TELECOM SudParis

Institut Mines-Télécom

High Tech Imaging

IMA 4509 | Visual Content Analysis

Interest point detection

Nicolas ROUGON

ARTEMIS Department

Nicolas.Rougon@telecom-sudparis.eu

Problem statement

- **We hereafter review methods for extracting local geometric features of interest in gray level images, useable in a variety of image matching problems**
  - Image registration
    - image stitching | augmented reality
  - Image retrieval & object recognition
    - image & video indexing
  - 3D scene reconstruction
    - vision-based 3D photogrammetry
  - Tracking
    - simultaneous localization and mapping (SLAM)

Problem statement

- **Requirements**

  Relevant features of interest should be distinctive, and satisfy properties ensuring stable and efficient detection / matching

  - Structural properties
    - ✓ generic
    - ✓ sparse ▶ compactness ▶ computational efficiency
    - ✓ numerous ▶ robustness
    - ✓ uniformly distributed ⇆ occlusions | clutter | cropping

Problem statement

- **Requirements**

  Relevant features of interest should be distinctive, and satisfy properties ensuring stable and efficient detection / matching

  - Invariance properties ► repeatability
    - ✓ contrast transforms
      - ⇆ sensor photometric calibration
      - ⇆ scene lighting

      monotonic luminance transforms
    - ✓ spatial transforms
      - ⇆ sensor geometric calibration
      - ⇆ viewpoint

      isometries | scalings | affine transforms

Problem statement

- **Requirements**

  Relevant features of interest should be distinctive, and satisfy properties ensuring stable and efficient detection / matching

  - Robustness properties ► repeatability ► accuracy
    - ✓ sampling & quantization ⇆ digital image acquisition
    - ✓ noise ⇆ sensor model

Problem statement

- **Candidate features**

Problem statement

- **Candidate features**

  Edges are **not eligible** as features of interest

  - Generic, sparse, uniformly distributed
  - Reasonably invariant to contrast changes
  - Not distinctive
    - ► matching ambiguity occurs along edges
  - Not invariant to spatial transforms

Problem statement

- **Candidate features**

  Corners provide interesting features of interest

  - Distinctive
  - Generic, sparse, numerous, uniformly distributed
  - Invariant to contrast changes
  - Invariant to spatial transforms except scalings

