



Đặc trưng SIFT

NHẬP MÔN THỊ GIÁC MÁY TÍNH

Trình bày: TS Trần Thái Sơn; Email: ttson@fit.hcmus.edu.vn

Các đặc tính cần có của đặc trưng

- Các đặc điểm:
 - Đặc trưng mạnh.
 - Bất biến với các sự biến đổi (affine và phép chiếu / che khuất).
 - Giải quyết được cái vấn đề:
 - Xác định được vị trí đối tượng trong nhiều ảnh (ví dụ trong video).
 - Theo vết được vị trí đối tượng, suy diễn ra chuyển động.



Scale Invariant Feature Transform (SIFT)

- Chọn một đặc trưng bất biến với các phép thu phóng và xoay hình ảnh.
- Bất biến một phần với thay đổi chiếu sáng và thay đổi góc nhìn 3D.
- Dễ dàng định vị trong miền không gian và tần số.
 - Ổn định với nhiễu, hỗn loạn và che khuất.
- Đặc trưng có tính phân biệt cao, được tìm thấy với xác suất cao.



Động lực

- Các phương pháp trước:
 - Phương pháp tìm đặc trưng góc Harris:
 - Nhạy với thay đổi trong thay đổi tỉ lệ kích cỡ ảnh.
 - Tìm vị trí trong ảnh dùng tất cả thông tin đạo hàm ảnh tại mọi hướng.
 - Không có phương pháp nào bất biến hoàn toàn với biến đổi affine.
 - Mặc dù phương pháp SIFT không bất biến hoàn toàn nhưng cũng bất biến đáng kể với biến đổi affine.
 - SIFT cũng bất biến với thay đổi góc nhìn 3D.



Thuật toán SIFT

1. Xây dựng không gian ảnh nhiều tỉ lệ.
2. Xấp xỉ Laplacian của Gaussian (đạo hàm bậc 2 ảnh).
 - Hàm DOG (Difference-of-Gaussian).
3. Xác định vị trí các điểm trọng yếu (keypoint).
 - Xác định ngưỡng để đảm bảo tính ổn định của keypoint.
4. Gán hướng cho điểm trọng yếu.
5. Xây dựng đặc trưng SIFT.



Chọn điểm trọng yếu

- Mục đích xây dựng đặc trưng có tính ổn định với tất cả các thay đổi về tỉ lệ ảnh.
 - Sử dụng hàm liên tục của tỉ lệ ảnh (scale space).
- Hàm nhân (kernel) được sử dụng cho không gian tỉ lệ:
 - Hàm Gaussian.



Không gian tỉ lệ ảnh

- Scale-Space của ảnh được tạo ra từ một dãy tỉ lệ của σ cho $G(x,y,k\sigma)$ trong một octave. Để có được octave thứ nhất ta tăng kích cỡ ảnh lên gấp đôi.
- Octave tiếp theo thì kích cỡ ảnh bằng $\frac{1}{2}$ octave trước.
- Một ảnh L trong không gian tỉ lệ ảnh:

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

$G(x, y, k\sigma)$ Variable-scale Gaussian

$I(x, y)$ Input image

- Để tìm những điểm keypoint ổn định, ta tìm điểm cực đại trong tập các hình trong scale-space sau khi qua hàm difference-of-Gaussian.

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

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



Difference of Gaussian

- Hiệu quả trong tính toán.
 - Ảnh làm trơn L cần ở bước sau nên D có thể tính đơn giản bằng phép trừ.
- Xấp xỉ gần đúng của không gian tỉ lệ chuẩn hóa đạo hàm (scale-normalized Laplacian of Gaussian).
 - Yêu cầu tỉ lệ bất biến chính xác.



Scale-normalized Laplacian of Gaussian

- Tìm sự liên quan giữa D và $\sigma^2 \nabla^2 G$
- Đạo hàm từng phần: $\frac{\partial G}{\partial \sigma} = \sigma \nabla^2 G$
- Tính xấp xỉ cho $\frac{\partial G}{\partial \sigma} \Rightarrow \sigma \nabla^2 G = \frac{\partial G}{\partial \sigma} \approx \frac{G(x, y, k\sigma) - G(x, y, \sigma)}{k\sigma - \sigma}$
- $G(x, y, k\sigma) - G(x, y, \sigma) \approx (k - 1)\sigma^2 \nabla^2 G$
- Phần $(k-1)$ của công thức:
 - Là hằng số, không ảnh hưởng đến vị trí của keypoint.
 - Lỗi xấp xỉ tiến dần đến 0 nếu $k \rightarrow 1$, và không ảnh hưởng đến sự ổn định của keypoint.

