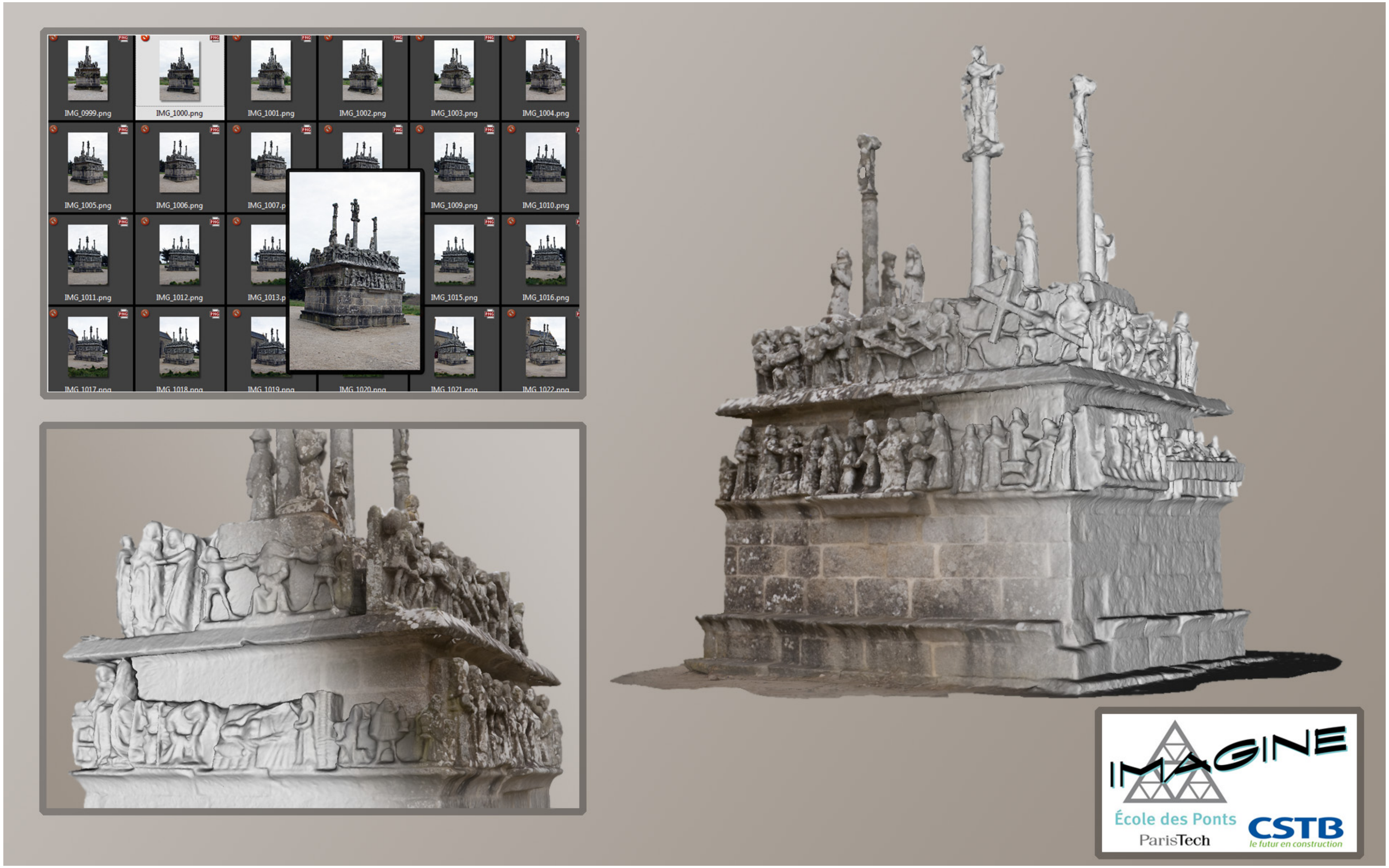
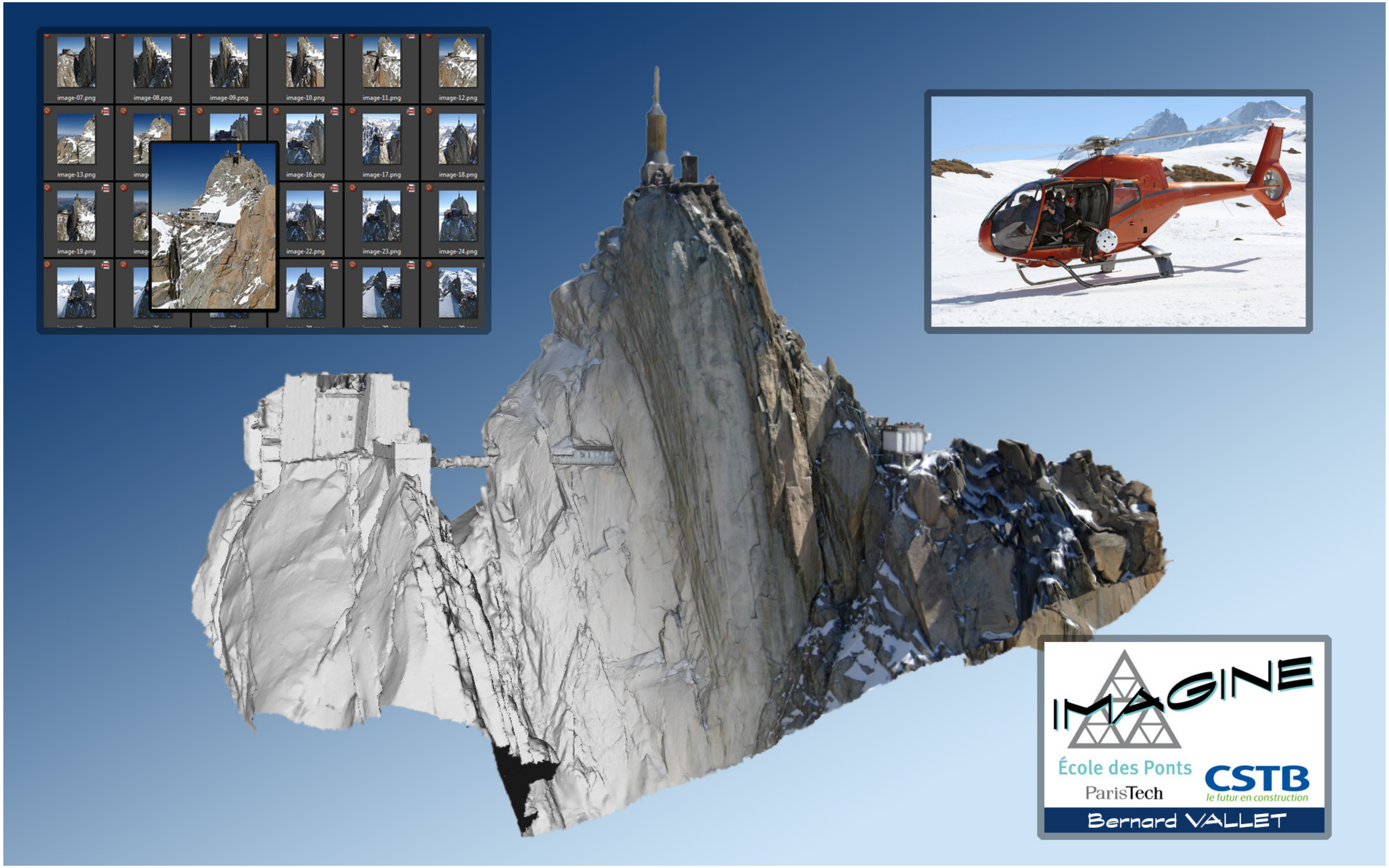
Side note: Cluny abbey

Applications to virtual tourism, video games, film industry...

Problem 4: 3D model construction (cont.)



Problem 4: 3D model construction (cont.)



