Distinctive Image Features from Scale-Invariant Keypoints (SIFT)

Mehmet Aygun, November 2014
Purpose

-We want to Recognize
  -For recognizing, need features

-Extract features from image
  -Edge detection (Sobel, Canny), Corner detection (Harris)...

-False features due to bad illuminations, different scales, or rotation

-This paper focus extracting distinctive invariant features
  -Invariant to: image scale and rotation, addition of noise or change in illumination
Steps

- Scale-space extrema detection
  - Difference of Gaussian over all scales and all locations

- Keypoint localization

- Orientation assignment
  - Image gradient directions

- Keypoint descriptor
Properties

-Near real-time

-For 500X500 pixel about 2000 stable features

-Key point descriptors are highly distinctive, this allow a feature enough to find correct match
Detection of scale-space extrema

- Identify locations and scales

\[-L(x,y,\sigma) = G(x,y,\sigma) * I(x,y) \quad \ast: \text{convolution}\]

- \(L(x,y,\sigma)\): scale space of an image
- \(G(x,y,\sigma)\): Gaussian
- \(I(x,y)\): Image

\[-G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}} \quad \text{Gaussian}\]
Detection of scale-space extrema

-Difference-Gaussian

- \( D(x,y,\sigma) = L(x,y,k\sigma) - L(x,y,\sigma) \)

\[ k = 2^{1/s} \text{ (generally } s = 2, k = \sqrt{2} \) \]

We divide each octave by \( s+3 \)
Each octave image = image/2
Detection of scale-space extrema

-Local extrema detection

-4 Scales for per octave (best result)
-Take 3 difference-of-gaussian.

-Compare 8 neighbour and 9 pixels at the top scales and at the bottom 9 pixels.

-If it is higher or less then all of the pixels, it is local extrema point.
Keypoint localization

- Using taylor series DoG can be expand

\[ D(x) = D + \frac{\partial D}{\partial x} x + \frac{1}{2} x^T \frac{\partial^2 D}{\partial^2 x^2} x \]

- \( x = (x, y, \sigma)^T \)

- Extrema can be found with take derivative and set to zero.

\[ \bar{x} = -\frac{\partial^2 D^{-1}}{\partial x^2} \frac{\partial D}{\partial x} \]

\[ D(\bar{x}) = D + \frac{1}{2} \frac{\partial D^T}{\partial x} \bar{x} \]

- If a extrema with \( |D(\bar{x})| \) less than 0.03 were discarded

- Range of image \([0, 1]\)

- With this operation, unstable extrema with low contrast can be eliminate

  in an 233X197 image 832 ->729 points
Keypoint localization

- Eliminating edge responses

- \( H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \)  \( H: \) Hessian Matrix

- \( \text{Tr}(H) = D_{xx} + D_{yy} = \lambda_1 + \lambda_2 \)  \( \lambda_1, \lambda_2 : \) eigenvalues

- \( \text{Det}(H) = D_{xx}D_{yy} - (D_{xy})^2 = \lambda_1 \lambda_2 \)

- \( \frac{\text{Tr}(H)^2}{\text{Det}(H)} = \frac{(r+1)^2}{r} \)  \( \lambda_1 = r \lambda_2 \)

If \( r > 10 \) eliminate this point

- 729->536 points
Orientation assignment

- Compute gradient magnitude and orientation using:
  
  \[ m(x,y) = \sqrt{(L(x + 1, y) - L(x - 1, y))^2 + (L(x, y + 1) - L(x, y - 1))^2} \]

- \[ \theta(x,y) = \tan^{-1} \left( \frac{(L(x, y + 1) - L(x, y - 1))}{(L(x + 1, y) - L(x - 1, y))} \right) \]

- The orientation histogram has 36 bins

- Highest peak selected from histogram

  - if other high nearly %80 of highest create multiple orientation
Image Descriptor

-Last step; Histogram to Descriptor
Image Descriptor

- For a key point we have 4X4X8 features vector

- Each vector can be in 8 direction with different magnitude

- Before the transformation we should apply Gaussian to Image Gradient Histogram
References

Distinctive Image Features from Scale-Invariant Keypoints by David G. Lowe taken from: