Scale Invariant Feature Transform by David Lowe

Presented by:
Jerry Chen   Achal Dave   Vaishaal Shankar

Some slides from Jason Clemons

Motivation

- Image Matching
  - Correspondence Problem
- Desirable Feature Characteristics
  - Scale Invariance
  - Rotation Invariance
  - Illumination invariance
  - Viewpoint invariance

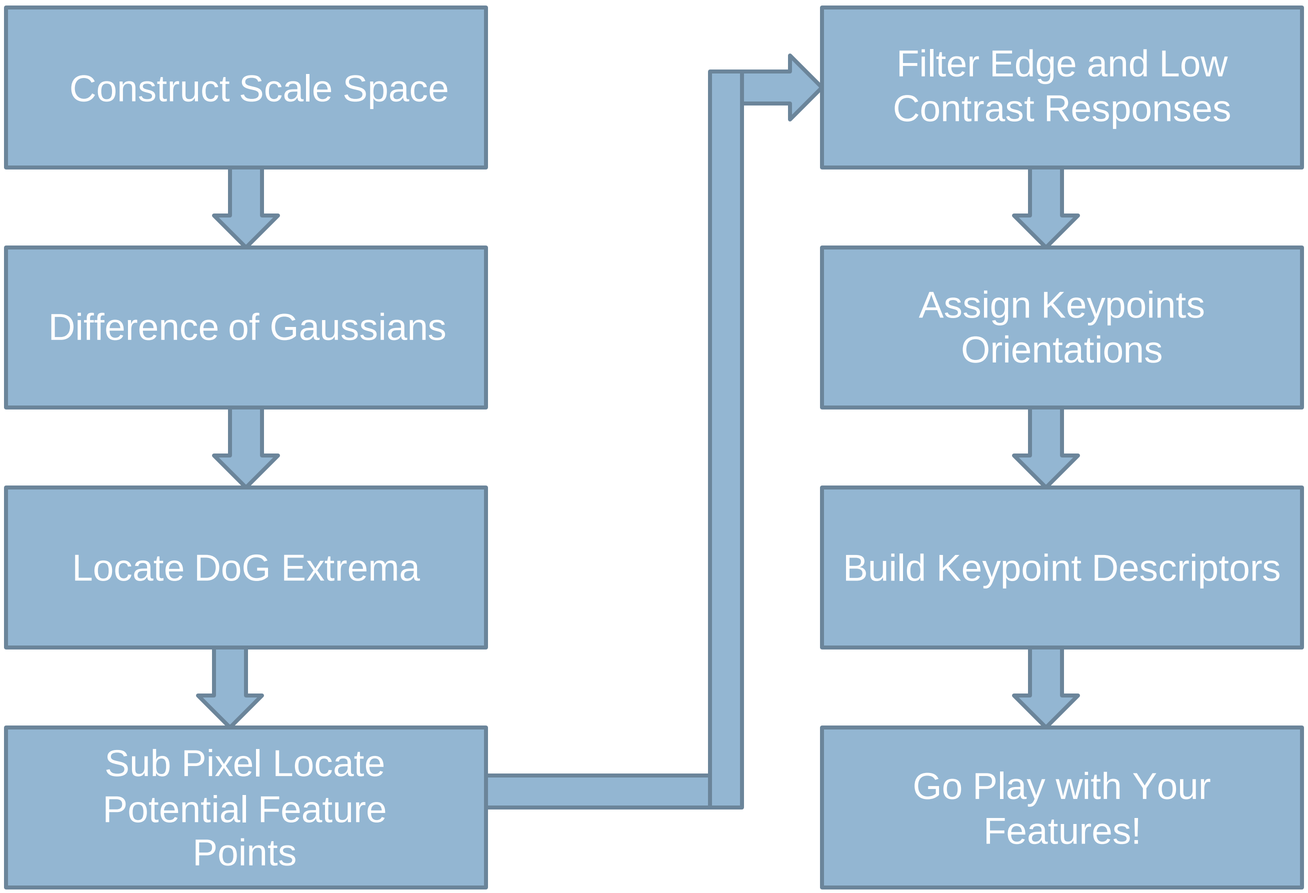
History

- Moravec (1981) – corner detector
- Harris (1988) – better corner detector, but still not scale invariant
- Lowe (1999) – local feature with scale invariance

Overview

- Scale Space, Difference of Gaussians
- Keypoint Localization
- Orientation Assignment
- Descriptor

Overview of Algorithm



Motivation for Scale Space

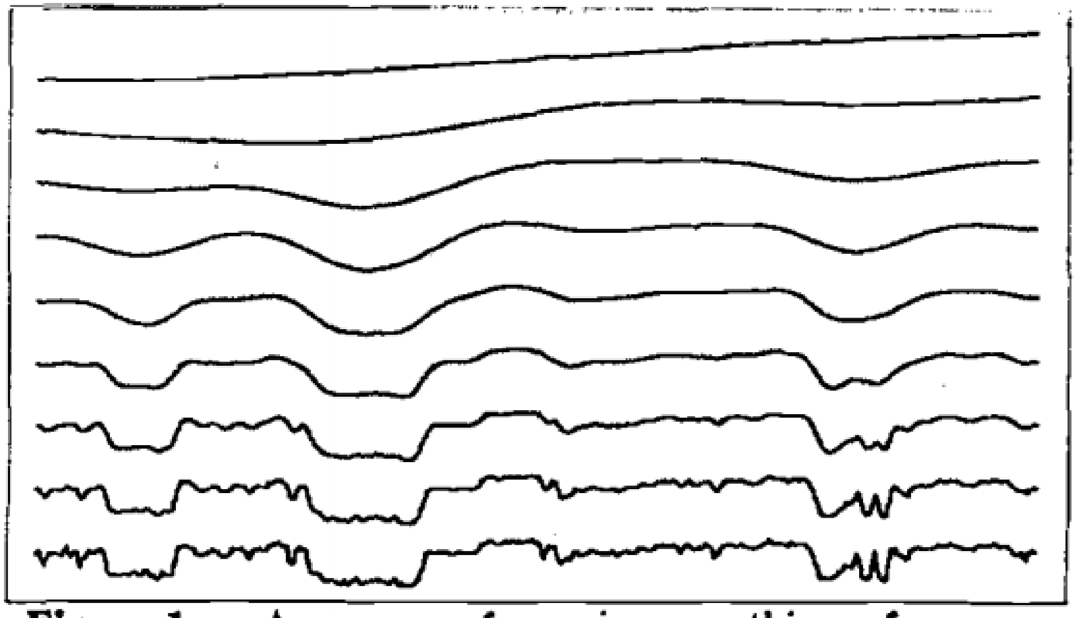


Figure 1. A sequence of gaussian smoothings of a waveform, with σ decreasing from top to bottom. Each graph is a constant-σ profile from the scale-space image.

Motivation for Scale Space

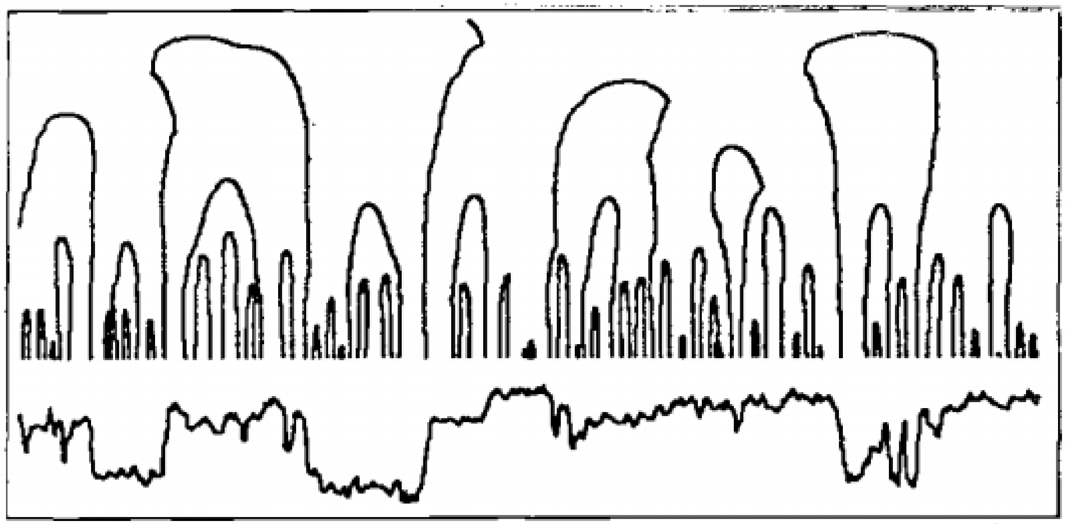


Figure 2. Contours of $F_{xx} = 0$ in a scale-space image. The x-axis is horizontal; the coarsest scale is on top. To simulate the effect of a continuous scale-change on the qualitative description, hold a straight-edge (or better still, a slit) horizontally. The intersections of the edge with the zero-contours are the extremal points at some single value of σ. Moving the edge up or down increases or decreases σ.

