



Thị giác máy tính

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

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, ...

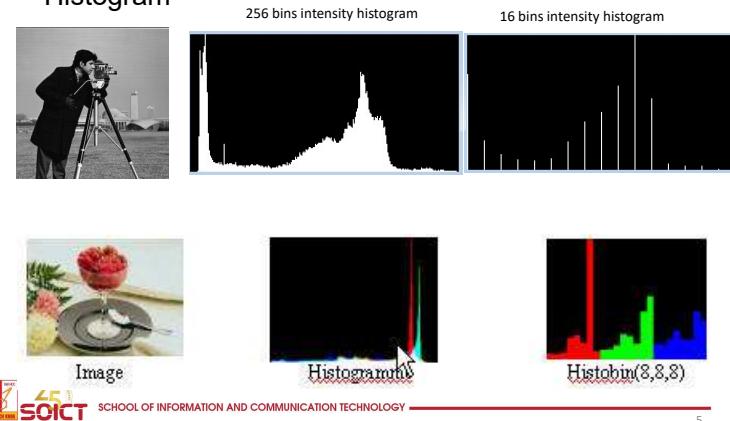
Đặc trưng toàn cục

How to distinguish these objects?



Đặc trưng màu

- Histogram

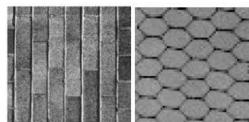


Đặc trưng kết cấu

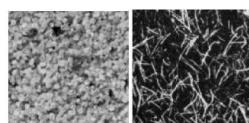
- #### • Kết cấu:

- 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)



Histogram

- Đơn giản
 - Bất biến đổi với các phép biến đổi hình học

– quay, dich, zoom

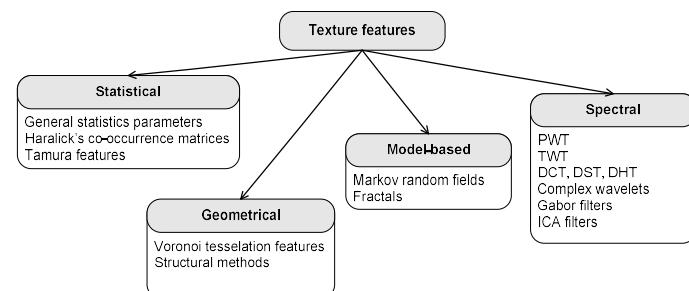
- Không tính đến sự tương đồng giữa các bin gần nhau

- Phân bố trên không gian ảnh không được tính đến

Ảnh hưởng bởi yếu tố nền

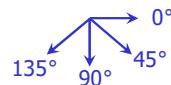


Đặc trưng kết cấu



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

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 β

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	1	1	2
3	0	1	0
4	0	1	1

Matrix for distance=1
and direction=0°

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

Matrix for distance=1
and direction=45°



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

10



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

11

GLCM

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

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

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

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

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



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

12

GLCM

- **Đặc trưng Haralick :**

- Mỗi ma trận GLCM → 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:**

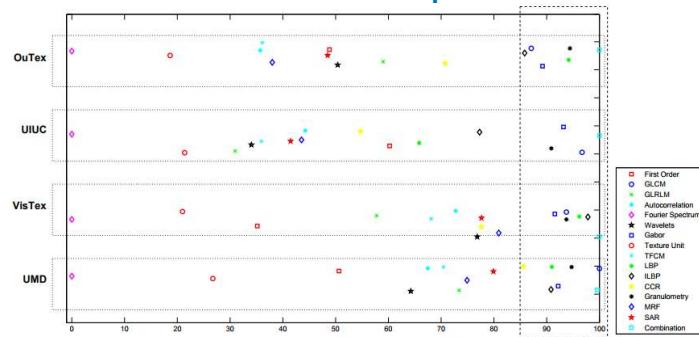
- Bất biến với phép quay ?
- Bất biến với phép biến đổi tỷ lệ?



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

13

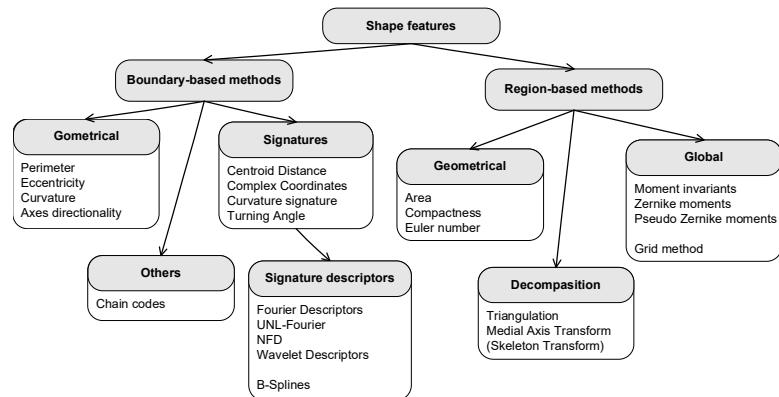
Texture features comparision



Source : William Robson Schwartz et al. Evaluation of Feature Descriptors for Texture Classification – 2012 JEI

14

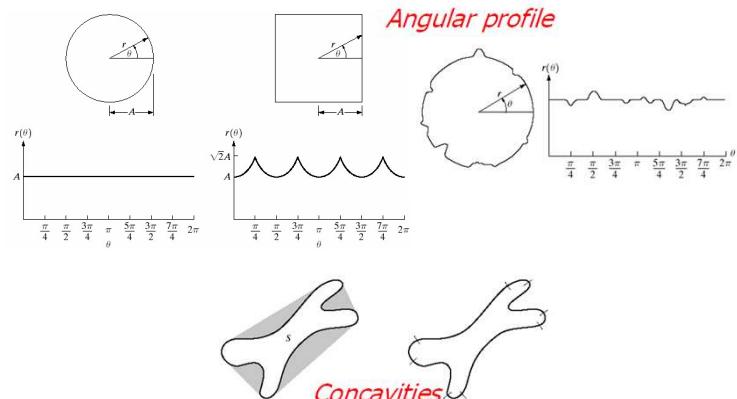
Đặc trưng hình dạng



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

15

Examples: angular profile, ...