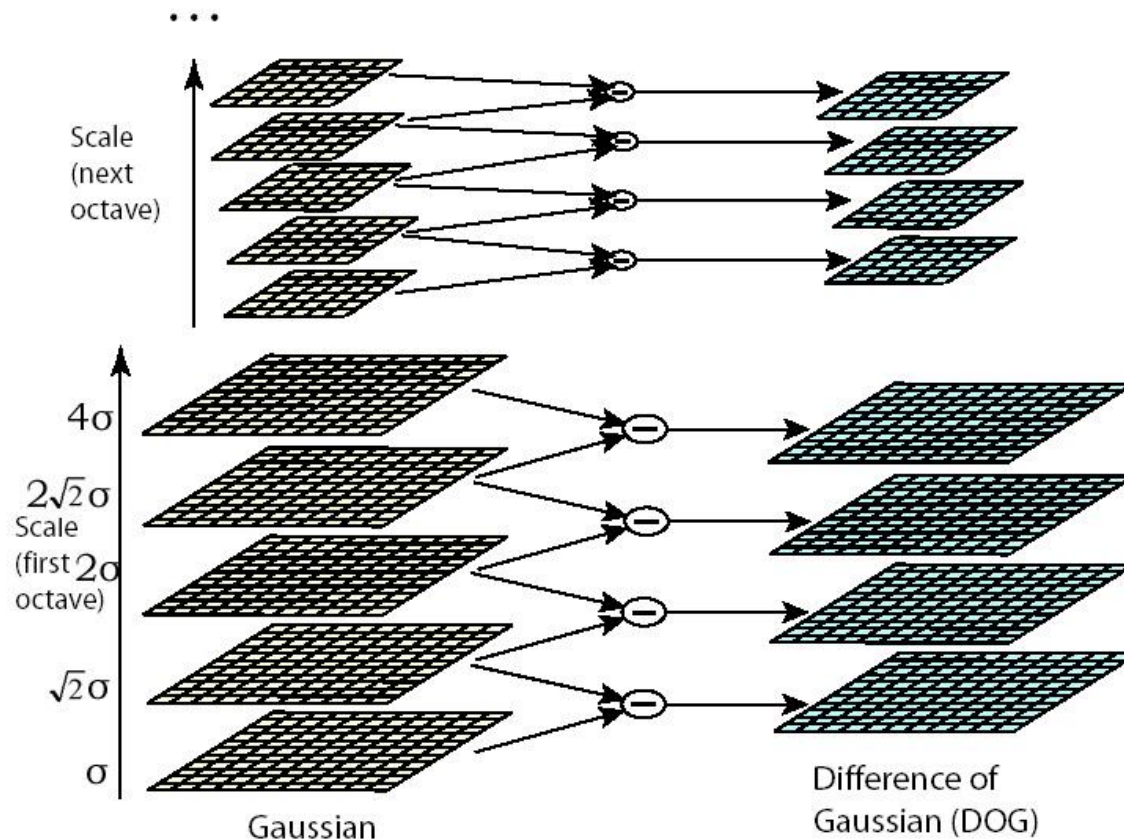


Lựa chọn điểm trọng yếu

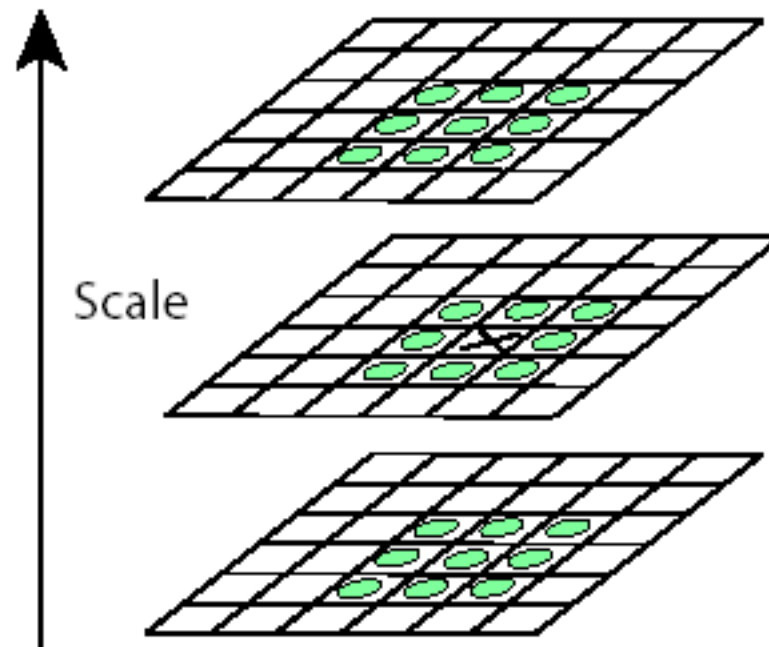
1. Supersample ảnh gốc.
2. Làm trơn ảnh trên nhiều tỉ lệ của σ trong 1 octave.
3. Tính ảnh DOG từ 2 ảnh tỉ lệ kề nhau trong 1 octave.
4. Subsample ảnh trên tỉ lệ 2σ trên octave đang xét và lặp lại việc này (2-3) cho octave kế tiếp.
5. Những điểm trọng yếu tách biệt sẽ được tìm ra bằng cách tìm cực đại trên ảnh DOG so với các điểm ảnh lân cận.



Ảnh ví dụ



Tìm kiếm keypoint



Nếu X là một keypoint nếu nó có giá trị lớn nhất hay bé nhất trong tất cả 26 điểm lân cận trong mỗi octave.



Tìm kiếm vị trí keypoint

- Từ tìm kiếm vị trí cực đại cục bộ của DOG, ta xấp xỉ được giá trị của keypoint.
- Nếu các xấp xỉ đó được sử dụng trực tiếp, số lượng keypoint lấy được sẽ rất lớn trong một ảnh.



Khai triển Taylor

- Khai triển Taylor cho hàm scale-space $D(x, y, \sigma)$.

- Khai triển đến bậc 2:

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

- Với $x = (x, y, \sigma)^T$

- Tính đạo hàm và gán bằng 0 để tìm vị trí điểm cực đại giá trị hàm số.

