

SCALE-INVARIANT FEATURE TRANSFORM

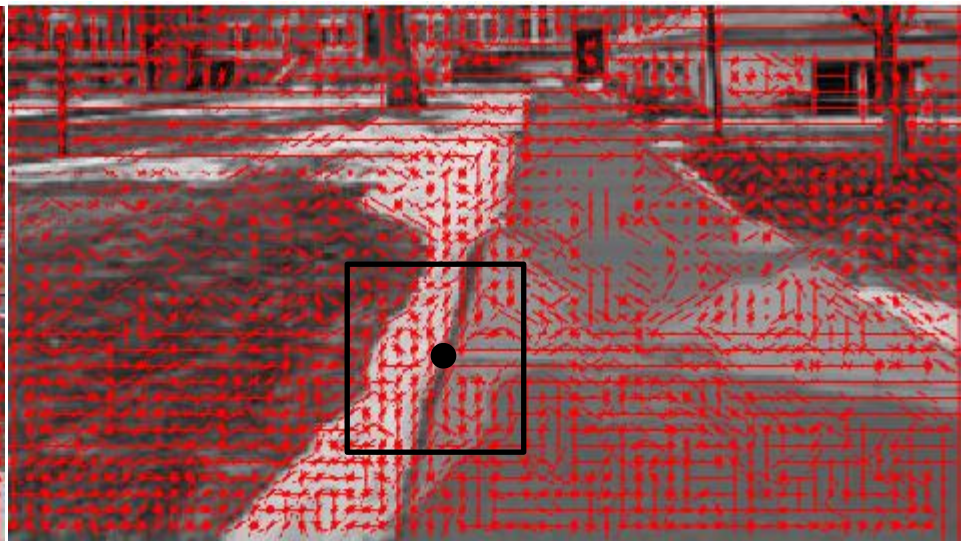
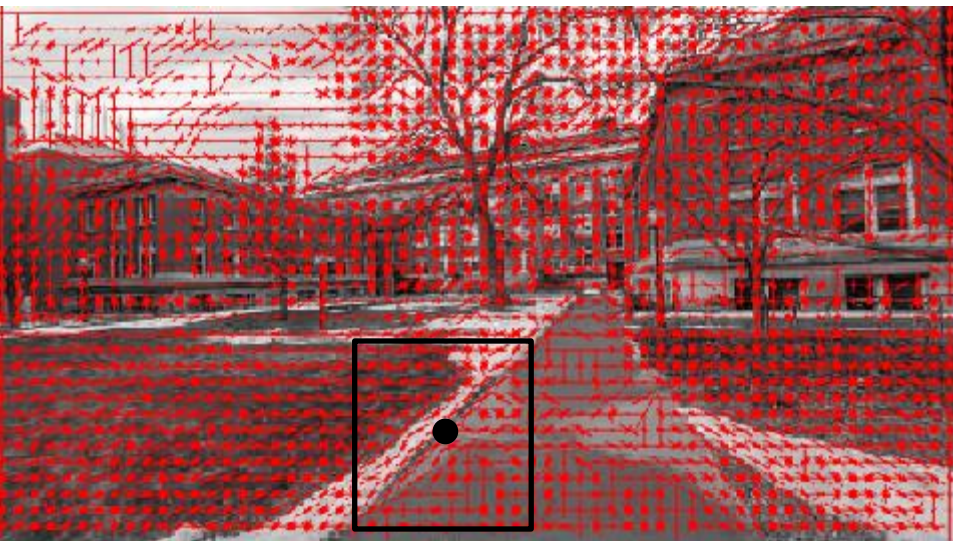
The image shows a brick building with a path and trees. Overlaid on the image is a grid of green squares, each containing a red dot. The grid is composed of many small squares and a few larger ones, illustrating the concept of scale-invariant feature transform. The text 'SCALE-INVARIANT FEATURE TRANSFORM' is written in a large, white, italicized font across the middle of the image. The name 'HYUN SOO PARK' is written in a smaller, orange font in the bottom left corner.

HYUN SOO PARK

FEATURE MATCHING



RECALL: HOG



$$\text{corr} \left[\begin{array}{c} \text{[Red HOG feature patch]} \\ \text{[Red HOG feature patch]} \end{array} \right] = 0.15 \quad \text{What's wrong?}$$

CHALLENGES OF FEATURE MATCHING

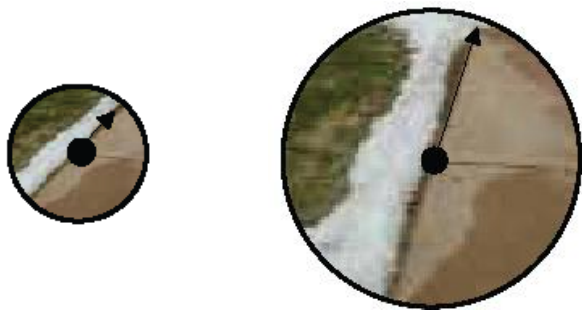


- Scale
- Orientation

FEATURE NORMALIZATION



FEATURE NORMALIZATION



FEATURE NORMALIZATION



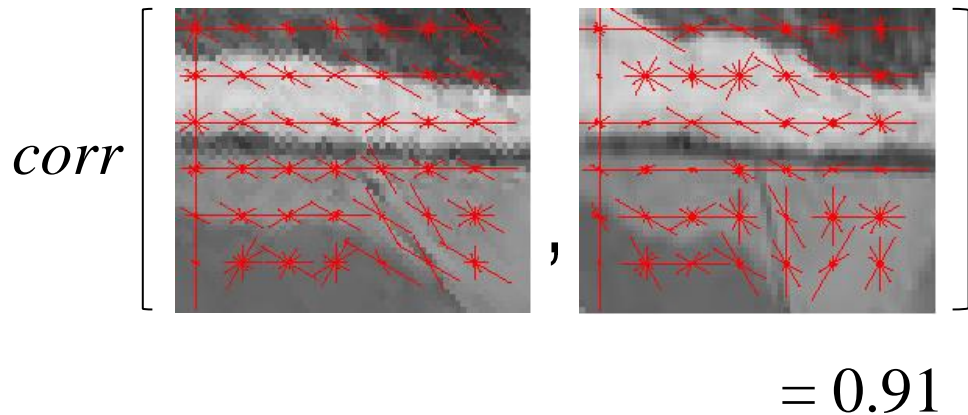
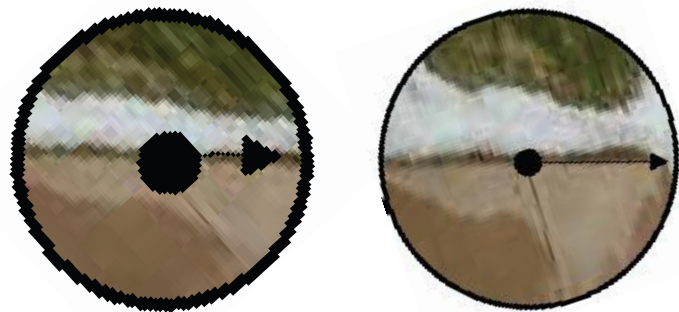
Scale normalization

FEATURE NORMALIZATION

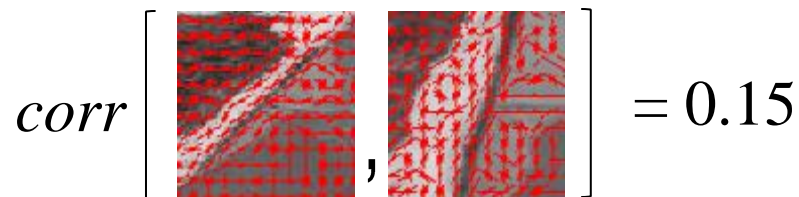


Orientation normalization

SCALE-INVARIANT FEATURE TRANSFORM (SIFT)



cf)



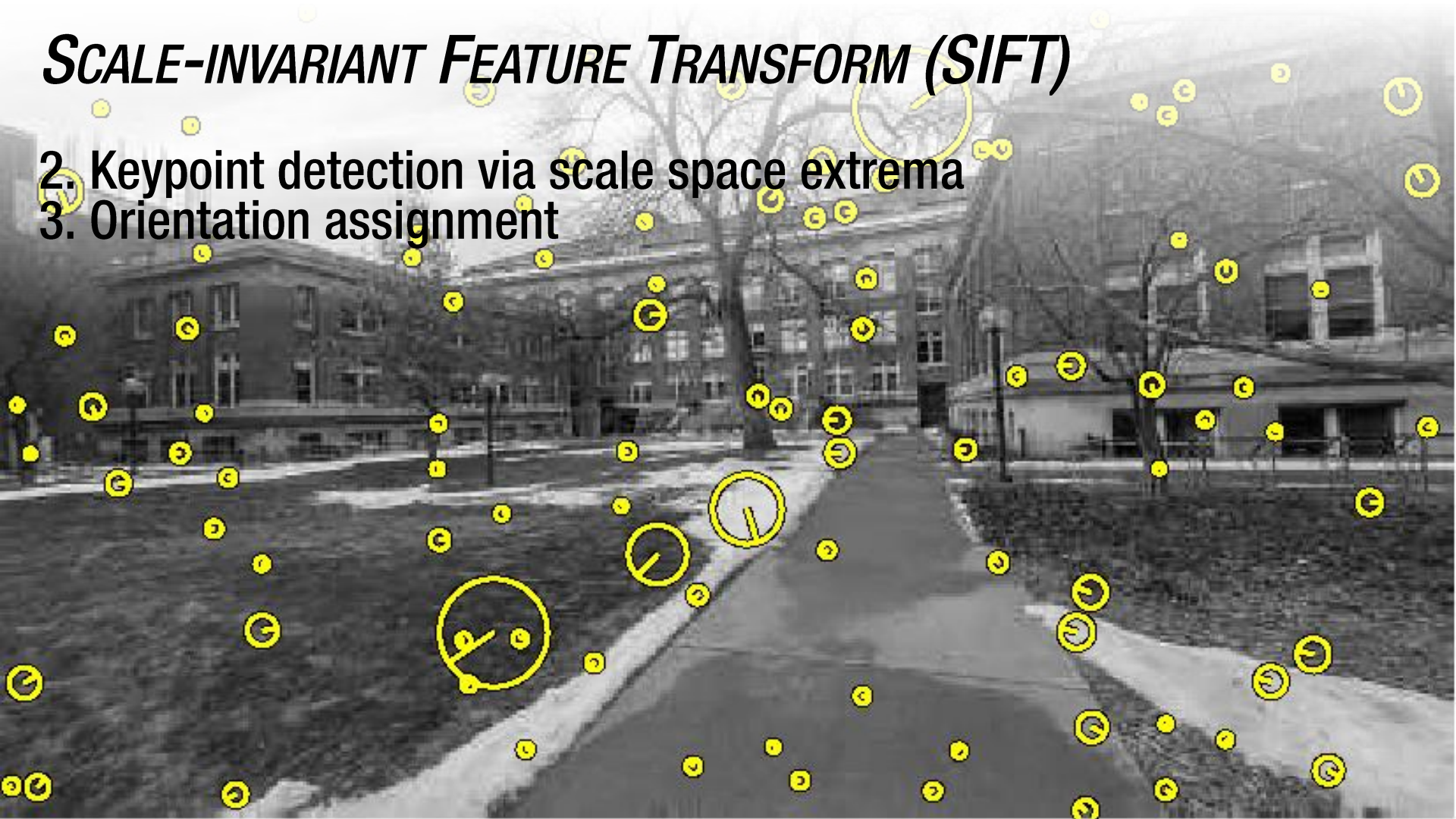
SCALE-INVARIANT FEATURE TRANSFORM (SIFT)

1. Input image



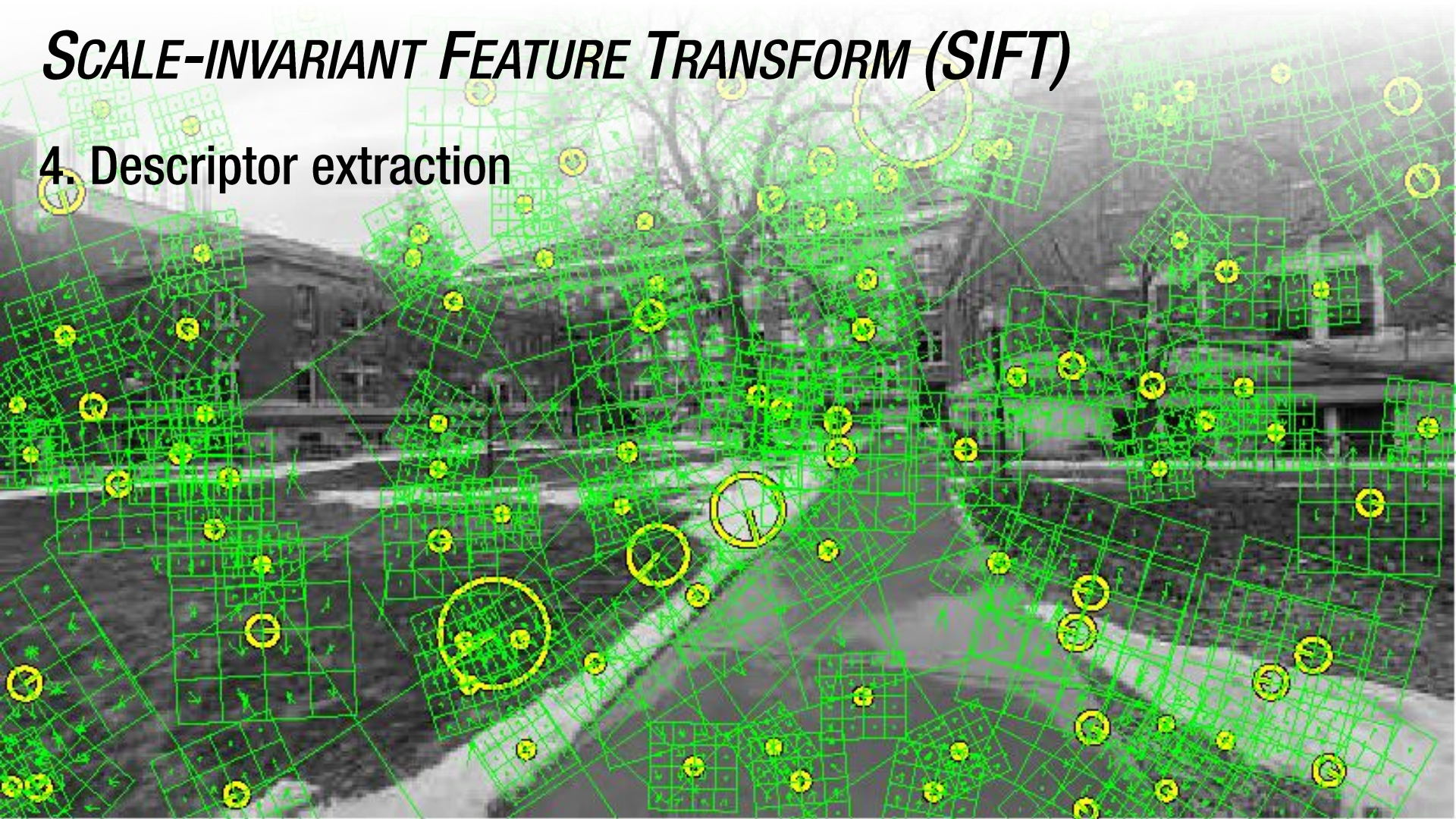
SCALE-INVARIANT FEATURE TRANSFORM (SIFT)

2. Keypoint detection via scale space extrema
3. Orientation assignment

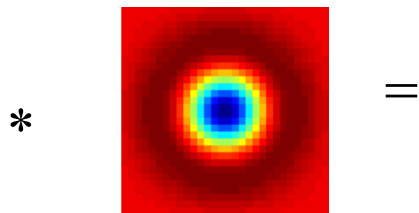


SCALE-INVARIANT FEATURE TRANSFORM (SIFT)

4. Descriptor extraction

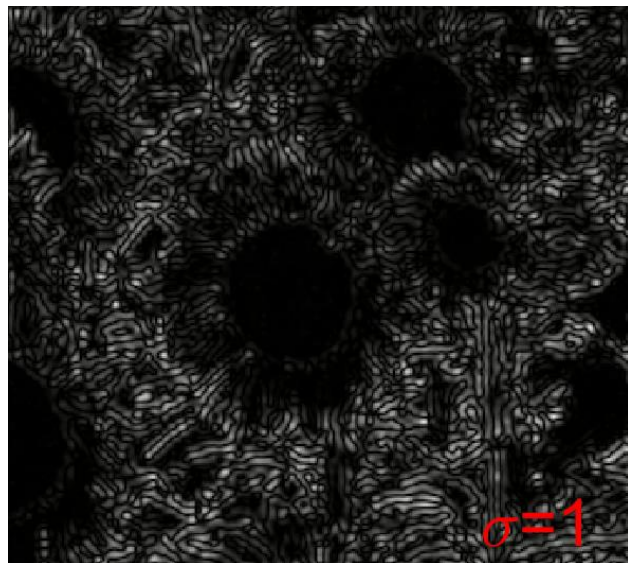


KEYPOINT DETECTION: SCALE SPACE EXTREMA



$$L = \sigma^2 \nabla^2 G$$

Laplacian of Gaussian (LoG)