Problem statement

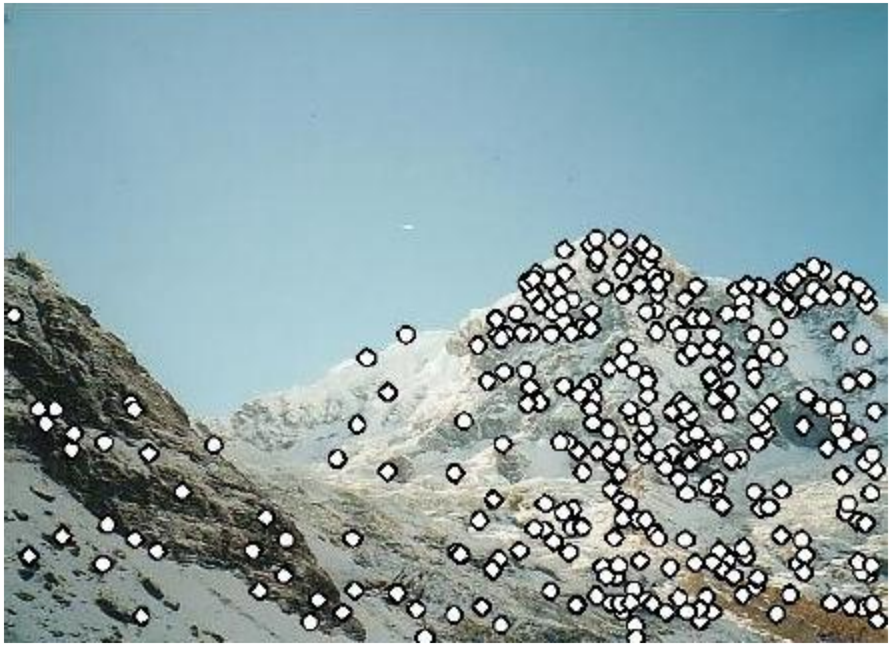
- **Image stitching**

View #1

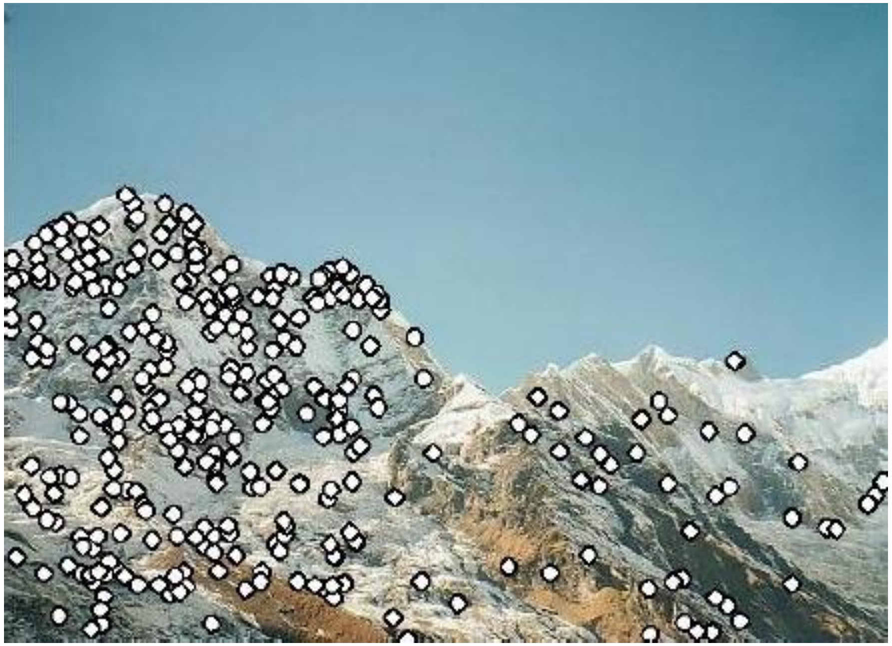
View #2

Problem statement

- **Image stitching**



Corners #1

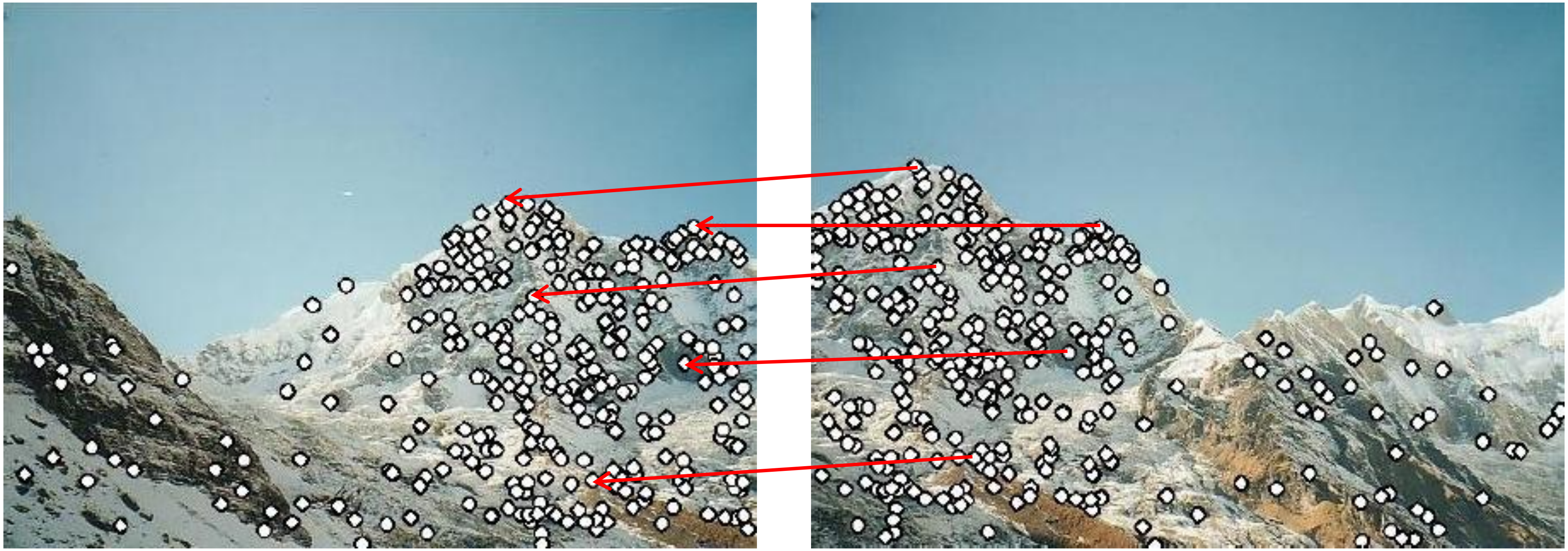


Corners #2

► Corners capture the geometry of textured shapes

Problem statement

- **Image stitching**



Corner matching

Problem statement

- **Image stitching**

View stitching

Problem statement

- **Multi-view 3D scene reconstruction**

Feature extraction
> corner points

▼

Feature matching
> motion vectors

▼

Camera + sparse depth estimation
> 3D point cloud

▼

Surface reconstruction + texturing
> 3D textured mesh

Problem statement

- **Requirements**

  Expected performances of relevant interest point detectors

  - Good detection
    - ✓ high sensitivity / recall
      = all (most) true positives
    - ✓ high specificity / precision
      = few false positives
  - Robustness against noise

  ► repeatability

  - Good localization

  ► accuracy

  - Computational efficiency

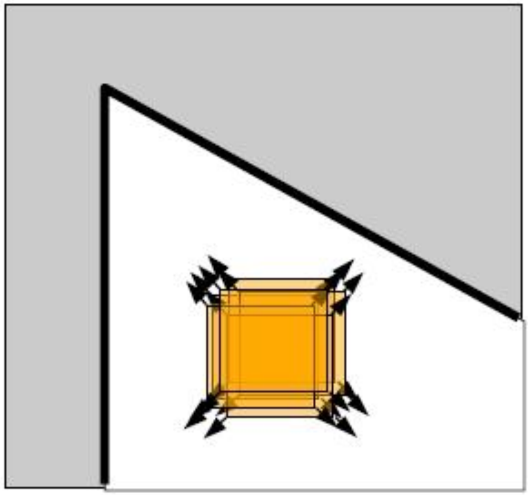
Differential corner detection

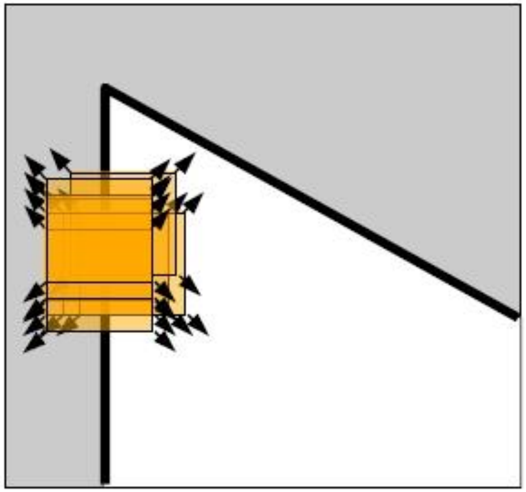
Corner detection

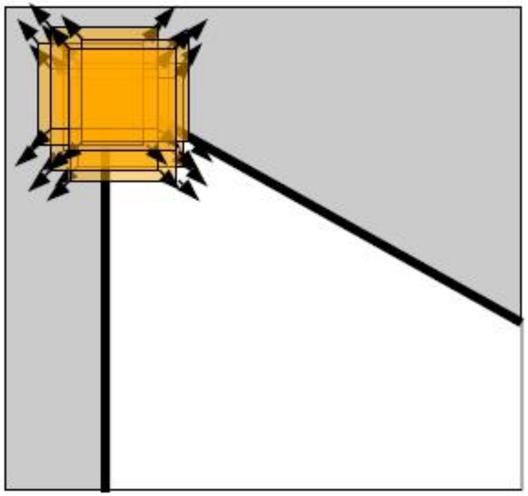
- **Basic idea**

  At corner points, comparing an image patch to its neighbors shows dissimilarity in all directions

- Low-texture region
- Edge
- Corner







- similar in all directions
- similar along edge direction
- dissimilar in all directions

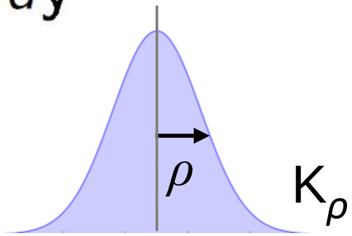
Corner detection

- **Basic idea**

  At corner points, comparing an image patch to its neighbors shows dissimilarity in all directions

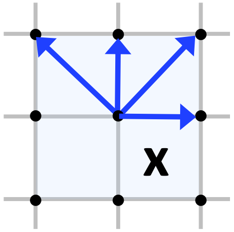
  - This requires defining a similarity metric between image patches, and search for its local maxima over a space of admissible shifts **u**
  - A natural choice is (weighted) autocorrelation

$$E(\mathbf{x}, \mathbf{u}) \;=\; \int \mathrm{K}_\rho(\mathbf{x}-\mathbf{y})\,\big(L(\mathbf{y}+\mathbf{u}) - L(\mathbf{y})\big)^2\, d\mathbf{y}$$



✓ K_ρ : kernel with extension ρ ► patch support

unit | Gaussian



✓ Opting for a unit kernel and 8-connected shifts yields the (early) Moravec detector ► not isotropic

Corner detection

- **Quadratic approximation of the autocorrelation metric**

$$E(\mathbf{x},\mathbf{u}) = \int \mathrm{K}_\rho(\mathbf{x}-\mathbf{y})\left(L(\mathbf{y}+\mathbf{u}) - L(\mathbf{y})\right)^2 dy$$

  - 1st-order Taylor expansion: $L(\mathbf{y}+\mathbf{u}) \approx L(\mathbf{y}) + \nabla L \cdot \mathbf{u}$

$$\begin{aligned} E(\mathbf{x},\mathbf{u}) &\approx \int \mathrm{K}_\rho(\mathbf{x}-\mathbf{y})\left(\nabla L \cdot \mathbf{u}\right)^2 dy \\ &= \mathrm{K}_\rho * \left(\mathbf{u}^T \left[\nabla L \, (\nabla L)^T\right] \mathbf{u}\right) \\ &= \mathbf{u}^T \left(\mathrm{K}_\rho * \left[\nabla L \, (\nabla L)^T\right]\right) \mathbf{u} \end{aligned}$$

► Quadratic approximation

  - The matrix $\mathrm{K}_\rho * \left[\nabla L \, (\nabla L)^T\right]$ is known as the structure tensor

Corner detection

- **Structure tensor**
  - A robust estimate of the structure tensor is obtained using regularized image derivatives ∇L_σ

    Gaussian: $\nabla L_\sigma = \nabla G_\sigma * L$ | Canny-Deriche

$$J(\mathbf{x}, \sigma, \rho) = K_\rho * \left[\nabla L_\sigma (\nabla L_\sigma)^T \right] (\mathbf{x})$$

    ✓ σ = local scale   ✓ ρ = integration scale

  - Expanded form

$$J(\mathbf{x}, \sigma, \rho) = K_\rho * \begin{pmatrix} L_x^2 & L_x L_y \\ L_x L_y & L_y^2 \end{pmatrix} = \begin{pmatrix} K_\rho * (L_x^2) & K_\rho * (L_x L_y) \\ K_\rho * (L_x L_y) & K_\rho * (L_y^2) \end{pmatrix}$$

Corner detection

- **Structure tensor**

  The information in the tensor $J(\mathbf{x}, \sigma, \rho)$ (symmetric, positive definite) is described by its eigenvectors ($\mathbf{d}_{max}$, $\mathbf{d}_{min}$) and eigenvalues (λ_{max}, λ_{min})