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



Đặt ngưỡng keypoint

- Giá trị hàm số ở điểm cực đại được sử dụng để loại bỏ các điểm cực đại không ổn định (tương phản thấp).
- Tính giá trị:

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

- Ngưỡng để loại bỏ điểm cực đại:
 - Giá trị tuyệt đối của D nếu nhỏ hơn 0.03.



Đặt ngưỡng keypoint

- Hàm DOG (Difference-of-Gaussian) là một đặc trưng mạnh dọc theo các cạnh.
- Một số vị trí ở trên biên cạnh sẽ khó xác định và sẽ không ổn định nếu có một nhiễu được thêm vào.
- Những vị trí này có hướng thay đổi giá trị lớn (principal curvatures) theo cạnh nhưng lại nhỏ ở hướng vuông góc với cạnh.
- Nên ta cần tính giá trị thay đổi tại vị trí đó tại các hướng và so sánh cả 2.



Tính Principal Curvatures

- Ma trận Hessian

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

- Trị riêng của H là thành phần của principal curvatures.

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

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

r là tỉ lệ giữa 2 trị riêng của H.

- Ta không quan đến giá trị của từng trị riêng mà chỉ quan tâm đến tỉ lệ giữ chúng.

$$\frac{Tr(\mathbf{H})^2}{Det(\mathbf{H})} = \frac{(\alpha + \beta)^2}{\alpha\beta} = \frac{(r\beta + \beta)^2}{r\beta^2} = \frac{(r + 1)^2}{r}$$

Thực tế, ta đặt $r = 10$ để loại trừ các keypoint nếu:

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



Ví dụ

Keypoint detection

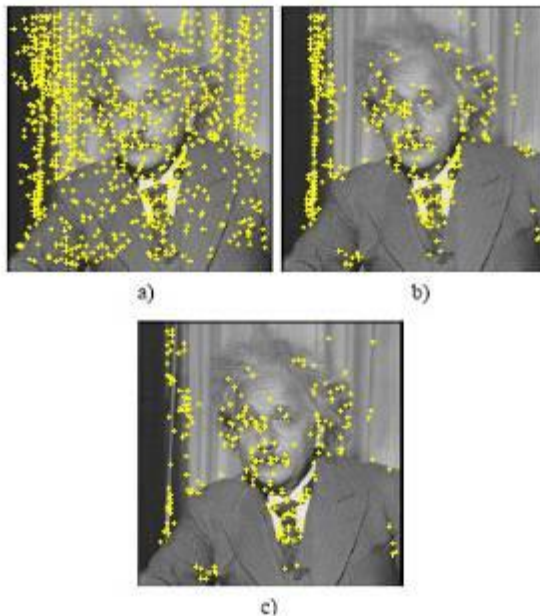


Figure 5: a) Maxima of DoG across scales. b) Remaining keypoints after removal of low contrast points. C) Remaining keypoints after removal of edge responses (bottom).

Final keypoints with selected orientation and scale

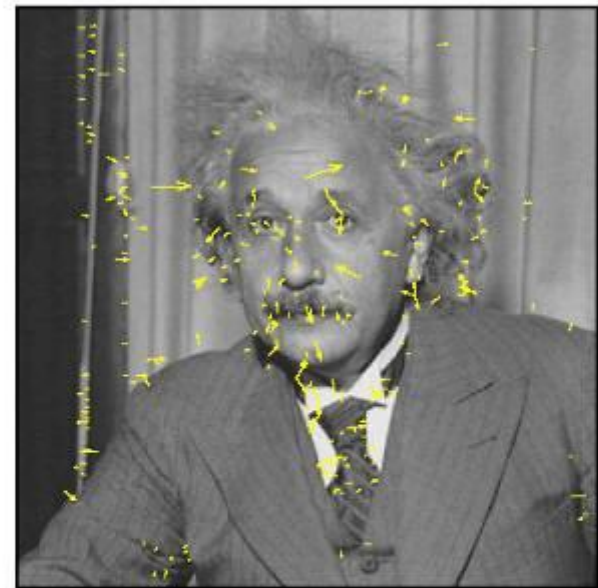


Figure 6: Extracted keypoints, arrows indicate scale and orientation.

Gán hướng cho đặc trưng SIFT

- Một khi ta tìm được tập các keypoint để biểu diễn lại ảnh.
- Bước tiếp là gán hướng cho mỗi keypoint.
 - Dùng cho các kỹ thuật so khớp mà bất biến với phép xoay.
 - Dùng cho xác định ảnh test có nằm trong tập train hay không.



Gán hướng cho đặc trưng SIFT

- Ảnh đã được làm trơn từ Gaussian L, với tỉ lệ được chọn là bất biến (scale invariance).
- Các điểm trong vùng xung quanh keypoint được chọn và tính giá trị và hướng đạo hàm các điểm này.

