



# Thị giác máy tính

Ch4.2: Trích chọn đặc trưng và so khớp ảnh

Nguyễn Thị Oanh  
oanhnt@soict.hust.edu.vn

1

## Nội dung

- Trích chọn đặc trưng
  - Global features
  - Local features
- So khớp ảnh
- Ứng dụng



## Đặc trưng toàn cục và đặc trưng cục bộ

- Hai loại đặc trưng được trích chọn từ ảnh:
  - Đặc trưng cục bộ và toàn cục
- **Đặc trưng toàn cục:**
  - Mô tả toàn bộ ảnh như 1 đối tượng
  - Đặc trưng đường biên, đặc trưng hình dạng, đặc trưng kết cấu
    - Ví dụ: Invariant Moments (Hu, Zernike), Histogram Oriented Gradients (HOG), PHOG, and Co-HOG,...
- **Đặc trưng cục bộ:**
  - Mô tả đặc trưng cục bộ mô tả từng vùng nhỏ trong ảnh, từng vùng cục bộ của đối tượng (điểm đặc trưng trong ảnh).
  - Biểu diễn đặc trưng kết cấu/màu sắc trong mỗi vùng cục bộ ảnh
    - Ví dụ: SIFT, SURF, LBP, BRISK, MSER và FREAK, ...

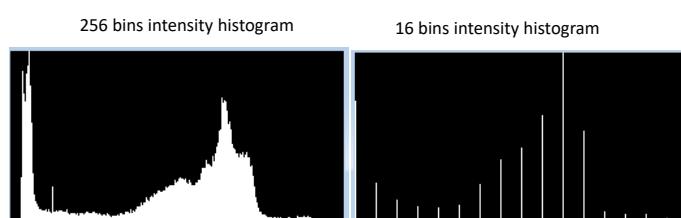
## Đặc trưng toàn cục

How to distinguish these objects?



## Đặc trưng màu

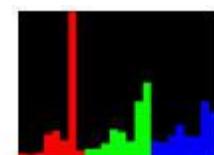
- Histogram



Image



Histogram



Histobin(8,8,8)

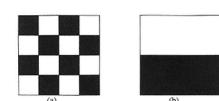
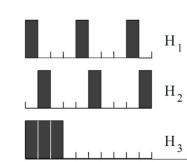


SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

5

## Histogram

- Đơn giản
- Bất biến đổi với các phép biến đổi hình học cơ bản:
  - quay, dịch, zoom
- Không tính đến sự **tương đồng giữa các bin gần nhau**
- Phân bố trên không gian ảnh không được tính đến
- Ảnh hưởng bởi yếu tố **nền**



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

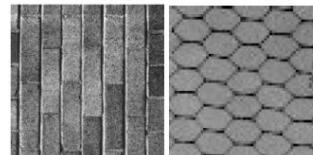
7

## Đặc trưng kết cấu

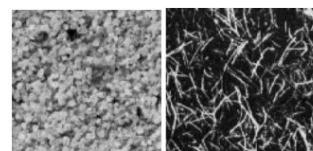
### • Kết cấu:

- Vùng có sự biến thiên về cường độ sáng
- Có tính đến sự sắp đặt về cường độ sáng trên bề mặt

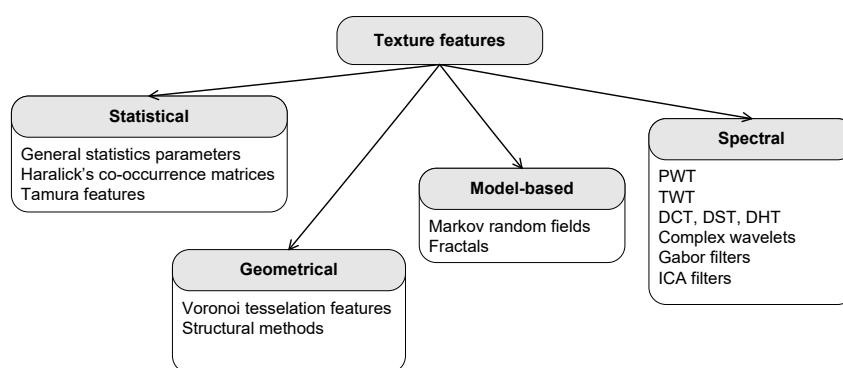
A texture can be **periodic** (a pattern that repeats itself) ...



...or **non-periodic** (no pattern, more disorganized)

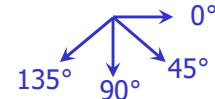


## Đặc trưng kết cấu



## GLCM (Grey Level Co-occurrence Matrices)

- Ý tưởng:
  - xác định mức lặp của cặp mức xám theo **hướng** và ở **1 khoảng cách** nào đó
    - Co-occurrence matrices (Haralick)
- Ma trận:  $Ng \times Ng$** 
  - $Ng$  : số mức xám của ảnh (256x256)
  - Số mức xám thường được giảm xuống: 8x8, 16x16 or 32x32
- Nhiều ma trận, mỗi ma trận cho 1 hướng và 1 khoảng cách
  - Distance** : 1, 2, 3 (4, ...)
  - Direction** :  $0^\circ, 45^\circ, 90^\circ, 135^\circ, \dots$
- Thời gian xử lý lâu



<http://haralick.org/journals/TexturalFeatures.pdf>

10



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

## GLCM

$$CM_{d,\beta}(c_i, c_j) = \frac{\text{card}(\{p_1, p_2 | I(p_1) = c_i, I(p_2) = c_j, N_{d,\beta}(p_1, p_2) = \text{true}\})}{\text{card}(\{p_1, p_2 | N_{d,\beta}(p_1, p_2) = \text{true}\})}$$

$N_{d,\beta}(p_1, p_2) = \text{true}$        $p_2$  is a neighbor of  $p_1$  at a distance  $d$  direction  $\beta$

1	4	4	3
4	2	3	2
1	2	1	4
1	2	2	3

Image

	1	2	3	4
1	0	2	0	2
2	1	1	2	0
3	0	1	0	0
4	0	1	1	1

Matrix for distance=1  
and direction= $0^\circ$

	1	2	3	4
1	0	2	1	0
2	1	1	0	0
3	0	0	0	1
4	0	2	1	0

Matrix for distance=1  
and direction= $45^\circ$



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

11

## GLCM

- Các tham số phổ biến/quan trọng được ước lượng từ GLCM:

$$Energy = \sum_i \sum_j CM_d^2(i, j) \quad \text{minimal when all elements are equal}$$

$$entropy = -\sum_i \sum_j CM_d(i, j) \log(CM_d(i, j)) \quad \begin{array}{l} \text{a measure of chaos,} \\ \text{maximal when all elements are equal} \end{array}$$

$$contrast = \sum_i \sum_j (i - j)^2 CM_d(i, j) \quad \begin{array}{l} \text{small values when big elements} \\ \text{are near the main diagonal} \end{array}$$

$$idm = \sum_i \sum_j \frac{1}{1 + (i - j)^2} CM_d(i, j) \quad \begin{array}{l} \text{idm (inverse differential moment) has small values} \\ \text{when big elements are far from the main diagonal} \end{array}$$



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

12

## GLCM

- Đặc trưng Haralick :

- Mỗi ma trận GLCM  $\rightarrow$  tính 14 (13) giá trị đặc trưng cho thông tin kết cấu, tham số quan trọng nhất : mean, variance, energy, inertia, entropy, inverse differential moment
- Ref: <http://haralick.org/journals/TexturalFeatures.pdf>

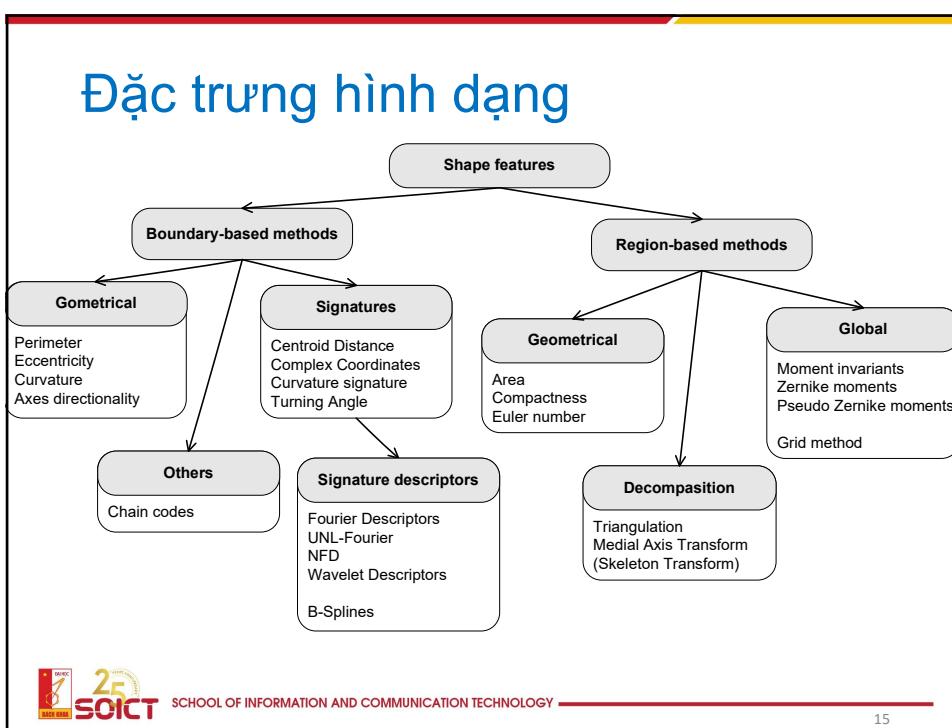
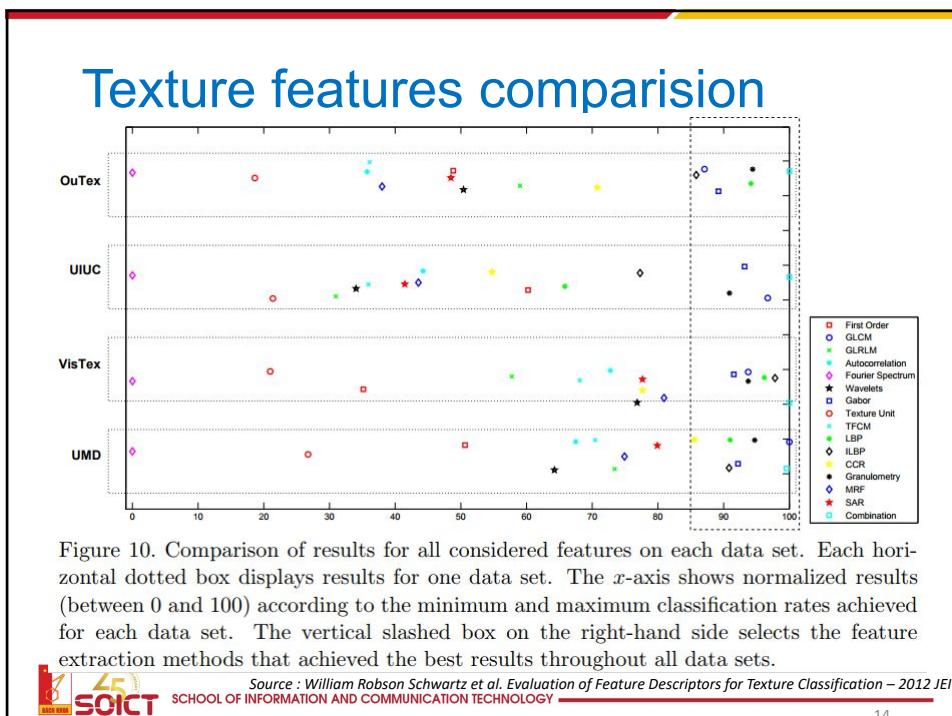
- GLCM:

- Bất biến với phép quay ?
- Bất biến với phép biến đổi tỷ lệ?