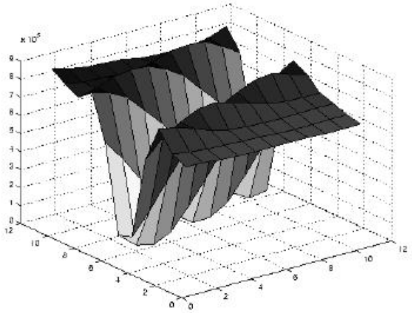
  - Pointwise patch ($\rho = 0$)
    - ✓ $\mathbf{d}_{max} = \frac{\nabla L}{|\nabla L|}$, $\lambda_{max} = |\nabla L|$
    - ✓ $\mathbf{d}_{min} = \frac{\nabla^{\perp} L}{|\nabla L|}$, $\lambda_{min} = 0$

    ► same information in $\nabla L (\nabla L)^T$ and ∇L

  - $J(\mathbf{x}, \sigma, \rho)$ describes average gradient properties over patch support
    - ✓ $\mathbf{d}_{max}$ ($\mathbf{d}_{min}$) : dominant (anti-dominant) orientation
    - ✓ λ_{max}, λ_{min} : directional contrast values

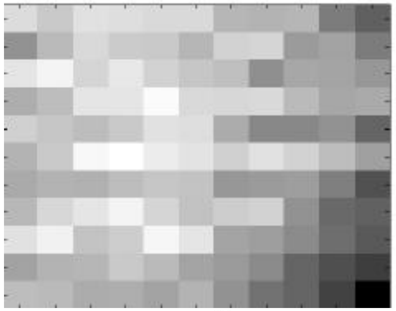
Corner detection

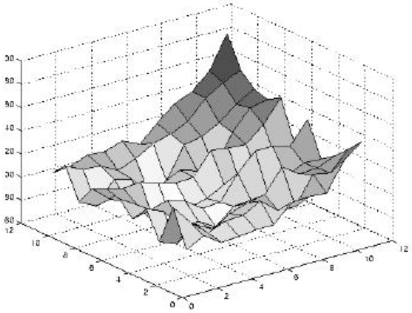
- **Structure tensor**
  - Edge
    - ✓ 1 dominant direction – 1 large directional gradient
    - ✓ λ_{max} large – λ_{min} small



Corner detection

- **Structure tensor**
  - Low-textured region
    - ✓ no dominant direction – no gradient information
    - ✓ λ_{max} small – λ_{min} small

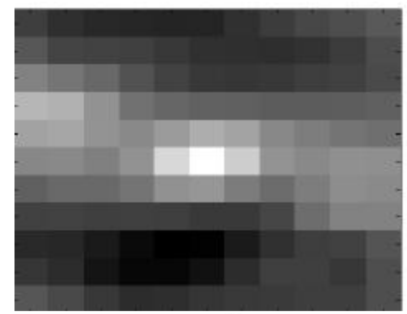


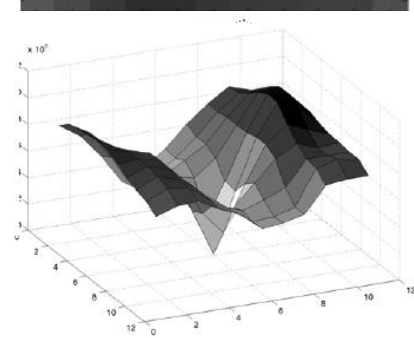


Corner detection

Structure tensor

- High-textured region | Corner
  - ✓ no dominant direction – large directional gradients
  - ✓ λ_{max} large – λ_{min} large





Corner detection

- **Structure tensor**
  - At corner points, the smallest eigenvalue λ_{min} of the structure tensor is large enough
  - This property is exploited by 2 widely-used corner detectors
    - ✓ Kanade-Lucas-Tomasi (KLT)
    - ✓ Harris-Förstner

The KLT detector

Kanade-Lucas-Tomasi (KLT) detector

Algorithm

Given a threshold $\underline{\lambda}$ on λ_{min} and the size D of a square neighborhood

- Compute image gradient ∇L_σ
- Initialize a point list L. For each $\mathbf{x} \in \Omega$
  - compute structure tensor using a unit (D x D) kernel
  - compute smallest eigenvalue λ_{min}
  - if $\lambda_{min} > \underline{\lambda}$, insert $\mathbf{x}$ into L
- Sort L in decreasing order of λ_{min} ► L_{sort}
- Scan L_{sort} from top to bottom
  - for each current point $\mathbf{x} \in \mathrm{L}_{sort}$, discard all points after $\mathbf{x}$ in L_{sort} located in the (D x D) neighborhood of $\mathbf{x}$

Kanade-Lucas-Tomasi (KLT) detector

- **Hyperparameters**
  - The threshold $\underline{\lambda}$ controls the sensitivity of the detector
    - ✓ $\underline{\lambda}$ can be estimated from the histogram of λ_{min} which has usually an obvious valley near 0
    - ✓ however, this valley does not always exist
  - The kernel / neighborhood size *D* is estimated empirically
    - ✓ in most cases: $D \in [2,10]$
    - ✓ large values of *D* induce delocalization artifacts and neighboring corner fusion

The Harris-Förstner detector

Harris-Förstner detector

Principles

- The Harris-Förstner detector makes use of both eigenvalues of the structure tensor via their ratio

$$r = \frac{\lambda_{\text{min}}}{\lambda_{\text{max}}} \in [0, 1]$$

- To avoid computing $(\lambda_{max}, \lambda_{min})$ explicitly, similitude invariants of $\mathsf{J}(\mathbf{x}, \sigma, \rho)$ are used

$$\begin{aligned} \det(\mathsf{J}) &= \lambda_{\text{min}} \lambda_{\text{max}} \\ \text{trace}(\mathsf{J}) &= \lambda_{\text{min}} + \lambda_{\text{max}} \end{aligned}$$

Harris-Förstner detector

Principles

- The latter are combined into a corner index

$$\alpha = \frac{\det(\mathrm{J})}{\mathrm{trace}^2(\mathrm{J})} = \frac{r}{(r+1)^2}$$

  ✓ setting a threshold on r induces a threshold on α

- This yields the following corner metric

$$R(\mathbf{x}) = \det(\mathrm{J}) - \alpha\, \mathrm{trace}^2(\mathrm{J})$$

  - large at corner points
  - small in low-texture regions
  - negative at edge points

  ✓ hyperparameter $\alpha \in [0, 0.25]$

Harris-Förstner detector

Algorithm

Given a value of α and a threshold $\underline{R}$ on R

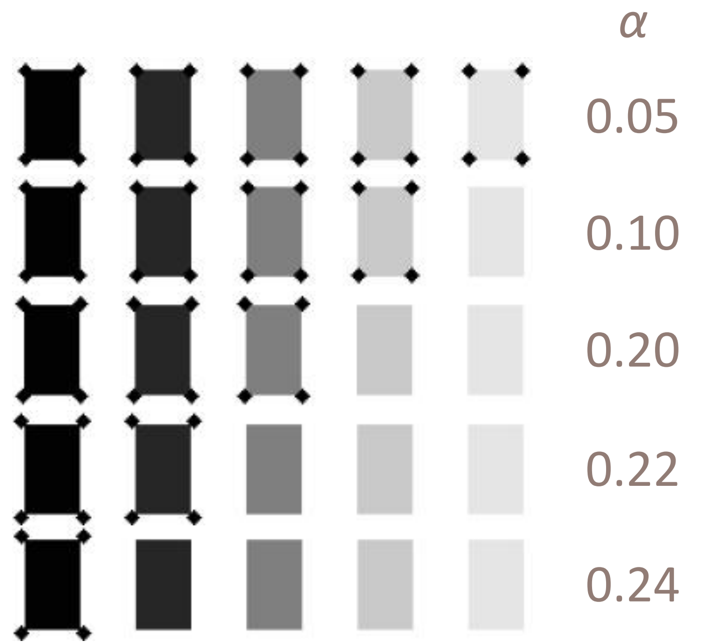
- Compute image gradient ∇L_σ
- For each $\mathbf{x} \in \Omega$
  - compute structure tensor using a Gaussian kernel G_ρ
    Standard choice: $\rho = 2\sigma$
  - compute corner metric $R(\mathbf{x})$
- Threshold corner map R above $\underline{R}$ and retain only local maxima
- Filter out weak* corners in the ρ-neighborhood of strong* corners in a way similar to KLT