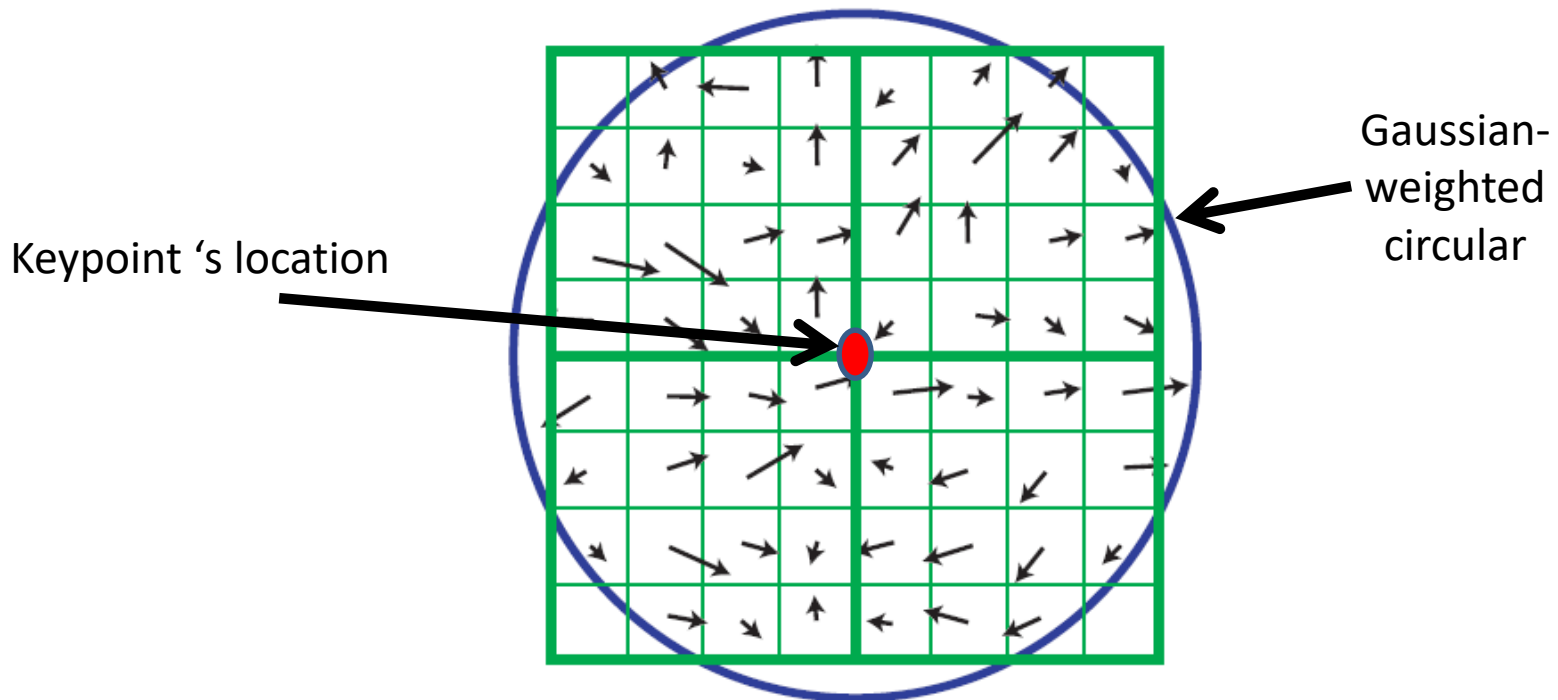
$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2}$$

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



Gán hướng cho đặc trưng SIFT

- Biểu đồ hướng được chia thành 36 hướng chính. Mỗi điểm với hướng đạo hàm của nó được thêm vào vị trí hướng tương ứng trên biểu đồ và giá trị thêm vào là độ lớn đạo hàm nhân với giá trị của số tròn Gaussian với $\sigma = 1.5$ lần scale của keypoint.



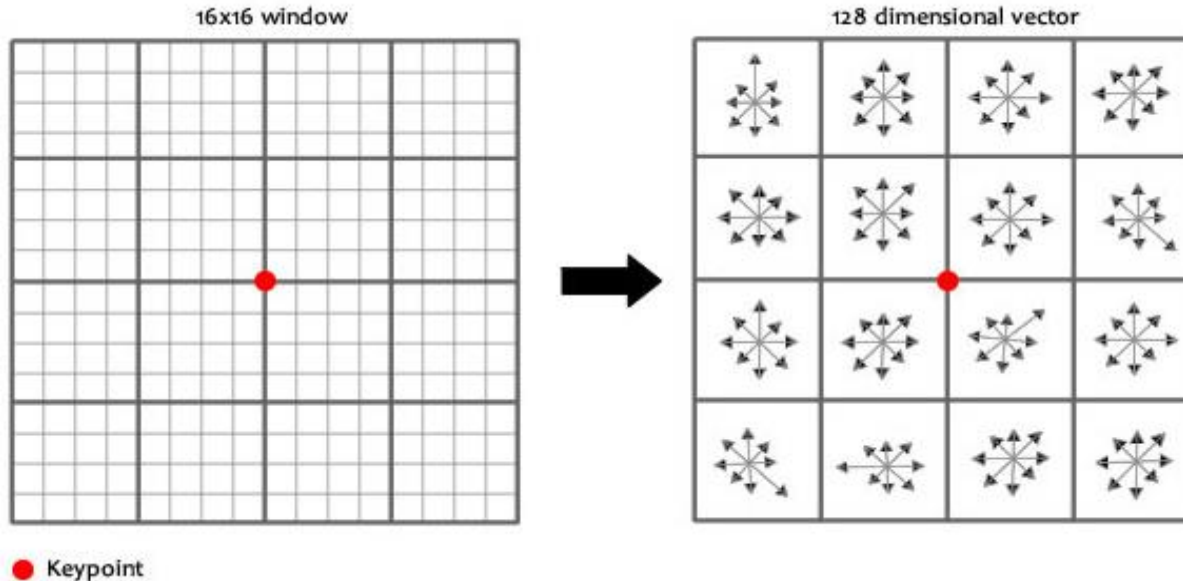
Gán hướng cho đặc trưng SIFT

- Giá trị hướng chủ yếu tại một điểm là hướng được tìm thấy với tỉ lệ 80% so với tất cả các hướng khác.
- 3 giá trị trên biểu đồ hướng gần nhất với hướng chủ yếu được sử dụng để nội suy ra hướng chủ yếu chính xác hơn.



Gán hướng cho đặc trưng SIFT

- Với mỗi keypoint có giá trị là 4 chiều: vị trí chiều x, chiều y, tỉ lệ (scale), và hướng.



Mảng 4x4 từ cửa sổ 16 x 16 với 8 hướng = vector đặc trưng 128 chiều



Miêu tả keypoint

- Mục đích để bất biến với các thay đổi về chiếu sáng và góc nhìn 3D trên ảnh.
- Được tính cho vùng xung quanh điểm trọng yếu (keypoint) , không tính trực tiếp trên keypoint.
- Như trước thì độ lớn và hướng đạo hàm được tính cho tập các điểm trong vùng xung quanh keypoint sử dụng L của scale gần nhất.
- Để chắc chắn tính bất biến về hướng, hướng đạo hàm và tọa độ hình học của miêu tả (descriptor) được xoay để tương ứng với hướng của điểm trọng yếu.



Miêu tả keypoint

- Độ lớn đạo hàm của mỗi điểm là độ lớn với σ của $1/2$ chiều rộng của cửa sổ miêu tả.
 - Tránh được việc thay đổi đột ngột trong miêu tả đặc trưng nếu có sự thay đổi nhỏ về vị trí cửa sổ.
 - Đặt trọng số nhỏ cho các giá trị đạo hàm ảnh ở các điểm xa keypoint.
- Vùng ảnh được chia 4×4 vùng con xung quanh keypoint. Cho phép tồn tại sự thay đổi trong 4 vị trí nhưng vẫn thuộc cùng một histogram.



Miêu tả keypoint

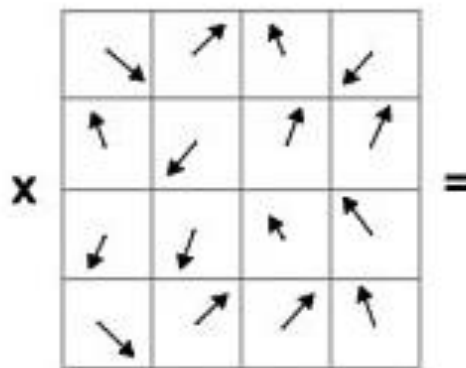
- Để tránh hiệu ứng lề trong histogram.
 - Nội suy tam tuyến (trilinear) để phân bố đều độ ảnh hưởng của giá trị đạo hàm của các pixel đến các bin kề nhau trong histogram.
 - Trọng số từ 1-d, trong đó d là khoảng cách của 1 mẫu xác định đến bin trung tâm.
- Chuẩn hóa vector:
 - Được thực hiện ở bước cuối để chắc chắn bất biến với thay đổi độ sáng (biến đổi affine).
 - Toàn bộ vector được chuẩn hóa thành chuẩn 1.
 - Để đối phó với trường hợp thay đổi độ sáng phi tuyến, vector đặc trưng được đặt ngưỡng không lớn hơn 0.2 và vector được chuẩn hóa lại.



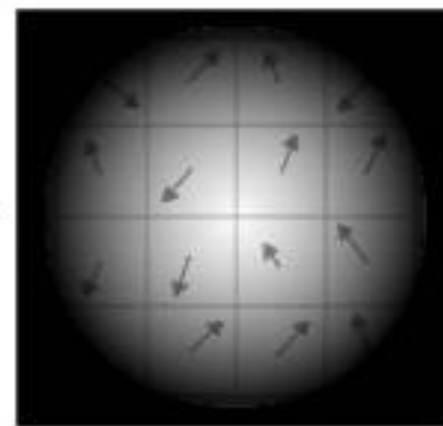
Miêu tả keypoint có trọng số



Normal distribution



Keypoint descriptor



Weighted keypoint descriptor



Độ phức tạp

- Kích cỡ vector = rn^2 .
- Kích cỡ vector tăng theo:
 - Để tăng khả năng phân biệt đối tượng trong cơ sở dữ liệu lớn.
 - Để tăng độ nhạy với sự méo hình từ các góc nhìn 3D và che khuất.
- Slide này trình bày đặc trưng với 4x4 mảng và 8 hướng cho ra vector đặc trưng 128 chiều.



Đặc trưng SIFT

- Ta có được:
 - Tìm được một xấp xỉ cho đặc trưng bởi sử dụng DOG (difference-of-Gaussian).
 - Xác định vị trí keypoint chính xác hơn.
 - Đặt ngưỡng các keypoint ít giá trị.
 - Xác định hướng của keypoint.
 - Tính toán được đặc trưng là 1 vector 128 chiều cho mỗi keypoint(keypoint descriptor).



So khớp keypoints giữa 2 ảnh

- Có 2 ảnh hay nhiều:
 - Một ảnh được sử dụng để học đối tượng ta chú ý.
 - Những ảnh khác là ảnh thực tế có thể chứa đối tượng ta quan tâm ở ảnh đã học.
- Cả 2 ảnh đều chứa đặc trưng liên quan nhau nhưng cách nào để tìm đặc trưng phù hợp với nhau.