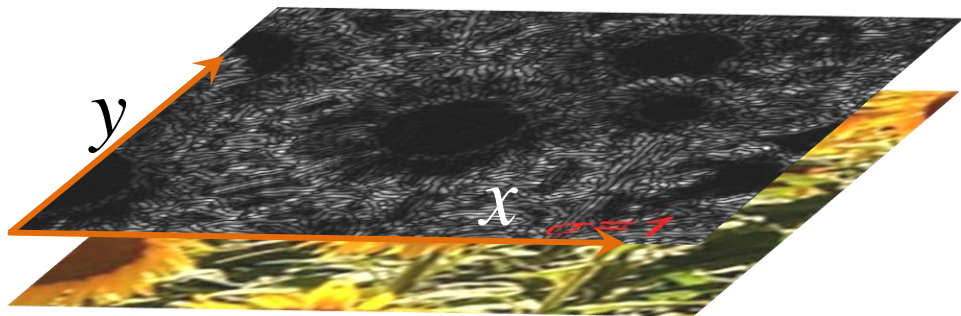


KEYPOINT DETECTION: SCALE SPACE EXTREMA



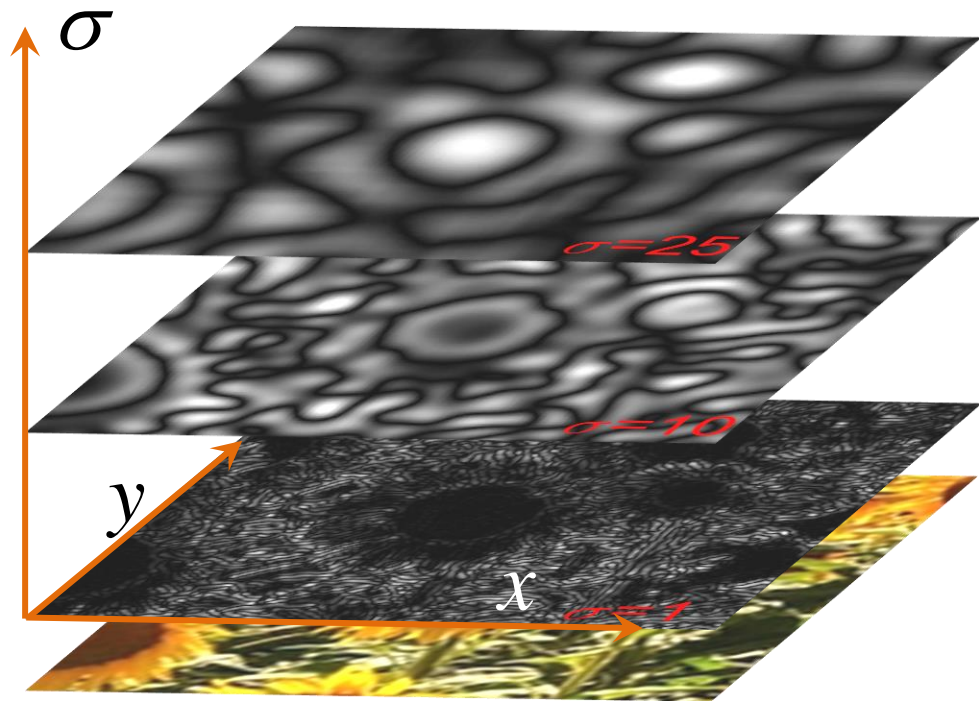
$I(x, y)$

KEYPOINT DETECTION: SCALE SPACE EXTREMA



$$I(x, y) * L(\sigma)$$

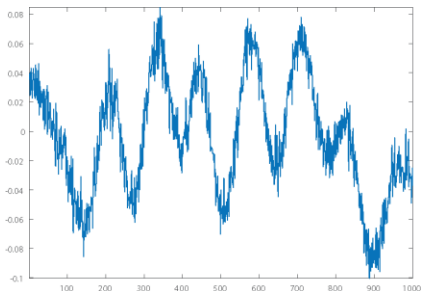
KEYPOINT DETECTION: SCALE SPACE EXTREMA



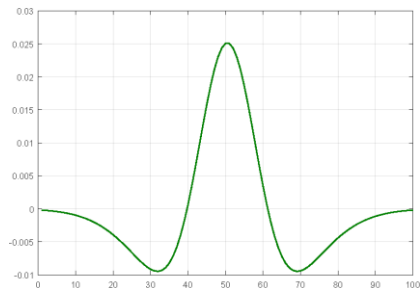
$$D(x, y, \sigma) = I(x, y) * L(\sigma)$$

Scale space response

RECALL: LAPLACIAN OF GAUSSIAN (LoG) ~ DoG

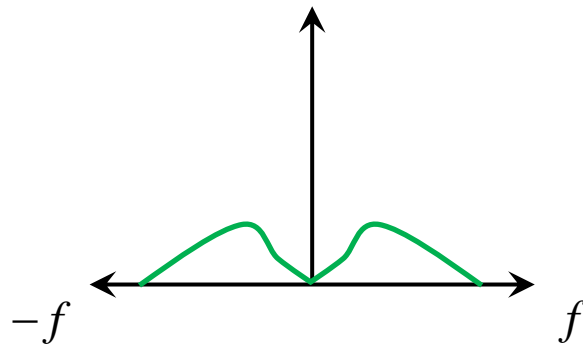


*



FT
→

←
Inverse FT



$x(t)$

*

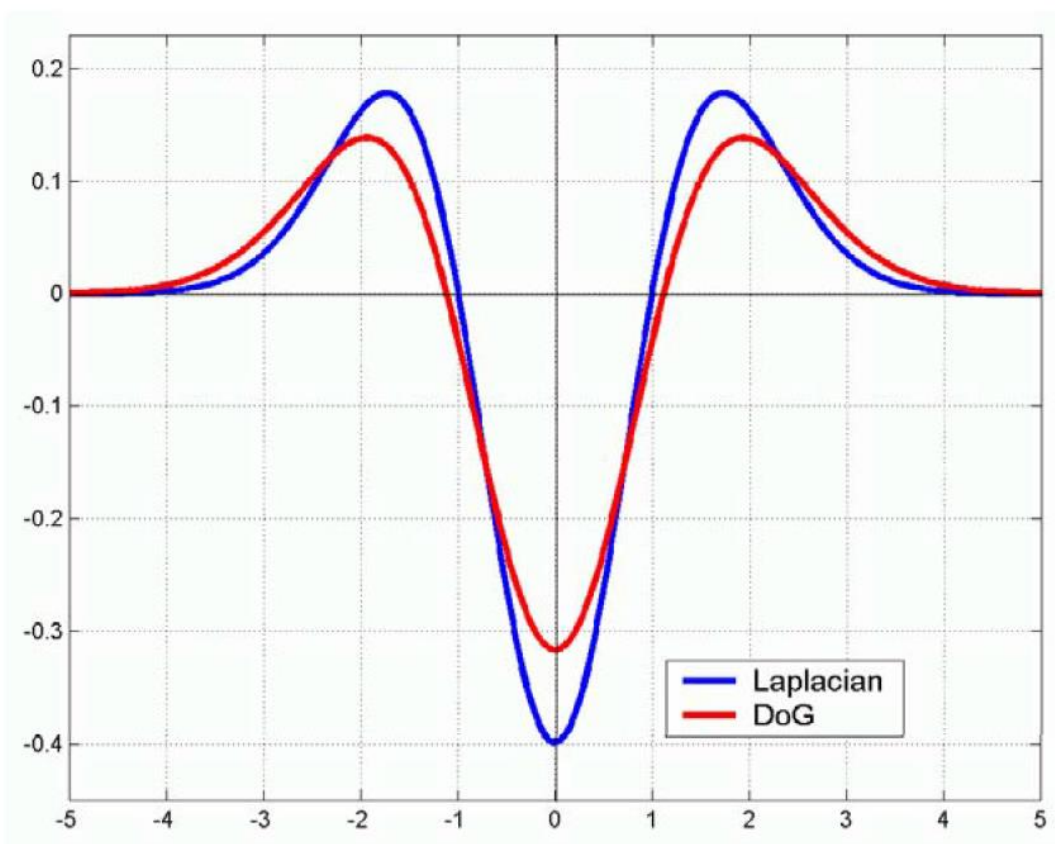
$$g(t; \sigma_1) - g(t; \sigma_2)$$

$$\approx \nabla \cdot \nabla g$$

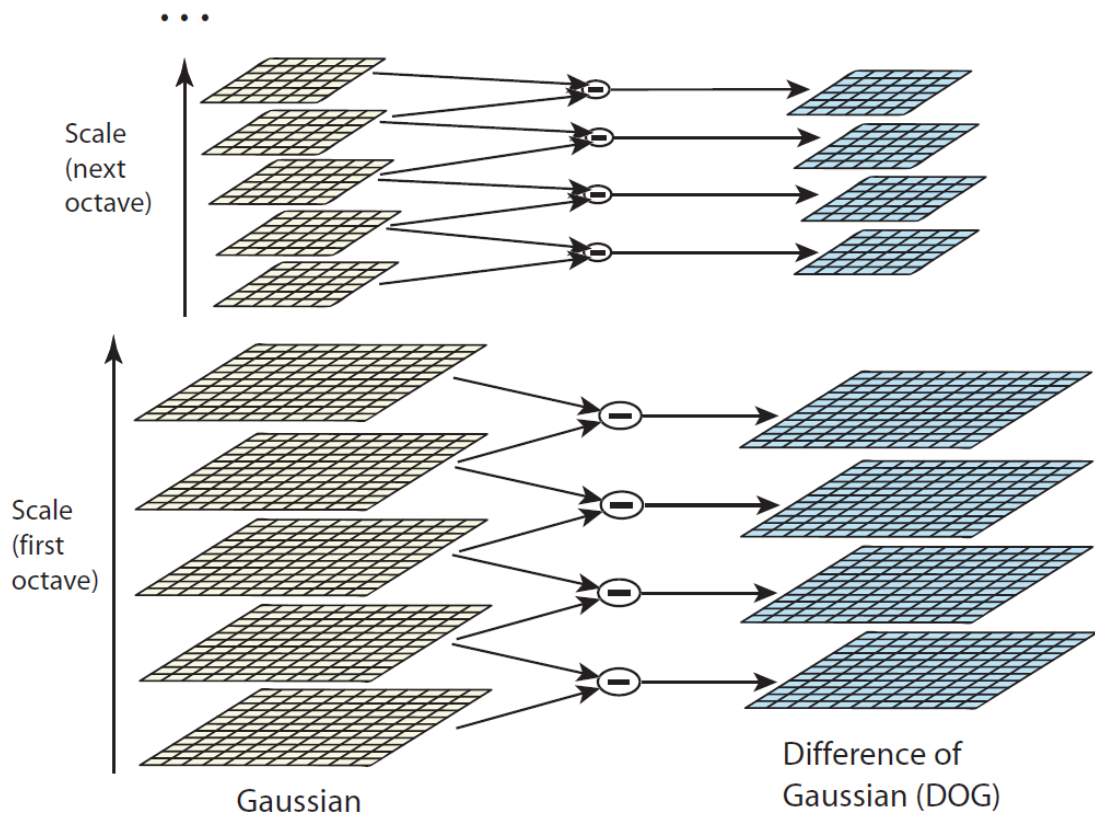
Laplacian of Gaussian

$$X(f)(G(f; \sigma_1) - G(f; \sigma_2))$$

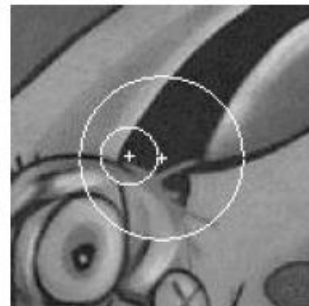
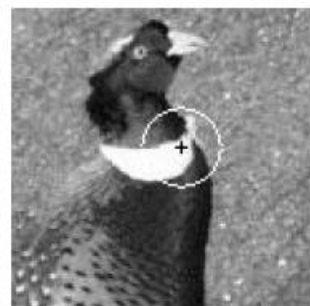
RECALL: LAPLACIAN OF GAUSSIAN (LoG) ~ DoG



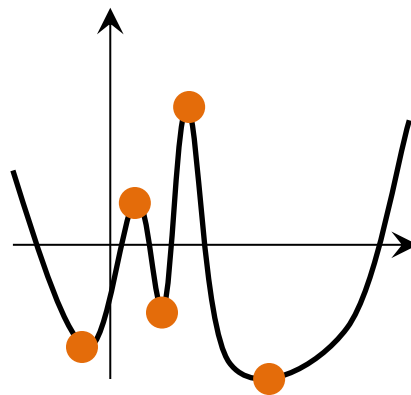
LAPLACIAN ~ DIFFERENCE OF GAUSSIAN



WHAT CHARACTERIZES KEYPOINT?