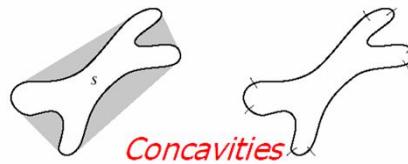
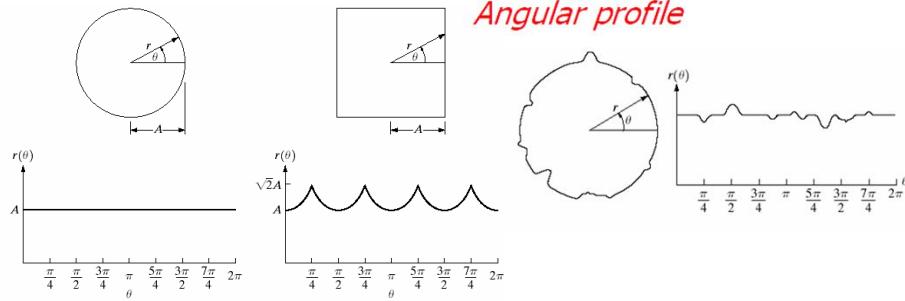


SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

13



## Examples: angular profile, ...



## Examples : Image moments

- Moment  $M_{ij} = \sum_x \sum_y x^i y^j I(x, y)$   
 $M_{0,0} = \text{area of the region } D$   
 $(M_{0,1}, M_{1,0}) = \text{centroid of } D$

- Central moments:

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q I(x, y)$$

$$\bar{x} = \frac{M_{10}}{M_{00}} \quad \bar{y} = \frac{M_{01}}{M_{00}}$$

Invariant to translation

## Invariant moments (Hu's moments)

invariant to  
translation,  
scale, and  
rotation, and  
reflection

$$\begin{aligned}
 I_1 &= \eta_{20} + \eta_{02} \\
 I_2 &= (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \\
 I_3 &= (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \\
 I_4 &= (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \\
 I_5 &= (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} \\
 &\quad - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \\
 I_6 &= (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \\
 I_7 &= (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} \\
 &\quad + \eta_{03})^2] + (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})[3(\eta_{30} \\
 &\quad + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]
 \end{aligned}$$

$$\eta_{ij} = \frac{\mu_{ij}}{\mu_{00}^{(1+\frac{i+j}{2})}}$$

Change for  
image  
reflection



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

18

## Hu's moments

6 images and their Hu Moments

id	Image	H[0]	H[1]	H[2]	H[3]	H[4]	H[5]	H[6]
K0	K	2.78871	6.50638	9.44249	9.84018	-19.593	-13.1205	19.6797
S0	S	2.67431	5.77446	9.90311	11.0016	-21.4722	-14.1102	22.0012
S1	S	2.67431	5.77446	9.90311	11.0016	-21.4722	-14.1102	22.0012
S2	S	2.65884	5.7358	9.66822	10.7427	-20.9914	-13.8694	21.3202
S3	S	2.66083	5.745	9.80616	10.8859	-21.2468	-13.9653	21.8214
S4	2	2.66083	5.745	9.80616	10.8859	-21.2468	-13.9653	-21.8214

<https://www.learnopencv.com/wp-content/uploads/2018/12/HuMoments-Shape-Matching.png>



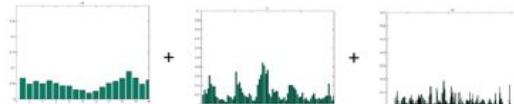
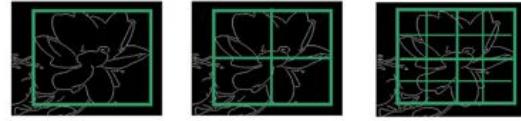
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

19

## PHOG

PHOG:  
Pyramid Histogram of Oriented Gradients

Input Image (image.jpg)



Output PHOG descriptor (image.jpg.txt)



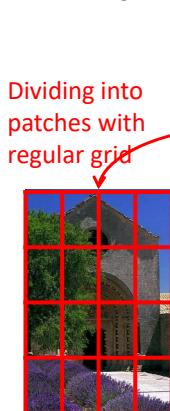
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Source: <http://www.robots.ox.ac.uk/~vgg/research/caltech/phog.html>

20

## Đặc trưng cục bộ

- Đặc trưng cục bộ: xác định vùng cục bộ như thế nào?



Không cần biết nội dung của ảnh



Dựa trên nội dung của ảnh

Keypoint detection



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

21

## Vai trò của đặc trưng cục bộ

- Image matching: a challenging problem



Source: CS131 - Juan Carlos Niebles and Ranjay Krishna

22



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

## Vai trò của đặc trưng cục bộ

- Image matching: a challenging problem



by [Diva Sian](#)



by [swashford](#)



by [scgbt](#)

Slide credit: Steve Seitz



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

23

## Vai trò của đặc trưng cục bộ

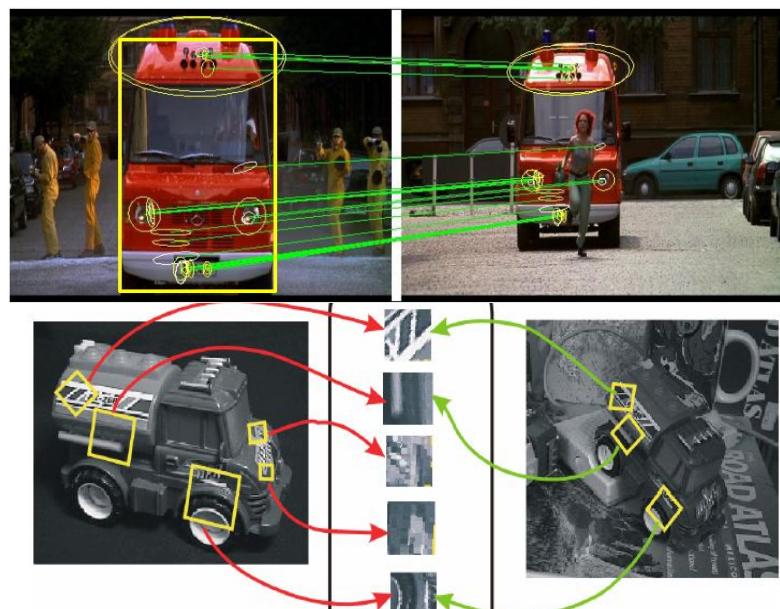
- Partial search
- Object detection/ recognition



Sivic and Zisserman, 2003



D. Lowe 2002



Source : Jim Little, Lowe: features, UBC.

## Yêu cầu quan trọng

- Problem 1:

- Detect the same point *independently* in both images



No chance to match!

We need a repeatable **detector**!

- Problem 2:

- For each point correctly recognize the corresponding one



We need a reliable and distinctive **descriptor**!



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Slide credit: Darya Frolova, Denis Simakov

27

## Yêu cầu

- Region extraction needs to be **repeatable** and **accurate**
  - **Invariant** to translation, rotation, scale changes
  - **Robust** or **covariant** to out-of-plane ( $\approx$ affine) transformations
  - **Robust** to lighting variations, noise, blur, quantization
- **Locality**: Features are local, therefore robust to occlusion and clutter
- **Quantity**: We need a sufficient number of regions to cover the object
- **Distinctiveness**: The regions should contain “interesting” structure
- **Efficiency**: Close to real-time performance



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

28

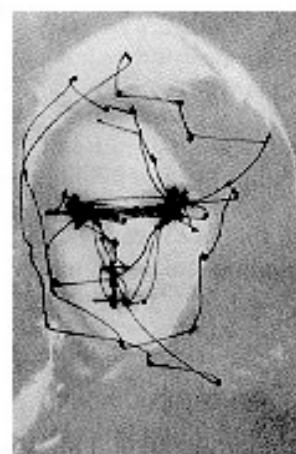
## Phát hiện các điểm đặc trưng



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

29

## Điểm quan trọng nằm ở đâu?



Yarbus eye tracking

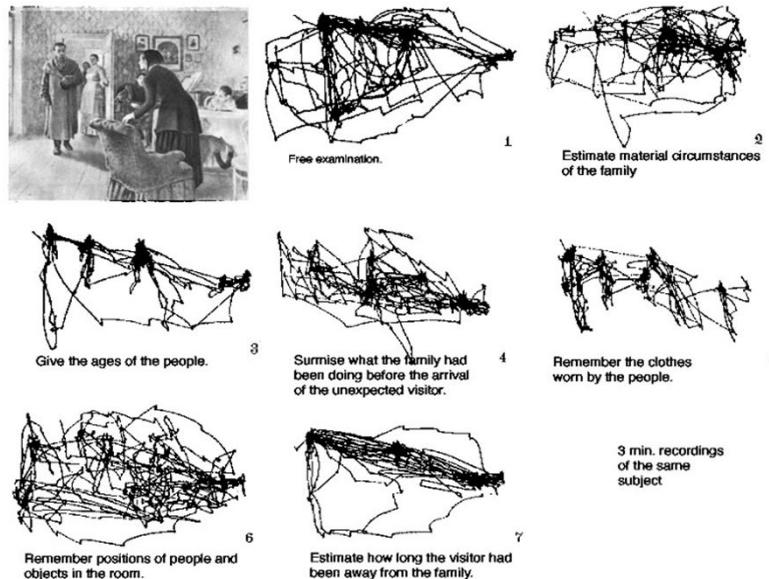
Source : Derek Hoiem, Computer Vision, University of Illinois.