Motivation for Scale Space

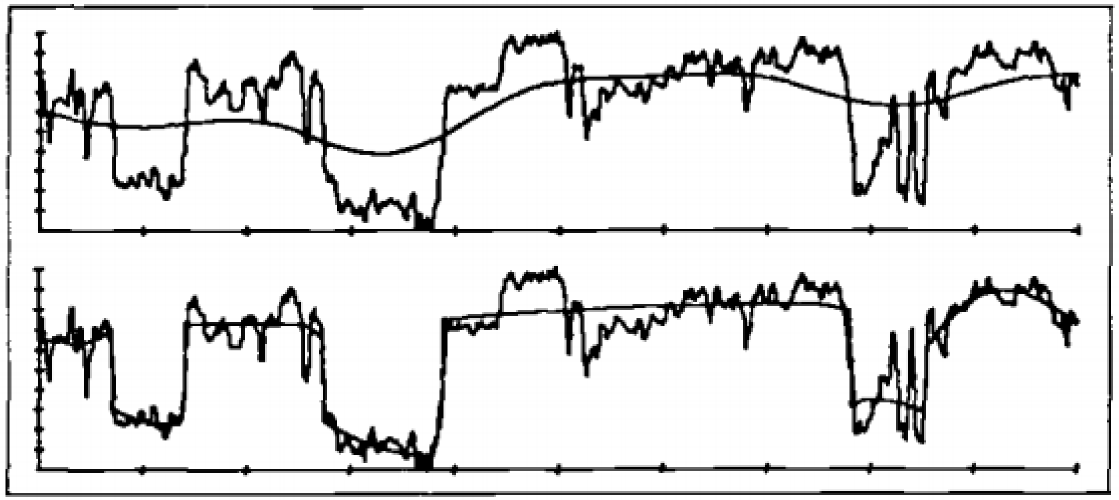
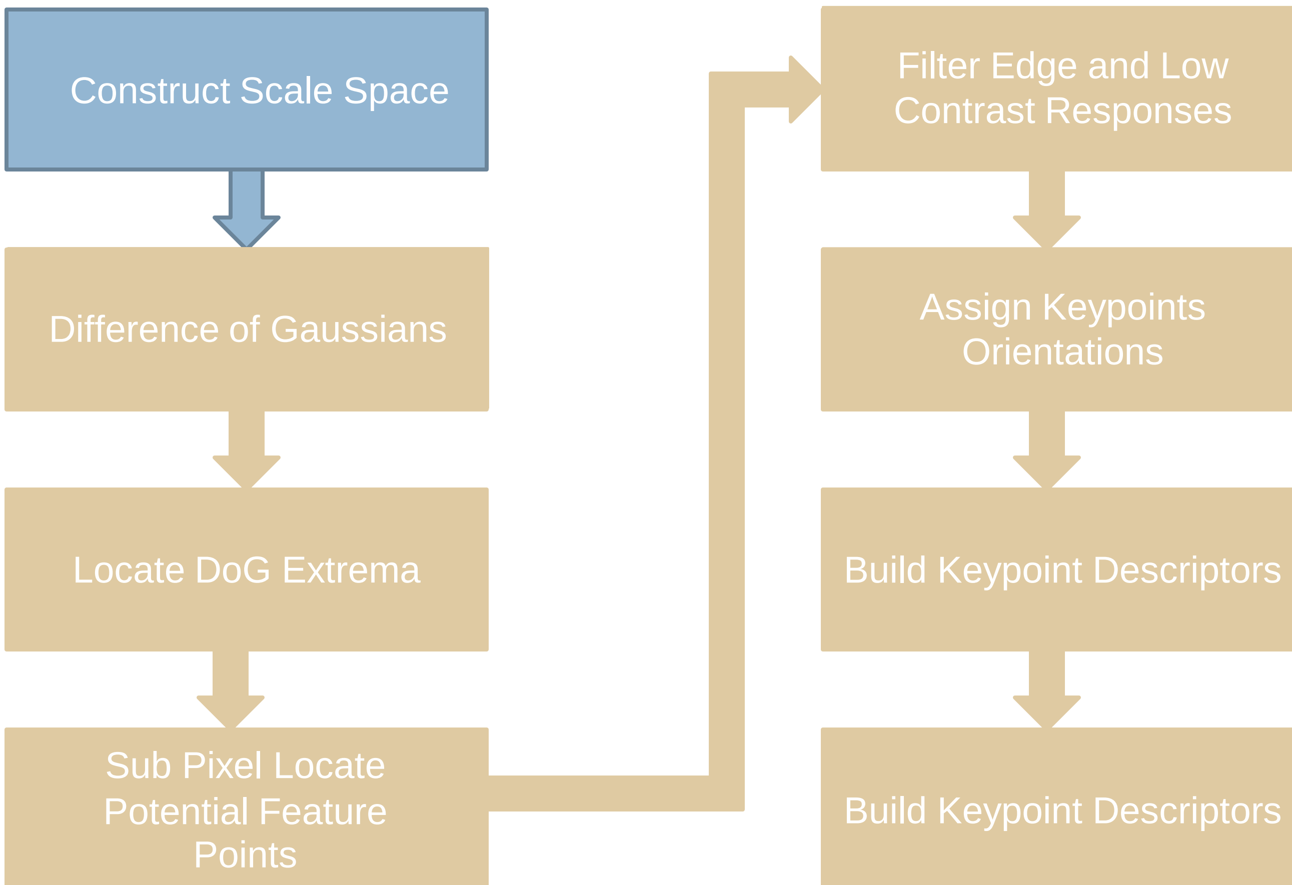
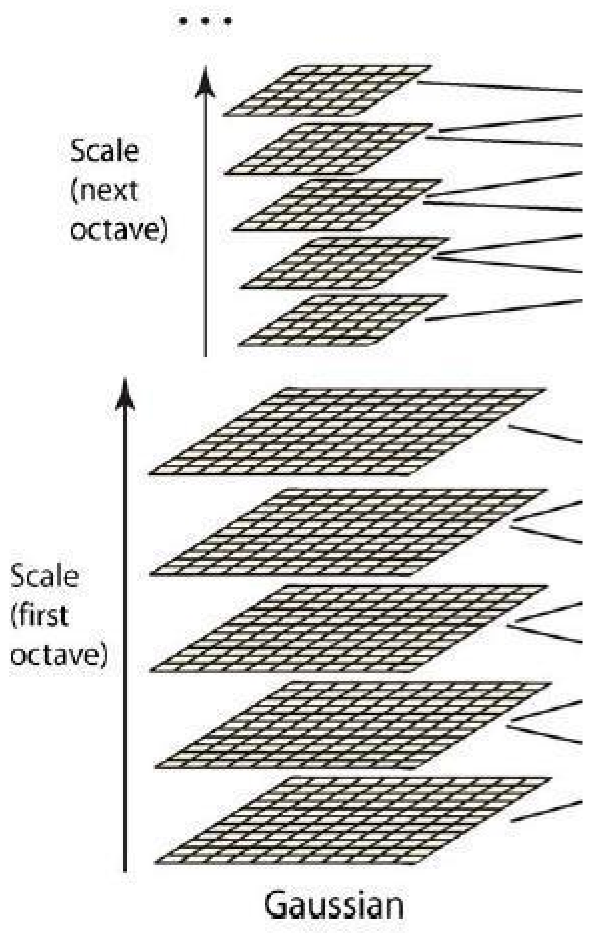


Figure 3. Below is shown a signal with a coarse-to-fine tracking approximation superimposed. The approximation was produced by independent parabolic fits between the localized inflections. Above is shown the corresponding (qualitatively isomorphic) gaussian smoothing.