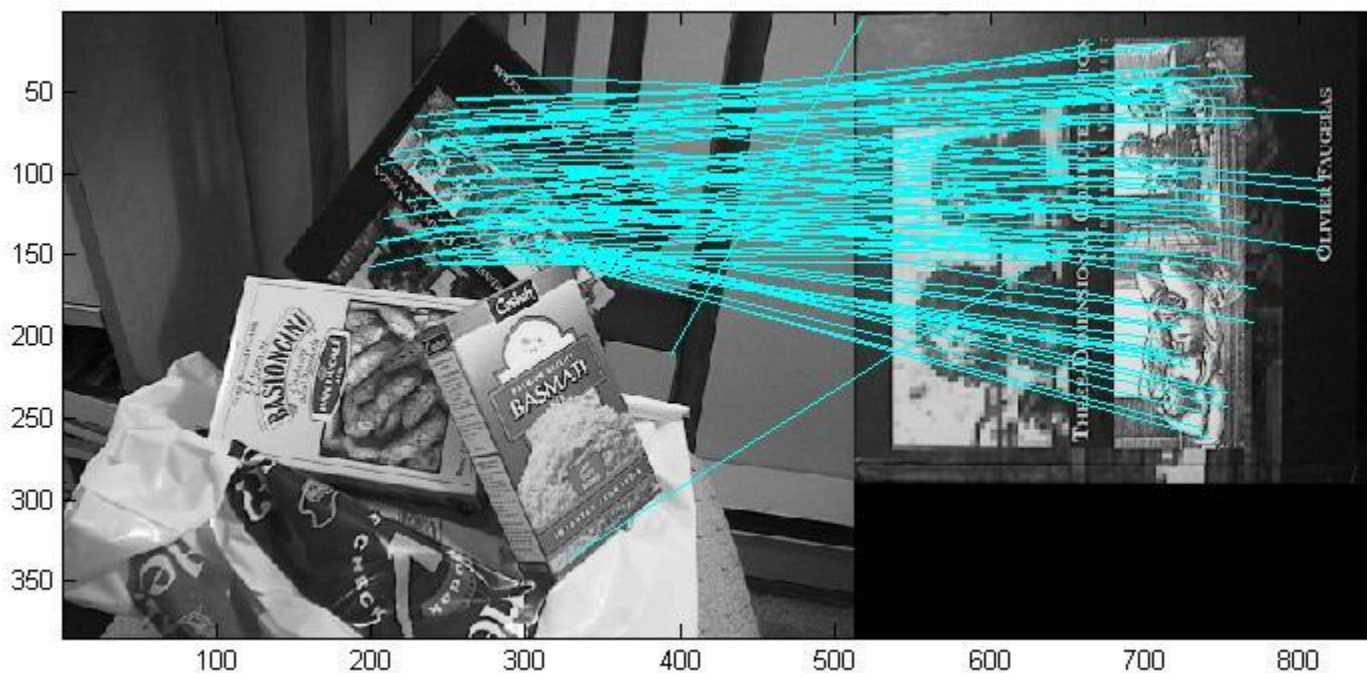


So khớp keypoints giữa 2 ảnh

- Nearest Neighbor Algorithm:
 - Độc lập tìm tất cả các keypoint liên quan đến nhau trong toàn bộ các octave trong 1 ảnh và các ảnh khác.
 - Để giải quyết vấn đề không điểm nào so khớp ở ảnh khác (nhiều nền ...).
 - Khoảng cách lớn nhất (hiệu suất không cao).
 - Tỷ lệ của điểm gần nhất với điểm gần thứ hai.



Ví dụ

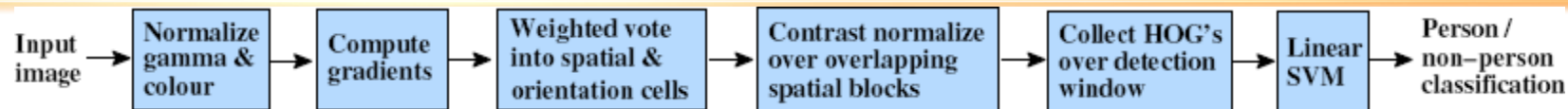


Tham khảo

- David G. Lowe, "**Distinctive image features from scale-invariant keypoints**," *International Journal of Computer Vision*, 60, 2 (2004), pp. 91-110.