Problem 4: 3D model construction (cont.) Two main tasks

- External camera calibration

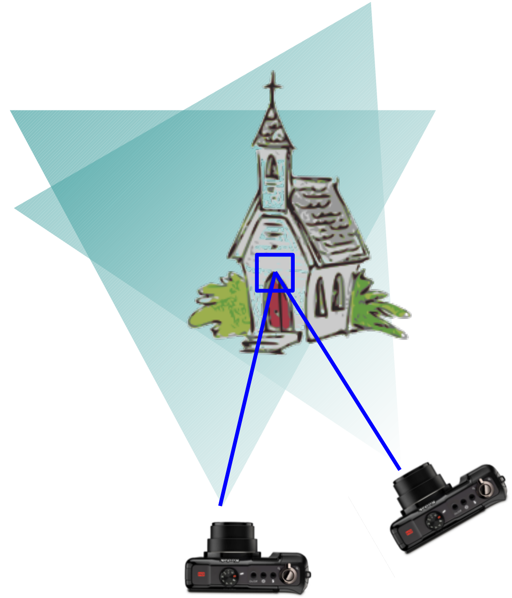
  = determination of pose (i.e., location and orientation) of each camera in a common coordinate system

  – requires corresponding points in several images → **detection** and **matching** of salient points

- Dense 3D reconstruction

  = by triangulation, given camera pose (!) not restricted to salient points only

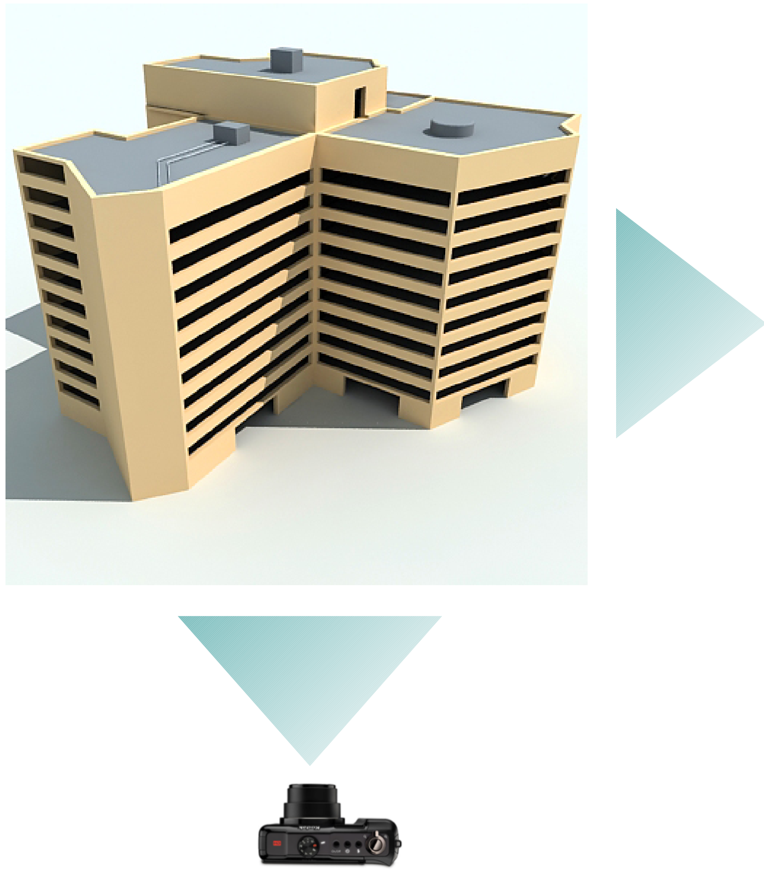
  – requires **matching** image patches in several images



Question

Suppose you are given two views of an object.

What can be obstacles to feature detection & matching?



Robustness / repeatability issues

Obstacles to detection and matching :

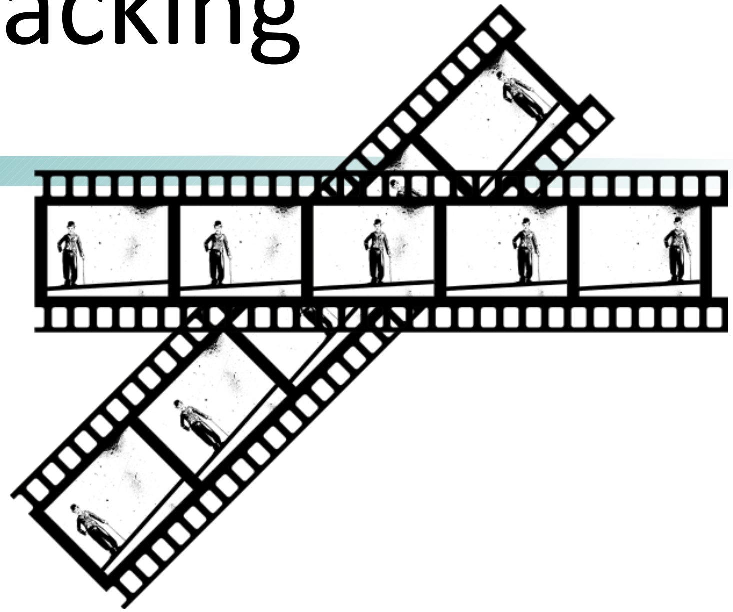
- change of scale
- change of orientation (rotation)
- change of viewpoint (affine, projective transformations)
- change of illumination
- noise
- clutter & occlusion
- repetitive patterns

☛ Design of robust similarity measures, detector and descriptors/matchers

Wrap-up: Problems to address

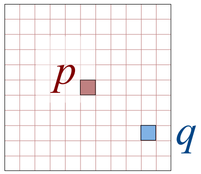
- Similarity measures
  - how to compare image patches?
- Salient point detection
  - what are singular patches?
- Salient point matching
  - how to abstract patches and compare abstraction?
- … in a robust way

A similar setting: tracking



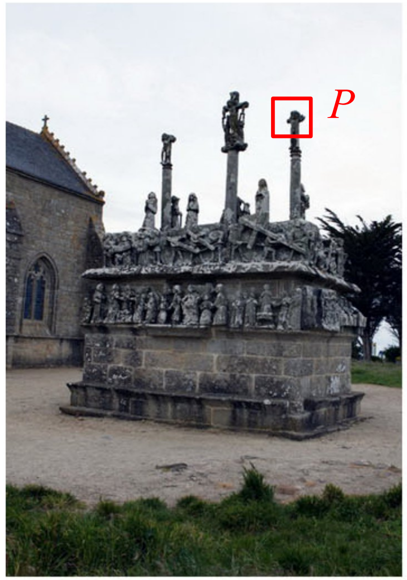
- Problem: in a **video**
  - maintain a set of correspondences
- Solution 1: naïve approach
  - detect features in all images (frames) and match them
- Solution 2: tracking approach
    - limited movement between successive frames
    - next displacement can be anticipated from previous motion
  - in frame 1, detect features
  - in following frames, look for corresponding features (or similar image patches) **only where expected**

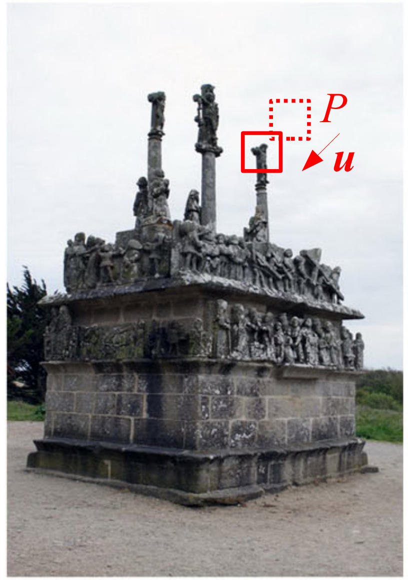
Common similarity measures



patch P
centered on p

- P (or P_p): patch of pixels around given point p in image I
- $\boldsymbol{x}_q = (x,y)$: position of pixel $q \in P$ in image I
- $\boldsymbol{u} = (u,v)$: displacement of patch P in image I'
- N.B. smaller value ↔ more similar (0 ↔ equal)