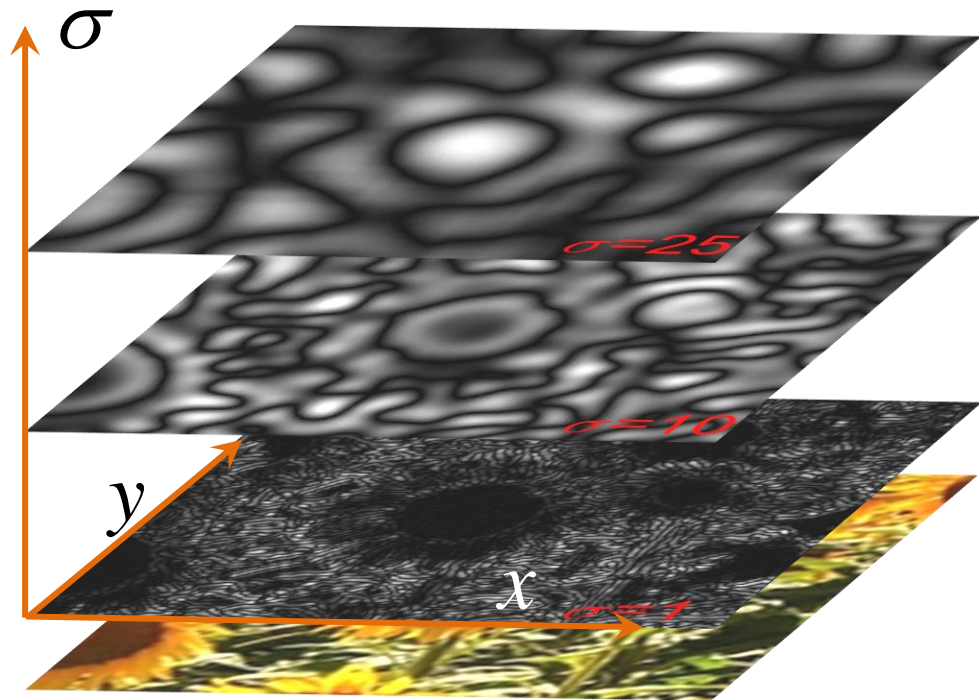


- Robust
- Repeatable
- Uniqueness



Local extrema (minima/maxma)

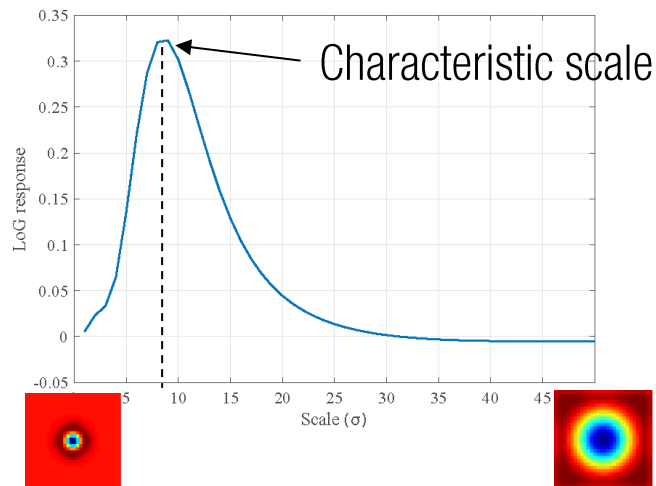
WHAT CHARACTERIZES KEYPOINT?



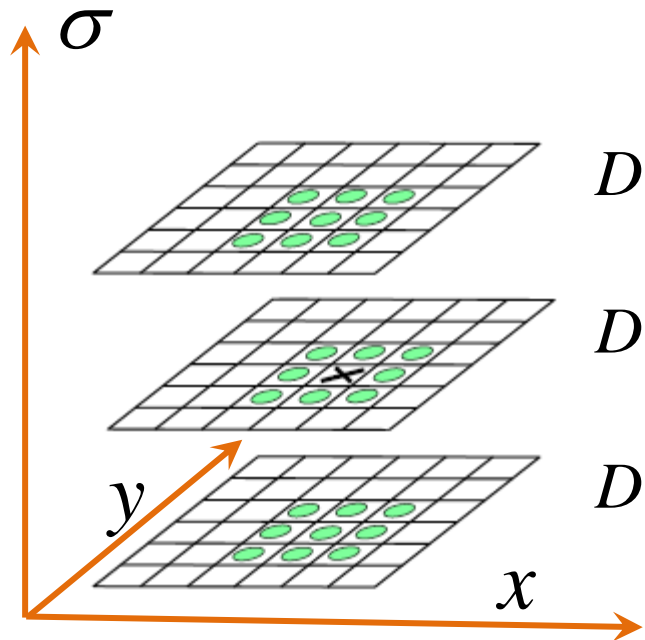
$$D(x, y, \sigma) = I(x, y) * L(\sigma)$$

Scale space response

Local extrema (minima/maxima)



WHAT CHARACTERIZES KEYPOINT?