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

30

### Same image with different questions



Change blindness : <http://www.simonslab.com/videos.html>

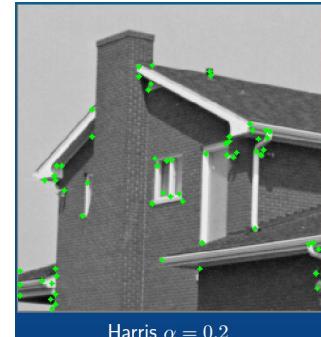
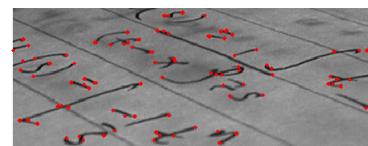


SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

31

## Đặc trưng cục bộ và các điểm đặc trưng

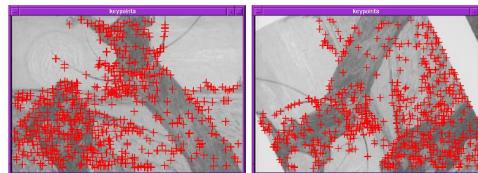
- Phát hiện các điểm đặc trưng (interest point, keypoint, point of interest):
  - DoG /SIFT detector
  - Harris corner detector
  - Moravec
  - ...
- Đặc trưng cục bộ: tính trên các vùng cục bộ xung quanh điểm đặc trưng:
  - SIFT,
  - SURF(Speeded Up Robust Features),
  - PCA-SIFT
  - LBP, BRISK, MSER and FREAK, ...



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

32

## Tìm góc – Harris corner detector



- Ý tưởng chính:
  - Trong vùng xung quan góc  $\rightarrow$  image gradient có ít nhất 2 hướng biến đổi mạnh
- Góc đảm bảo: *repeatable* và *distinctive*

C.Harris and M.Stephens. ["A Combined Corner and Edge Detector."](#) *Proceedings of the 4th Alvey Vision Conference*, 1988.



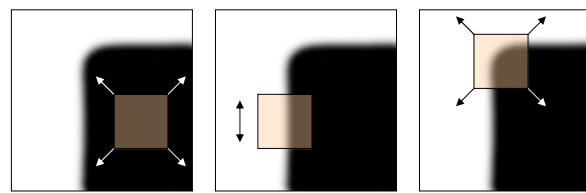
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Slide credit: Svetlana Lazebnik

33

## Góc coi như điểm đặc trưng

- Cách thiết kế
  - Tìm điểm góc = quan sát sự thay đổi trong cửa sổ nhỏ (*locality*)
  - Dịch chuyển cửa sổ *theo mọi hướng* để xác định có vùng có *thay đổi lớn* về cường độ (*good localization*)



**"flat"** region:  
no change in  
all directions

**"edge"**:  
no change along  
the edge  
direction

**"corner"**:  
significant change  
in all directions

Slide credit: Alyosha Efros



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

34

## Góc vs cạnh



$$\begin{aligned}\sum I_x^2 &\rightarrow \text{Large} \\ \sum I_y^2 &\rightarrow \text{Large}\end{aligned}$$

Corner



$$\begin{aligned}\sum I_x^2 &\rightarrow \text{Small} \\ \sum I_y^2 &\rightarrow \text{Large}\end{aligned}$$

Edge



$$\begin{aligned}\sum I_x^2 &\rightarrow \text{Small} \\ \sum I_y^2 &\rightarrow \text{Small}\end{aligned}$$

Nothing

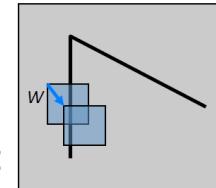


SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

35

## Harris detector: formulation

Sự thay đổi khi dịch cửa sổ 1 khoảng  $[u, v]$ :

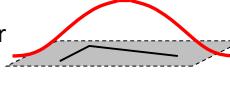


$$E(u, v) = \sum_{x, y} w(x, y) [I(x+u, y+v) - I(x, y)]^2$$

Window function

Shifted intensity

Intensity

Window function  $W(x, y) =$   or 

1 in window, 0 outside

Gaussian



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

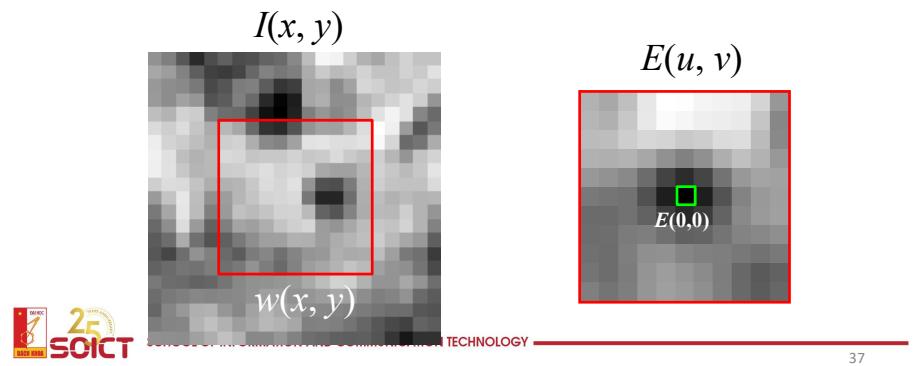
Source: R. Szeliski

36

## Phát hiện góc = phép auto-correlation

Sự thay đổi khi dịch cửa sổ 1 khoảng  $[u, v]$ :

$$E(u, v) = \sum_{x, y} w(x, y) [I(x+u, y+v) - I(x, y)]^2$$

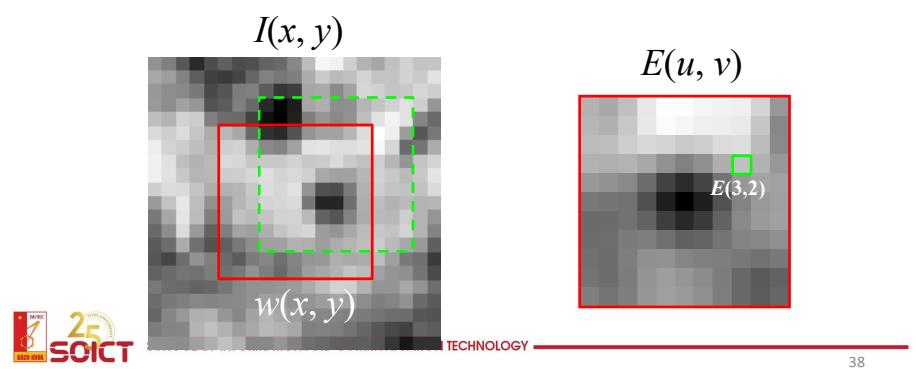


37

## Phát hiện góc = phép auto-correlation

Sự thay đổi khi dịch cửa sổ 1 khoảng  $[u, v]$ :

$$E(u, v) = \sum_{x, y} w(x, y) [I(x+u, y+v) - I(x, y)]^2$$



38

## Phát hiện góc = phép auto-correlation

Sự thay đổi khi dịch cửa sổ 1 khoảng  $[u, v]$ :

$$E(u, v) = \sum_{x, y} w(x, y) [I(x+u, y+v) - I(x, y)]^2$$

Cần đánh giá sự thay đổi E với các dịch chuyển nhỏ

→ **very slow :**

$O(\text{window\_width}^2 * \text{shift\_range}^2 * \text{image\_width}^2)$

$O(11^2 * 11^2 * 600^2) = 5.2 \text{ billion of these}$



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

39

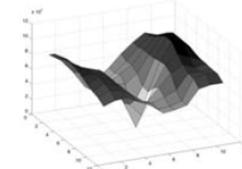
## Phát hiện góc = phép auto-correlation

Sự thay đổi khi dịch cửa sổ 1 khoảng  $[u, v]$ :

$$E(u, v) = \sum_{x, y} w(x, y) [I(x+u, y+v) - I(x, y)]^2$$

Cần đánh giá sự thay đổi E với các dịch chuyển nhỏ

→ **Tìm nơi có sự thay đổi mạnh → Xấp xỉ E**  
**(chuỗi Taylor)**



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

40

## Phát hiện góc = phép auto-correlation

$$E(u, v) = \sum_{x, y} w(x, y) [I(x+u, y+v) - I(x, y)]^2$$

$$E(u, v) \approx [u \ v] \begin{bmatrix} \sum_{x, y} w(x, y) I_x^2(x, y) & \sum_{x, y} w(x, y) I_x(x, y) I_y(x, y) \\ \sum_{x, y} w(x, y) I_x(x, y) I_y(x, y) & \sum_{x, y} w(x, y) I_y^2(x, y) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}$$



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

41

## Harris detector formulation

- Sự thay đổi được xấp xỉ thành:

$$E(u, v) \approx [u \ v] M \begin{bmatrix} u \\ v \end{bmatrix}$$

trong đó  $M$  :  $2 \times 2$  tính từ đạo hàm bậc 1:

$$M = \sum_{x, y} w(x, y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}$$

Đạo hàm theo  $x$  nhân với đạo  
hàm theo  $y$

Tính tổng trên cửa sổ để tìm góc

$$M = \begin{bmatrix} \sum I_x I_x & \sum I_x I_y \\ \sum I_x I_y & \sum I_y I_y \end{bmatrix} = \sum \begin{bmatrix} I_x \\ I_y \end{bmatrix} [I_x \ I_y]$$



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

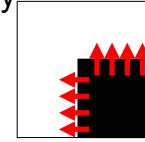
Slide credit: Rick Szeliski

42

## M ?

- Cạnh của góc song song với các trục x, y :

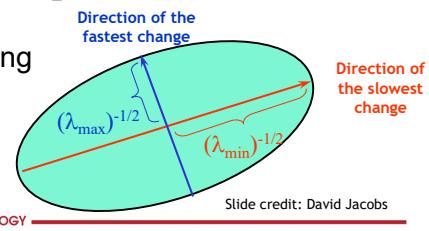
$$M = \begin{bmatrix} \sum I_x^2 & \sum I_x I_y \\ \sum I_x I_y & \sum I_y^2 \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$



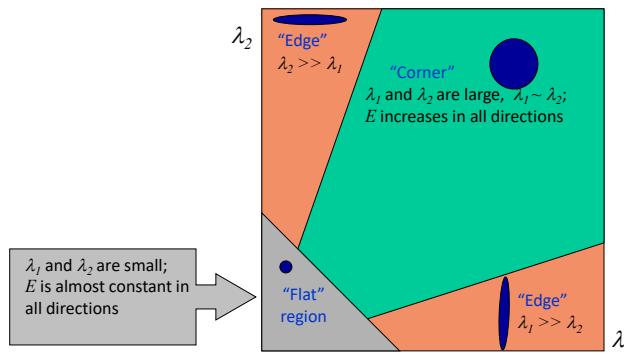
- Trường hợp tổng quát:

– M đối xứng  $\rightarrow M = R^{-1} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} R$  (Eigenvalue decomposition)

- Độ dài trục elipse: theo trị riêng
- R: ma trận quay

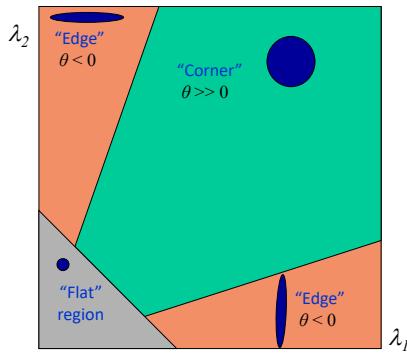


## Ý nghĩa của trị riêng



## Đáp ứng góc

$$\theta = \det(M) - \alpha \operatorname{trace}(M)^2 = \lambda_1 \lambda_2 - \alpha(\lambda_1 + \lambda_2)^2$$

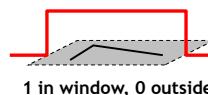


- Xấp xỉ nhanh
  - Tránh phải tính trị riêng
  - $\alpha$ : constant  
(0.04 to 0.06)

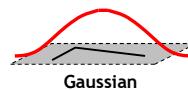
## Hàm cửa sổ $w(x,y)$

$$M = \sum_{x,y} w(x,y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}$$

- Option 1: uniform window
  - Sum over square window
  - Problem: not rotation invariant
- Option 2: Smooth with Gaussian
  - Gaussian already performs weighted sum
  - Result is rotation invariant



$$M = \sum_{x,y} \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}$$

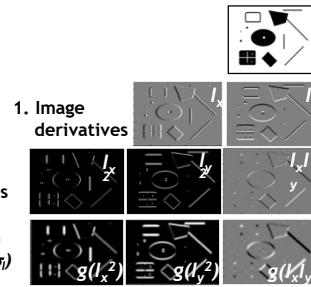


$$M = g(\sigma) * \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}$$

## Tổng kết: Harris Detector [Harris88]

- Tính ma trận moment bậc 2 (ma trận tự tương quan)

$$M(\sigma_I, \sigma_D) = g(\sigma_I) * \begin{bmatrix} I_x^2(\sigma_D) & I_x I_y(\sigma_D) \\ I_x I_y(\sigma_D) & I_y^2(\sigma_D) \end{bmatrix}$$



- Tính đáp ứng góc

### 4. Cornerness function - two strong eigenvalues

$$\begin{aligned} \theta &= \det[M(\sigma_I, \sigma_D)] - \alpha[\text{trace}(M(\sigma_I, \sigma_D))]^2 \\ &= g(I_x^2)g(I_y^2) - [g(I_x I_y)]^2 - \alpha[g(I_x^2) + g(I_y^2)]^2 \end{aligned}$$



### 5. Perform non-maximum suppression

C. Harris and M. Stephens. "A Combined Corner and Edge Detector." *Proceedings of the 4th Alvey Vision Conference*: pages 147—151, 1988.

Slide credit: Krystian Mikolajczyk

47

## Harris Detector



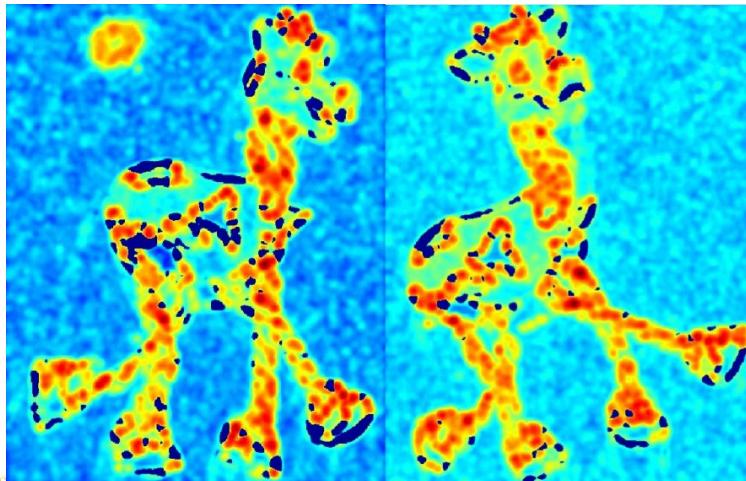
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Slide adapted from Darya Frolova, Denis Simakov

48

## Harris Detector

computer corner responses  $\theta$

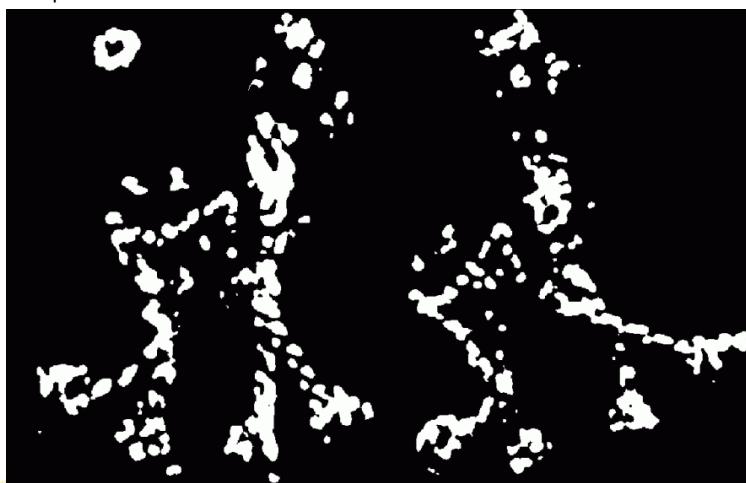


Slide adapted from Darya Frolova, Denis Simakov

49

## Harris Detector

Take points where  $\theta > \text{threshold}$



Slide adapted from Darya Frolova, Denis Simakov

49

## Harris Detector

Take only the local maxima of  $\theta$ , where  $\theta > \text{threshold}$



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Slide adapted from Darya Frolova, Denis Simakov

51

## Harris Detector

Resulting Harris points

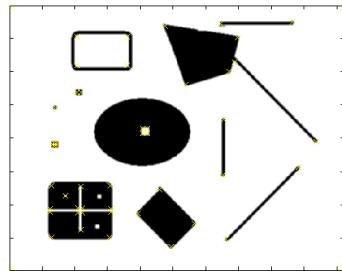


SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

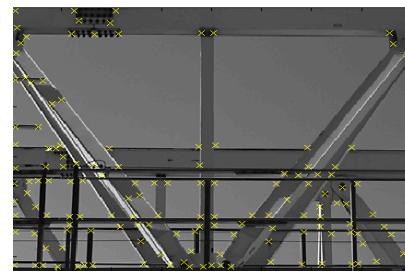
Slide adapted from Darya Frolova, Denis Simakov

52

## Harris Detector – Responses [Harris88]

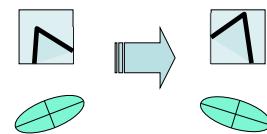


Effect: A very precise corner detector.



## Đặc điểm của Harris detector

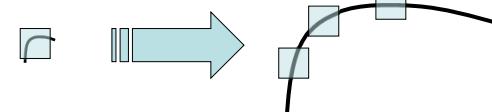
- Translation invariance
- Rotation invariance?



Ellipse rotates but its shape (i.e. eigenvalues) remains the same

**Corner response  $\theta$  is invariant to image rotation**

- Scale invariance?



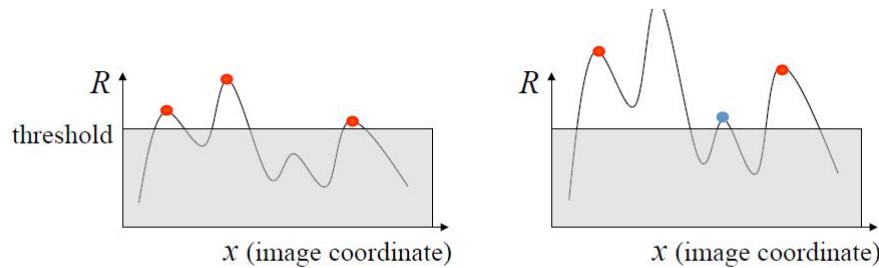
Corner

All points will be classified as edges!

**Not invariant to image scale!**

## Đặc điểm của Harris detector

- Bất biến một phần theo sự thay đổi cường độ sáng



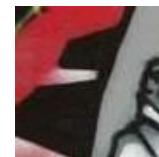
*Partially invariant to affine intensity change*



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

55

## Bất biến theo scale ?



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Slide adapted from T. Tuytelaars ECCV 2006 tutorial

56

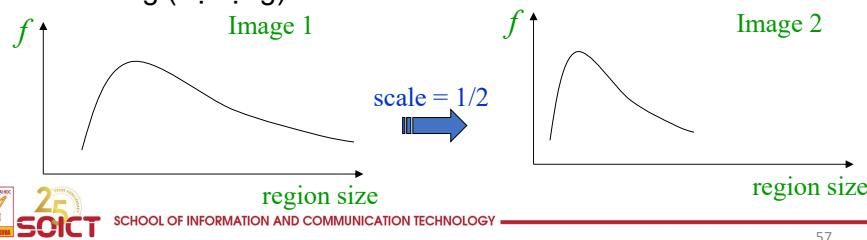
## Chọn kích thước cửa sổ (scale) tự động

- Thiết kế hàm theo độ rộng của vùng, hàm đảm bảo “*scale invariant*”:

- 2 vùng tương đồng (dù ở size khác nhau)  $\rightarrow$  kết quả hàm giống nhau

Ví dụ: trung bình cường độ sáng.