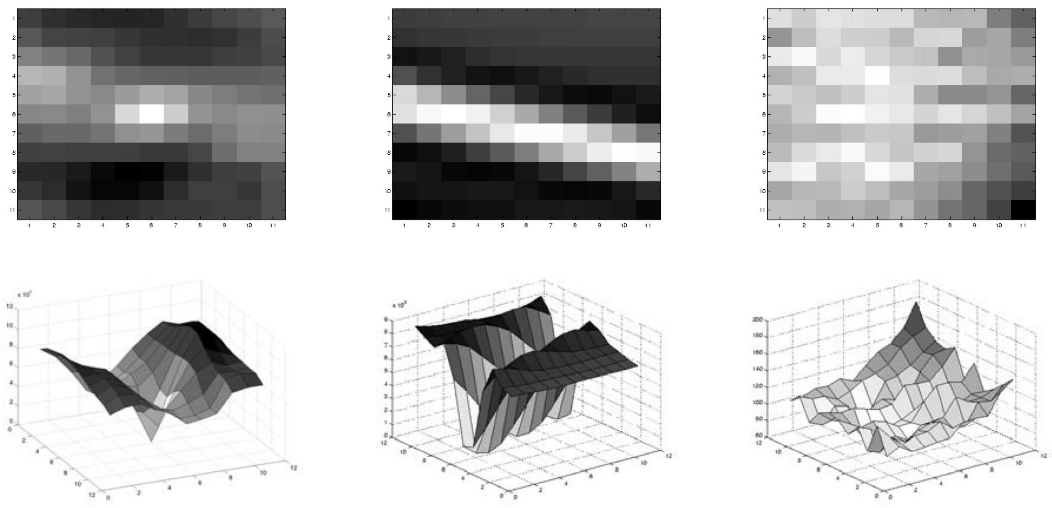
- Sum of square difference (SSD) [similar ↘]
  - $E_{SSD}(P\,;\boldsymbol{u}) = \sum_{q \in P} [I'(\boldsymbol{x}_q+\boldsymbol{u}) - I(\boldsymbol{x}_q)]^2$
- Cross correlation (CC) [similar ↗]
  - $E_{CC}(P\,;\boldsymbol{u}) = \sum_{q \in P} [I'(\boldsymbol{x}_q+\boldsymbol{u})\, I(\boldsymbol{x}_q)]$

    meaningful mainly if normalized (see below)
- Auto-correlation (AC): single image $I = I'$
  - $E_{AC}(P\,;\boldsymbol{u})$: applies to $E_{SSD}(P\,;\boldsymbol{u})$ or $E_{CC}(P\,;\boldsymbol{u})$

Auto-correlation surfaces

- AC surface = P fixed: $E_{AC}(\boldsymbol{u})$
- original image:

  red crosses = locations of AC surface computation

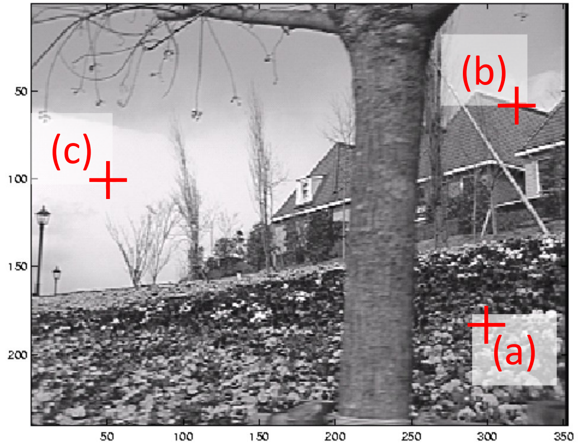


(a) (b) (c)

Q1: Which AC surface corresponds to which cross ?

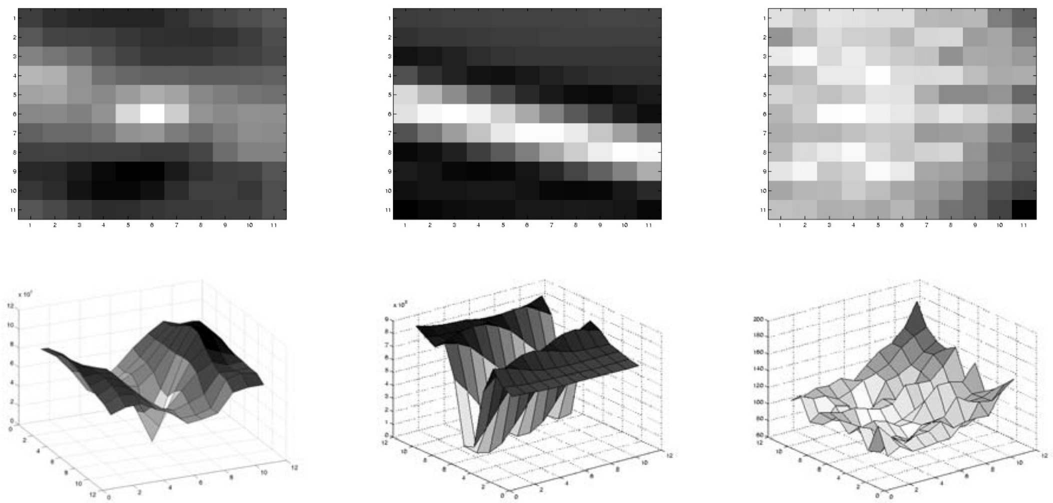
Q2: Which surface corresponds to a distinctive feature ?

Auto-correlation surfaces



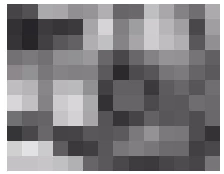
- AC surface = P fixed: $E_{AC}(\boldsymbol{u})$
- original image:

  red crosses = locations of
  AC surface computation



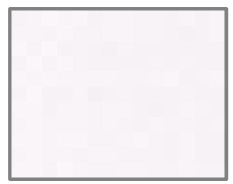
(a) (b) (c)

- (a): textured patch, good unique minimum



- (b): patch with edge, 1D aperture problem (↔ barber-pole illusion)

- (c): textureless, no peak



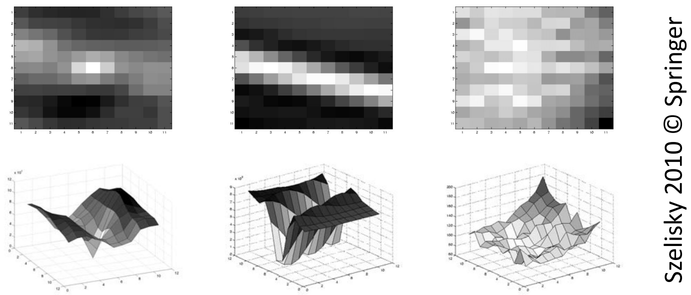
Two uses of local similarity measures

- Correspondence assessment
  - If a patch P_1 around point p_1 in image I is similar to a patch P_2 around point p_2 in image I', then p_1 and p_2 are potential matches.
- Saliency for detection
  - A point that is dissimilar to other points in its neighborhood is salient, and thus “detected”.

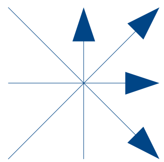
Auto-correlation for detection

(Moravec 1980)

threshold = seuil



- Directional variance
  - $E_{AC}(P\,;\boldsymbol{u}) = \sum_{q \;\; P} [I(\boldsymbol{x}_q+\boldsymbol{u})-I(\boldsymbol{x}_q)]^2$
    - patch P: square window (typ. 4x4 to 8x8)
    - 4 directions: $\boldsymbol{u} \in U = \{(0,1),(1,0),(1,1)(1,\text{-}1)\}$

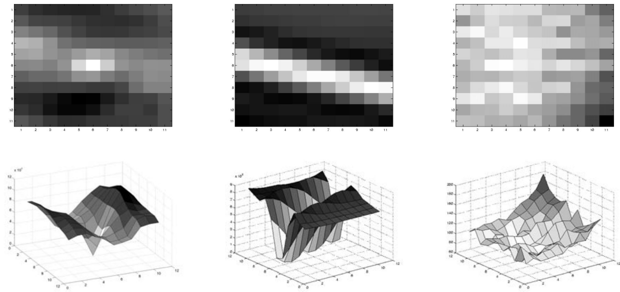


- Interest points
  - s.t. $\min_{\boldsymbol{u} \;\; U}(E_{AC}(P\,;\boldsymbol{u}))$ above threshold and local maximum (typ. 8 neighbors)

- Why is called a “corner” detector?

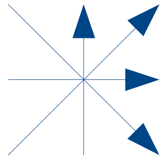
Auto-correlation for detection
(Moravec 1980)

threshold = seuil

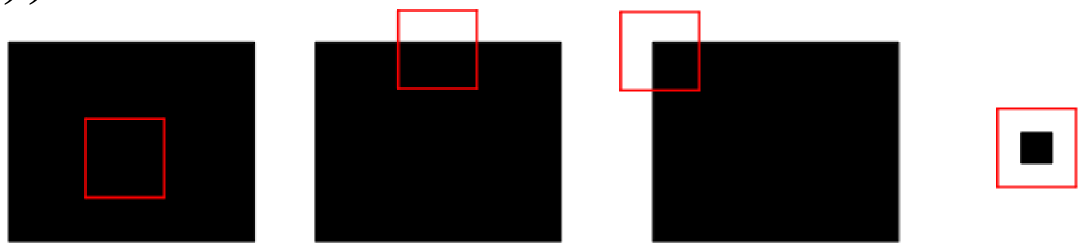


Szelisky 2010 © Springer

- Directional variance
  - $E_{AC}(P\,;\boldsymbol{u}) = \sum_{q\;\;P}[I(\boldsymbol{x}_q+\boldsymbol{u})-I(\boldsymbol{x}_q)]^2$
    - patch P: square window (typ. 4x4 to 8x8)
    - 4 directions: $\boldsymbol{u} \in U = \{(0,1),(1,0),(1,1)(1,-1)\}$



- Interest points
  - s.t. $\min_{\boldsymbol{u}\;\;U}(E_{AC}(P\,;\boldsymbol{u}))$ above threshold and local maximum (typ. 8 neighbors)
  - "corner" detector



A. Interior Region
Little intensity variation in any direction

B. Edge
Little intensity variation along edge, large variation perpendicular to edge

C. Edge
Large intensity variation in all directions

D. Edge
Large intensity variation in all directions

Parks & Gravel @ McGill U.

Fair performance, some problems...

Auto-correlation for detection
(Harris-Stephens 1988)

- Pb 1 (in Moravec): discrete set of shifts → anisotropic
- Solution: analytic expansion (Taylor, 1st order)

$$I(\boldsymbol{x}_q + \Delta\boldsymbol{u}) \approx I(\boldsymbol{x}_q) + \nabla I(\boldsymbol{x}_q)\, \Delta\boldsymbol{u}$$

$$E_{AC}(P;\Delta\boldsymbol{u}) = \sum_{q\in P} [I(\boldsymbol{x}_q + \Delta\boldsymbol{u}) - I(\boldsymbol{x}_q)]^2$$
$$\approx \sum_{q\in P} [\nabla I(\boldsymbol{x}_q)\, \Delta\boldsymbol{u}]^2 = \Delta\boldsymbol{u}^T \boldsymbol{A}_P\, \Delta\boldsymbol{u}$$

$$\boldsymbol{A}_P = \begin{bmatrix} \sum_{q\in P} I_x^2(\boldsymbol{x}_q) & \sum_{q\in P} I_x(\boldsymbol{x}_q) I_y(\boldsymbol{x}_q) \\ \sum_{q\in P} I_x(\boldsymbol{x}_q) I_y(\boldsymbol{x}_q) & \sum_{q\in P} I_y^2(\boldsymbol{x}_q) \end{bmatrix} \text{ with } \begin{aligned} I_x(\boldsymbol{x}_q) &= \frac{\partial I}{\partial x}(\boldsymbol{x}_q) \\ I_y(\boldsymbol{x}_q) &= \frac{\partial I}{\partial y}(\boldsymbol{x}_q) \end{aligned}$$

$\boldsymbol{A}_P$: auto-correlation matrix

(cf. second-moment matrix, structure tensor)

Auto-correlation for detection
(Harris-Stephens 1988)

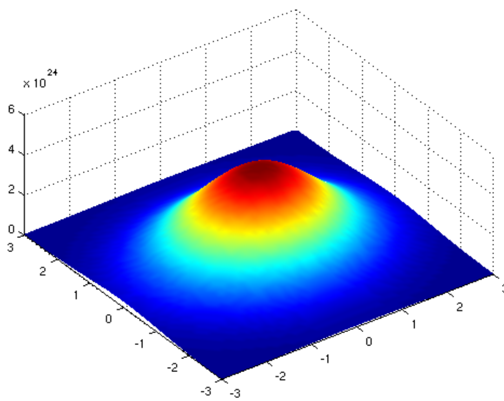
- Pb 2: rectangular binary window → noisy, anisotropic
- Solution: use smooth circular window, e.g., Gaussian

→ insensitive to in-plane rotation

$$E_{AC}(P;\Delta \boldsymbol{u}) = \sum_{q\in P} w(\boldsymbol{x}_q)\,[I(\boldsymbol{x}_q+\Delta \boldsymbol{u}) - I(\boldsymbol{x}_q)]^2$$
$$\approx \sum_{q\in P} w(\boldsymbol{x}_q)\,[\nabla I(\boldsymbol{x}_q)\,\Delta \boldsymbol{u}]^2 = \Delta \boldsymbol{u}^T \boldsymbol{A}_P\, \Delta \boldsymbol{u}$$

$$\boldsymbol{A}_P = \sum_{q\in P} \left(w * \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \right)(\boldsymbol{x}_q)$$

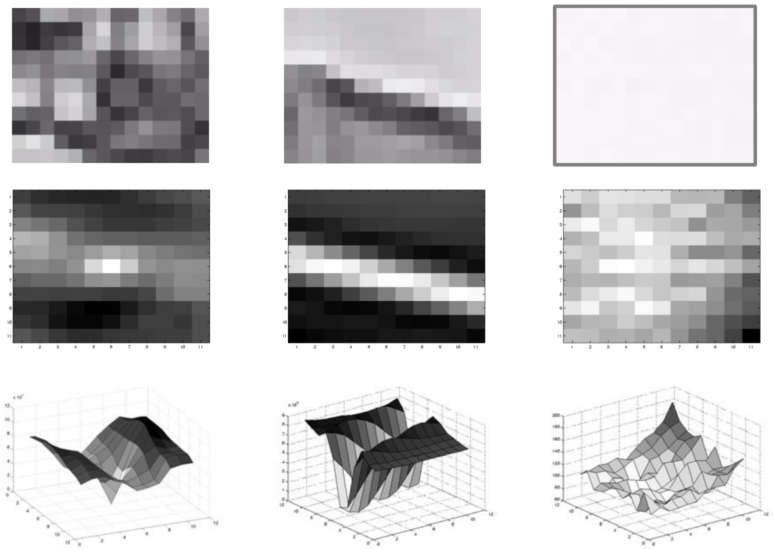
e.g., $w(\boldsymbol{x}_q) = G(x,y;\sigma) = \dfrac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$



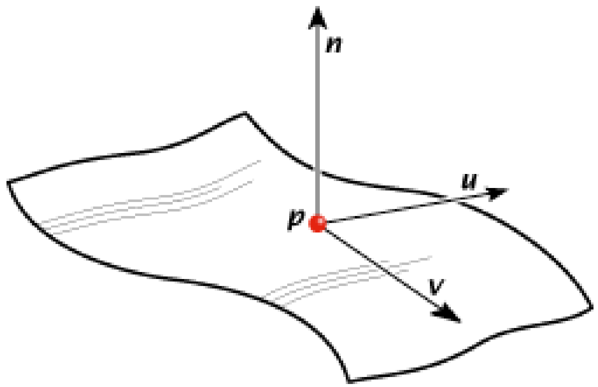
Auto-correlation for detection

(Harris-Stephens 1988)

- Pb 3: $\min_{\boldsymbol{u}}(E_{AC}(P\,;\boldsymbol{u})) \rightarrow$ too many edge responses



- Solution:
  - keep only marked peaks
  - for this, look at local curvature

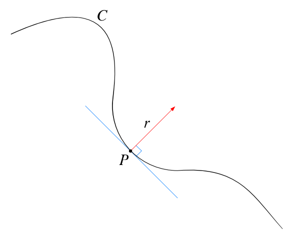


curve = courbe
curvature = courbure
osculating circle = cercle osculateur
principal curvature = courbure principale
eigenvalue = valeur propre
Hessian (matrix) = (matrice) hessienne

Curvature

- Curvature k of plane curve C at point P

  = $1/r$ where r radius of osculating circle



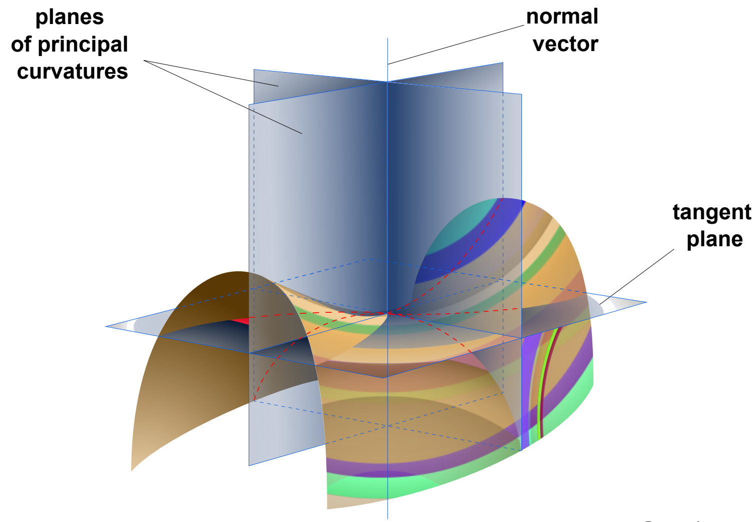
- Principal curvatures k_1 and k_2 of surface $S(u,v)$ at P