  * w.r.t. the corner metric R

Harris-Förstner detector

Hyperparameters

- The parameter α controls the sensitivity of the detector and is tuned empirically
  - ✓ sensitivity ↘ when α ↗
  - ✓ in most cases: $\alpha \in [0.04, 0.06]$

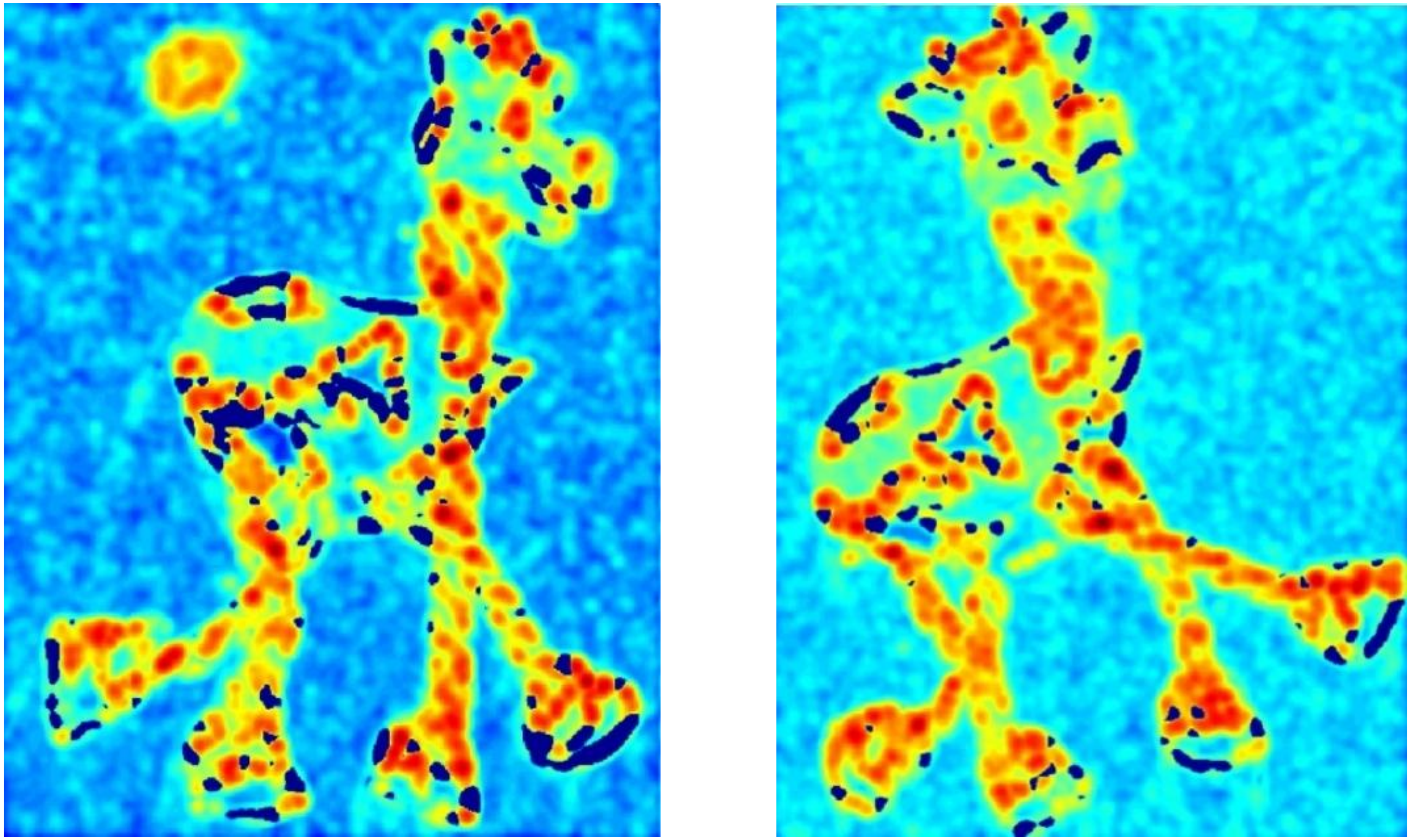


- The threshold $\underline{R}$ is tuned empirically
  - ✓ usually, $\underline{R}$ is set close to 0