$$D(x, y, \sigma + \Delta\sigma)$$

$$D(x, y, \sigma)$$

$$D(x, y, \sigma - \Delta\sigma)$$

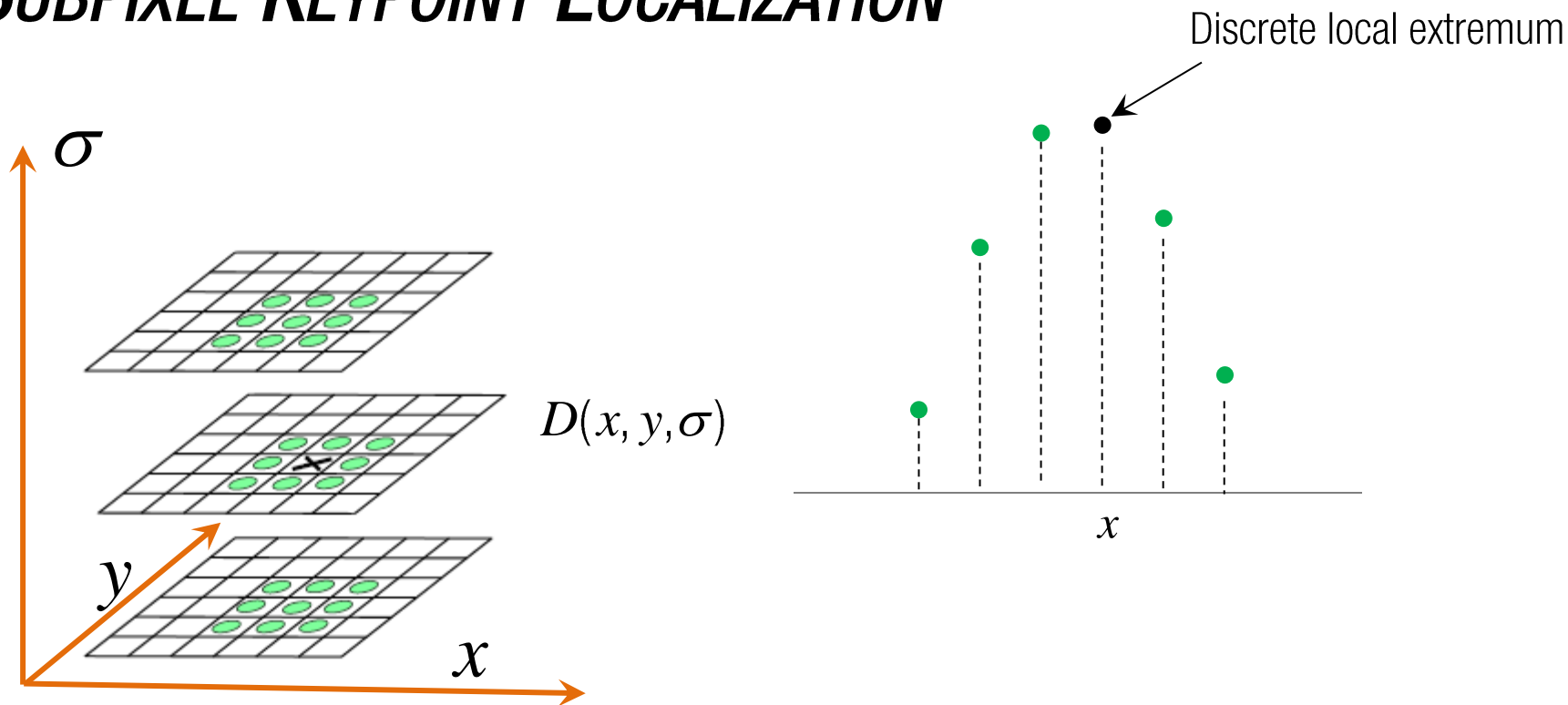
(x, y) is keypoint if

$$D(x, y, \sigma) > D(x \pm \Delta x, y \pm \Delta y, \sigma \pm \Delta\sigma)$$

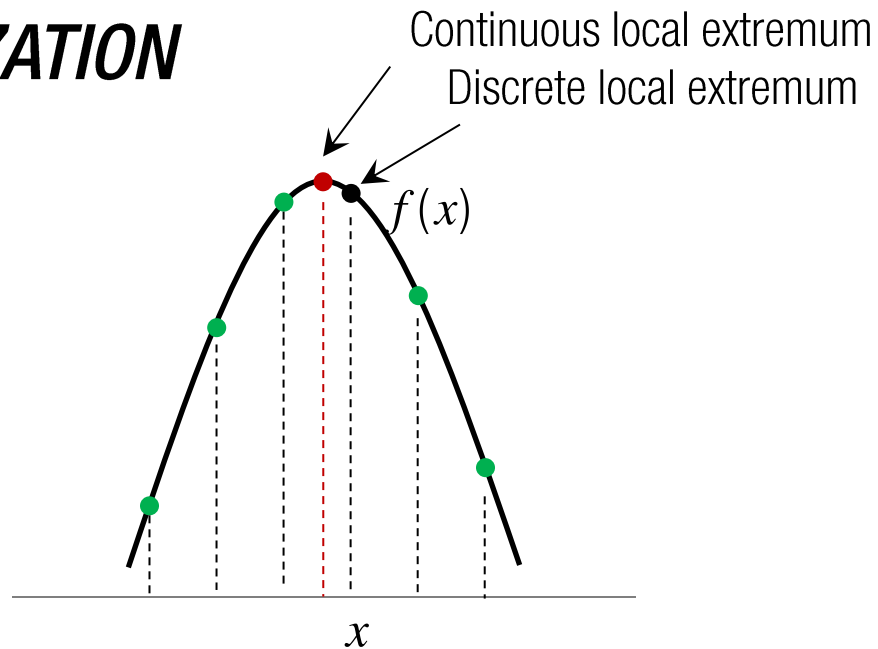
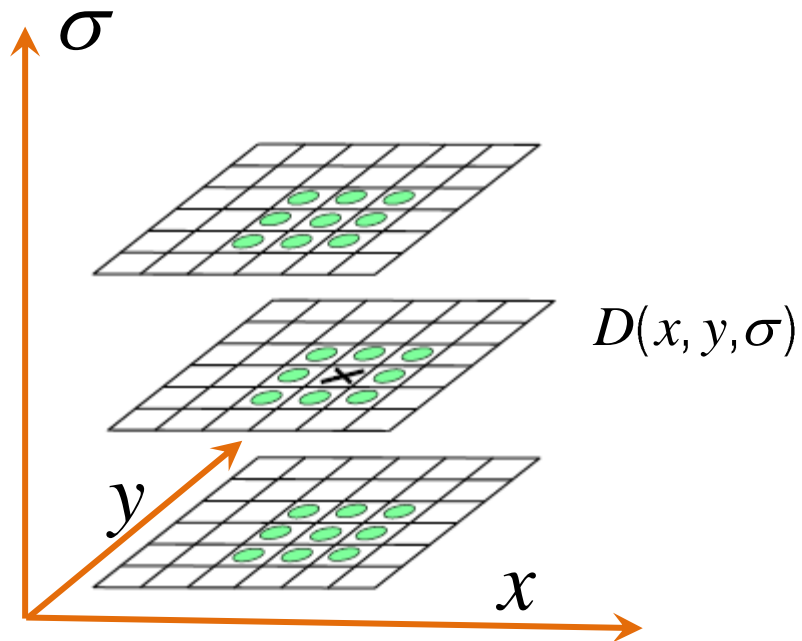
or

$$D(x, y, \sigma) < D(x \pm \Delta x, y \pm \Delta y, \sigma \pm \Delta\sigma)$$

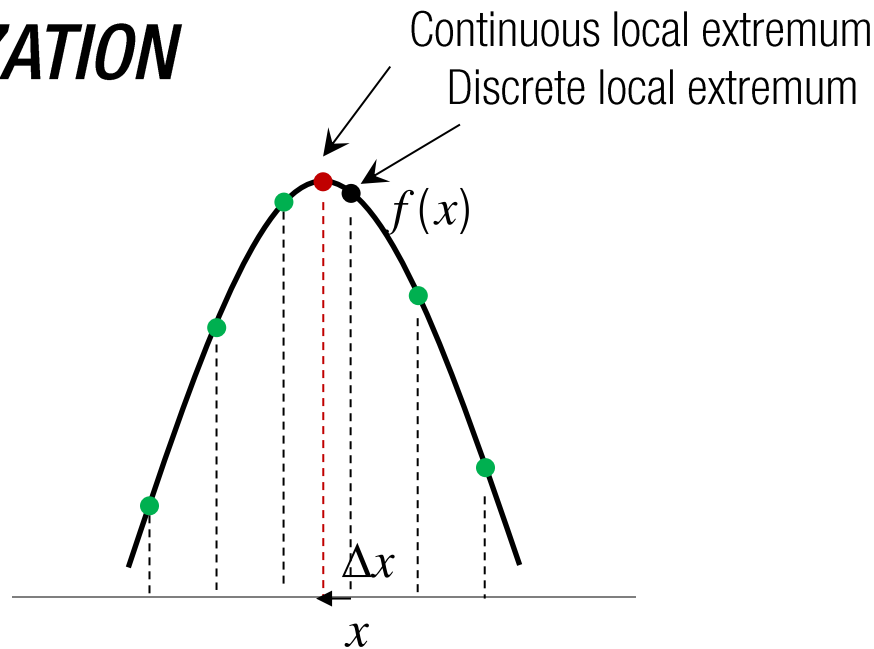
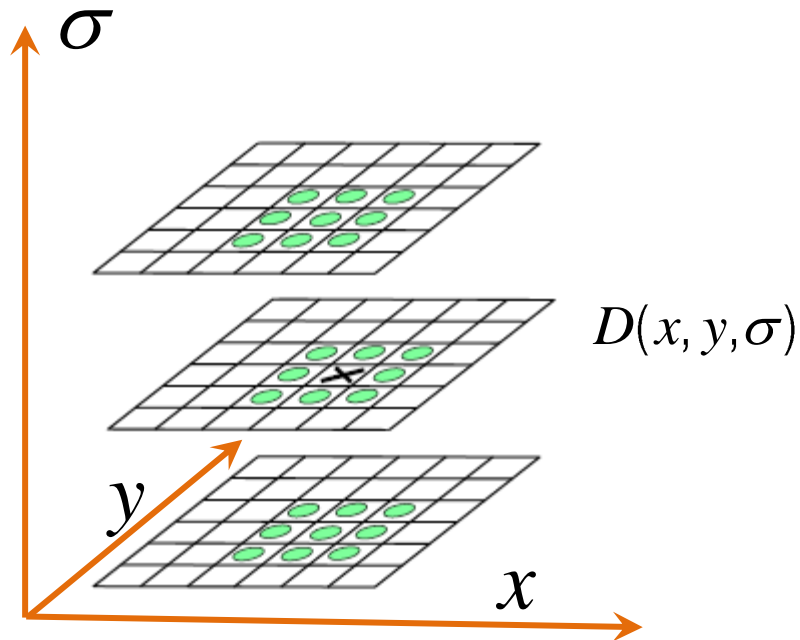
SUBPIXEL KEYPOINT LOCALIZATION



SUBPIXEL KEYPOINT LOCALIZATION



SUBPIXEL KEYPOINT LOCALIZATION

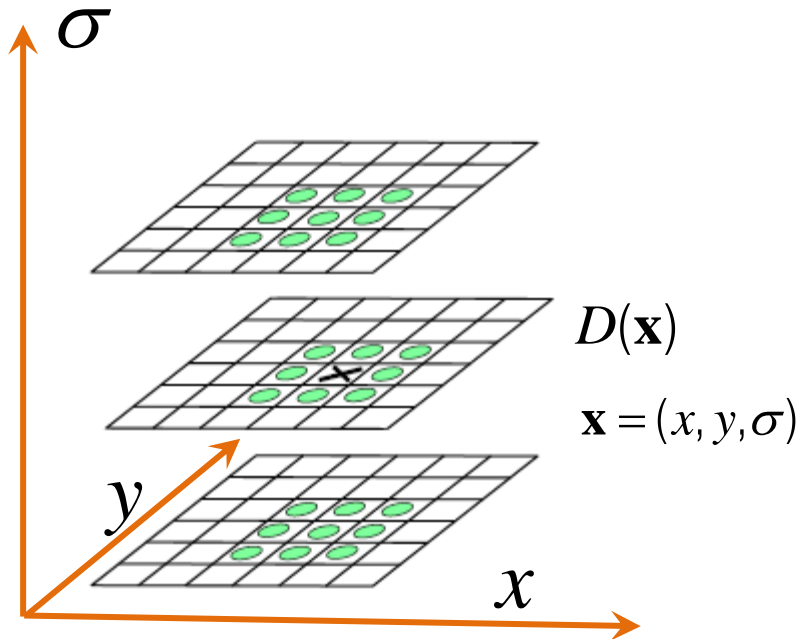


Newton's method

$$f(x + \Delta x) = f(x) + f' \Delta x + \frac{1}{2} f'' \Delta x^2$$

$$\Delta x = -\frac{f'}{f''}$$

SUBPIXEL KEYPOINT LOCALIZATION



Newton's method

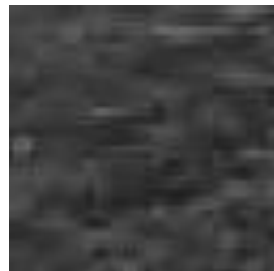
$$f(x + \Delta x) = f(x) + f' \Delta x + \frac{1}{2} f'' \Delta x^2$$

$$\Delta x = -\frac{f'}{f''}$$

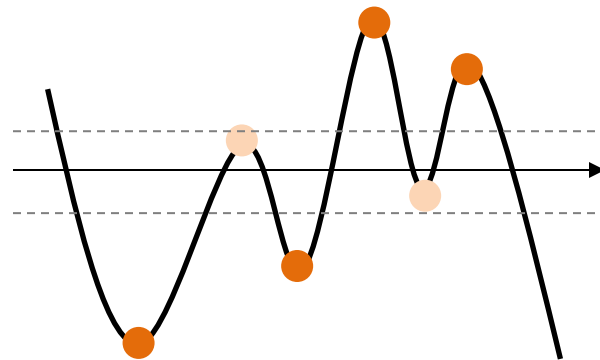
$$D(\mathbf{x} + \Delta \mathbf{x}) = D(\mathbf{x}) + \underbrace{\frac{\partial D}{\partial \mathbf{x}}}_{\text{Gradient}} \Delta \mathbf{x} + \frac{1}{2} \Delta \mathbf{x}^\top \underbrace{\frac{\partial^2 D}{\partial \mathbf{x}^2}}_{\text{Hessian}} \Delta \mathbf{x}$$

$$\Delta \mathbf{x} = -\left(\frac{\partial^2 D}{\partial \mathbf{x}^2} \right)^{-1} \frac{\partial D}{\partial \mathbf{x}}$$