  = max & min value of curvature for different normal planes

  – sign convention: + if turns in same direction as chosen normal

  = eigenvalues of Hessian of S (shape operator) at P

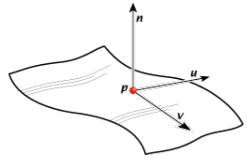
$$H(S) = \begin{bmatrix} \frac{\partial^2 S}{\partial u^2} & \frac{\partial^2 S}{\partial u \partial v} \\ \frac{\partial^2 S}{\partial u \partial v} & \frac{\partial^2 S}{\partial v^2} \end{bmatrix}$$



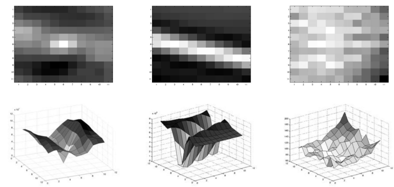
Auto-correlation for detection
(Harris-Stephens 1988)

peak = pic
sharp = tranchant, aigu, marqué...
ridge = crête

- Pb 3: $\min_{\boldsymbol{u}}(E_{AC}(P\,;\boldsymbol{u})) \to$ too many edge responses
- Solution: look at local curvature of E_{AC}



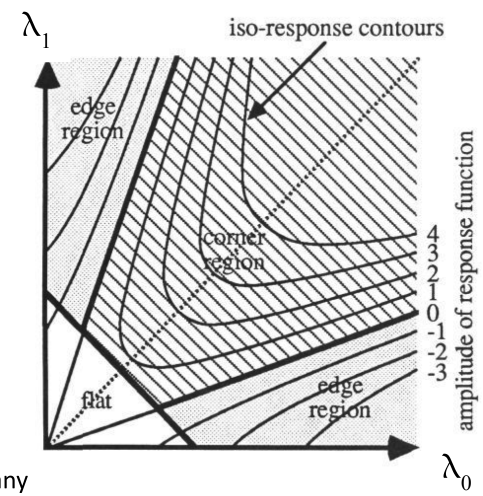
  - $E_{AC}(P\,;\boldsymbol{u}) \approx \boldsymbol{u}^T \boldsymbol{A}_P\, \boldsymbol{u}$ for $\boldsymbol{u}$ small (2nd order discarded)
  - $(H(E_{AC}))(P) \approx \boldsymbol{A}_P$
  - principal curvatures: eigenvalues λ_0, λ_1 of $\boldsymbol{A}_P$
    (→ rotational invariance description of $\boldsymbol{A}_P$)



(a) λ_0, λ_1 large: E_{AC} sharply peaked $\to$ corner

(b) λ_0 small, λ_1 large: E_{AC} ridged shape $\to$ edge

(c) λ_0, λ_1 small: E_{AC} flat $\to$ +/- constant intensity

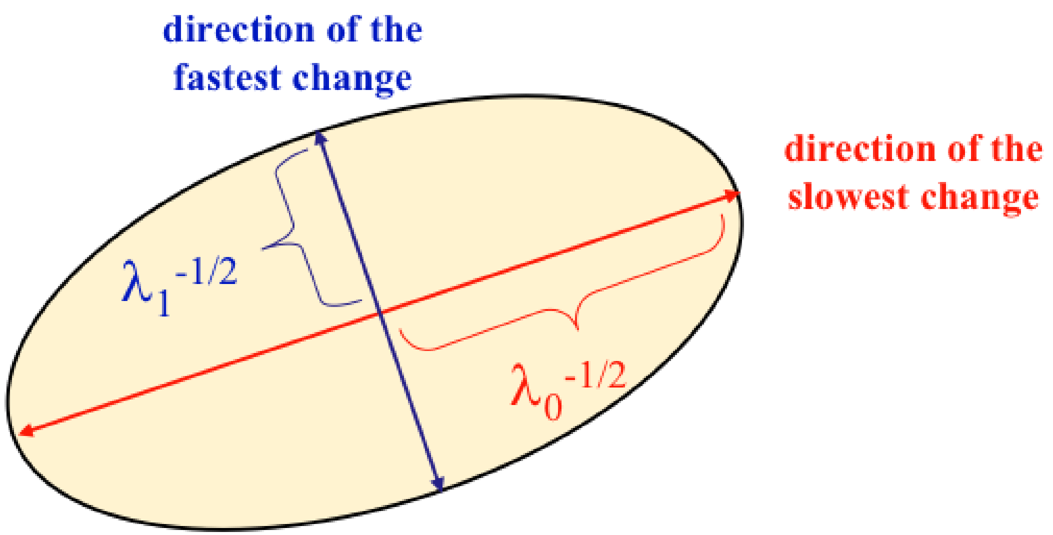


Eigenvalue-based criteria

- Good features to track (Shi & Tomasi 1994)

  larger uncertainty ↔ smaller eigenvalue λ_0

  ☛ look for maxima in smaller eigenvalue λ_0



Szelisky 2010 © Springer

Eigenvalue-based criteria (cont.)

Avoid explicit eigenvalue decomposition (square root)

☛ only use determinant and trace of $\boldsymbol{A}$

- Corner response (Harris-Stephens 1988)

$$R = \det(\boldsymbol{A}) - \alpha\ \mathrm{tr}(\boldsymbol{A})^2 = \lambda_0\lambda_1 - \alpha\,(\lambda_0 + \lambda_1)^2$$

with $\alpha = 0{,}06$ (common: $0{,}04 \leq \alpha \leq 0{,}15$)

- Corner strength (Brown et al. 2005): harmonic mean

$$f = \frac{\det(\boldsymbol{A})}{\mathrm{tr}(\boldsymbol{A})} = \frac{\lambda_0\lambda_1}{\lambda_0+\lambda_1}$$

→ smoother response in the region where $\lambda_0 \approx \lambda_1$

[see also SIFT detector below]

Computations for the so-called "Harris corner detector"

- Compute for each point p and corresponding patch P:

$$\boldsymbol{A}_P = \sum_{q \in P} \left(w * \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \right) (\boldsymbol{x}_q)$$

$$\text{where} \quad I_x(\boldsymbol{x}_q) = \frac{\partial I}{\partial x}(\boldsymbol{x}_q), \quad I_y(\boldsymbol{x}_q) = \frac{\partial I}{\partial y}(\boldsymbol{x}_q)$$

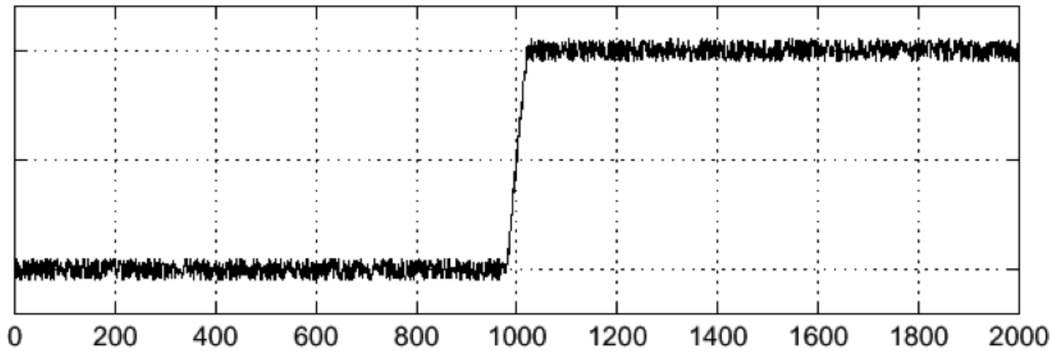
$$w(\boldsymbol{x}_q) = G(x, y; \sigma) = \frac{1}{2\pi\sigma^2} \mathrm{e}^{-\frac{x^2+y^2}{2\sigma^2}}$$

then consider criterion based on $\det(\boldsymbol{A})$, $\operatorname{tr}(\boldsymbol{A})$

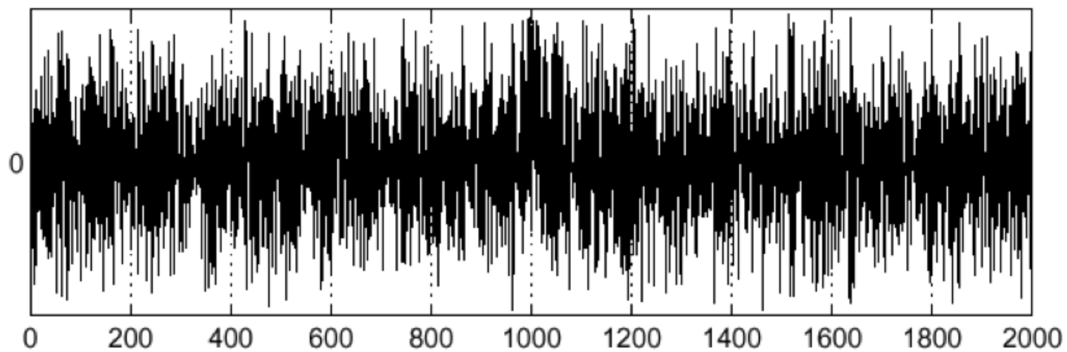
- Is this computation efficient?
  How to compute it efficiently?

Differentiating in the presence of noise

$f(x)$



$\frac{d}{dx} f(x)$

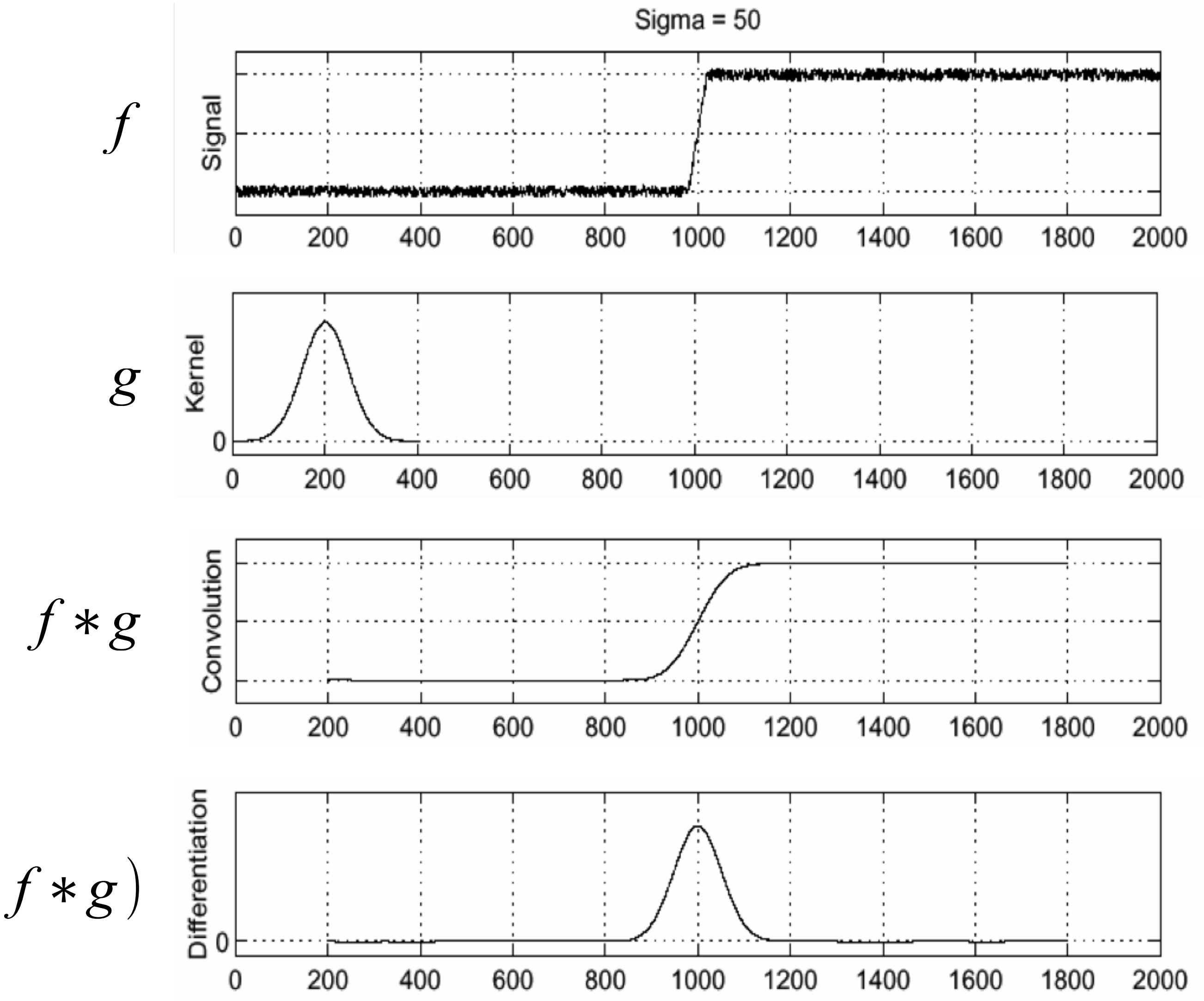


Seitz @ U. Washington

Edge not noticeable because of noise

Differentiating in the presence of noise

- Smooth first
- Look for peak of $\frac{d}{dx}(f * g)$



Seitz @ U. Washington

Efficient Differentiation

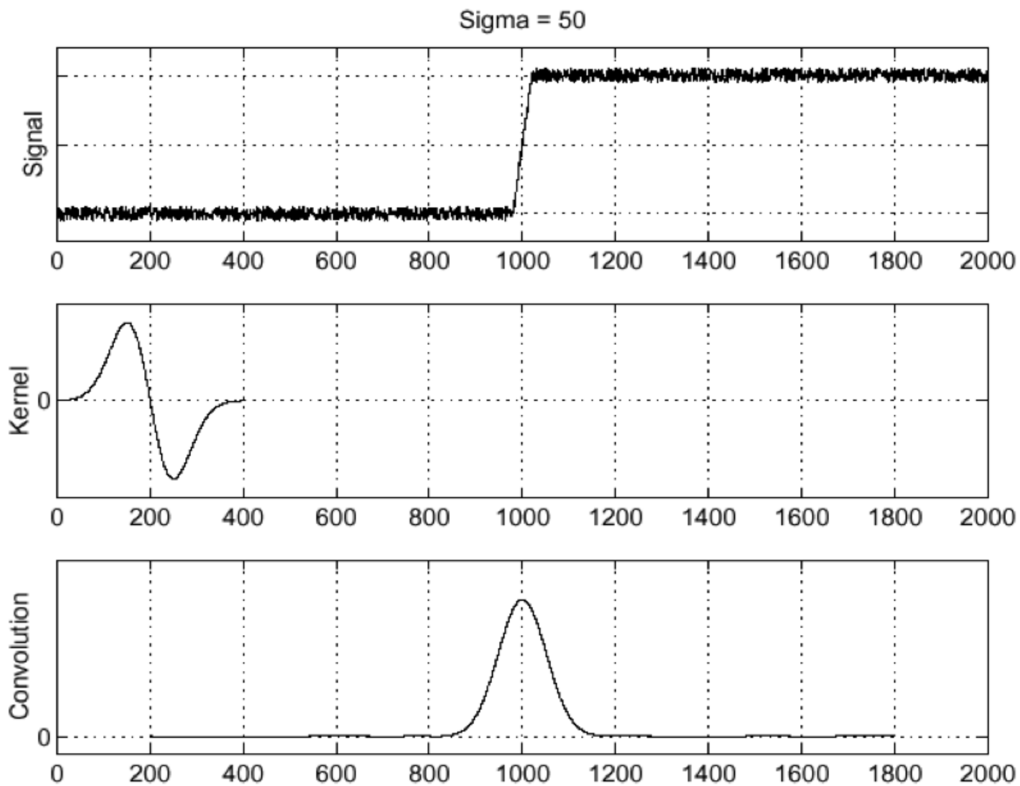
- Property of differentiation and convolution

$$\frac{d}{dx}(f*g) = \frac{d}{dx}(g*f) = (\frac{d}{dx}g)*f = f*\frac{d}{dx}g$$

f

$\frac{d}{dx}g$

$f*\frac{d}{dx}g$



- Associativity → smaller size masks → less computations (similar to clever associativity for efficient matrix multiplication)