Đặc trưng HOG



$$\begin{bmatrix} -1 & 0 & 1 \end{bmatrix}$$

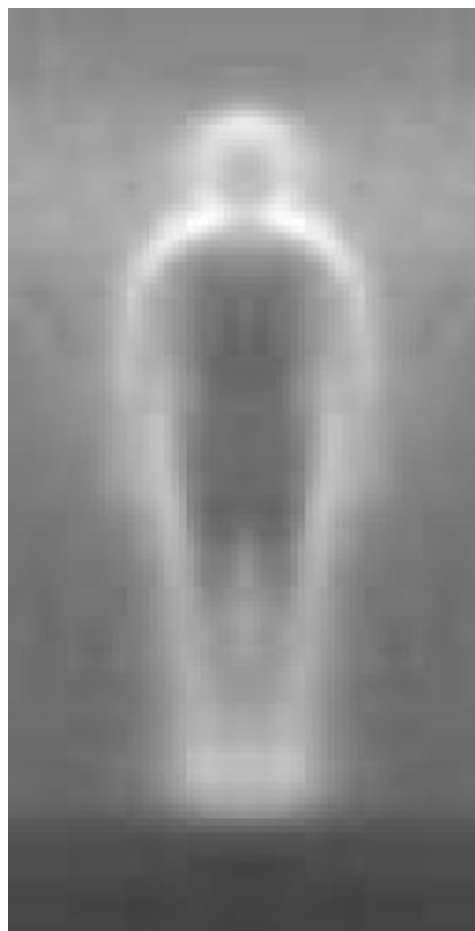
centered

$$\begin{bmatrix} -1 & 1 \end{bmatrix}$$

uncentered

$$\begin{bmatrix} 1 & -8 & 0 & 8 & -1 \end{bmatrix}$$

cubic-corrected



$$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

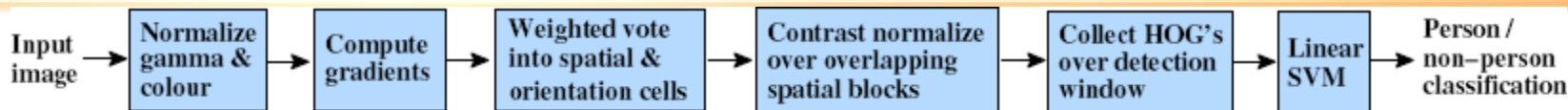
diagonal

$$\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$

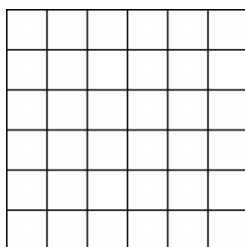
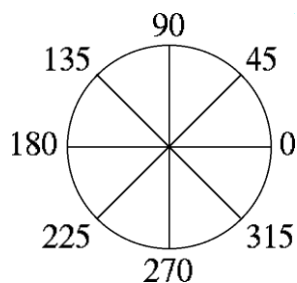
Sobel



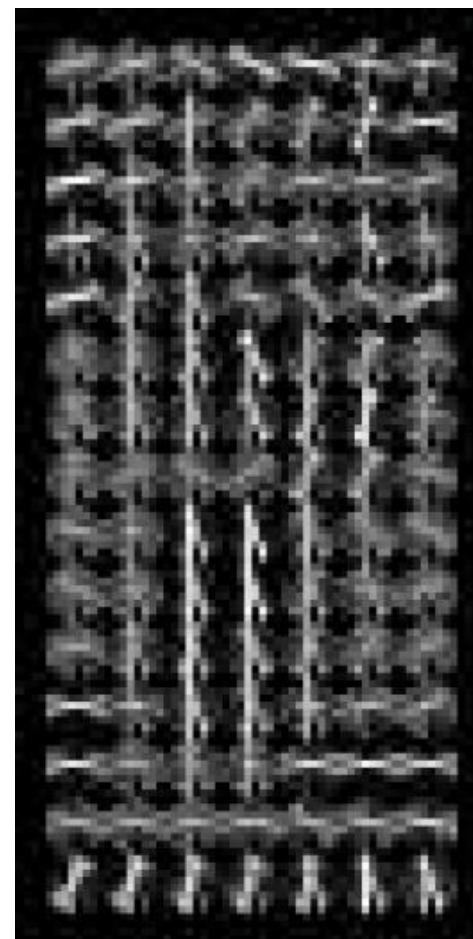
Đặc trưng HOG



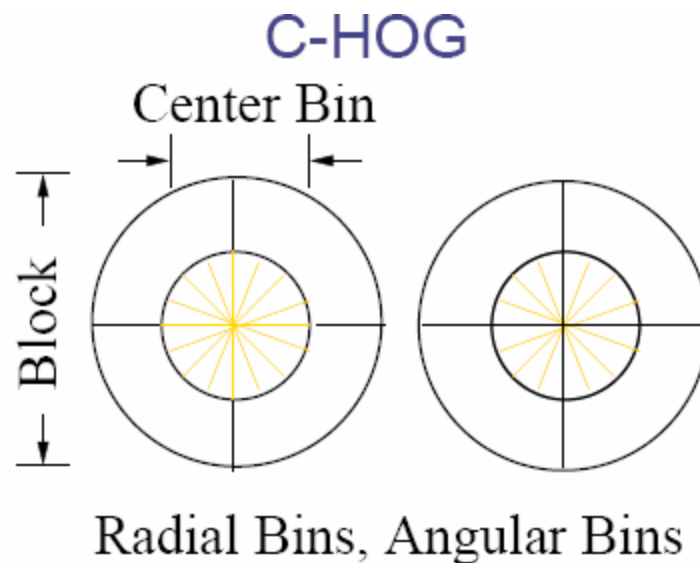
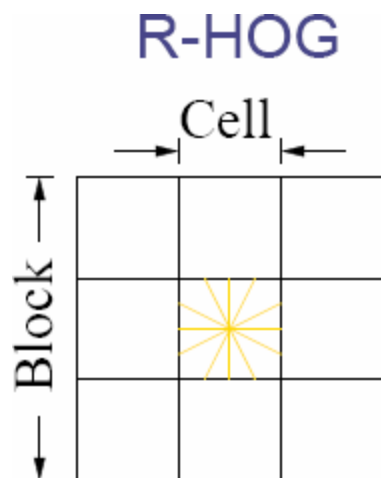
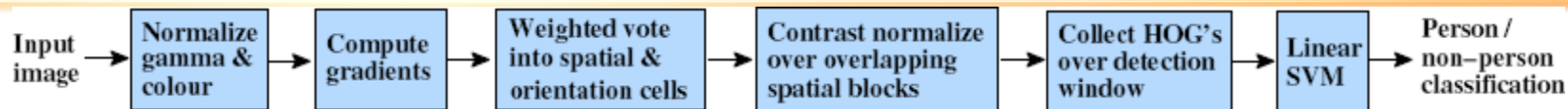
- Histogram of gradient orientations
 - Hướng
 - Vị trí