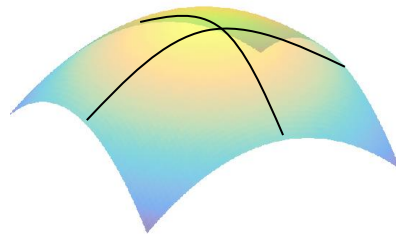
THRESHOLDING: LOW CONTRAST



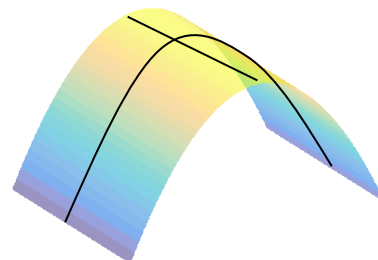
$$|D(\mathbf{x})| < 0.03$$



THRESHOLDING: EDGE



$$\lambda_1 \approx \lambda_2$$

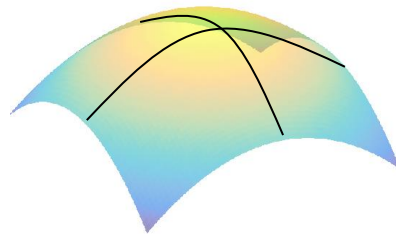


$$\lambda_1 \gg \lambda_2$$

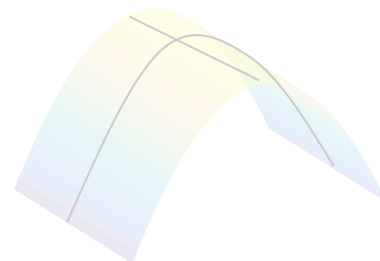
Principal curvatures are eigenvalues of Hessian matrix:

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

THRESHOLDING: EDGE



$$\lambda_1 \approx \lambda_2$$



$$\lambda_1 \gg \lambda_2$$

Principal curvatures are eigenvalues of Hessian matrix:

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

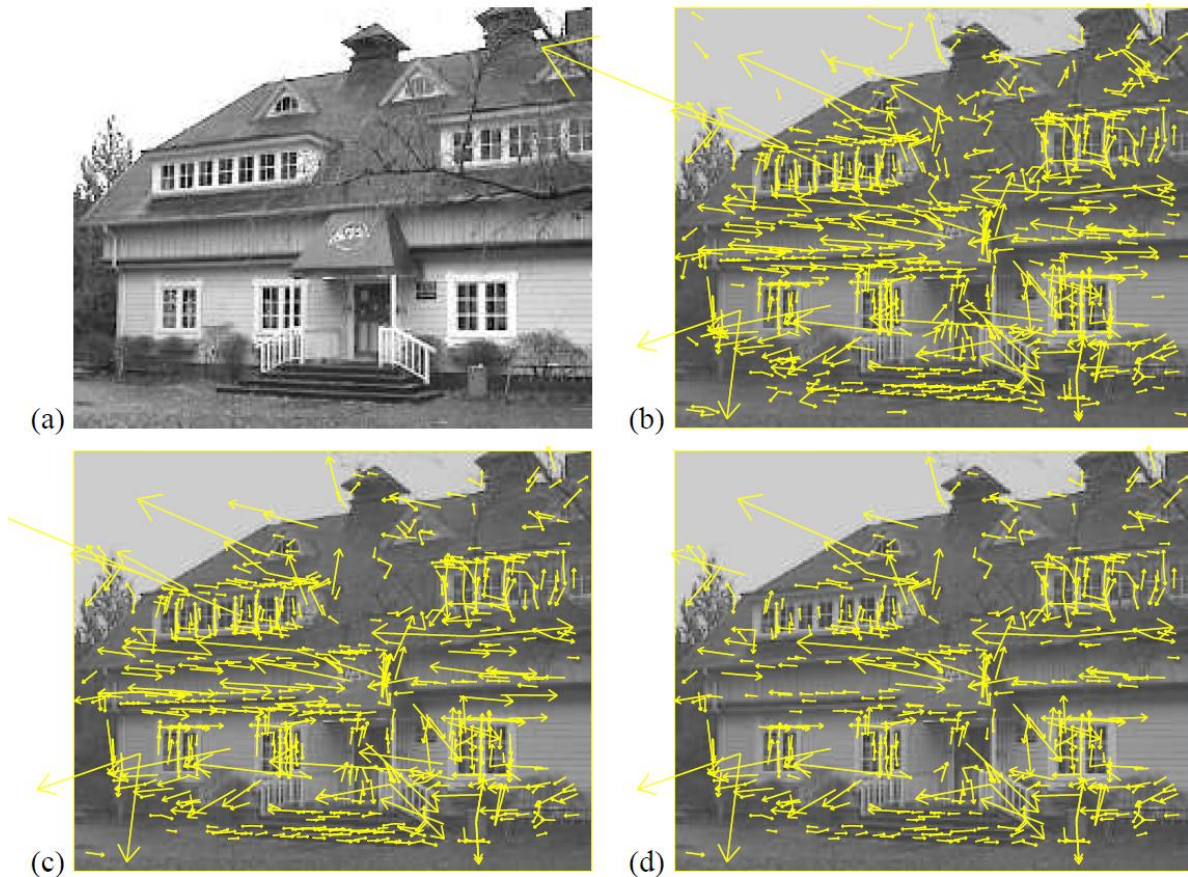
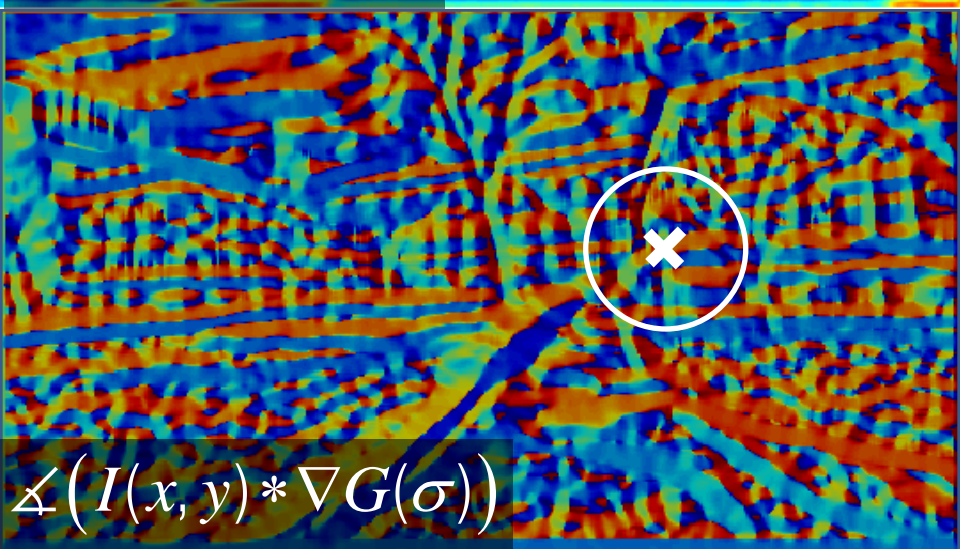
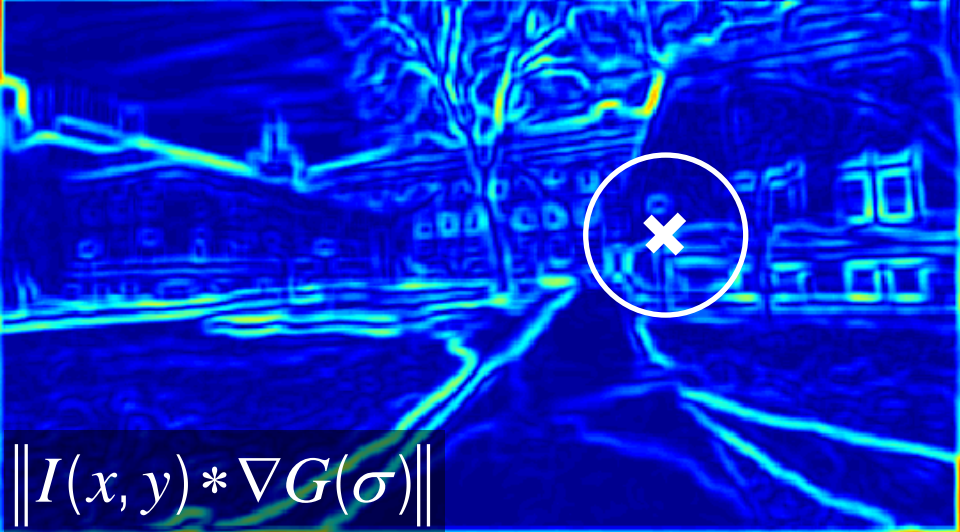
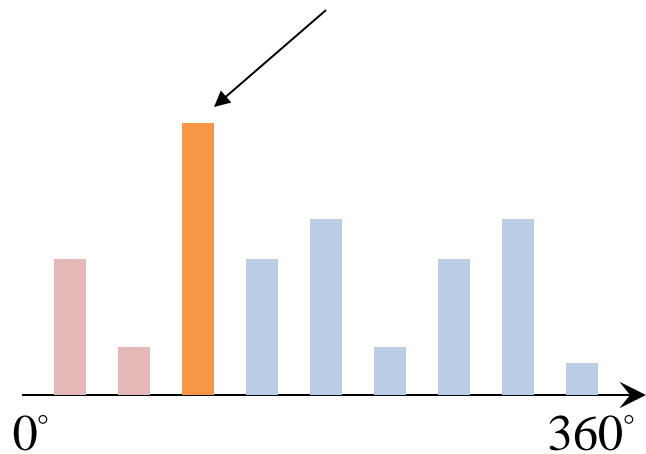


Figure 5: This figure shows the stages of keypoint selection. (a) The 233x189 pixel original image. (b) The initial 832 keypoints locations at maxima and minima of the difference-of-Gaussian function. Keypoints are displayed as vectors indicating scale, orientation, and location. (c) After applying a threshold on minimum contrast, 729 keypoints remain. (d) The final 536 keypoints that remain following an additional threshold on ratio of principal curvatures.