Constructing Scale Space



Scale Space



Scale Space

Constructing Scale Space

- Gaussian kernel used to create scale space
  - Only possible scale space kernel (Lindberg „94)

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

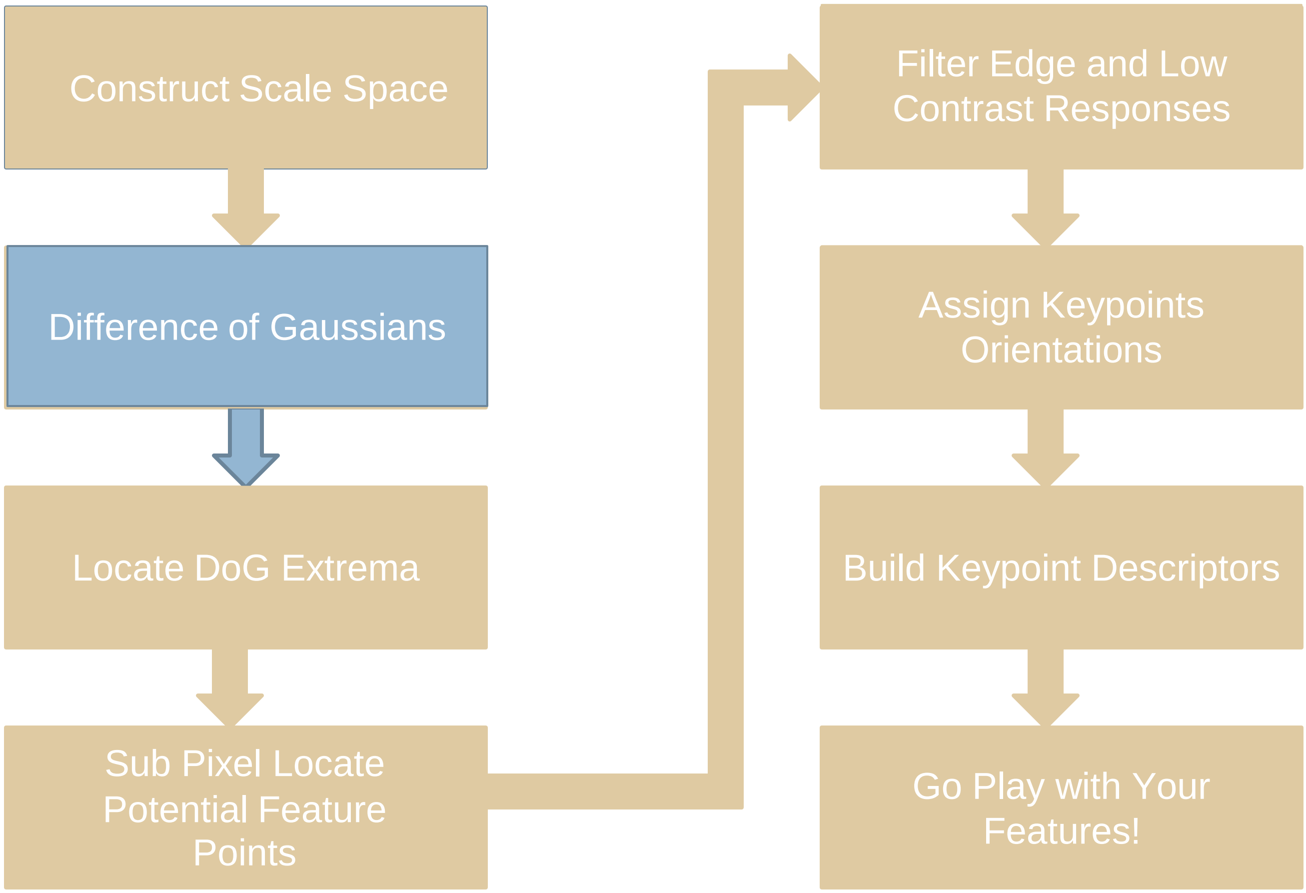
where

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}$$

Laplacian of Gaussians

- LoG - $\sigma^2 \nabla^2 G$
- Extrema Useful:
  - Found to be stable features
  - Gives excellent notion of scale
- Calculation costly...

Take DoG


Construct Scale Space
Difference of Gaussians
Locate DoG Extrema
Sub Pixel Locate Potential Feature Points
Filter Edge and Low Contrast Responses
Assign Keypoints Orientations
Build Keypoint Descriptors
Go Play with Your Features!

Difference of Gaussian

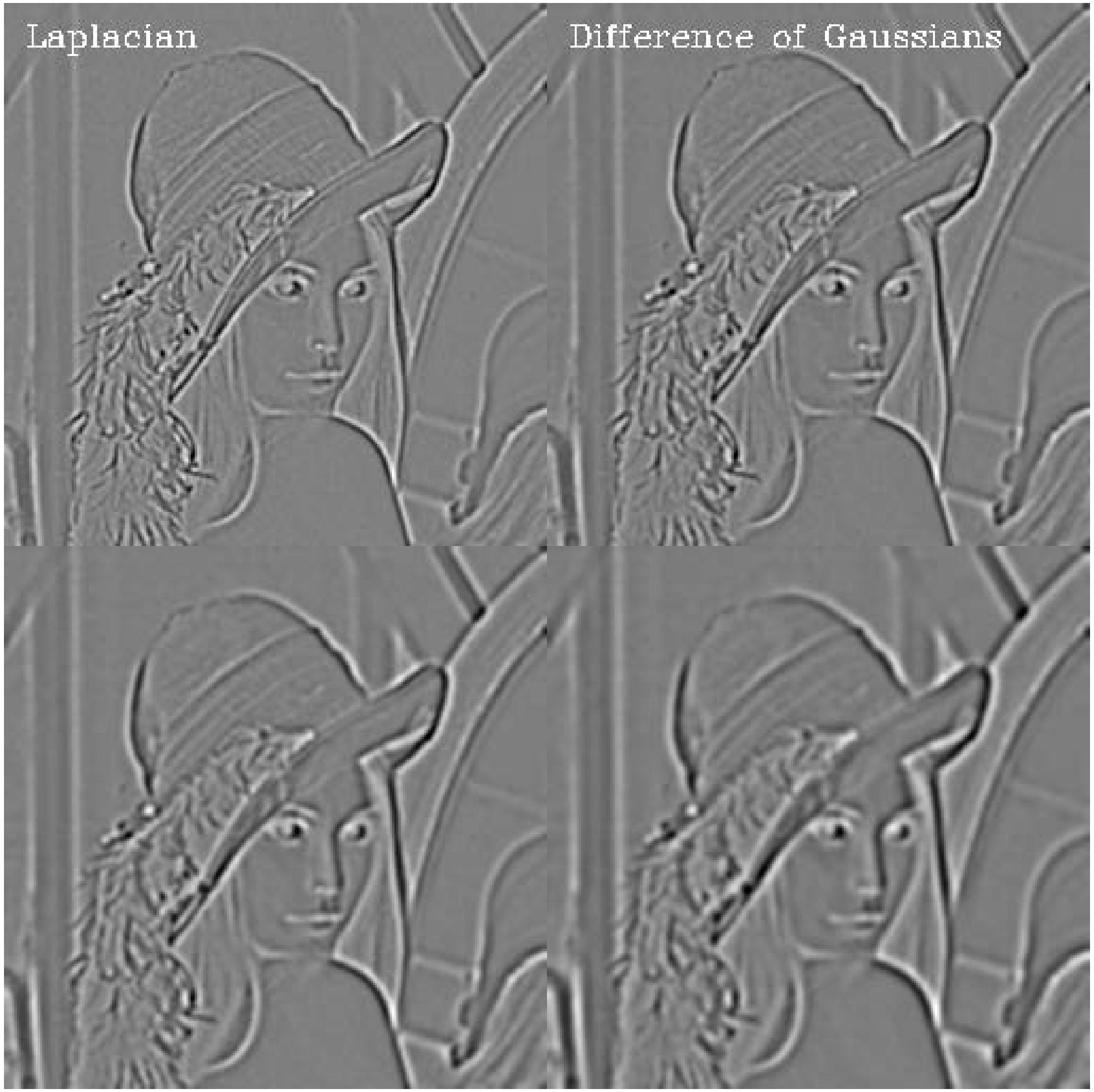
Approximation of Laplacian of Gaussians

$$\sigma \nabla^2 G = \frac{\partial G}{\partial \sigma} \approx \frac{G(x,y,k\sigma) - G(x,y,\sigma)}{k\sigma - \sigma}$$