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

16

Examples : Image moments

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



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

17

Invariant moments (Hu's moments)

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

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

 $I_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]$

invariant to translation, scale, and rotation, and reflection
Change for image reflection



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

18

Hu's moments

6 images and their Hu Moments

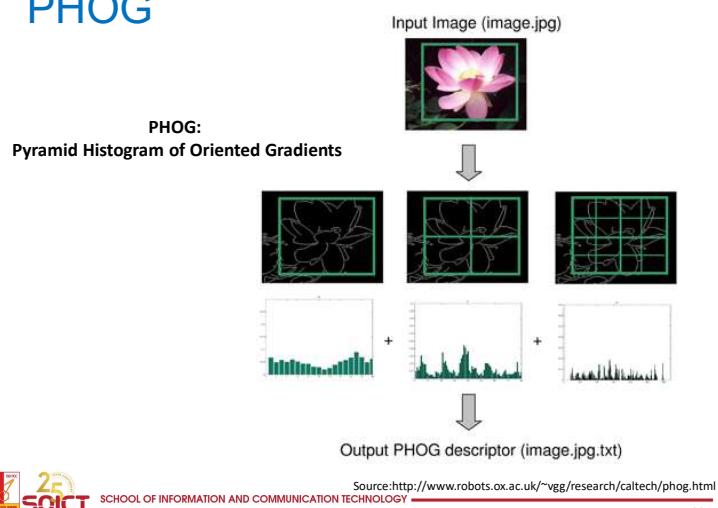
ID	Image	H[0]	H[1]	H[2]	H[3]	H[4]	H[5]	H[6]
K0	K	2.78871	6.50638	9.44249	9.84018	-19.593	-13.1205	19.6797
S0	S	2.67431	5.77446	9.90311	11.0016	-21.4722	-14.1102	22.0012
S1	S	2.67431	5.77446	9.90311	11.0016	-21.4722	-14.1102	22.0012
S2	S	2.65884	5.7358	9.66822	10.7427	-20.9914	-13.8694	21.3202
S3	S	2.66083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

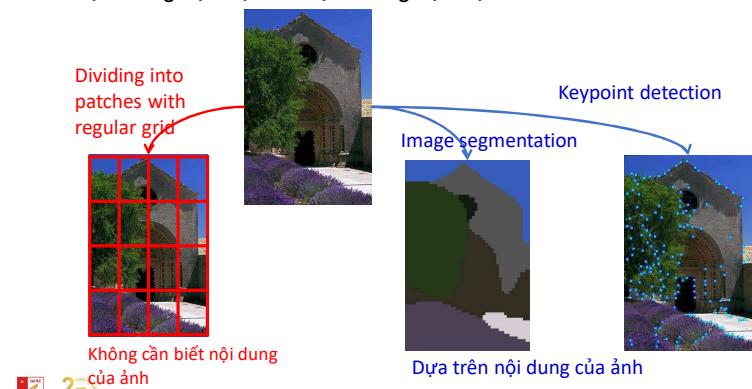
19

PHOG



Đặc trưng cục bộ

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



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Vai trò của đặc trưng cục bộ

- Image matching: a challenging problem



Source: CS131 - Juan Carlos Niebles and Ranjay Krishna

22

Vai trò của đặc trưng cục bộ

- Image matching: a challenging problem



by Diva Sian



by swashford



Slide credit: Steve Seitz

Source: CS131 - Juan Carlos Niebles and Ranjay Krishna

23

Vai trò của đặc trưng cục bộ

- Partial search
- Object detection/ recognition



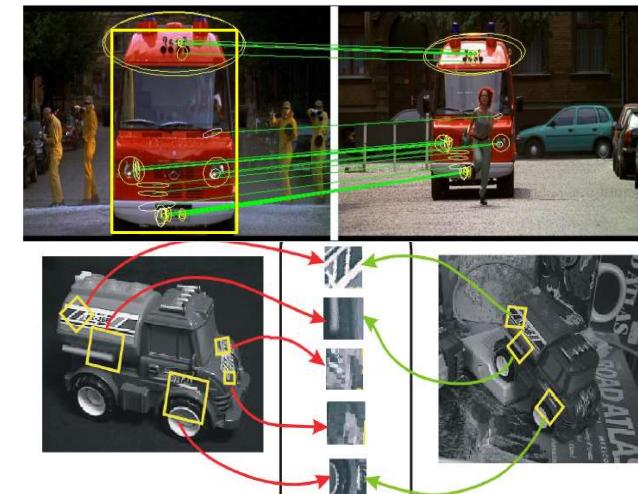
Sivic and Zisserman, 2003



D. Lowe 2002

Source: CS131 - Juan Carlos Niebles and Ranjay Krishna

24



Source : Jim Little, Lowe; features, UBC.

26

Yêu cầu quan trọng

- Problem 1:
 - Detect the same point *independently* in both images



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 (~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

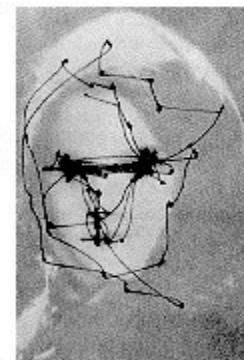
Phát hiện các điểm đặc trưng



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

29

Điểm quan trọng nằm ở đâu?



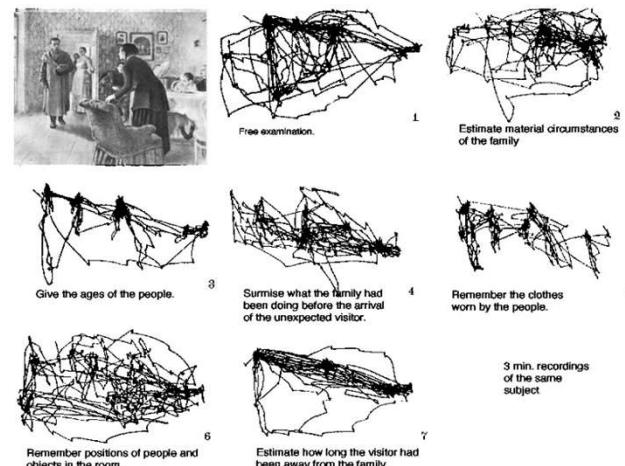
Yarbus eye tracking



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

30

Same image with different questions



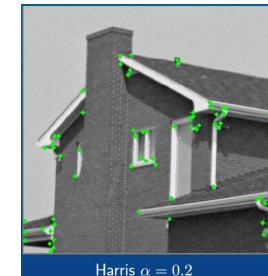
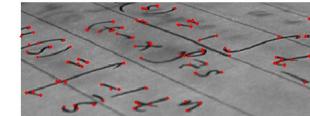
Change blindness : <http://www.simonslab.com/videos.html>

31

Đặc trưng cục bộ và các điểm đặc trưng

- Phát hiện các điểm đặc trưng (interest point, keypoint, point of interest):

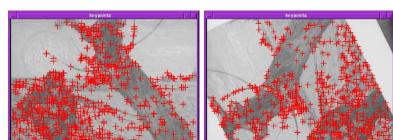
- DoG /SIFT detector
- Harris corner detector
- Moravec
- ...



 SOICT SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

32

Tìm góc – Harris corner detector



- Ý tưởng chính:
 - Trong vùng xung quanh 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

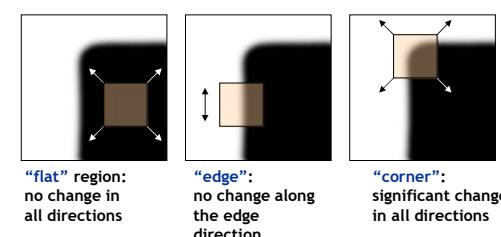
Slide credit: Svetlana Lazebnik

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*)



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



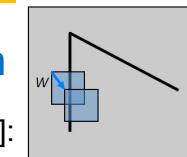
$$\begin{aligned} \sum I_x^2 &\rightarrow \text{Small} \\ \sum I_y^2 &\rightarrow \text{Small} \end{aligned}$$

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

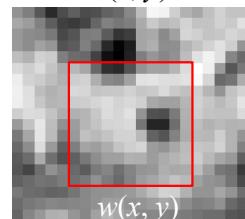
Source: R. Szeliski

36

Phát hiện góc = phép auto-correlation

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

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

 $I(x, y)$  $E(u, v)$  $E(0,0)$ 

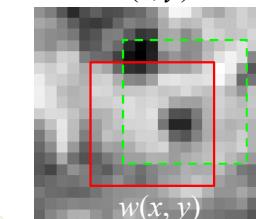
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

37

Phát hiện góc = phép auto-correlation

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

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

 $I(x, y)$  $E(u, v)$  $E(3,2)$ 

SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

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

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

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



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

43

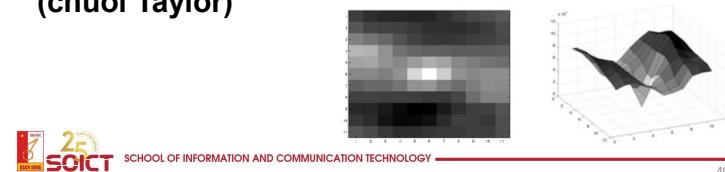
Phát hiện góc = phép auto-correlation

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)



40

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

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



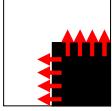
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

SCHILLER IN PICTURES

42

M ?

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

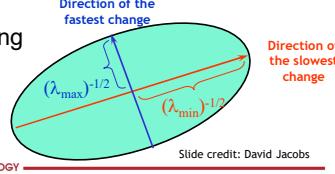
$$M = \begin{bmatrix} \sum I_x^2 & \sum I_x I_y \\ \sum I_x I_y & \sum I_y^2 \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$


- 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 ellipse: theo trị riêng

– R: ma trận quay



Slide credit: David Jacobs

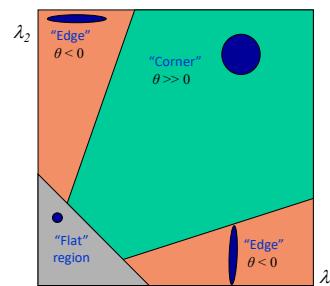
43



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Đáp ứng góc

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



- Xấp xỉ nhanh

- Tránh phải tính trị riêng
- α : constant (0.04 to 0.06)

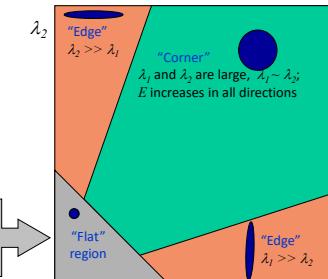
Slide credit: Kristen Grauman

45



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Ý nghĩa của trị riêng



Slide credit: Kristen Grauman

44

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

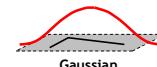


$$M = \sum_{x,y} \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}$$

- Problem: not rotation invariant

- Option 2: Smooth with Gaussian

- Gaussian already performs weighted sum



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

- Result is rotation invariant

Slide credit: Bastian Leibe

46

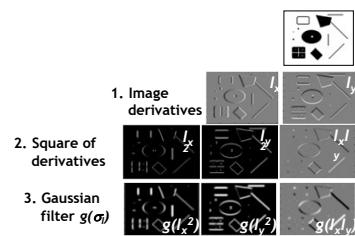


SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

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

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



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



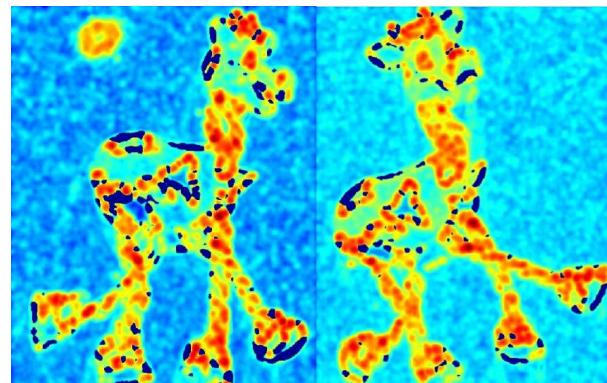
SOICT SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Slide adapted from Darya Frolova, Denis Simakov

48

Harris Detector

computer corner responses θ



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Slide adapted from Darya Frolova, Denis Simakov

49

Harris Detector

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



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Slide adapted from Darya Frolova, Denis Simakov

50

Harris Detector

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



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Slide adapted from Darya Frolova, Denis Simakov

51

Harris Detector

Resulting Harris points

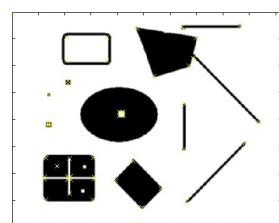


SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Slide adapted from Darya Frolova, Denis Simakov

52

Harris Detector – Responses_[Harris88]



Effect: A very precise corner detector.



Slide credit: Krystian Mikolajczyk

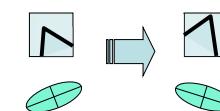
53



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Đặc điểm của Harris detector

- Translation invariance
- Rotation invariance?



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

Corner response θ is invariant to image rotation

- Scale invariance?



Corner
All points will be classified as edges!

Not invariant to image scale!

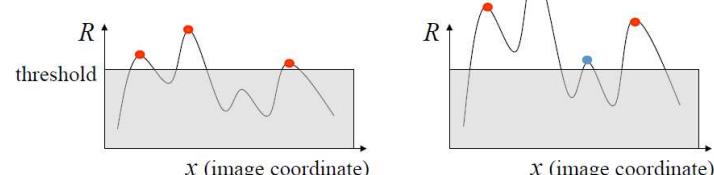


SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

54

Đặc điểm của Harris detector

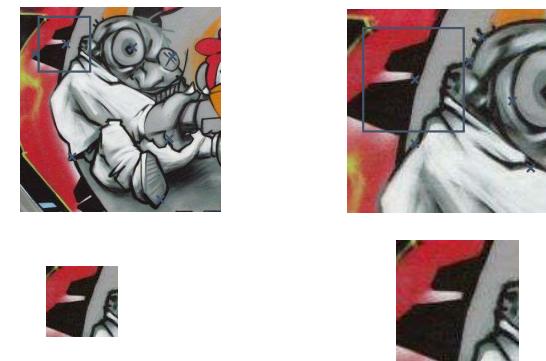
- Bắt biến một phần theo sự thay đổi cường độ sáng



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

55

Bắt biến theo scale ?



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Slide adapted from T. Tuytelaars ECCV 2006 tutorial

56

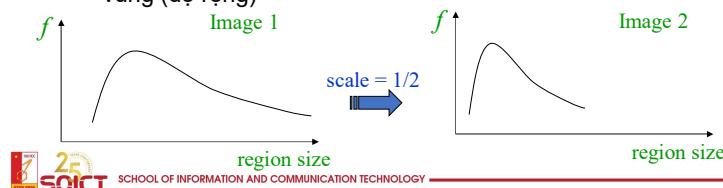
Chọn kích thước cửa sổ (scale) tự động

- 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)



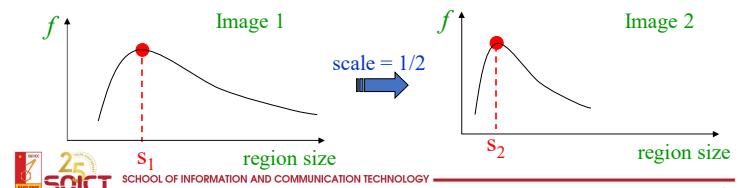
Chọn kích thước cửa sổ (scale) tự động

- Các bước chính:

Tìm cục đại địa phương trên hàm

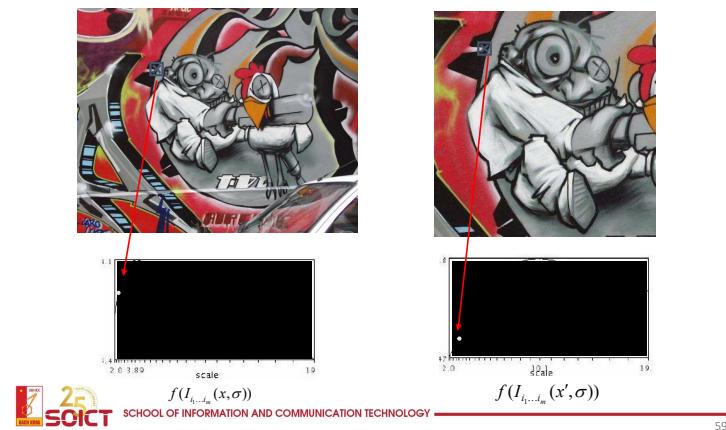
Kích thước vùng tương ứng điểm cục đại địa: đảm bảo tính bất biến (invariant) theo scale của ảnh

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



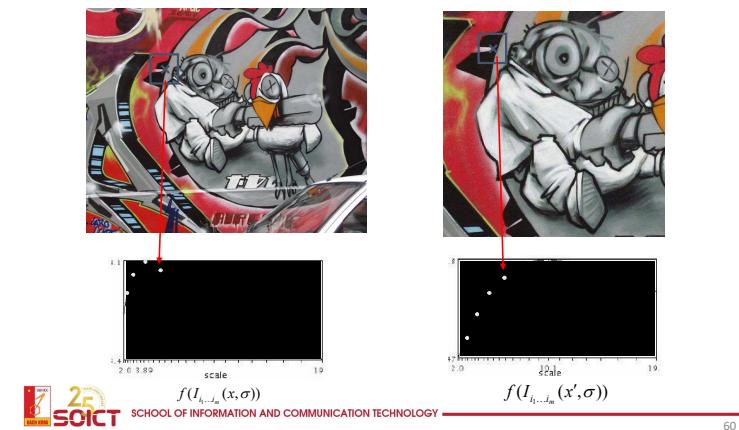
Ví dụ

Function responses for increasing scale (scale signature)



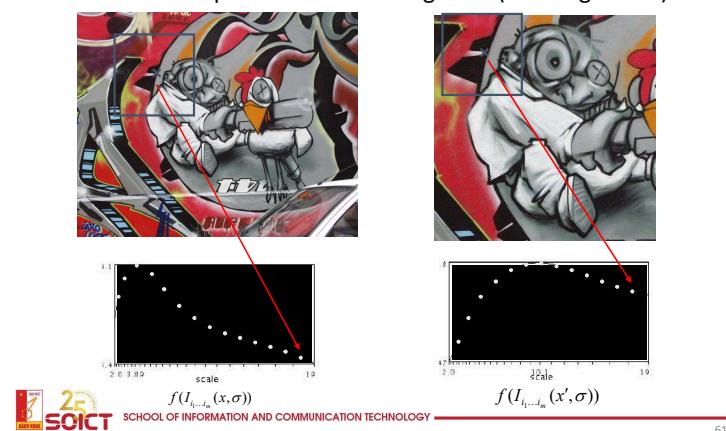
Ví dụ

Function responses for increasing scale (scale signature)



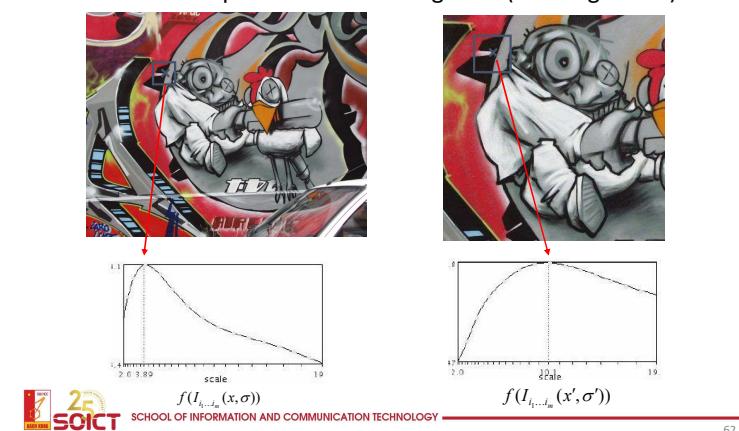
Ví dụ

Function responses for increasing scale (scale signature)



Ví dụ

Function responses for increasing scale (scale signature)



Hàm tốt?

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



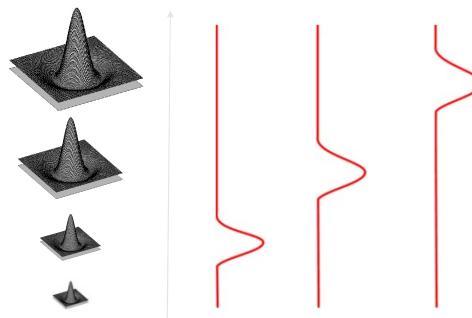
- Ảnh thông thường: hàm tốt có thể là hàm tương ứng độ tương phản (thể hiện sự thay đổi cục bộ rõ rệt)



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

LoG – DoG

- Laplacian-of-Gaussian = “blob” detector



Source: K. Grauman, B. Leibe



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

LoG – DoG

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

Kernels:

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

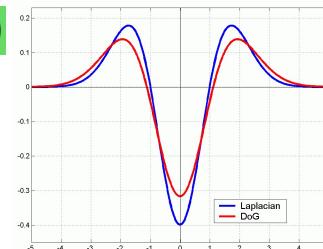
(Laplacian)

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

(Difference of Gaussians)

where Gaussian

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



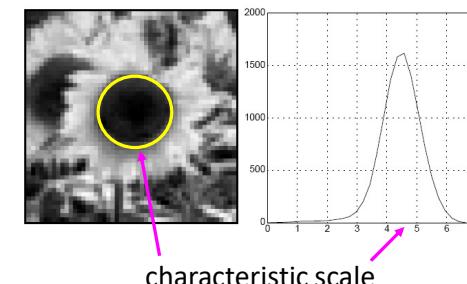
Note: both kernels are invariant to scale and rotation



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.

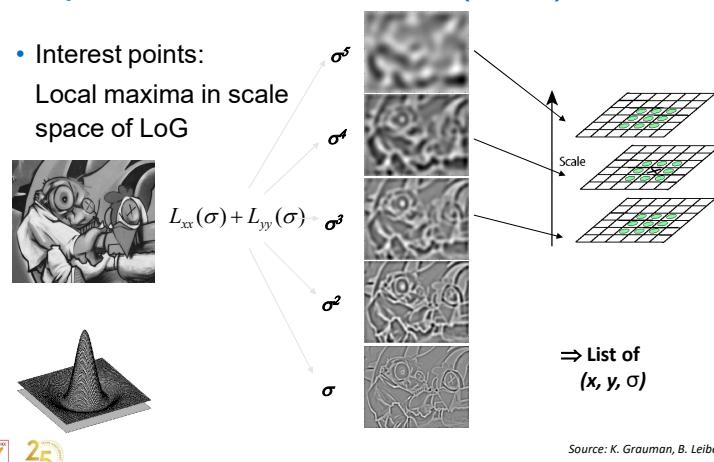
Source: Lana Lazebnik



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Laplacian-of-Gaussian (LoG)

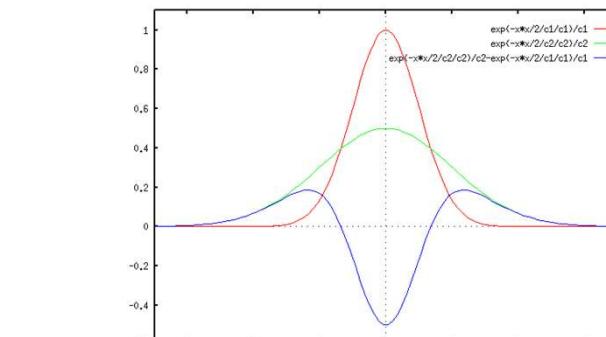
- Interest points:
Local maxima in scale space of LoG



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

LoG-DoG

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

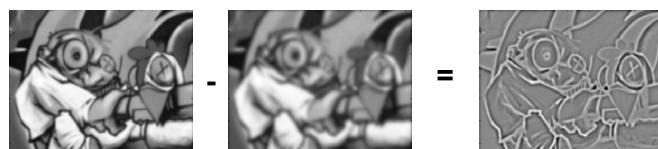


SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

LoG-DoG

- Approximate LoG with Difference-of-Gaussian (DoG):
 - Blur image with σ Gaussian kernel
 - Blur image with $k\sigma$ Gaussian kernel
 - 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, $k = \sqrt{2}$)

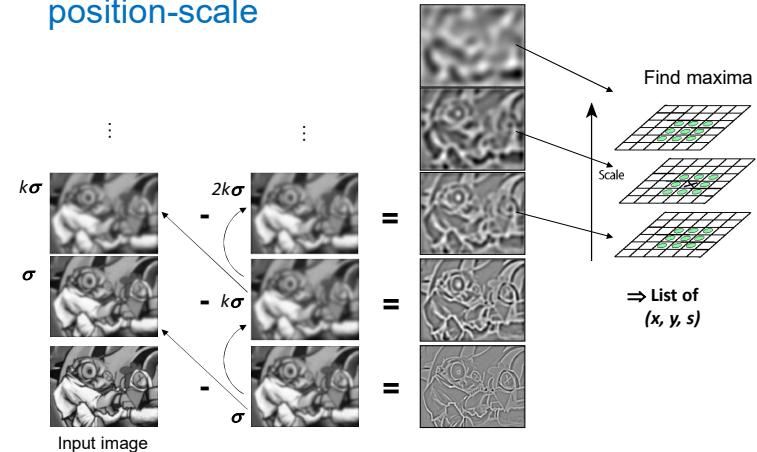


Source: K. Grauman, B. Leibe



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Tìm cực đại địa phương của DoG theo position-scale