- Tại mỗi điểm trên ảnh, tính giá trị hàm theo kích thước vùng (độ rộng)



57

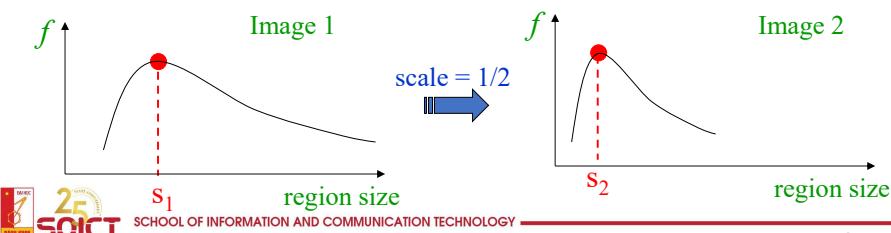
## Chọn kích thước cửa sổ (scale) tự động

- Các bước chính:

Tìm cục đại địa phương trên hàm

Kích thước vùng tương ứng điểm cực đại địa: đảm bảo tính bất biến (*invariant*) theo scale của ảnh

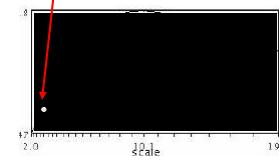
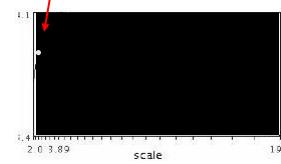
**Quan trọng:** kích thước vùng tìm được là **độc lập ở từng ảnh**



58

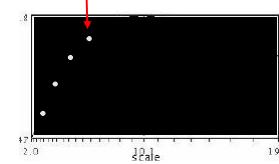
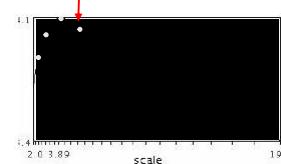
## Ví dụ

Function responses for increasing scale (scale signature)



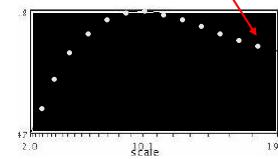
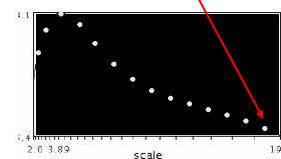
## Ví dụ

Function responses for increasing scale (scale signature)



## Ví dụ

Function responses for increasing scale (scale signature)

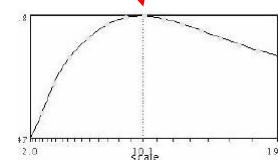
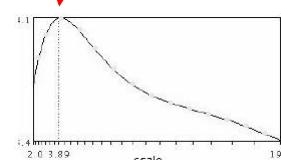


SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

61

## Ví dụ

Function responses for increasing scale (scale signature)



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

62

## Hàm tốt?

- Hàm “tốt” để xác định scale tự động: có 1 đỉnh rõ



- Ảnh thông thường: hàm tốt có thể là hàm tương ứng độ tương phản (thể hiện sự thay đổi cục bộ rõ rệt)

## LoG – DoG

- Functions for determining scale  $f = \text{Kernel} * \text{Image}$

Kernels:

$$L = \sigma^2 (G_{xx}(x, y, \sigma) + G_{yy}(x, y, \sigma))$$

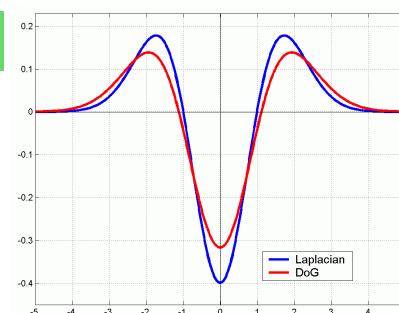
(Laplacian)

$$DoG = G(x, y, k\sigma) - G(x, y, \sigma)$$

(Difference of Gaussians)

where Gaussian

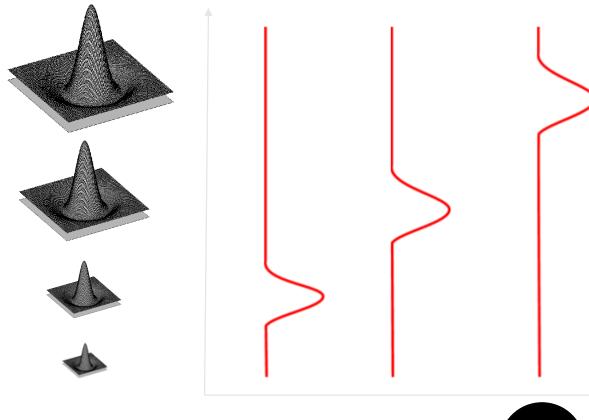
$$G(x, y, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2+y^2}{2\sigma^2}}$$



Note: both kernels are invariant to scale and rotation

## LoG – DoG

- Laplacian-of-Gaussian = “blob” detector



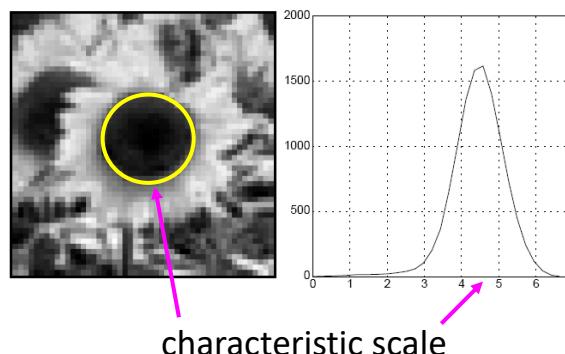
Source: K. Grauman, B. Leibe



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

## Characteristic scale

- **characteristic scale**: the scale that produces peak of Laplacian response



T. Lindeberg (1998). ["Feature detection with automatic scale selection."](#) *IJCV* 30 (2): pp 77--116.



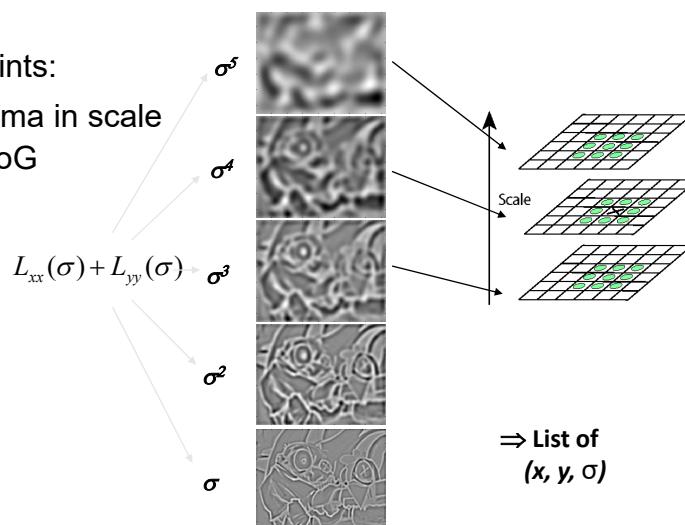
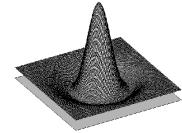
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Source: Lana Lazebnik

## Laplacian-of-Gaussian (LoG)

- Interest points:

Local maxima in scale space of LoG



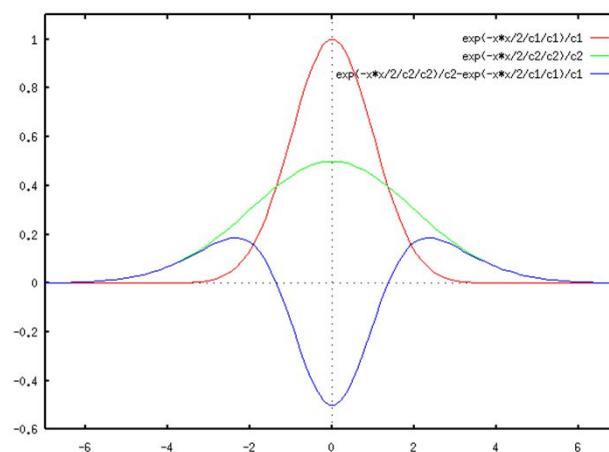
Source: K. Grauman, B. Leibe



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

## LoG-DoG

Approximate LoG with Difference-of-Gaussian (DoG).

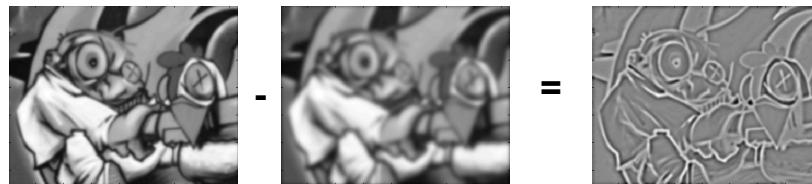


SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

## LoG-DoG

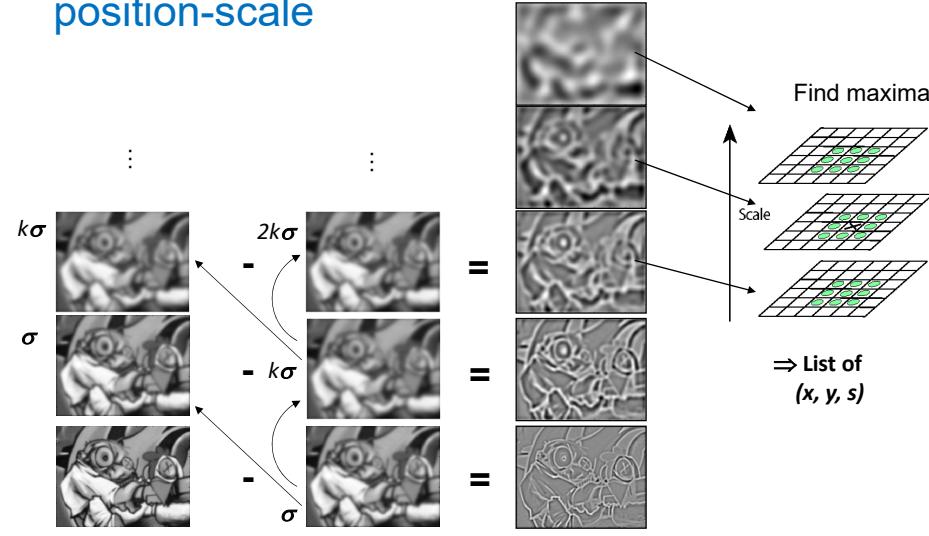
- Approximate LoG with Difference-of-Gaussian (DoG):
  - 1. Blur image with  $\sigma$  Gaussian kernel
  - 2. Blur image with  $k\sigma$  Gaussian kernel
  - 3. Subtract 2. from 1.

Small  $k$  gives a closer approximation to LoG, but usually we want to build a scale space quickly out of this.  $k = 1.6$  gives an appropriate scale space,  $k = \sqrt{2}$



Source: K. Grauman, B. Leibe

## Tìm cực đại địa phương của DoG theo position-scale

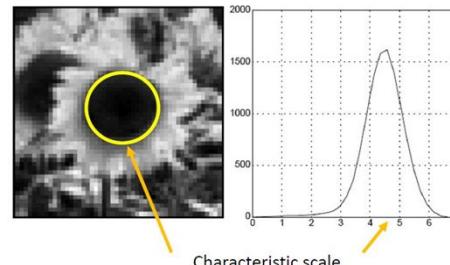
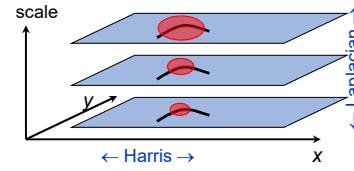


## Harris-Laplacian

- **Harris-Laplacian<sup>1</sup>**

*Find local maximum of:*

- Harris corner detector in space (image coordinates)
- Laplacian in scale



<sup>1</sup> K.Mikolajczyk, C.Schmid. "Indexing Based on Scale Invariant Interest Points". ICCV 2001

<sup>2</sup> D.Lowe. "Distinctive Image Features from Scale-Invariant Keypoints". IJCV 2004



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

71

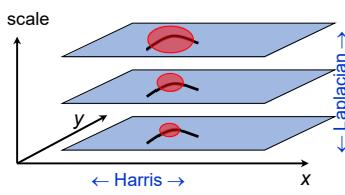
## Scale Invariant Detectors

72

- **Harris-Laplacian<sup>1</sup>**

*Find local maximum of:*

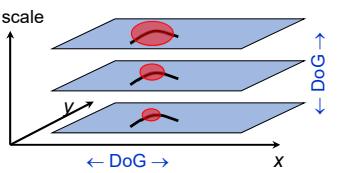
- Harris corner detector in space (image coordinates)
- Laplacian in scale



- **SIFT (D.Lowe)<sup>2</sup>**

*Find local maximum of:*

- Difference of Gaussians in space and scale



<sup>1</sup> K.Mikolajczyk, C.Schmid. "Indexing Based on Scale Invariant Interest Points". ICCV 2001

<sup>2</sup> D.Lowe. "Distinctive Image Features from Scale-Invariant Keypoints". IJCV 2004



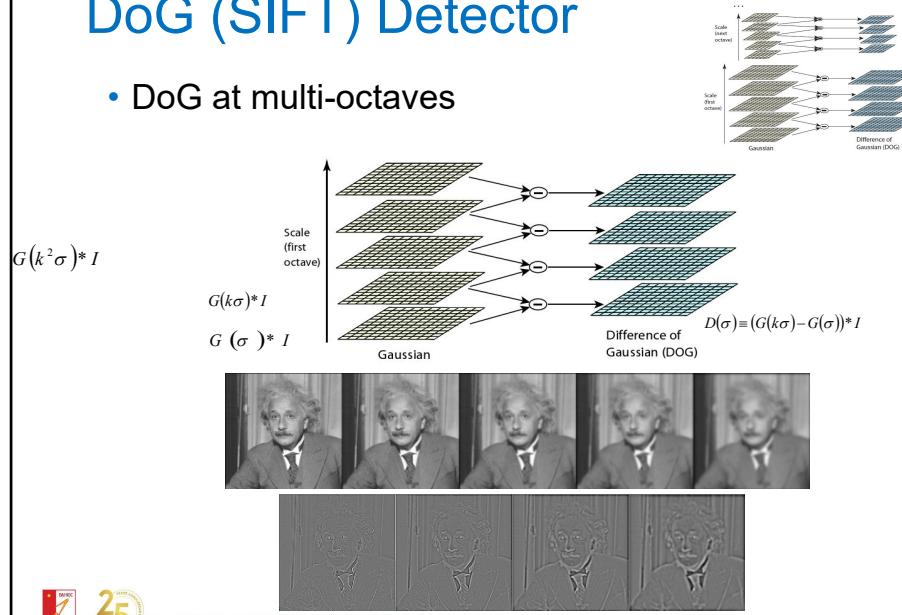
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

## Phát hiện điểm đặc trưng DoG (SIFT)

- DoG tính ở nhiều độ phân giải
- Phát hiện các cực trị trên miền không gian-scale (space – scale)
- Xác định vị trí điểm
  - Nội suy
  - Loại điểm không ổn định
- Gán hướng chủ đạo cho điểm đặc trưng

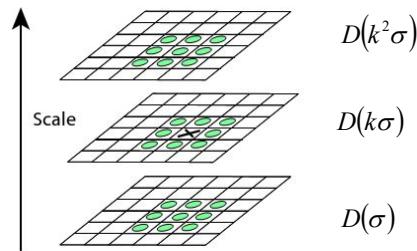
## DoG (SIFT) Detector

- DoG at multi-octaves



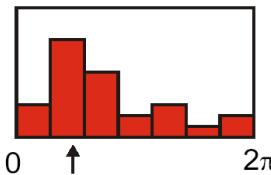
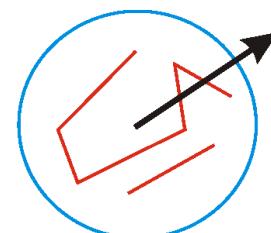
## DoG (SIFT) Detector

- Scale-Space Extrema: Chọn cực trị khi xét lân cận  $3 \times 3 \times 3$



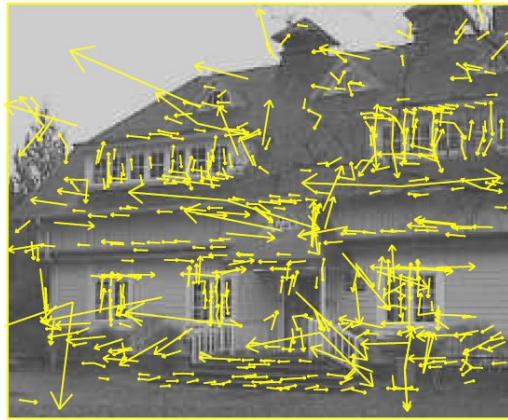
## DoG (SIFT) Detector

- Hướng chủ đạo
  - Tạo 1 histogram của hướng đạo hàm ở vùng được xác định (theo scale)
  - Chọn hướng cho giá trị lớn nhất trên histogram (đã đc làm mịn)
- Mỗi điểm đặc trưng được xác định bởi **(x, y, scale, orientation)**



If 2 major orientations, use both.

## DoG (SIFT) Detector



A SIFT keypoint : {x, y, scale, dominant orientation}



Source: [Distinctive Image Features from Scale-Invariant Keypoints](#) – IJCV 2004

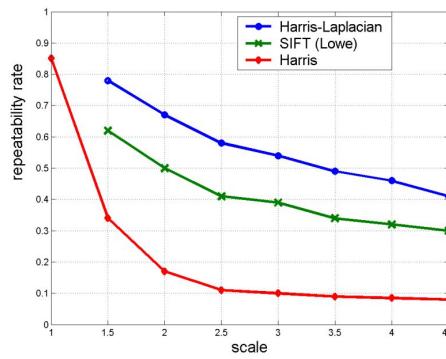
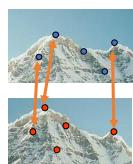
77

## So sánh

- Đánh giá về mức độ phát hiện được các điểm khi scale thay đổi

Repeatability rate:

$$\frac{\# \text{ correspondences}}{\# \text{ possible correspondences}}$$



K.Mikolajczyk, C.Schmid. "Indexing Based on Scale Invariant Interest Points". ICCV 2001



Slide credit: CS131 -Juan Carlos Niebles and Ranjay Krishna

78

## Các bộ phát hiện khác

- Hessian & Harris [Beaudet '78], [Harris '88]
- Laplacian, DoG [Lindeberg '98], [Lowe '99]
- Harris-/Hessian-Laplace [Mikolajczyk & Schmid '01]
- Harris-/Hessian-Affine [Mikolajczyk & Schmid '04]
- EBR and IBR [Tuytelaars & Van Gool '04]
- MSER [Matas '02]
- Salient Regions [Kadir & Brady '01]
- Others...
- *Those detectors have become a basic building block for many recent applications in Computer Vision.*

Slide credit: Bastian Leibe

79

## Feature extraction

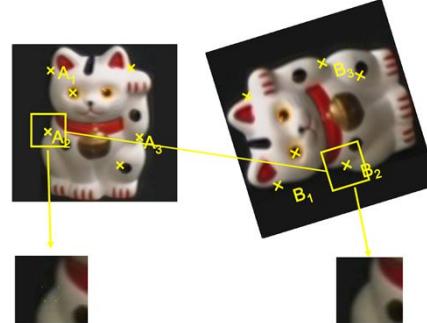
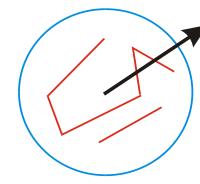
- Global features
- **Local features**
  - Interest point detector
  - **Local descriptor**
  - Matching

## Trích chọn đặc trưng: Đặc trưng tốt?

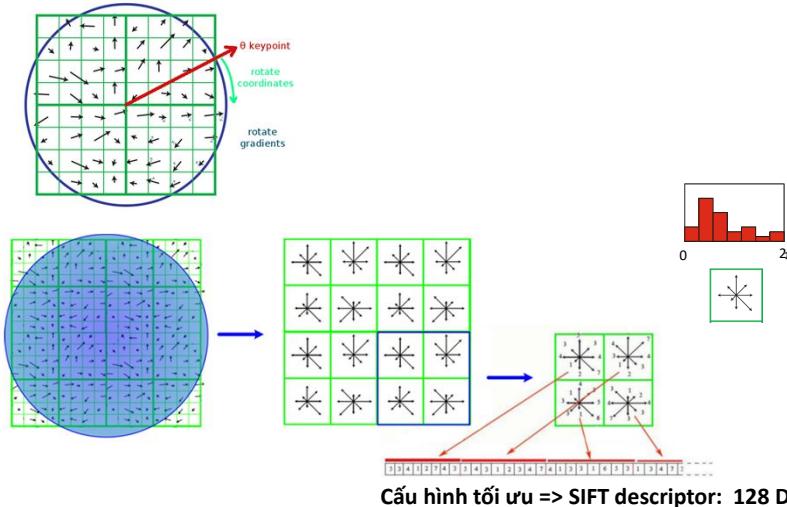
- Gọn (Compact)
- Bất biến với
  - Một số phép biến đổi hình học
  - Góc nhìn camera
  - Điều kiện chiếu sáng
- Ví dụ:
  - **SIFT**, SURF([Speeded Up Robust Features](#)), [PCA-SIFT](#), ...
  - LBP, BRISK, MSER and FREAK, ...