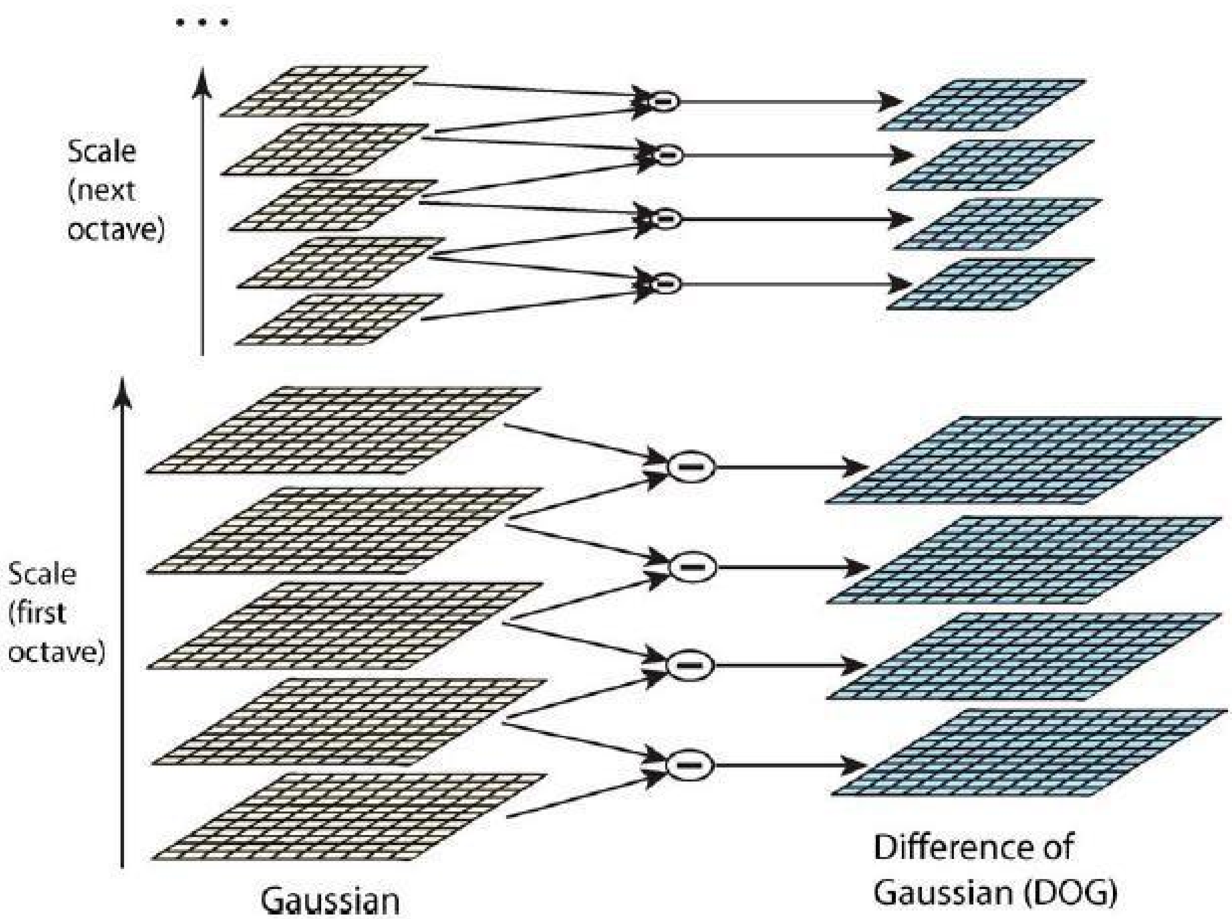
$$G(x,y,k\sigma) - G(x,y,\sigma) \approx (k-1)\sigma^2 \nabla^2 G$$

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

Difference of Gaussian

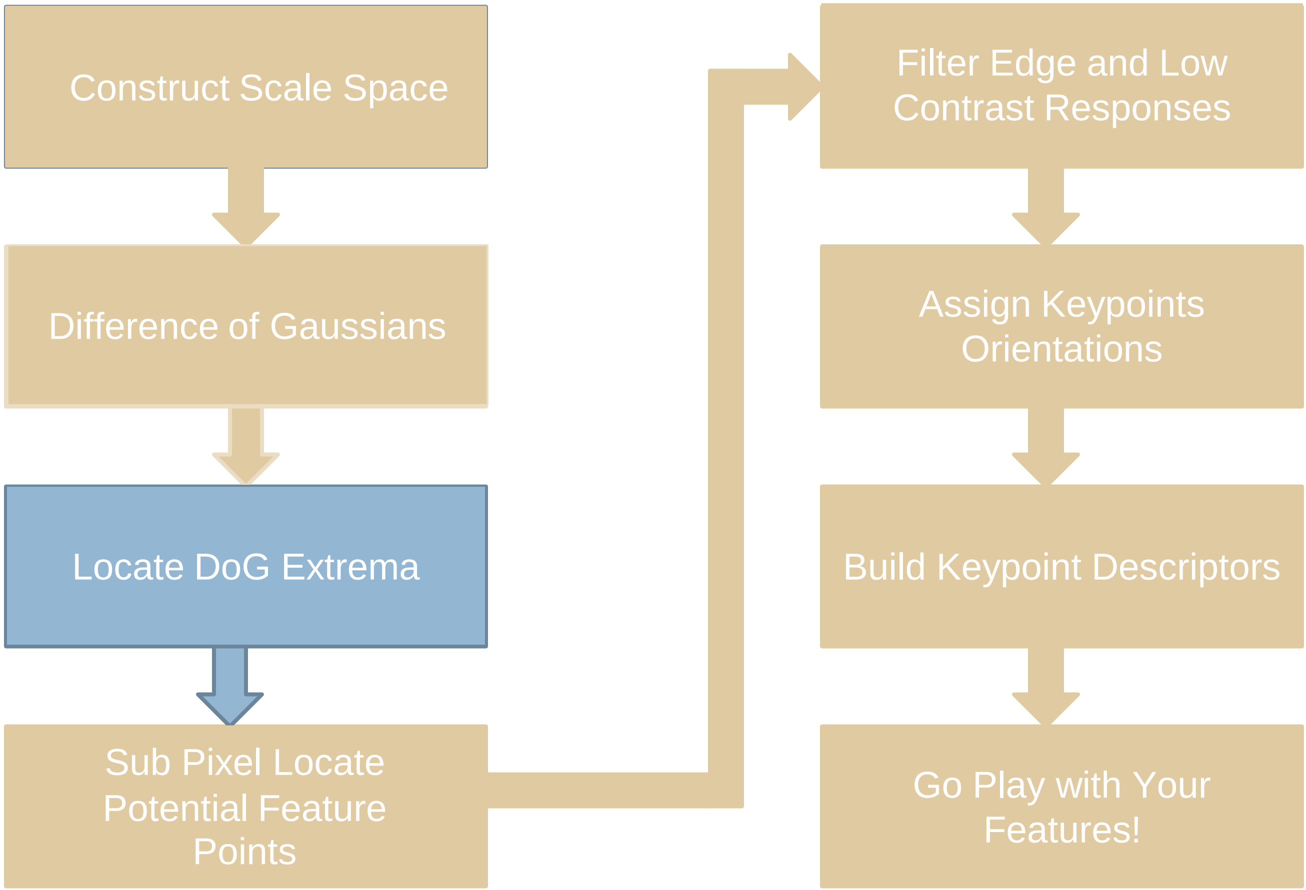


DoG Pyramid



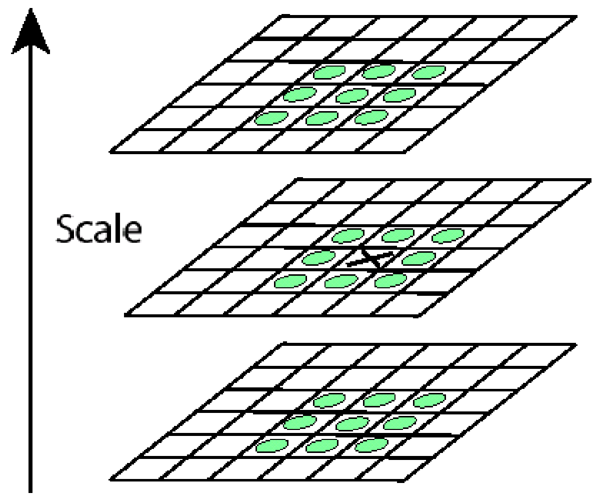
Scale
(next
octave)
Scale
(first
octave)
Gaussian
Difference of
Gaussian (DOG)