ORIENTATION ASSIGNMENT

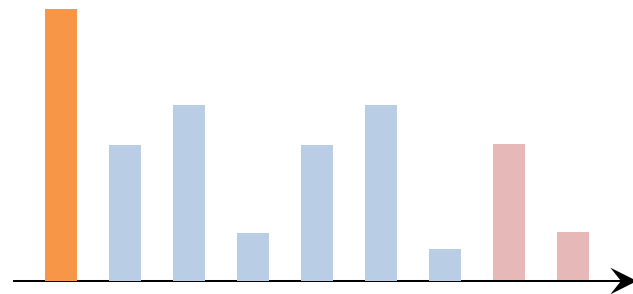
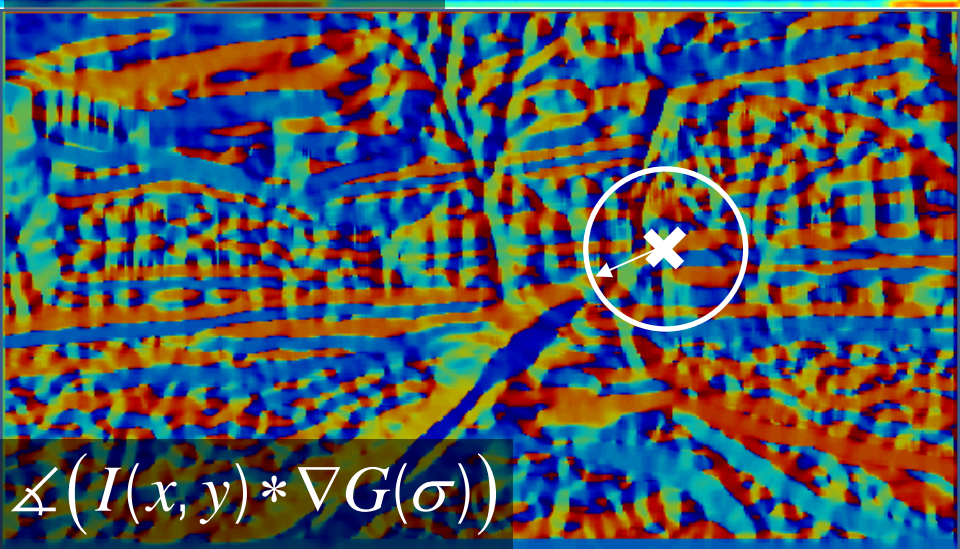
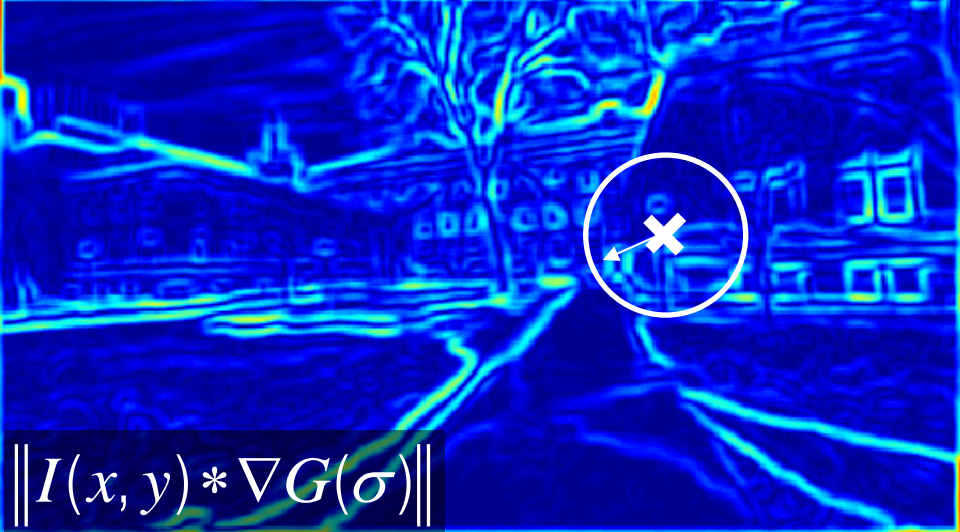


Peak orientation ~ dominant orientation



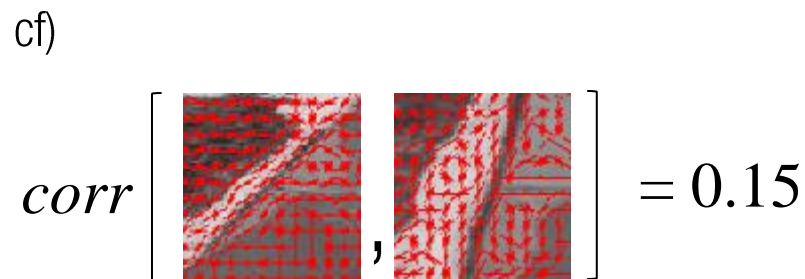
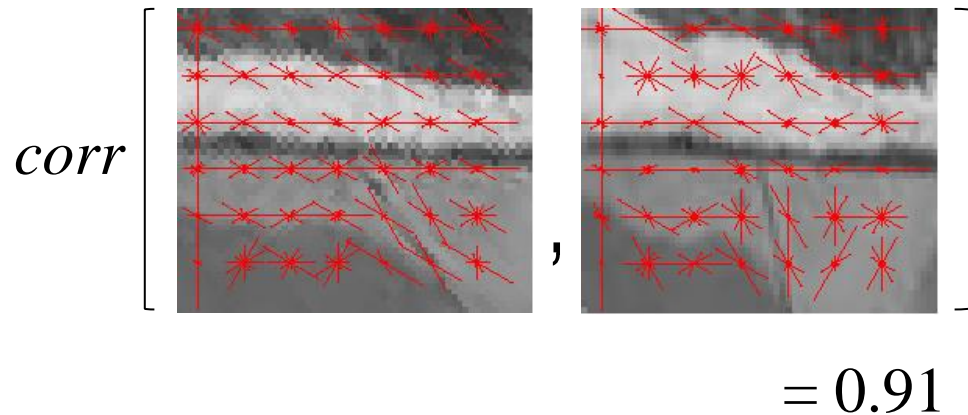
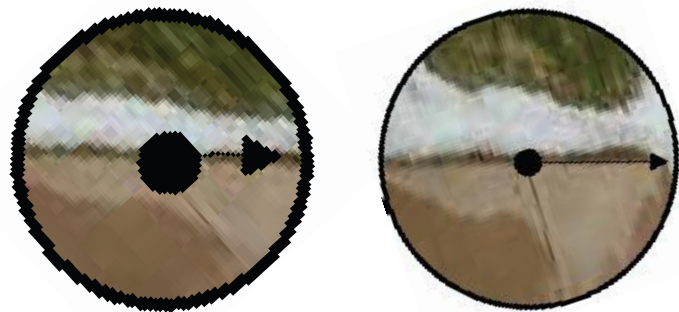
Histogram of orientation
weighted by gradient magnitude and Gaussian

ORIENTATION NORMALIZATION



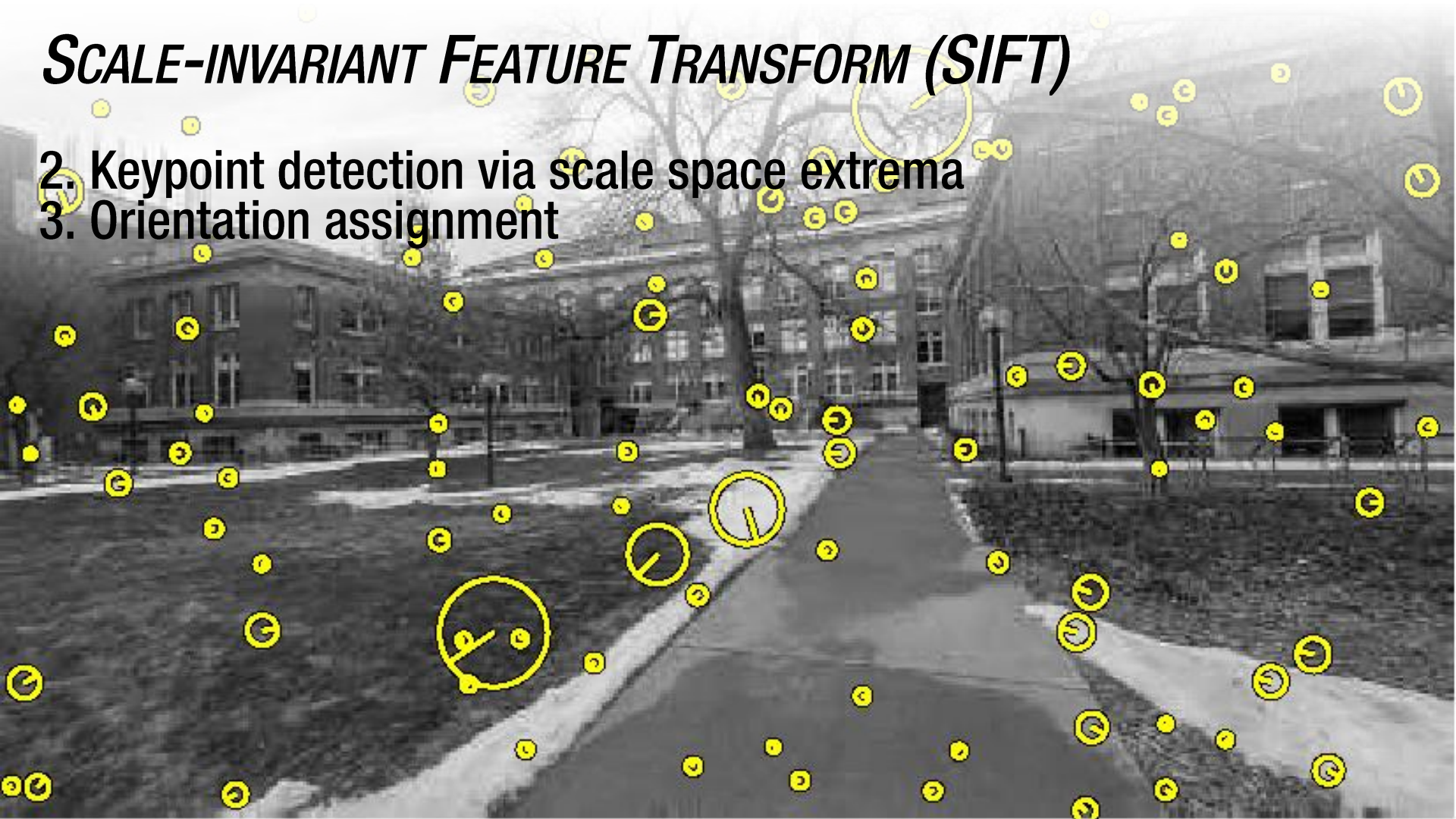
Histogram of orientation
weighted by gradient magnitude and Gaussian