## Bất biến với phép quay

- Sử dụng hướng chủ đạo (**a characteristic orientation/ dominant orientation**) tại mỗi điểm đặc trưng
- Bộ mô tả được chuẩn hóa theo hướng này



## Bộ mô tả SIFT



Source: [Distinctive Image Features from Scale-Invariant Keypoints – IJCV 2004](http://campar.in.tum.de/twiki/pub/Chair/TeachingWs13TDCV/feature_descriptors.pdf)  
[http://campar.in.tum.de/twiki/pub/Chair/TeachingWs13TDCV/feature\\_descriptors.pdf](http://campar.in.tum.de/twiki/pub/Chair/TeachingWs13TDCV/feature_descriptors.pdf)



## Bộ mô tả SIFT

- Hiệu quả khi **điều kiện chiếu sáng thay đổi**:
  - Bộ mô tả sử dụng đạo hàm bậc 1
    - $\rightarrow$  **đảm bảo bất biến với độ sáng thay đổi**
  - Độ tương phản thay đổi  $\rightarrow$  thay đổi độ lớn gradient
    - $\rightarrow$  **chuẩn hóa vector** (độ lớn vector = 1.0)
  - Độ lớn gradient quá lớn do hiệu ứng 3D khác (chói, etc)
    - $\rightarrow$  để giảm hiệu ứng, **giới hạn giá trị** trong vector  $\leq 0.2$  (giá trị thực nghiệm).  $\rightarrow$  Chuẩn hóa lại vector
- $\rightarrow$  **Đảm bảo bất biến với sự thay đổi điều kiện chiếu sáng**



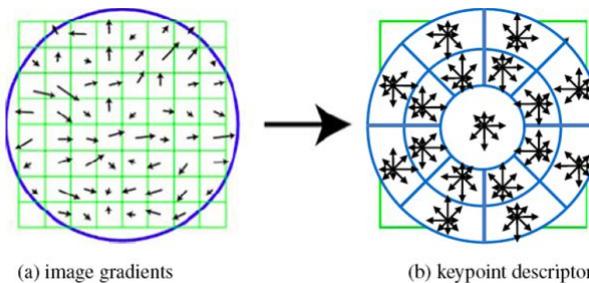
SIFT: [Distinctive Image Features from Scale-Invariant Keypoints – IJCV 2004](http://campar.in.tum.de/twiki/pub/Chair/TeachingWs13TDCV/feature_descriptors.pdf)

## Một số bộ mô tả khác

- GLOH, SURF([Speeded Up Robust Features](#)), [PCA-SIFT](#), ...
- LBP, ..
- BREF, BRISK, ORB, FREAK, KAZE, LDAHash

## GLOH (Gradient Local Orientation Histogram)

- Tương tự SIFT
  - Sử dụng phân chia không gian (bin) theo log-polar
  - Chia thành 17 ô trong không gian (16 spatial bins)
  - 16 hướng gradient (16 orientation bins)
  - Vector 272 D  $\rightarrow$  PCA  $\rightarrow$  128D



(a) image gradients

(b) keypoint descriptor

[Krystian Mikolajczyk and Cordelia Schmid, 2005](#)

# So khớp ảnh



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

91

## So khớp đặc trưng cục bộ

Với 1 keypoint trong ảnh  $I_1$ , tìm keypoints tương ứng trong ảnh  $I_2$ ?

1. Xác định khoảng cách sử dụng:
  - Use L1, L2, cosine, Mahalanobis, ... distance
2. Với mỗi đặc trưng trong  $I_2$ , tìm điểm có khoảng cách (dựa trên đặc trưng) nhỏ nhất

OpenCV:

- Brute force matching
- Flann Matching: Fast Library for Approximate Nearest Neighbors

[Muja and Lowe, 2009]

Marius Muja and David G Lowe. Fast approximate nearest neighbors with automatic algorithm configuration. In VISAPP (1), pages 331–340, 2009



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

## So khớp đặc trưng cục bộ

- Khoảng cách giữa  $f_1, f_2$ ?
  - Đơn giản: **sử dụng giá trị độ đo khoảng cách ( $f_1, f_2$ )**
    - $\rightarrow$  có thể đưa ra cả kết quả ghép cặp tốt và không tốt
  - Giải pháp tốt hơn: thêm **ràng buộc bổ sung**
    - Tỷ lệ khoảng cách (Nearest neighbor distance ratio)
    - Thêm ràng buộc không gian dựa trên hàng xóm
    - Kiểm tra chéo (Cross check test)
    - Tìm phép biến đổi giữa 2 tập điểm, lọc lại các cặp so khớp (RANSAC)



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

93

## Nearest neighbor distance ratio

- **radio of distance =  $d(f_1, f_2) / d(f_1, f_2')$** 
  - $f_2$ : đặc trưng gần nhất với  $f_1$  trong số các đặc trưng trong ảnh  $I_2$ ;
  - $f_2'$  : đặc trưng gần thứ 2 với  $f_1$  trong số các đặc trưng trong ảnh  $I_2$ ;
  - Ghép cặp không tốt  $\rightarrow$  ratio gần 1
  - Chỉ giữ lại những cặp có ratio nhỏ



94

## Ratio of distances reliable for matching

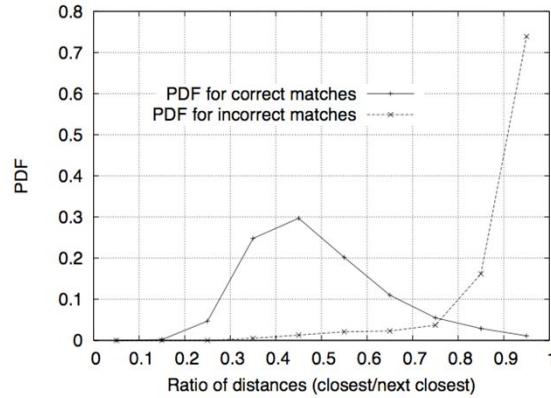


Figure 11: The probability that a match is correct can be determined by taking the ratio of distance from the closest neighbor to the distance of the second closest. Using a database of 40,000 keypoints, the solid line shows the PDF of this ratio for correct matches, while the dotted line is for matches that were incorrect.

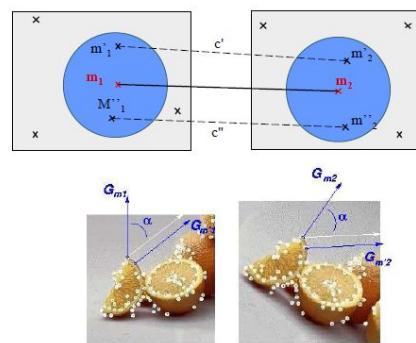


David G. Lowe, "Distinctive image features from scale-invariant keypoints," IJCV, 60, 2 (2004), pp. 91-110

SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

## Thêm ràng buộc không gian

- Spatial constraints between neighborhood pixels



Source: from slides of Valérie Gouet-Brunet



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

## Kiểm tra chéo

- Ghép cặp (fa, fb) nếu:
  - fb là điểm ghép cặp tốt nhất trong ảnh I2 với fa của I1
  - fa là điểm ghép cặp tốt nhất trong ảnh I1 với fb của I2



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

97

## Tìm mô hình biến đổi

- Mục đích: kiểm tra các sự ghép cặp giữa 2 ảnh đã phù hợp
- Fitting the transformation (RANSAC alg.)
  - Fitting 2D transformation matrix
  - Six variables
  - Each point give two equations
  - → at least three points
  - Least squares
- RANSAC: refinement of matches
  - Compute error:

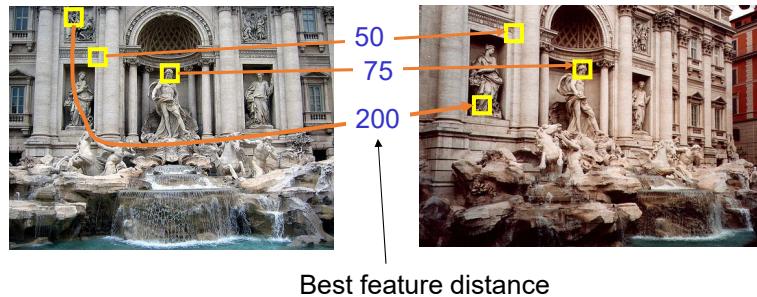


SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

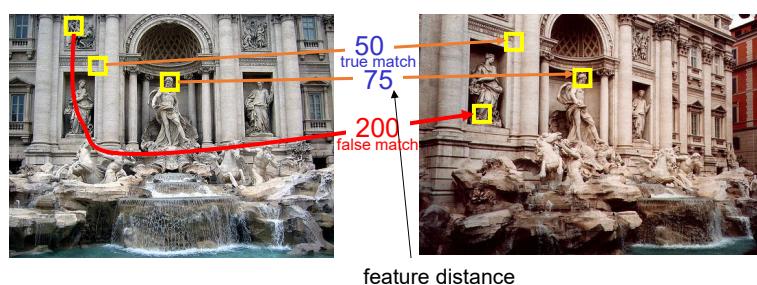
98

## Đánh giá kết quả ghép cặp

How can we measure the performance of a feature matcher?