DoG Extrema

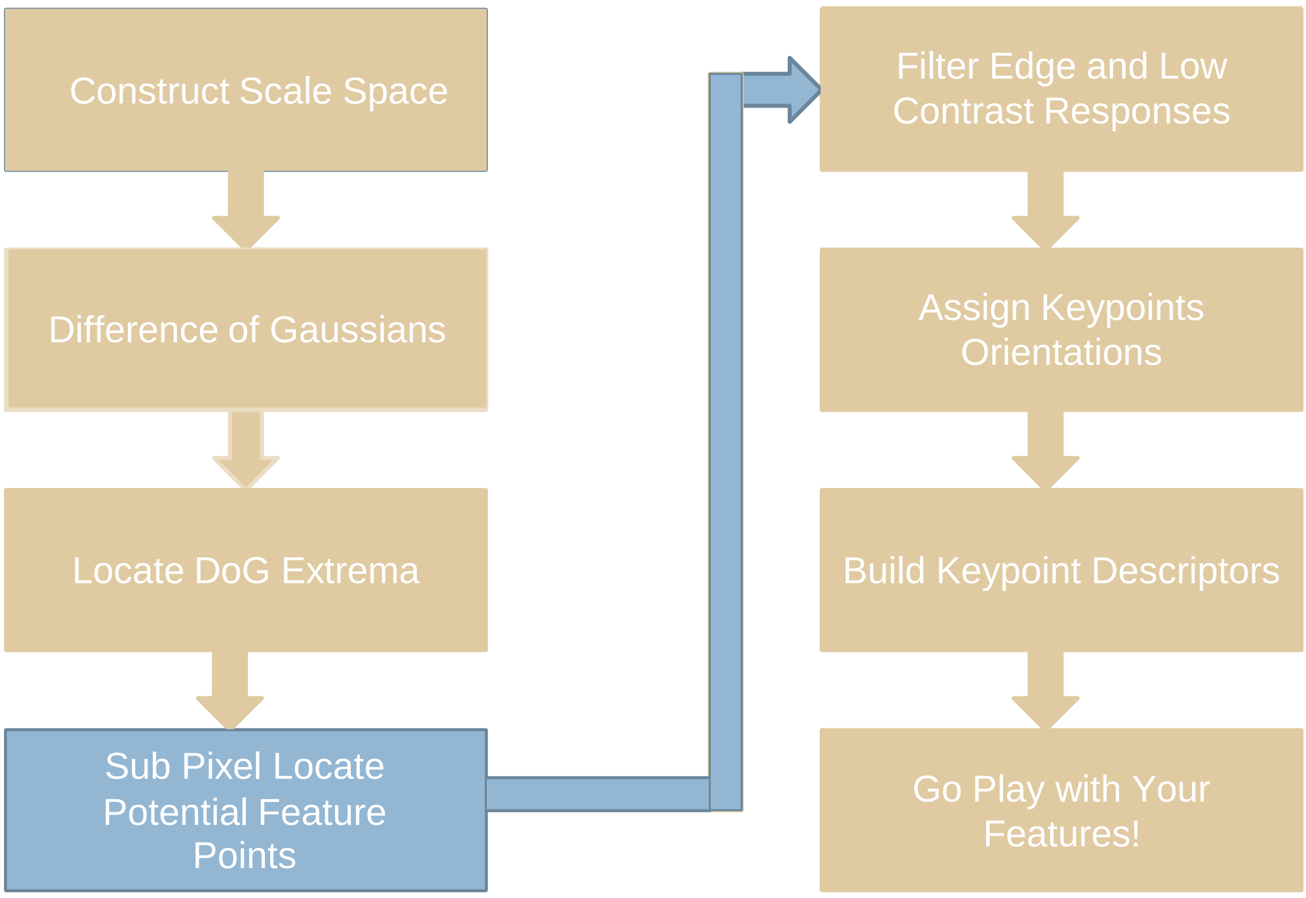

Construct Scale Space
Difference of Gaussians
Locate DoG Extrema
Sub Pixel Locate Potential Feature Points
Filter Edge and Low Contrast Responses
Assign Keypoints Orientations
Build Keypoint Descriptors
Go Play with Your Features!

Locate the Extrema of the DoG

- Scan each DOG image
  - Look at all neighboring points (including scale)
  - Identify Min and Max
    - 26 Comparisons



Sub-pixel Localization



Sub-pixel Localization

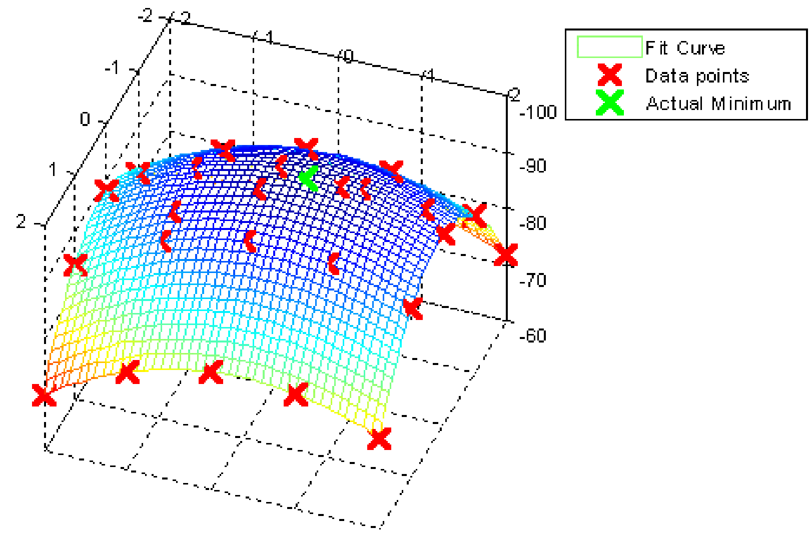
3D Curve Fitting

Taylor Series Expansion

$$D(\mathbf{x}) = D + \frac{\partial D}{\partial \mathbf{x}}^T \mathbf{x} + \frac{1}{2}\mathbf{x}^{\mathrm{T}} \frac{\partial^2 D}{\partial \mathbf{x}^2}\mathbf{x}$$