Harris-Förstner detector

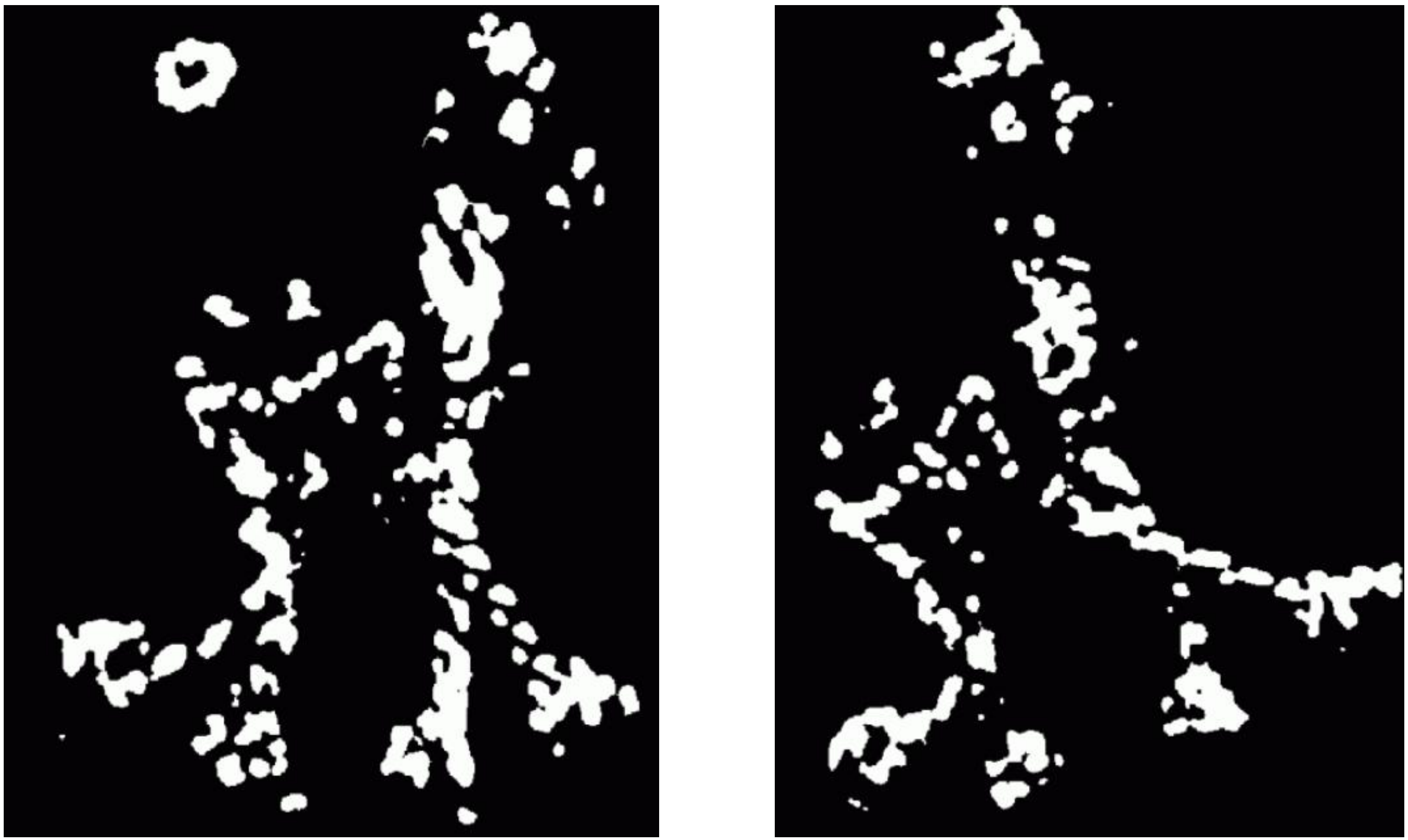
original

Harris-Förstner detector



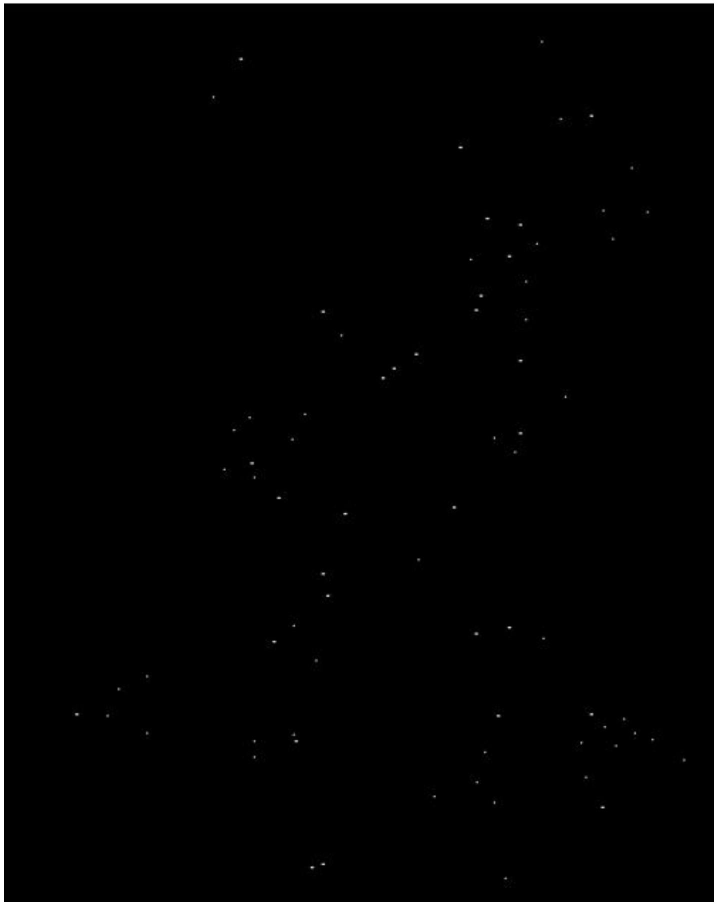
Corner metric *R*

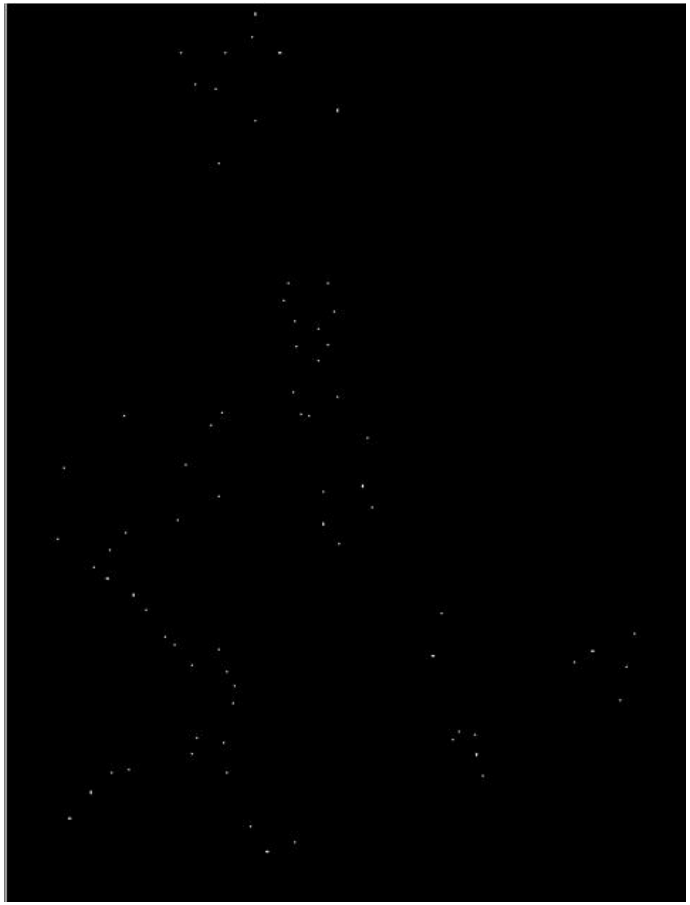
Harris-Förstner detector



Thresholded corner metric $R > \underline{R}$

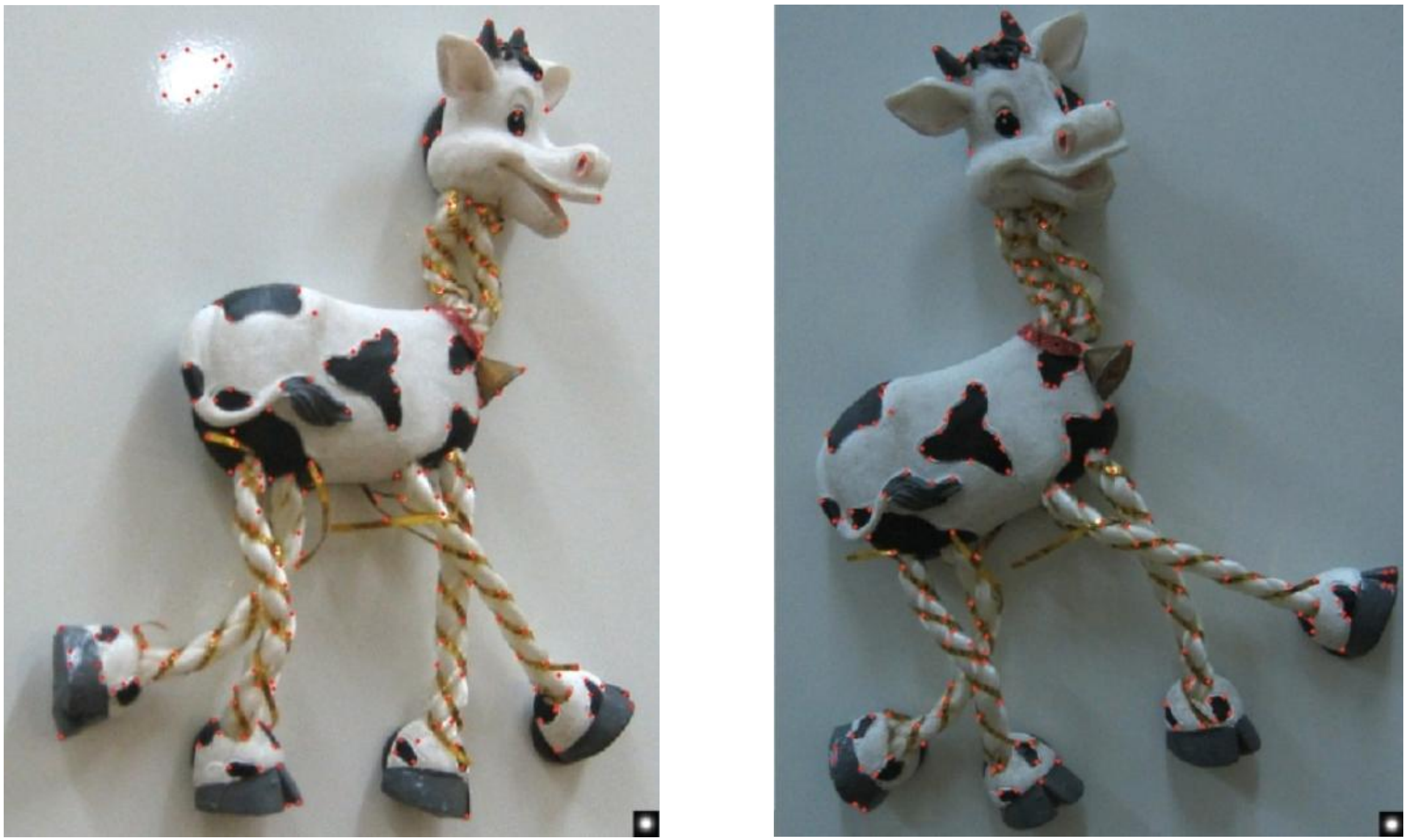
Harris-Förstner detector





Corner metric local maxima

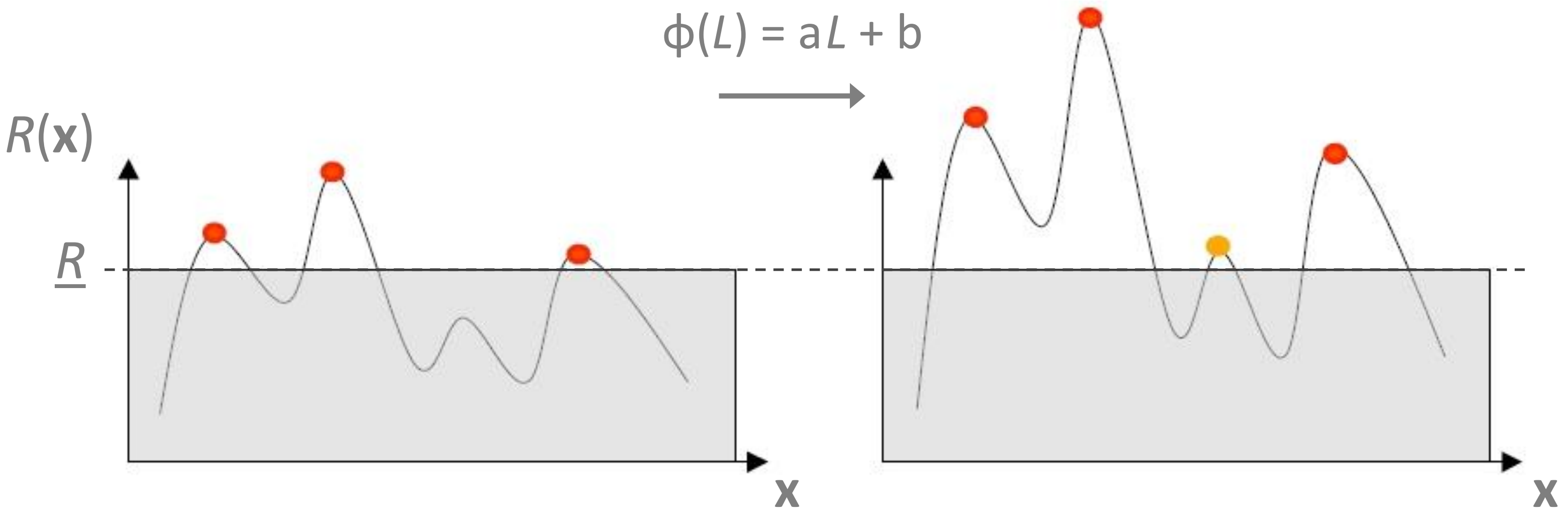