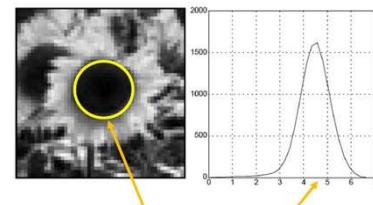
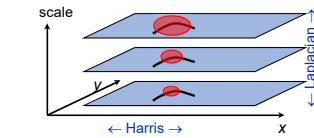


Harris-Laplacian

- Harris-Laplacian¹**

Find local maximum of:

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



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

²D Lowe. "Distinctive Image Features from Scale-Invariant Keypoints". IJCV 2004



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

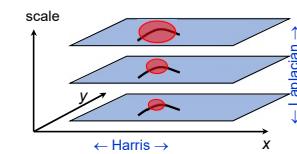
71

Scale Invariant Detectors

- Harris-Laplacian¹**

Find local maximum of:

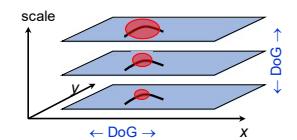
- Harris corner detector in space (image coordinates)
- Laplacian in scale



- SIFT (D.Lowe)²**

Find local maximum of:

- Difference of Gaussians in space and scale



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

²D Lowe. "Distinctive Image Features from Scale-Invariant Keypoints". IJCV 2004



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

72

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

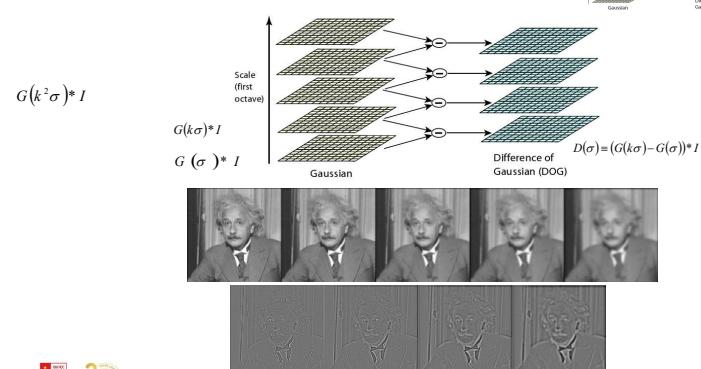


SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

73

DoG (SIFT) Detector

- DoG at multi-octaves**

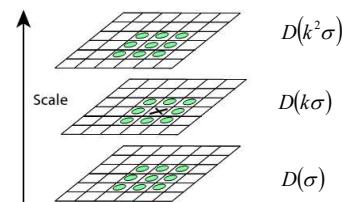


SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

74

DoG (SIFT) Detector

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

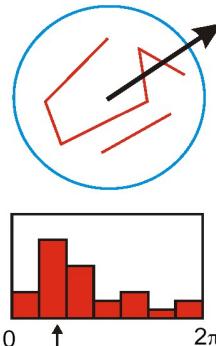


SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

75

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 $(x, y, scale, orientation)$



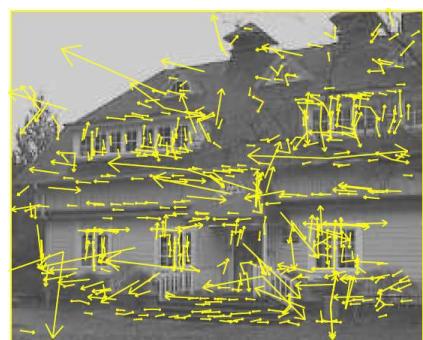
If 2 major orientations, use both.



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

76

DoG (SIFT) Detector

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

SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

77

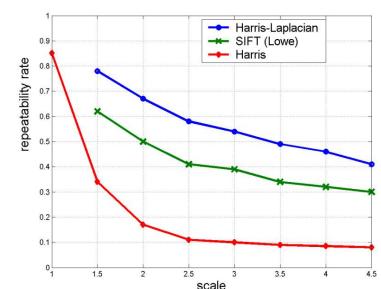
Source: Distinctive Image Features from Scale-Invariant Keypoints – IJCV 2004

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



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

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.*

Trích chọn đặc trưng: Đặc trưng tốt?

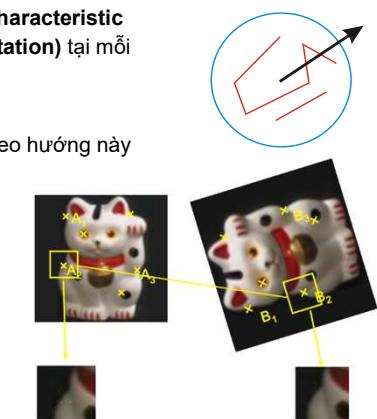
- Gọn (Compact)
- Bất biến với
 - Một số phép biến đổi hình học
 - Góc nhìn camera
 - Điều kiện chiếu sáng
- Ví dụ:
 - **SIFT**, SURF([Speeded Up Robust Features](#)), [PCA-SIFT](#), ...
 - LBP, BRISK, MSER and FREAK, ...

Feature extraction

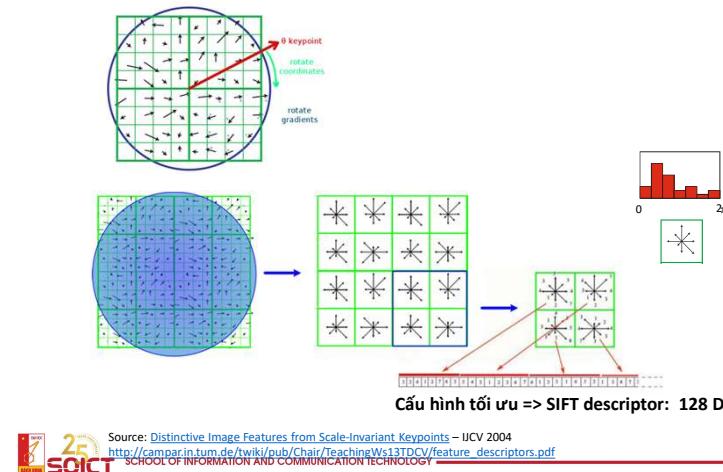
- Global features
- **Local features**
 - Interest point detector
 - Local descriptor
 - Matching

