector with Error Resilience)

FRIF (Fast Robust Invariant Feature)

Color detectors and descriptors

Artificial markers

QR codes

ARToolKit/ARToolKitPlus

ARTag