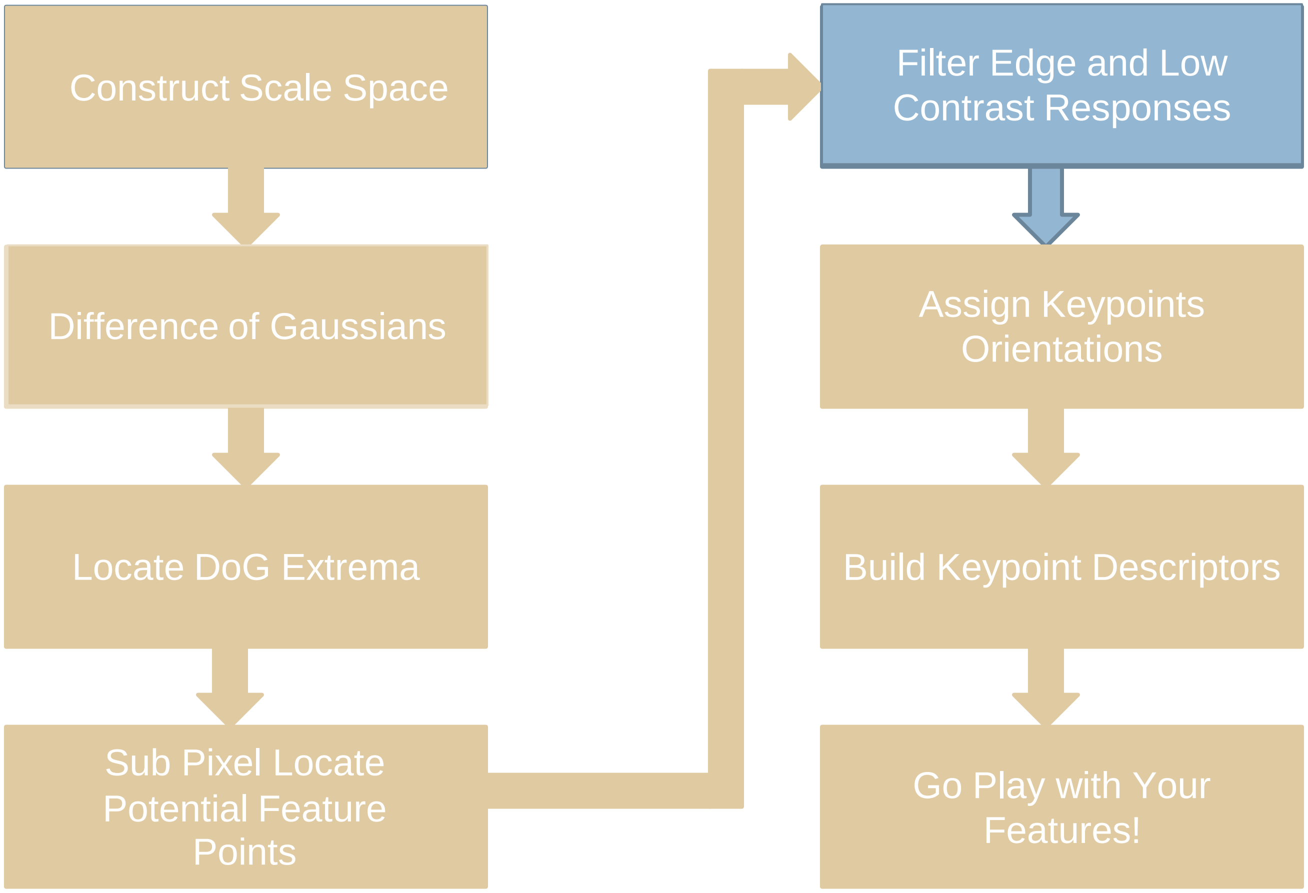
Differentiate and set to 0

$$\hat{\mathbf{x}} = -\frac{\partial^2 D}{\partial \mathbf{x}^2}^{-1} \frac{\partial D}{\partial \mathbf{x}}.$$



to get location in terms of (x,y,σ)

Filter Responses



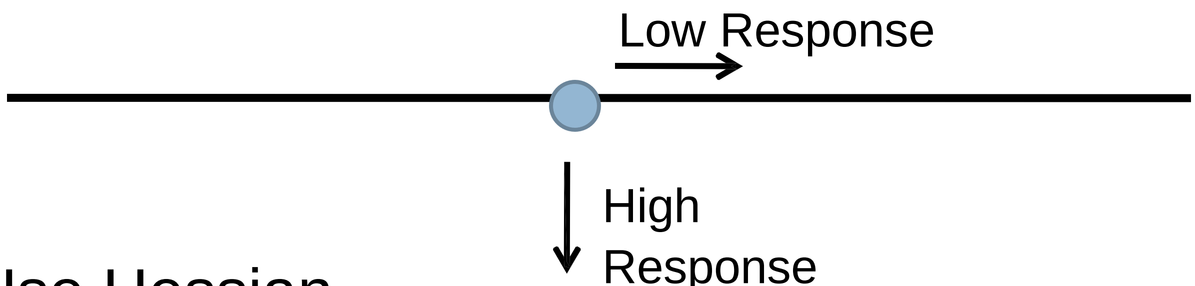
Filter Low Contrast Points

- Low Contrast Points Filter
  - Use Scale Space value at previously found location

$$D(\hat{\mathbf{x}}) = D + \frac{1}{2}\frac{\partial D}{\partial \mathbf{x}}^T \hat{\mathbf{x}}.$$

Edge Response Elimination

- Peak has high response along edge, poor other direction

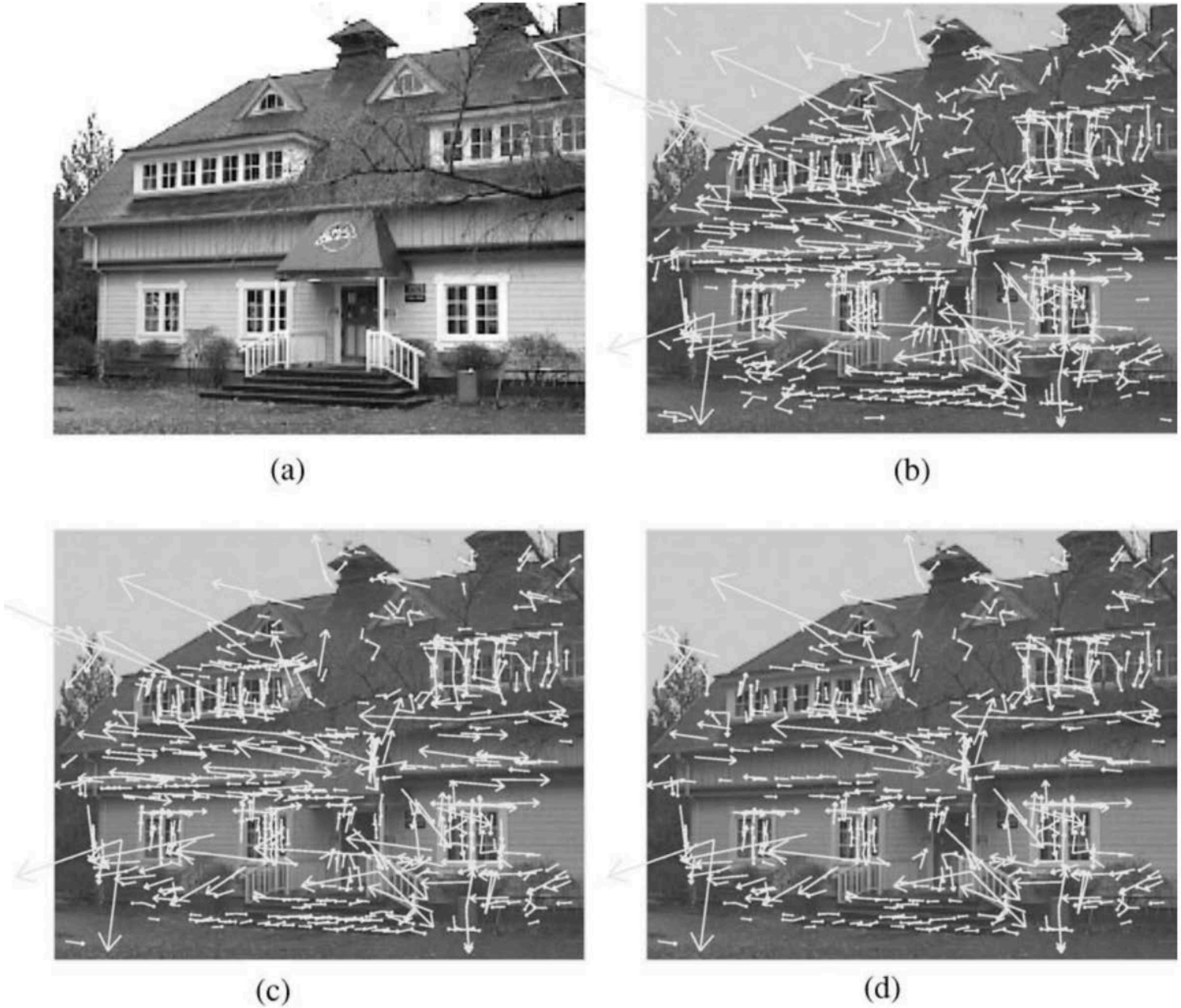


- Use Hessian
  - Eigenvalues proportional to principal curvatures
  - Use trace and determinant