## True/false positives

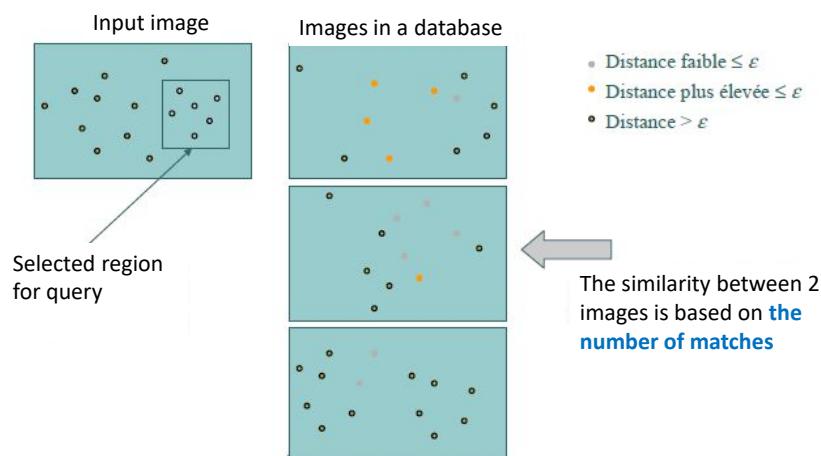


- **True positives** = số cặp ghép chính xác
  - Cần tối đa → chọn ngưỡng ?
- **False positives** = số cặp ghép sai
  - Cần tối thiểu → chọn ngưỡng ?

## So khớp ảnh

- Tính khoảng cách giữa 2 ảnh  $I_1, I_2$ ?
  - Sử dụng đặc trưng toàn cục: đơn giản  
 $d(I_1, I_2) = d(\text{feature of } I_1, \text{feature of } I_2)$
  - Sử dụng đặc trưng cục bộ:
    - Bỏ phiếu
    - Giải bài toán tối ưu (time consuming)
    - Xây dựng vector toàn cục từ các đặc trưng cục bộ : BoW (bag-of-words, bag-of-features), VLAD, ..

## Chiến lược bỏ phiếu



Source: Modified from slides of Valérie Gouet-Brunet

## Bài toán tối ưu

- Transportation problem

$I_1 : \{(r_i, w_i), i=1, N\}$  Provider

$I_2 : \{(r'_j, w_j), j=1, M\}$  Consommert

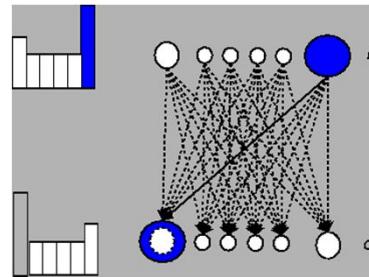
$d(I_1, I_2) = ???$

$$d(I_1, I_2) = \min \sum_i \sum_j f_{ij} \times d(r_i, r'_j)$$

$$f_{ij} \geq 0$$

$$\sum_i f_{ij} \leq w_j, \sum_j f_{ij} \leq w_i$$

$$\sum_i \sum_j f_{ij} = \min(\sum_i w_i, \sum_j w_j)$$



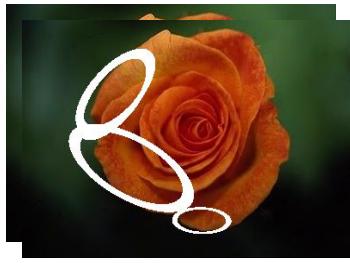
$$d_{EMD}(I_1, I_2) = \frac{\sum_i \sum_j f_{ij}^* \times d(r_i, r'_j)}{\sum_i \sum_j f_{ij}^*}$$

## Mô hình túi từ

- Đặc trưng cục bộ ~ 1 từ
- 1 ảnh ~ 1 văn bản
- Sử dụng mô hình vector: ảnh = tần suất xuất hiện của các từ trực quan



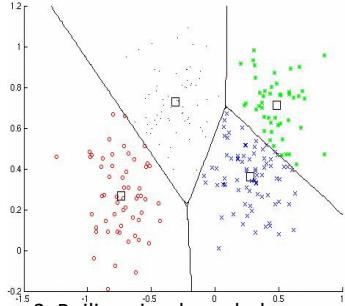
## Từ điển



1. Extracting local features from a set of images



2. Builing visual vocabulary (dictionary) using a clustering method



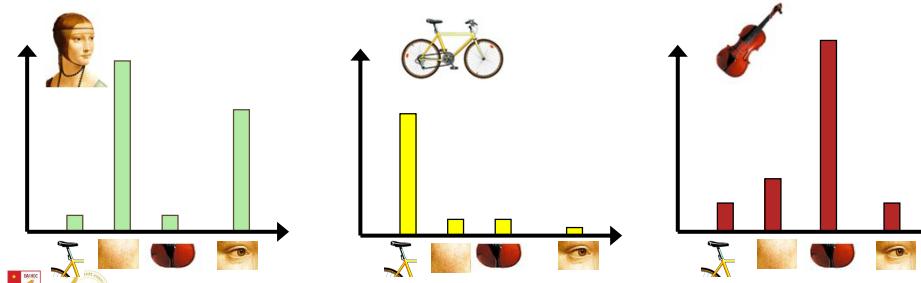
3. An image is represented by a bag of words  
→ can be represented by tf.idf vector

MARCO 25 SOICT SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

105

## Mô hình túi từ: các bước chính

1. Trích chọn đặc trưng
2. Tạo từ điển
3. Gán đặc trưng cục bộ thành các từ
4. Biểu diễn ảnh = vector tần suất xuất hiện



MARCO 25 SOICT SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

# Một số ứng dụng



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

107

## Object detection/recognition/search



Sivic and Zisserman, 2003



Rothganger et al. 2003



Lowe 2002



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

## Object detection/recognition

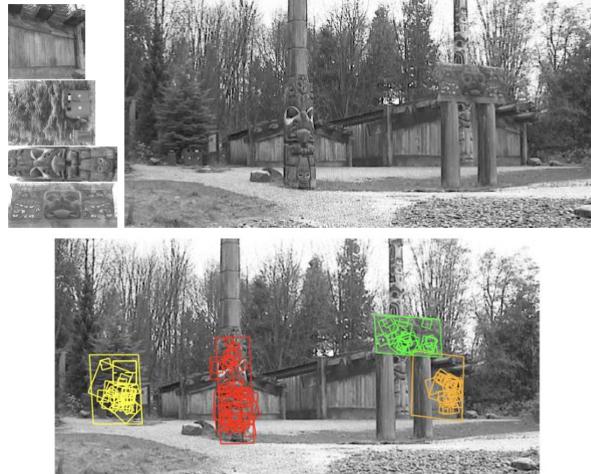


Figure 13: This example shows location recognition within a complex scene. The training images for locations are shown at the upper left and the 640x315 pixel test image taken from a different viewpoint is on the upper right. The recognized regions are shown on the lower image, with keypoints shown as squares and an outer parallelogram showing the boundaries of the training images under the affine transform used for recognition.



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

David Low, Distinctive Image Features from Scale-Invariant Keypoints, IJCV 2004

## Image Panoramas



Slide credit: Darya Frolova, Denis Simakov



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

## Image Panoramas



Slide credit: Darya Firova, Denis Simakov

- Procedure:

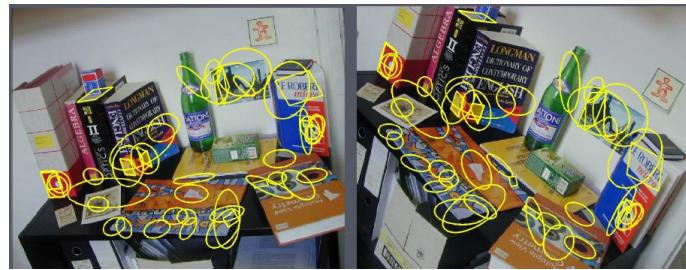
- Detect feature points in both images
- Find corresponding pairs
- Use these pairs to align the images

## Automatic mosaicing



<http://www.cs.ubc.ca/~mbrown/autostitch/autostitch.html>

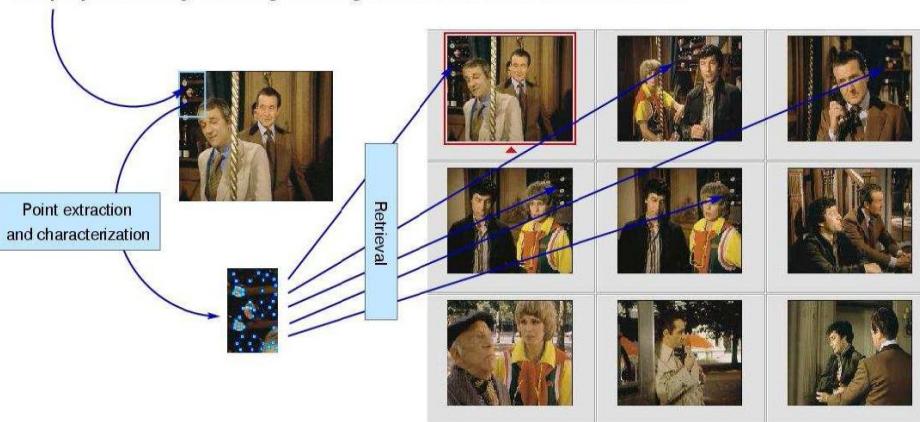
## Wide baseline stereo



[Image from T. Tuytelaars ECCV 2006 tutorial]

## CBIR: partial retrieval

The query : "I am looking for the images involving the room where there is this wine storeroom".



## References

- Lecture 5,6: CS231 - Juan Carlos Niebles and Ranjay Krishna, Stanford Vision and Learning Lab
- Vision par Ordinateur, Alain Boucher, IFI
- [SIFT: Keypoint detector \(ubc.ca\)](#)
- [SURF: <https://people.ee.ethz.ch/~surf/eccv06.pdf>](https://people.ee.ethz.ch/~surf/eccv06.pdf)
- Harris corner detector:  
[https://home.cis.rit.edu/~cnspec/references/dip/feature\\_extraction/harris1988.pdf](https://home.cis.rit.edu/~cnspec/references/dip/feature_extraction/harris1988.pdf)
- [CBIR: J. Sivic & A. Zisserman \(2003\). "Video Google: A Text Retrieval Approach to Object Matching in Videos" \(PDF\). Proc. of ICCV](#)
- [HOG for Human detection: <https://lear.inrialpes.fr/people/triggs/pubs/Dalal-cvpr05.pdf>](#)



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

115

VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG  
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

25  
SOICT

VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG  
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Thank you for  
your attention!