Seitz @ U. Washington

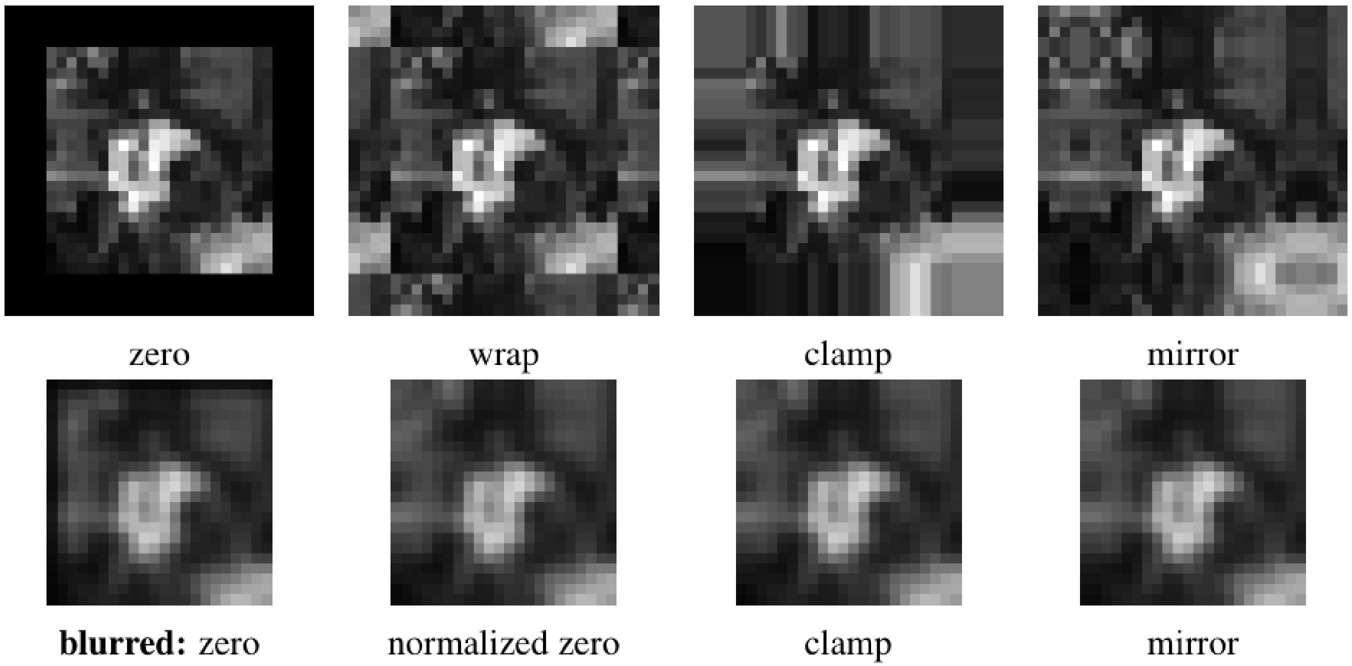
Algorithm for the so-called "Harris corner detector"

- For each p of I, compute the derivatives $I_x(\boldsymbol{x}_p)$ and $I_y(\boldsymbol{x}_p)$
  - convolve operators $d_x = [\text{-}½\ 0\ ½]$ and $d_y = [\text{-}½\ 0\ ½]^T$ with smoothing "derivation" Gaussian (e.g., $\sigma_d = 1$) → derivative masks
  - convolve I with the derivative masks → I_x and I_y
  - using 1D-convolutions only (1D-Gaussian and 1D-derivation), not 2D-convolutions → more efficient [see slides on convolution]
- For each p, compute product of derivatives $I_x^{\,2}$, I_xI_y, $I_y^{\,2}$
  - and extra smoothing with an "integration" Gaussian (e.g., $\sigma_i = 2$)
- For each p, consider auto-correlation matrix
  - compute "corner response"
  - response above threshold and local maximum → detection
  - possibly: only keep locally significant responses (see ANMS below)

Image boundary effects

- Padding strategies (aka wrapping mode, texture addressing mode)
  - pad with 0 (or constant), wrap (loop around), clamp (replicate edge pixel), mirror (reflect pixels across edge)

Blurring examples :



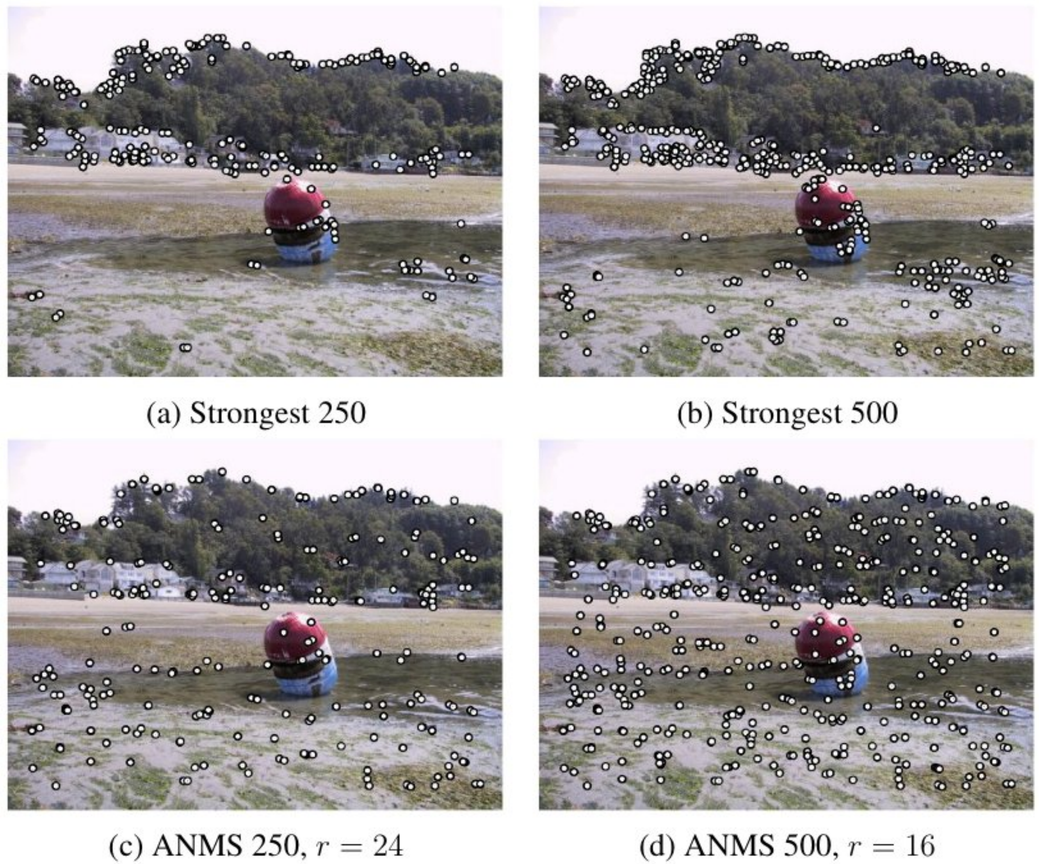
  - or discard results close to boundary...

Szelisky 2010 © Springer

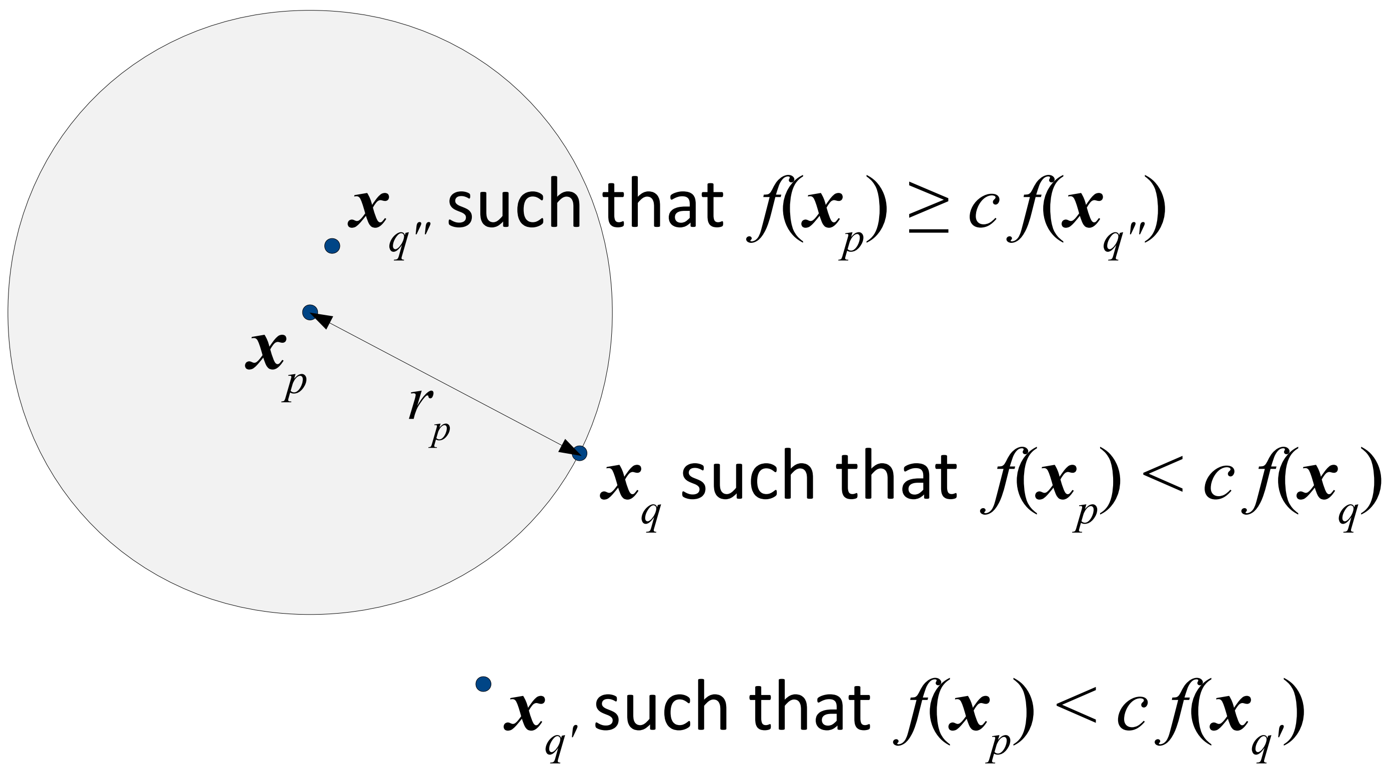
Adaptive non-maximal suppression (ANMS)