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



Bộ mô tả SIFT

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

→ Đảm bảo bất biến với sự thay đổi điều kiện chiếu sáng

SIFT: Distinctive Image Features from Scale-Invariant Keypoints – IJCV 2004

Sensitivity to number of histogram orientations

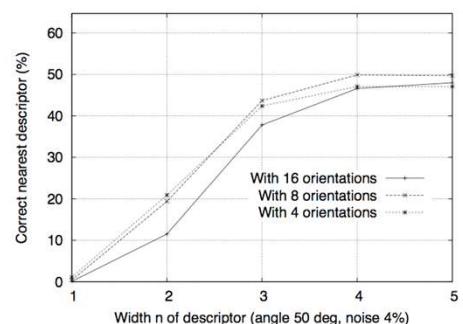
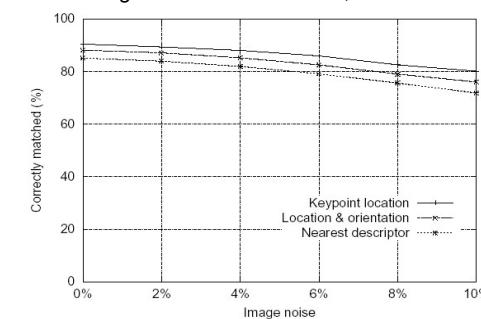


Figure 8: This graph shows the percent of keypoints giving the correct match to a database of 40,000 keypoints as a function of width of the $n \times n$ keypoint descriptor and the number of orientations in each histogram. The graph is computed for images with affine viewpoint change of 50 degrees and addition of 4% noise.

SOICT 25 SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Feature stability to noise

- Match features after random change in image scale & orientation, with differing levels of image noise
- Find nearest neighbor in database of 30,000 features



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

Feature stability to affine change

- Match features after random change in image scale & orientation, with 2% image noise, and affine distortion
- Find nearest neighbor in database of 30,000 features

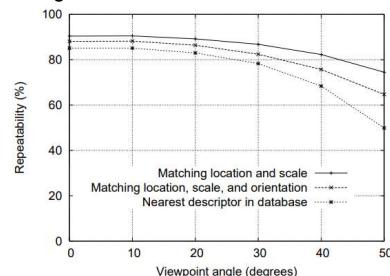


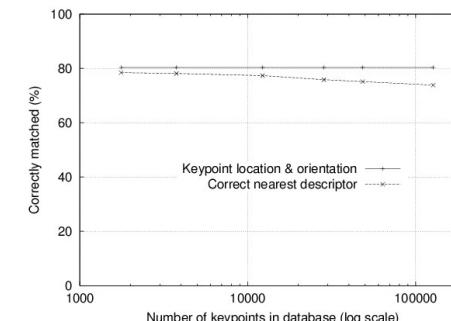
Figure 9: This graph shows the stability of detection for keypoint location, orientation, and final matching to a database as a function of affine distortion. The degree of affine distortion is expressed in terms of the equivalent viewpoint rotation in depth for a planar surface.



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Distinctiveness of features

- Vary size of database of features, with 30 degree affine change, 2% image noise
- Measure % correct for single nearest neighbor match



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Một số bộ mô tả khác

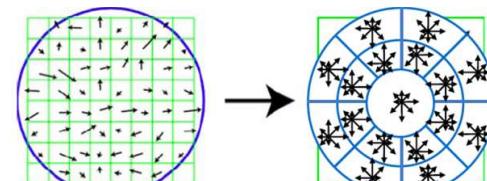
- GLOH, SURF([Speeded Up Robust Features](#)), [PCA-SIFT](#), ...
- LBP, ..
- BREF, BRISK, ORB, FREAK, KAZE, LDAHash



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

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 L_1 , L_2 , 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.



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)
 - → 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 → ratio gần 1
 - Chỉ giữ lại những cặp có ratio nhỏ



94

Ratio of distances reliable for matching

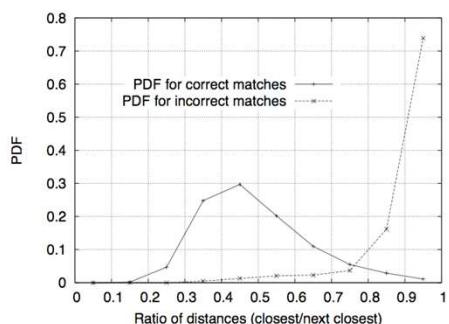


Figure 11: The probability that a match is correct can be determined by taking the ratio of distance from the closest neighbor to the distance of the second closest. Using a database of 40,000 keypoints, the solid line shows the PDF of this ratio for correct matches, while the dotted line is for matches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



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

Kiểm tra chéo

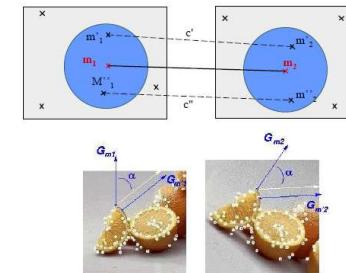
- Ghép cặp (fa, fb) nếu:
 - fb là điểm ghép cặp tốt nhất trong ảnh I2 với fa của I1
 - fa là điểm ghép cặp tốt nhất trong ảnh I1 với fb của I2



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Thêm ràng buộc không gian

- Spatial constraints between neighborhood pixels



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Source: from slides of Valérie Gouet-Brunet

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

$$H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ 0 & 0 & 1 \end{bmatrix}$$

- Six variables
 - Each point give two equations
 - → at least three points
- Least squares

- RANSAC: refinement of matches

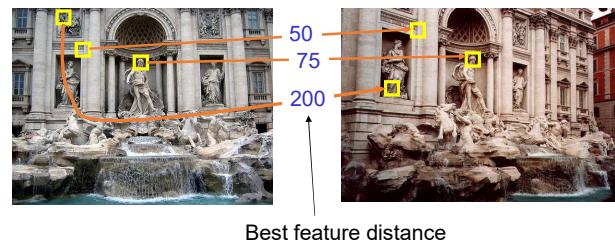
$$\left\| \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} - H \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} \right\|_2$$



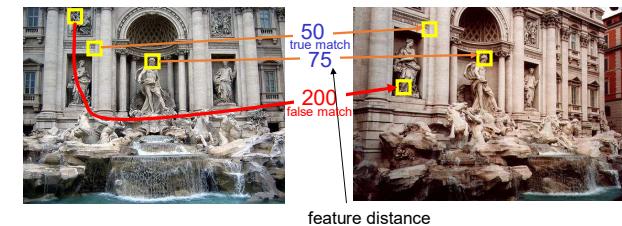
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

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 ?