SCALE-INVARIANT FEATURE TRANSFORM (SIFT)

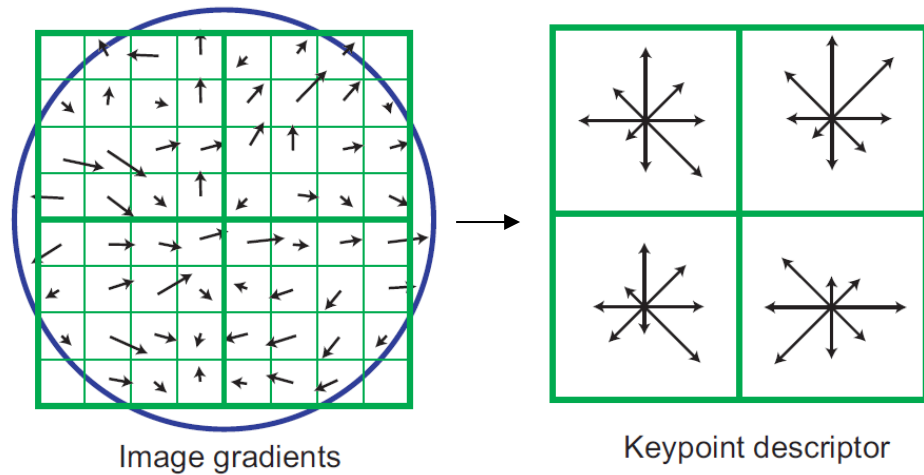
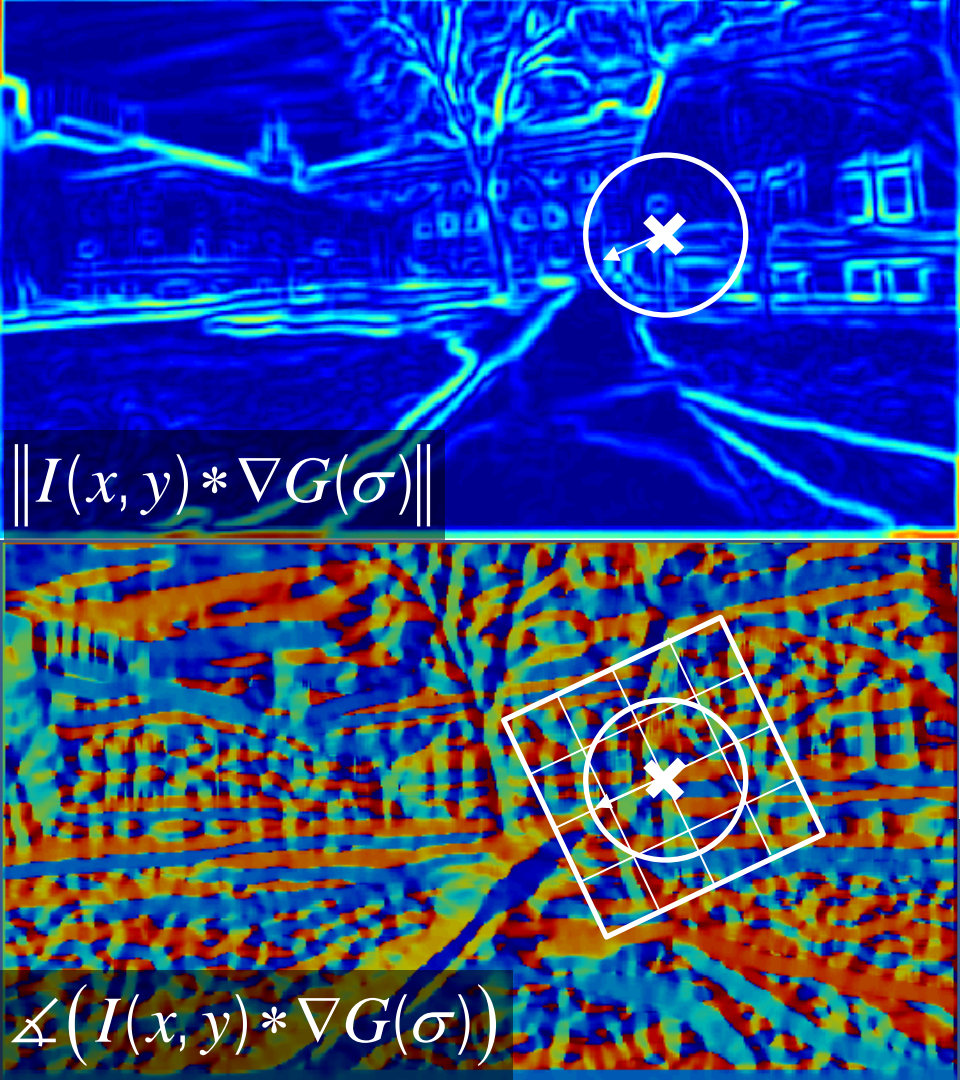


SCALE-INVARIANT FEATURE TRANSFORM (SIFT)

2. Keypoint detection via scale space extrema
3. Orientation assignment

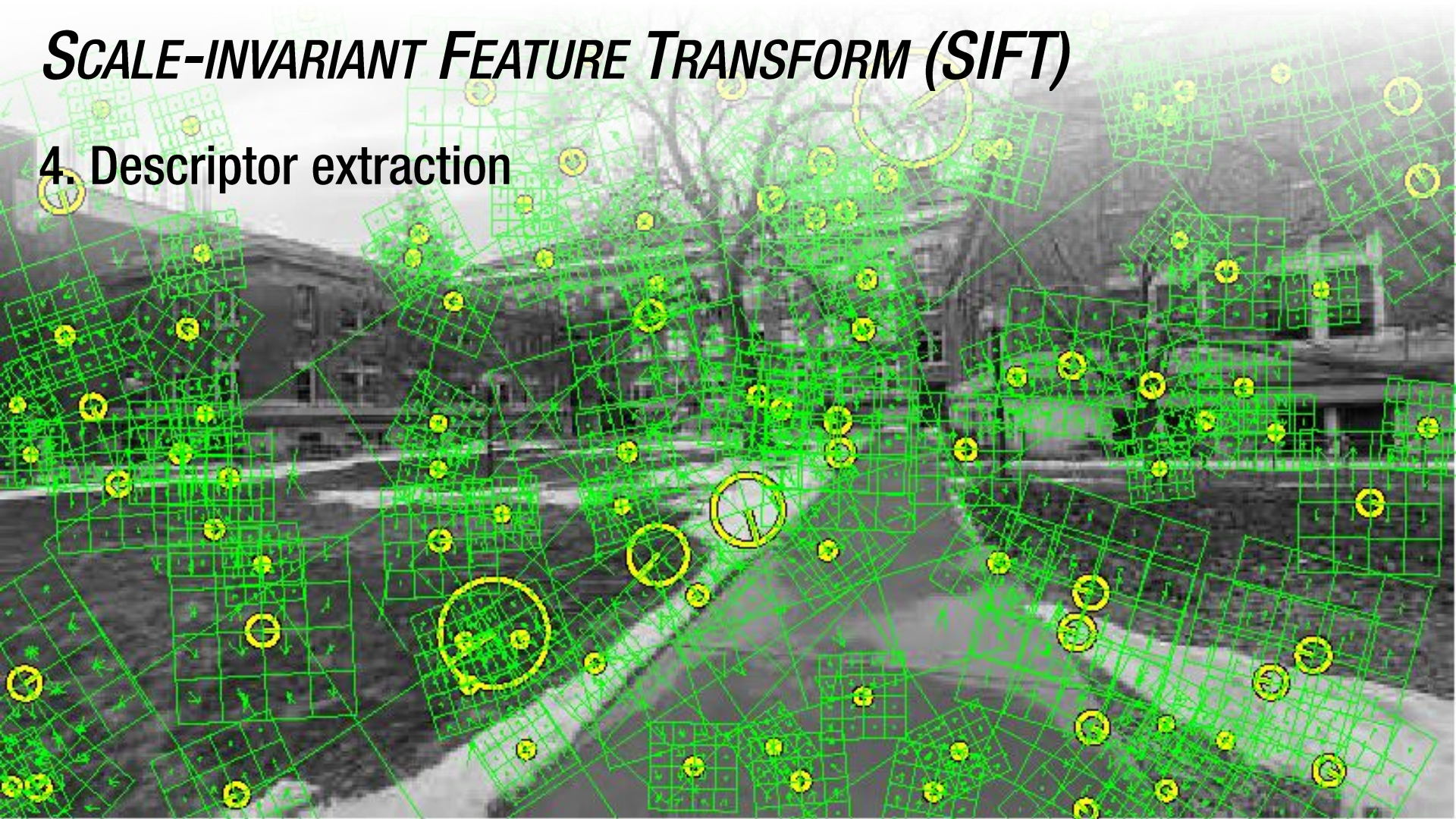


DESCRIPTOR COMPUTATION



SCALE-INVARIANT FEATURE TRANSFORM (SIFT)

4. Descriptor extraction





Local visual descriptor



Local visual descriptor



$$\left\| \begin{array}{c} \text{descriptor1} \\ \text{descriptor2} \end{array} \right\| = 0$$