- Problem: local maxima
  → uneven distribution
  - denser in regions of higher contrast
- Sol.: only keep locally significant responses
  - greater (e.g. 10%+) than all neighbors within given radius r
  - choose r such that n detections only:
    $r_p = \min_{q\ detection} \| \boldsymbol{x}_p - \boldsymbol{x}_q \|$ such that $f(\boldsymbol{x}_p) < 0.9\, f(\boldsymbol{x}_q)$

Brown et al. 2005



Adaptive non-maximal suppression (ANMS)



An Algorithm for ANMS

$r_{\min} = \infty$

$ProcessedPoints = $

sort detections by decreasing strength

for each detection p, in decreasing strength order

$r_p = \min_{q \in ProcessedPoints} \| \boldsymbol{x}_p - \boldsymbol{x}_q \|$ such that $f(\boldsymbol{x}_p) < c\, f(\boldsymbol{x}_q)$

(= suppression radius w.r.t. $ProcessedPoints$)

if $r_p < r_{\min}$ then

$r_{\min} = r_p$

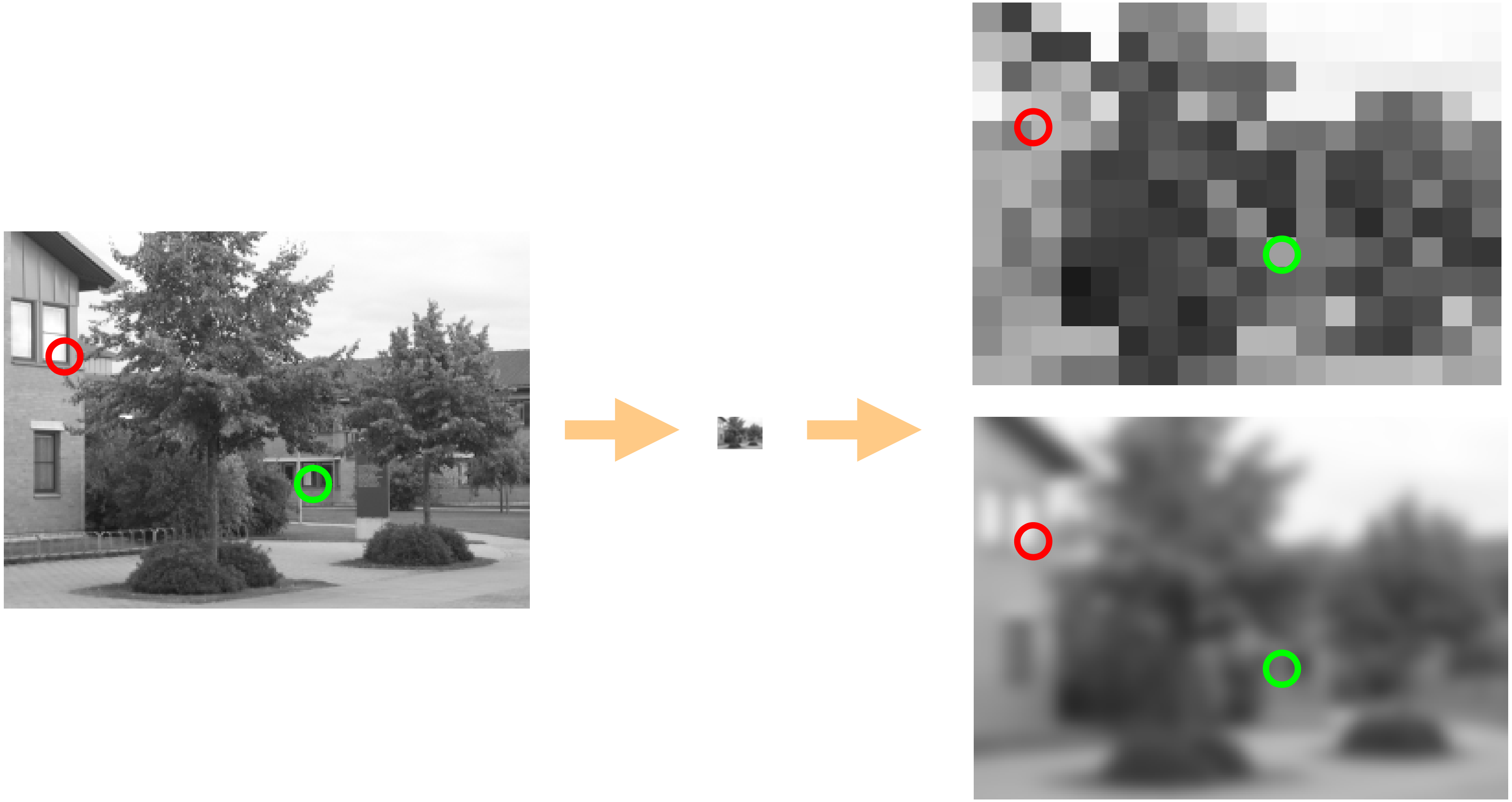
add p to $ProcessedPoints$

stop when $|ProcessedPoints| =$ number of requested detections

// Quadratic in number of points. (There are subquadratic algorithms.)

Sensitivity to change of scale

What is salient at some scale is not at another scale



Robustness / repeatability issues

- Obstacles to detection and matching :
  - change of scale
  - change of orientation (rotation)
  - change of viewpoint (affine, projective transformations)
  - change of illumination
  - noise
  - clutter & occlusion
  - repetitive patterns

What is the expected repeatability of Harris corner ?

☛ Design of robust similarity measures, detector and descriptors/matchers

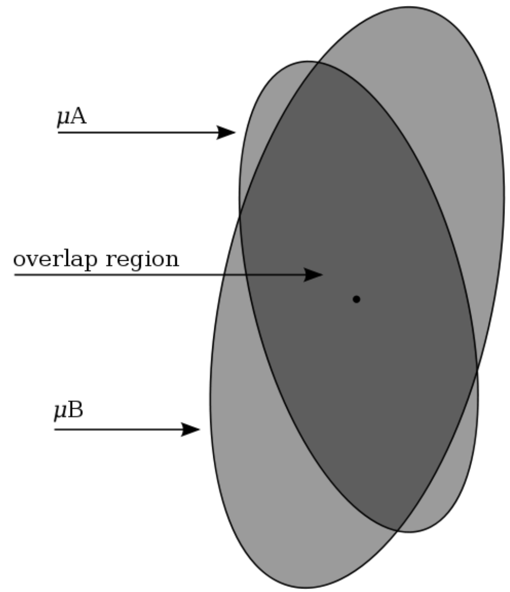
Some repeatability measures

- Setting (Schmid et al. 2000, Mikolajczyk & Schmid 2001, 2002)
  - images of planar scenes
  - known homography and scale transformations
- Location error
  - detected points $\boldsymbol{x}_a$ in I, $\boldsymbol{x}_b$ in I'
  - I and I' related by homography H: $I = H(I')$
  - $\epsilon_{\text{pos}} = \| \boldsymbol{x}_a - H \boldsymbol{x}_b \| < 1.5$ (e.g.) means success
- Scale error
  - scale ratio within given factor, e.g. 1.2, means success

Some repeatability measures (cont.)

- Affinity error
  - $\hat{H}$ local affine approximation of H at point $\boldsymbol{x_b}$
  - μ_A and μ_B elliptical regions defined by $x^T M x \leq 1$ corresponding to Harris correlation matrices A and B
  - Jaccard distance

$$\epsilon_{\text{surf}} = 1 - \frac{\mu_A \cap (\hat{H}^T \mu_B \hat{H})}{\mu_A \cup (\hat{H}^T \mu_B \hat{H})}$$



  - $\epsilon_{\text{surf}} < 0.2$ (e.g.) means success