Harris-Förstner detector



Harris points

KLT *vs.* Harris-Förstner detector

- **Properties**
  - Isometry-invariance
    Inherited from eigenvalues properties
  - Insensitive to affine intensity transforms
    Local maxima of λ_{min} / R are preserved



KLT *vs.* Harris-Förstner detector

- **Limitations**
  - Not invariant to **scaling**

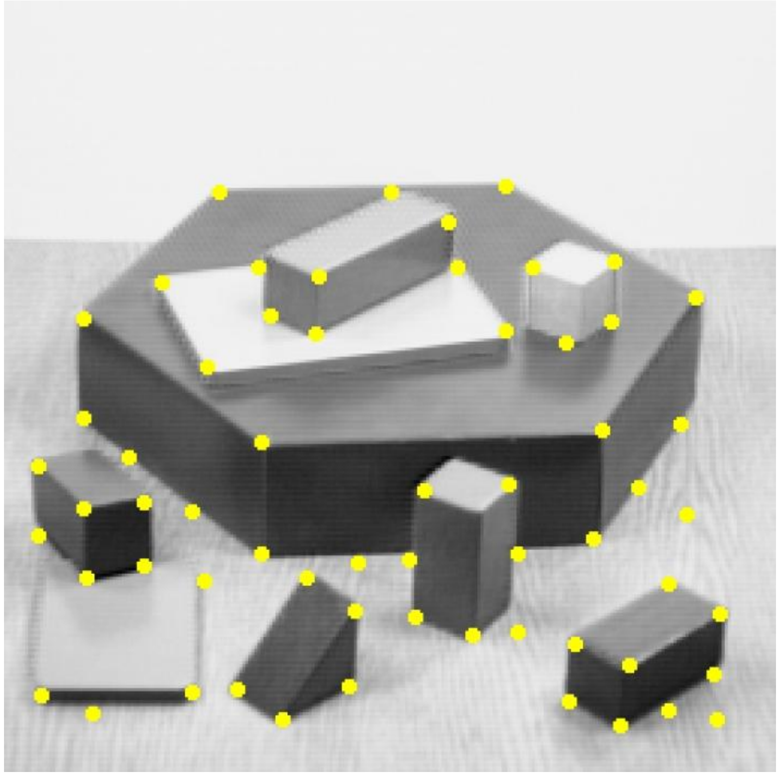
    Using J at fixed scale ρ in both views