$$Tr(H) = D_{xx} + D_{yy} = \alpha + \beta$$

$$Det(H) = D_{xx} D_{yy} - (D_{xy})^2 = \alpha \beta$$

$$\frac{Tr(H)^2}{Det(H)} < \frac{(r+1)^2}{r}$$



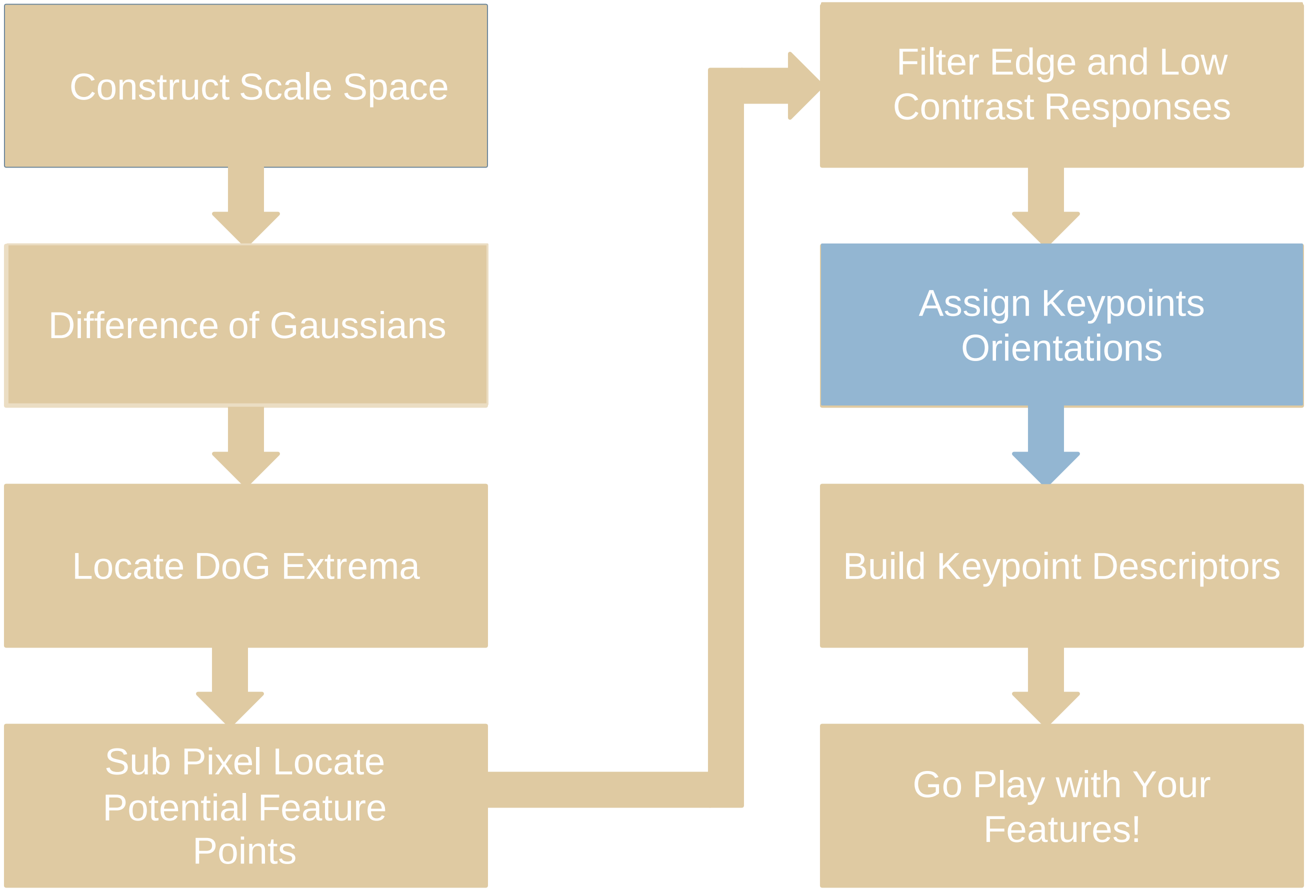
(a) (b)

(c) (d)

Apply Contrast Limit

Apply Contrast and Edge Response Elimination

Assign Keypoint Orientations



Orientation Assignment

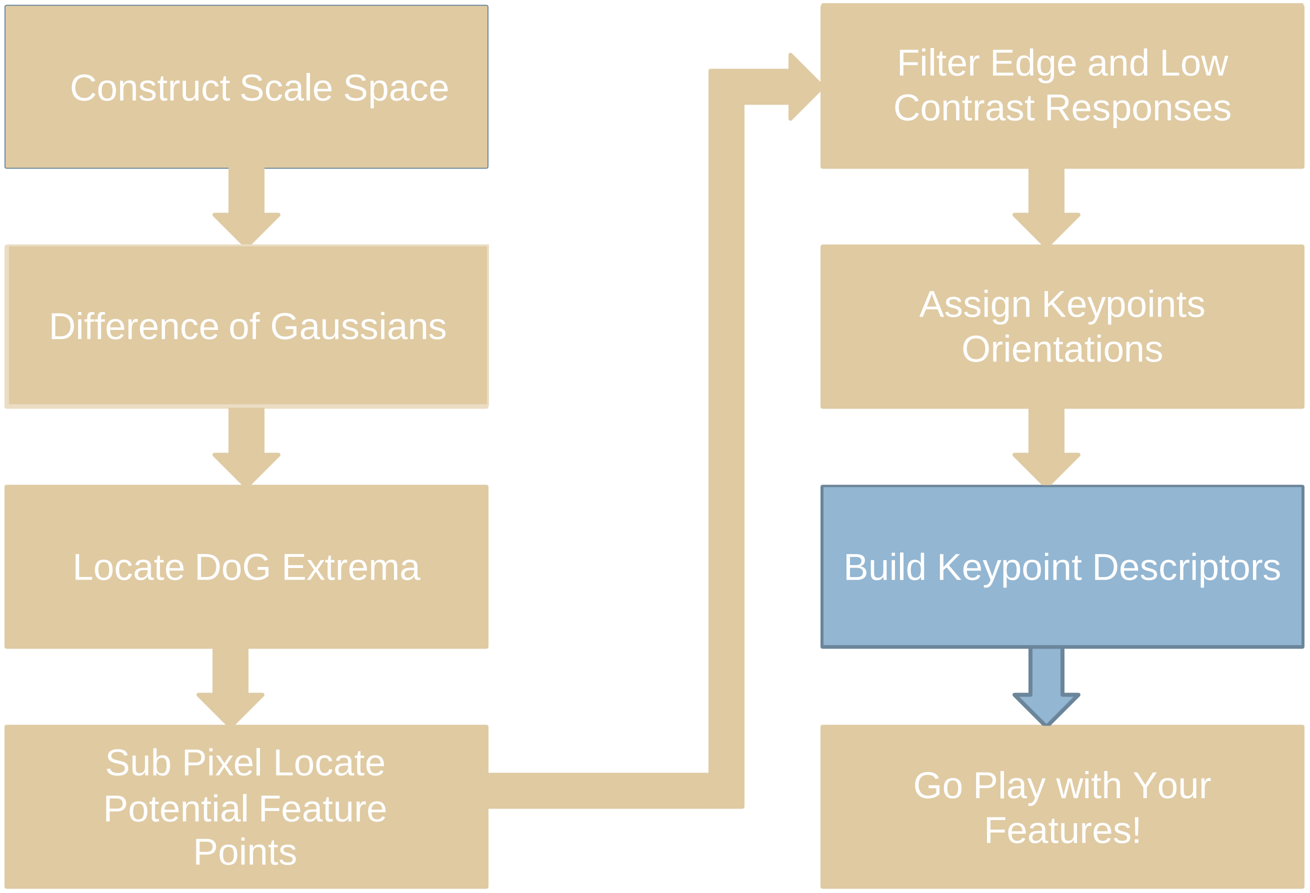
- Compute Gradient for each blurred image

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2}$$
$$\theta(x, y) = \tan^{-1}((L(x, y+1) - L(x, y-1))/(L(x+1, y) - L(x-1, y)))$$

- For region around keypoint
  - Create Histogram with 36 bins for orientation
  - Weight each point with Gaussian window of 1.5σ
  - Create keypoint for all peaks with value>=.8*(max bin))

Note that a parabola is fit to better locate each max (least squares)