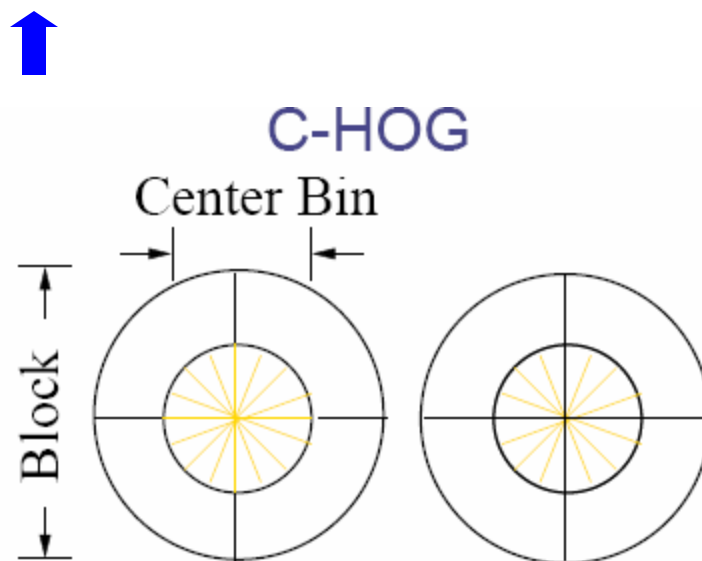
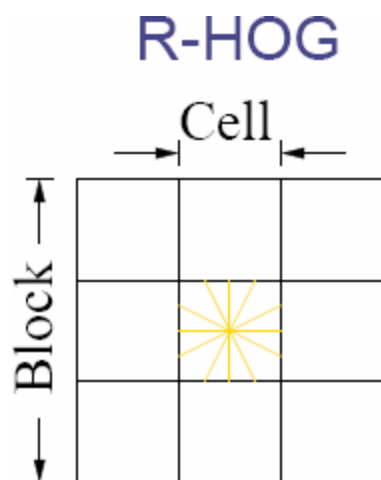
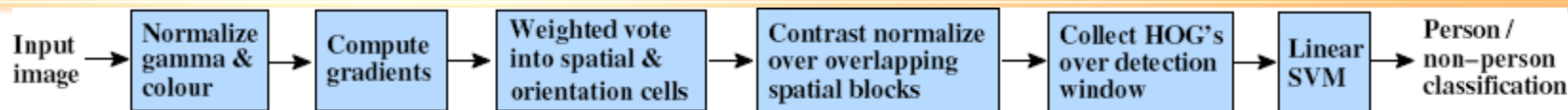
– Weighted by magnitude



Đặc trưng HOG



Đặc trưng HOG



Radial Bins, Angular Bins

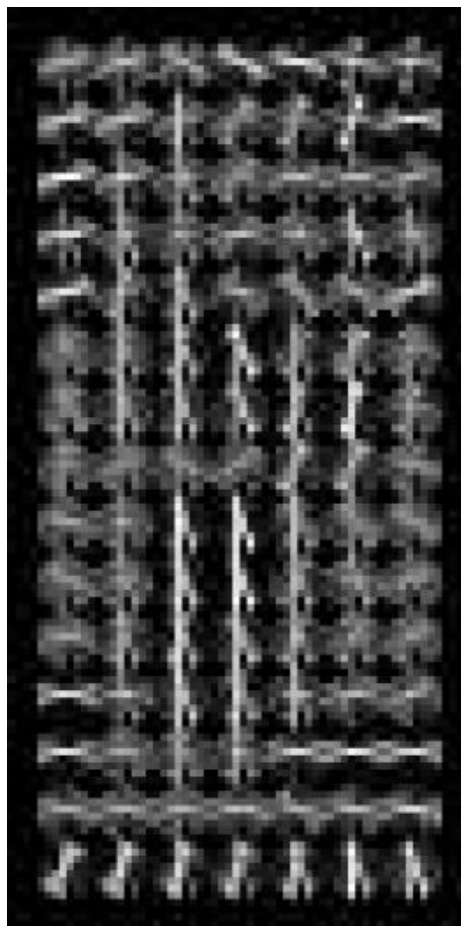
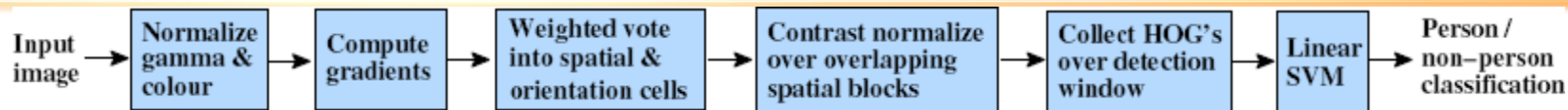
$$L1 - norm : v \longrightarrow v / (||v||_1 + \epsilon)$$

$$L1 - sqrt : v \longrightarrow \sqrt{v / (||v||_1 + \epsilon)}$$

$$L2 - norm : v \longrightarrow v / \sqrt{||v||_2^2 + \epsilon^2}$$

$$L2 - hys : L2\text{-norm, plus clipping at .2 and renormalizing}$$

Đặc trưng HOG



Tham khảo

- N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," CVPR 05, pp. 886-893.