▶ multiple points detected as edges

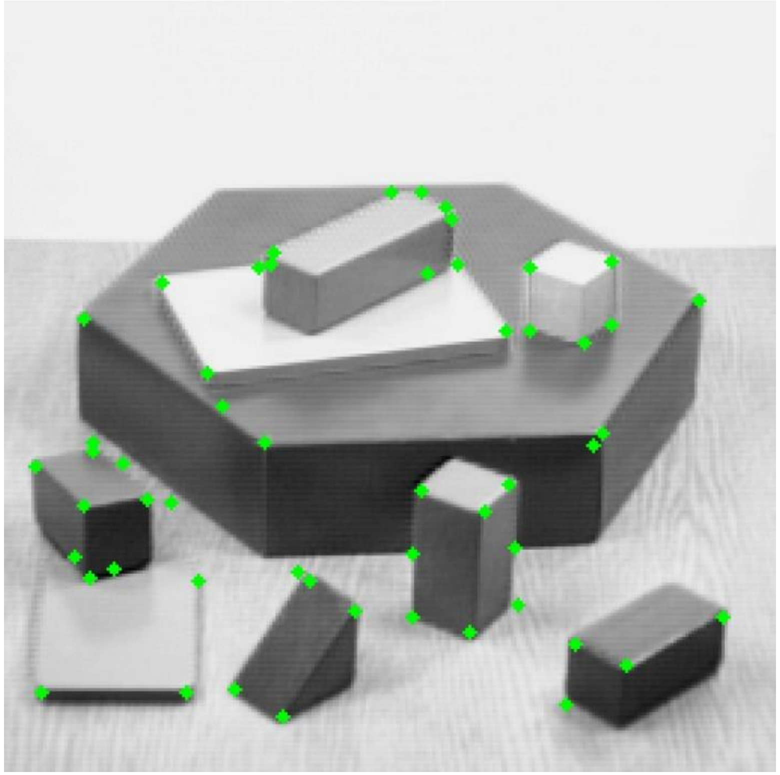
▶ single point detected as a corner

KLT *vs.* Harris-Förstner detector

- **Performance comparison**



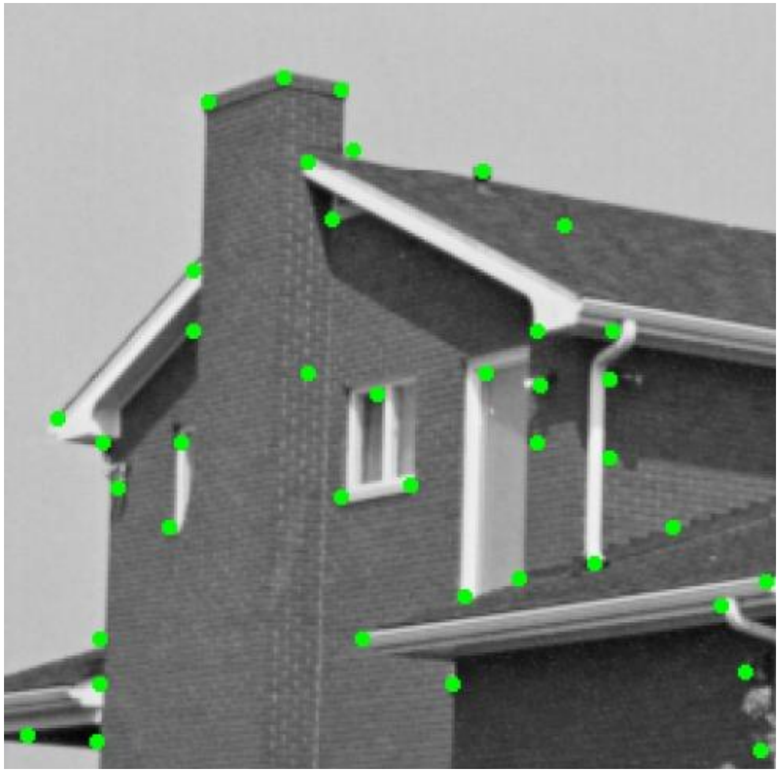
KLT



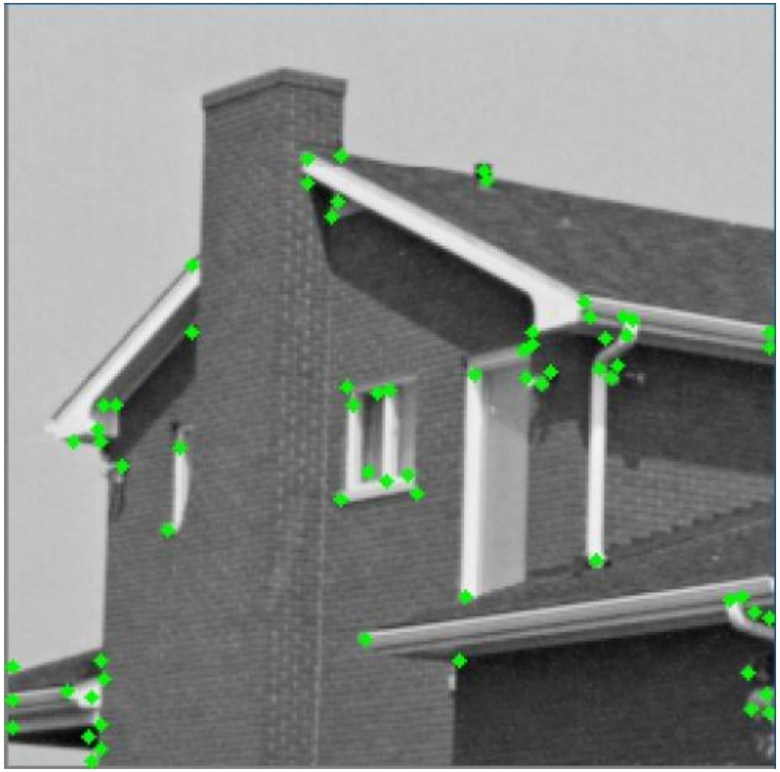
Harris

KLT *vs.* Harris-Förstner detector

- **Performance comparison**



KLT



Harris

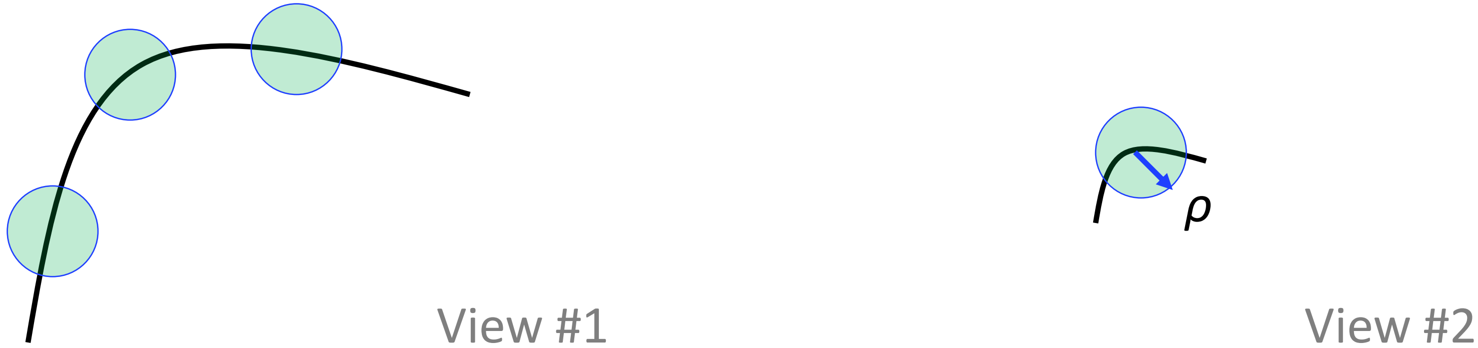
KLT *vs.* Harris-Förstner detector

- **KLT detector**
  - Output is usually closer to human perception of corners
  - Often used for motion tracking
    - ► widespread KLT Tracker
  - Mostly used in the US
- **Harris-Förstner detector**
  - Good repeatability under varying rotation and lighting
  - Often used for 3D scene reconstruction and image retrieval
  - Mostly used in Europe

The Harris-Laplace & SIFT detectors

Scale-invariant interest point detection

- **Key ideas**
  - When zooming occurs, computing J at fixed scale ρ in both views leads to matching ambiguities



▶ Harris / KLT detector repeatability degrades with scale changes

Scale-invariant interest point detection

Key ideas

- This issue is fixed by computing J at view-specific scales $\rho_1(\mathbf{x})$, $\rho_2(\mathbf{x'})$ so that corresponding pixel neighborhoods look the same

- Issue: how to determine these local characteristic scales independently in each view? ► scale-selection mechanism

Scale-invariant interest point detection

- **Automated scale-selection**

  For each view

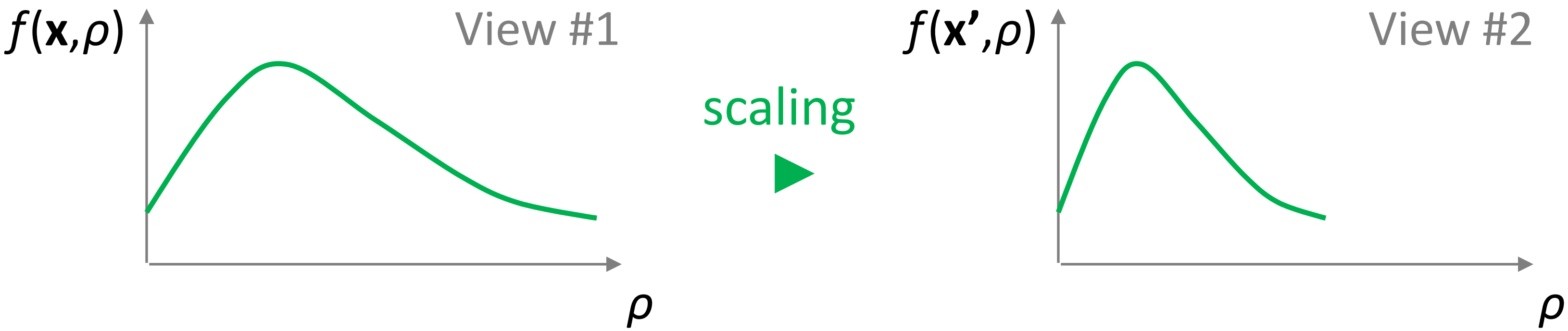
  - Build a multiscale (scale-space) image representation $(L(\mathbf{x},t))_t$

▶ for genericity purpose, the universal Gaussian scale-space is used (i.e. $\rho^2 = 2t$)