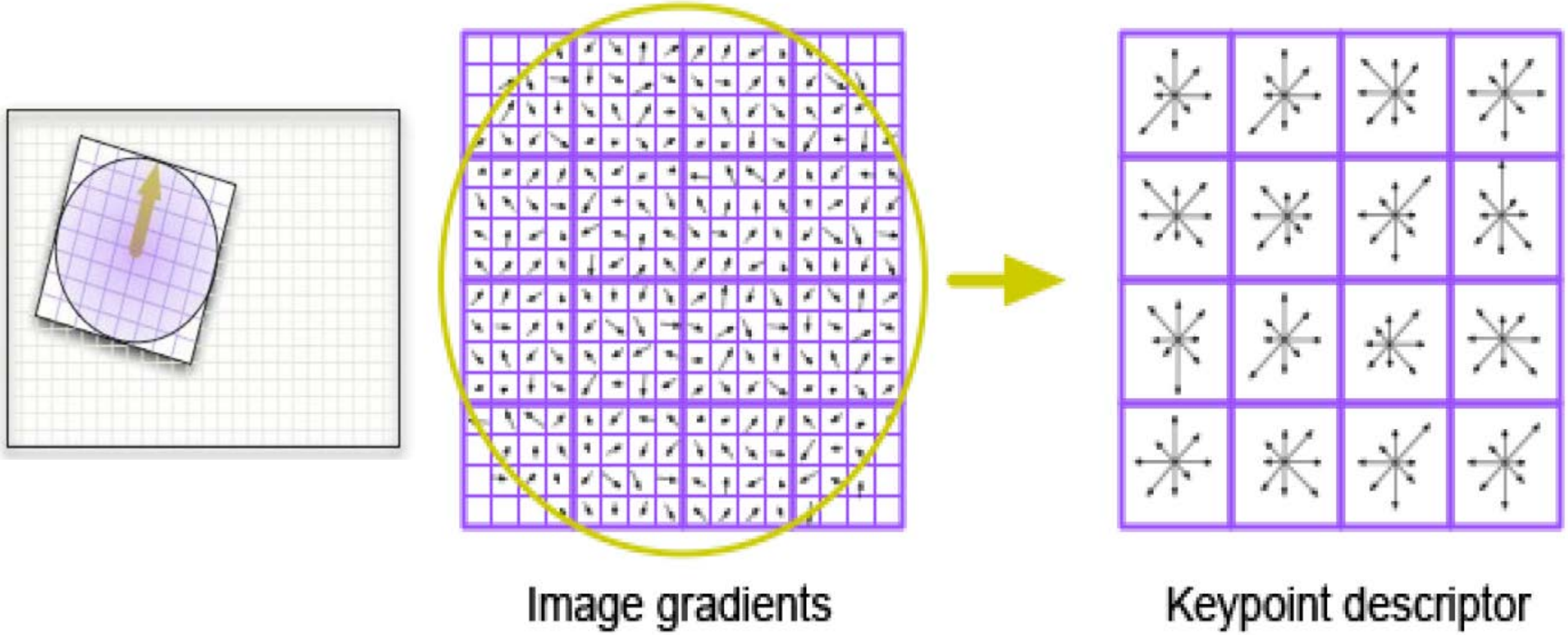
Build Keypoint Descriptors



Building the Descriptor

- Find the blurred image of closest scale
- Sample the points around the keypoint
- Rotate the gradients and coordinates by previously computed orientation
- Separate the region into subregions
  - Create histogram for each sub region with 8 bins
  - Weight the samples with $N(\sigma)$ = 1.5 Region width

Building a Descriptor

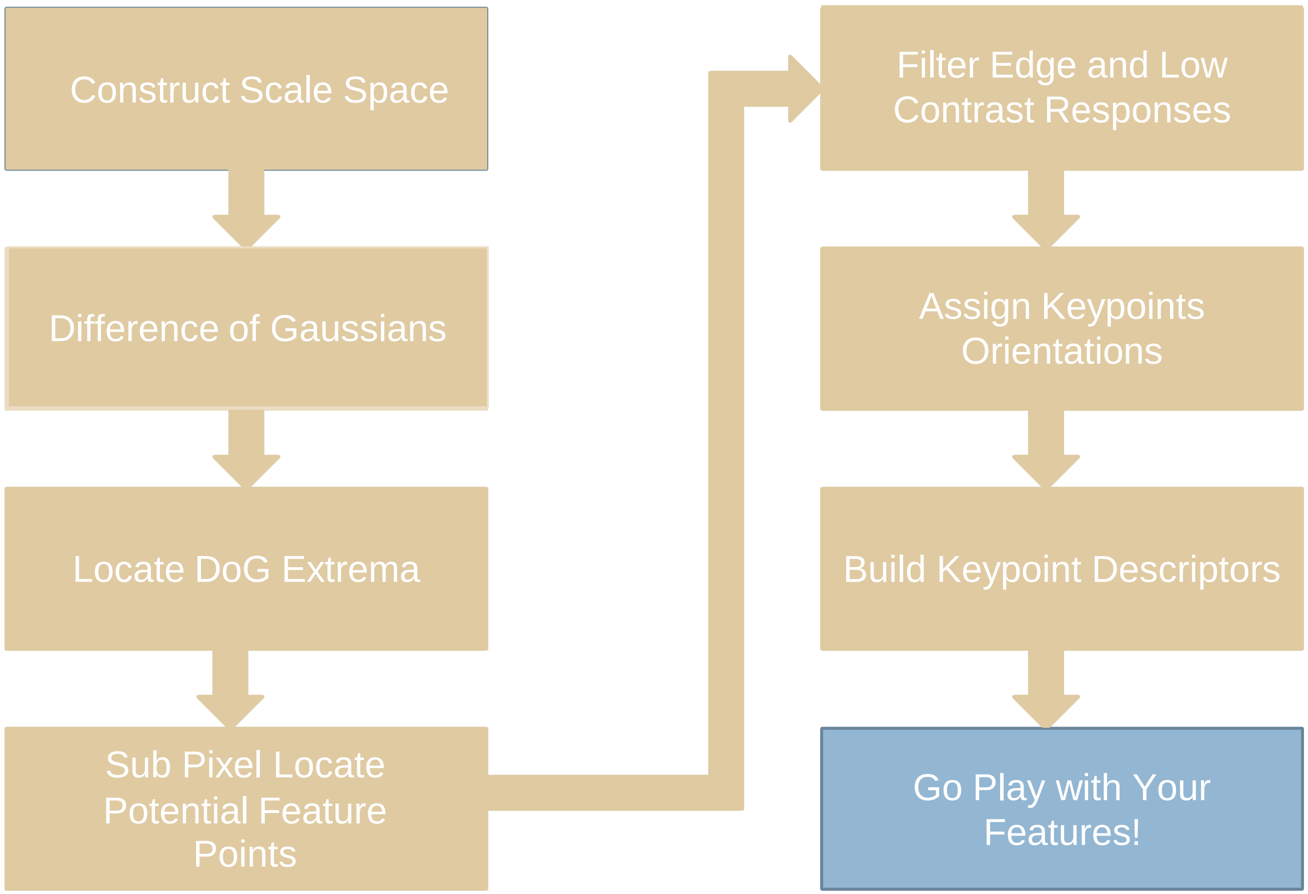


Actual implementation uses 4x4 descriptors from 16x16
which leads to a 4x4x8=128 element vector

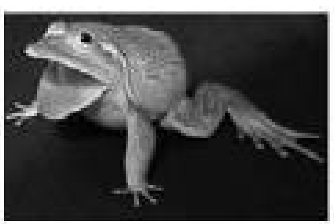
About matching…

- Can be done with as few as 3 features.
- Use Hough transform to cluster features in pose space
- Have to use broad bins since 4 items but 6 dof
    - Match to 2 closest bins
- After Hough finds clusters with 3 entries

Play with Features!



Playing with our Features: Where's Traino and Froggy?



Here's Traino and Froggy!

Credits

- Lowe, D. “Distinctive image features from scale-invariant keypoints” International Journal of Computer Vision, 60, 2 (2004), pp. 91-110
- Pele, Ofir. SIFT: Scale Invariant Feature Transform.Sift.ppt
- Lee, David. Object Recognition from Local Scale-Invariant Features (SIFT). O319.Sift.ppt
- Some Slide Information taken from Silvio Savarese