Scale-invariant interest point detection

- **Automated scale-selection**
  - Design a local image signature function $f(\mathbf{x},\rho)$ operating on pixel neighborhood at scale ρ
    - ✓ scale-invariance

      at corresponding pixels, $f(\mathbf{x},\rho)$ and $f(\mathbf{x}',\rho)$ have similar shapes

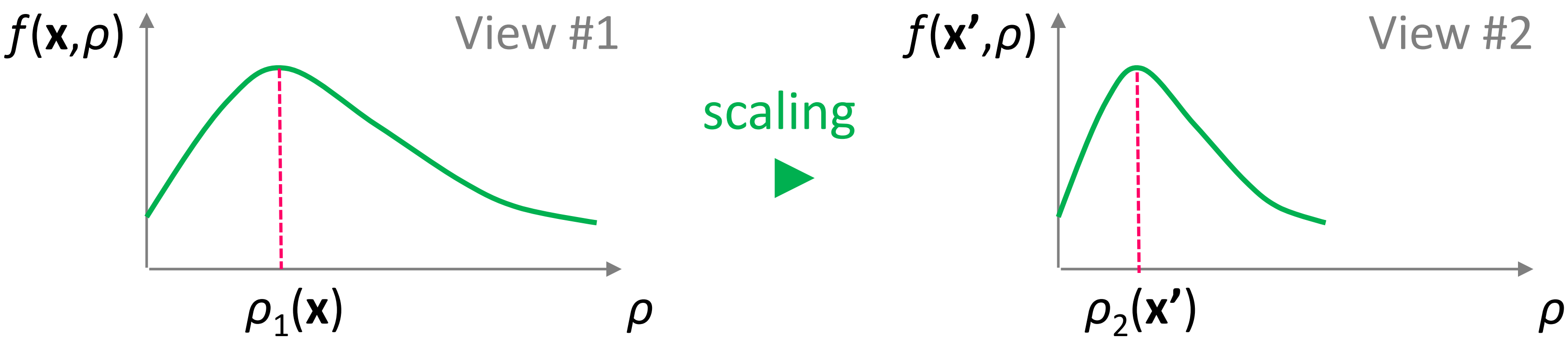


▶ f should involve scale-normalized coordinates $\tilde{\mathbf{x}} = \dfrac{\mathbf{x}}{\rho}$

Scale-invariant interest point detection

Automated scale-selection

- Corresponding neighborhood sizes can be detected by searching for local extremum of $f(\mathbf{x},\rho)$ w.r.t. scale
  - ✓ scale-invariance
    corresponding neighborhood size is invariant to image scalings



▶ The (local) scaling factor between the views is $\frac{\rho_2}{\rho_1}$

Scale-invariant interest point detection

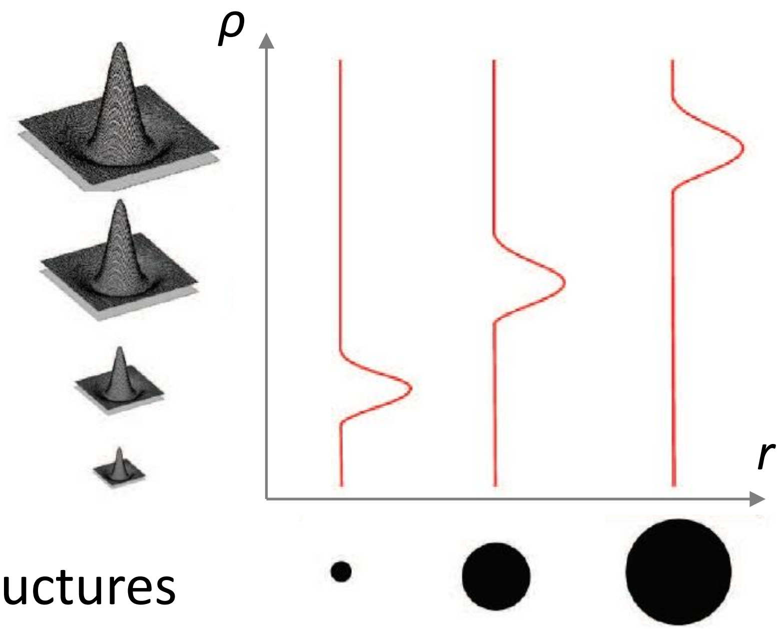
- **Automated scale-selection**
  - Admissibility conditions for local image signature function $f(\mathbf{x},\rho)$
    - ✓ single stable sharp peak

    - ✓ should respond to luminance contrast (= image structure)
      - ► 2^{nd}-order edge kernels

Scale-invariant interest point detection

- **Scale-selection kernels**
  - Scale-normalized Laplacian of Gaussian (LoG)

$$\mathcal{L}(\mathbf{x}, \rho) = \rho^2 \Delta G_\rho * L(\mathbf{x})$$



✓ scale extrema arise at image blobs with radius r equal to filter scale ρ

✓ searching for scale-space extrema allows for detecting scale-invariant circular blob structures

✓ blob centers provide repeatable key points

Scale-invariant interest point detection

- **Scale-selection kernels**
  - Difference of Gaussians (DoG)

$$D(\mathbf{x}, \rho) = \left(G_{k\rho} - G_{\rho}\right) * L(\mathbf{x})$$

    - ✓ accurately approximates the LoG kernel for constant k
    - ✓ incorporates scale normalization [Lowe 2004]

Scale-Invariant Feature Transform (SIFT)

Principles

- Interest points are searched as scale-space extrema of the DoG function, *i.e.* simultaneous extrema in the image plane and along the scale axis
  - ✓ detection is performed by comparing the DoG function values in the 26-connected scale-space neighborhood of $(\mathbf{x},\rho)$

Scale-Invariant Feature Transform (SIFT)

To probe further

D. Lowe | *Distinctive image features from scale-invariant keypoints*
International Journal of Computer Vision, 60(2):91-110, November 2004

Harris-Laplace detector

- **Principles**

  Combine Harris-Förstner detector specificity for corner-like structures with LoG-based scale selection mechanism

  - The scale-normalized structure tensor $\sigma^2 \mathsf{J}(\mathbf{x}, \sigma, \rho)$ is used to define a scale-invariant Harris corner metric $R(\mathbf{x}, \sigma, \rho)$
  - Scale-spaces are built for the Harris metric $R(\mathbf{x}, s\rho, \rho)$ (s = 0.7) and the Laplacian-of-Gaussian $\mathcal{L}(\mathbf{x}, \rho)$
    - ✓ at each scale ρ, candidate points are detected as local maxima of the Harris metric
    - ✓ candidate points for which the LoG simultaneously attains an extremum over scale are retained

Harris-Laplace detector

To probe further

K. Mikolajczyk, C. Schmid | *Indexing based on scale-invariant interest points*
8th IEEE International Conference on Computer Vision (ICCV'2001), Vancouver, Canada
Vol. 1, 525-531, July 2001

K. Mikolajczyk, C. Schmid | *Scale & affine invariant interest point detectors*
International Journal of Computer Vision, 60(1):63-86, October 2004

Harris-Laplace detector

Performances

- Harris-Laplace interest points are highly-discriminative
  ► increased repeatability
- Returns a much smaller number of points than the LoG / DoG detectors
  ► reduced robustness to partial occlusions
  problematic for object recognition applications (in particular, object categorization)
- A modified version of the Harris-Laplacian detector using a less strict criterion has been proposed
  - ✓ selects scale extrema of the LoG at locations for which the Harris corner metric also attains a maximum at any scale

Temporary conclusion

- **Topics to be further addressed**
  - Detection
    - ✓ 2nd-order detectors
      - ▶ Hessian detector
      - ▶ Hessian-Laplace detector
  - Description

    Define / extract feature vector descriptors around interest points
  - Matching

    Estimate correspondence between descriptors in each view

▶ To be continued in the HTI / Multimedia indexing 3rd-year course

Institut Mines-Télécom

High Tech Imaging

IMA 4509 | Visual Content Analysis

Interest point detection

Nicolas ROUGON

ARTEMIS Department

Nicolas.Rougon@telecom-sudparis.eu