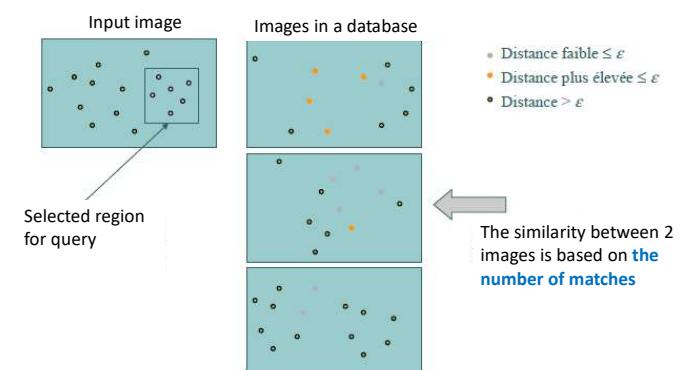
– **False positives** = số cặp ghép sai

- 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



Bài toán tối ưu

- Transportation problem

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

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

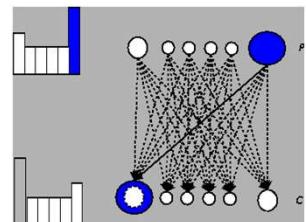
$d(I_1, I_2) = ???$

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

$$f_{ij} \geq 0$$

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

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



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



<http://vellum.cz/~mikc/oss-projects/CarRecognition/doc/dp/node29.html>

103

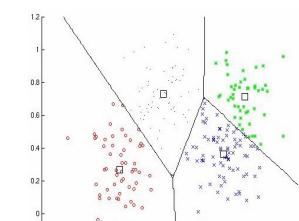
Từ điển



1. Extracting local features from a set of images



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



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



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

105

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

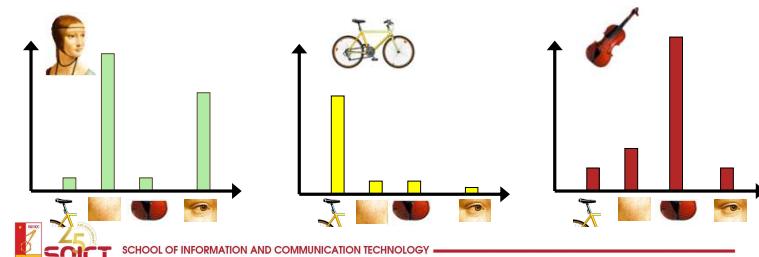


SOICT SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

104

Mô hình túi từ: các bước chính

- Trích chọn đặc trưng
- Tạo từ điển
- Gán đặc trưng cục bộ thành các từ
- Biểu diễn ảnh = vector tần suất xuất hiện



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

Object detection/recognition

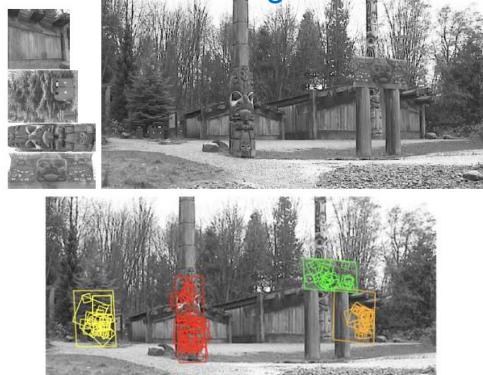


Figure 13: This example shows location recognition within a complex scene. The training images for locations are shown at the upper left and the 640x315 pixel test image taken from a different viewpoint is on the upper right. The recognized regions are shown on the lower image, with keypoints shown as 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 Lowe, Distinctive Image Features from Scale-Invariant Keypoints, IJCV 2004

Image Panoramas



Side credit: Darya Frolova, Denis Simakov



Image Panoramas



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

 SOICT SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Slide credit: Darya Frolova, Denis Simakov

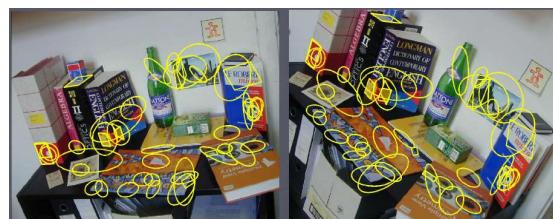
Automatic mosaicing



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

 SOICT SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Wide baseline stereo



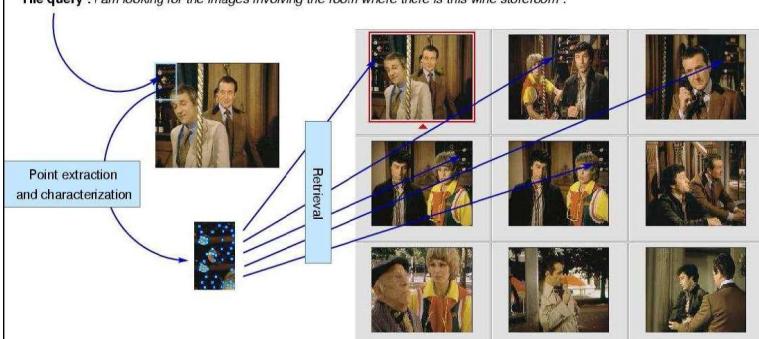
[Image from T. Tuytelaars ECCV 2006 tutorial]

 SOICT

SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

CBIR: partial retrieval

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



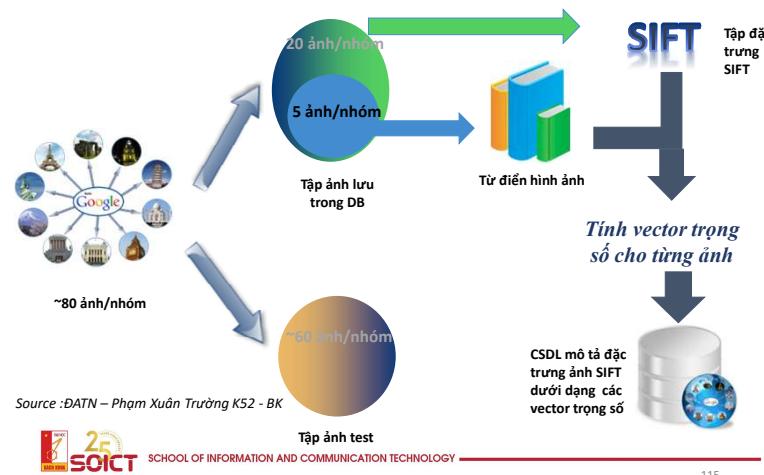
 SOICT

SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

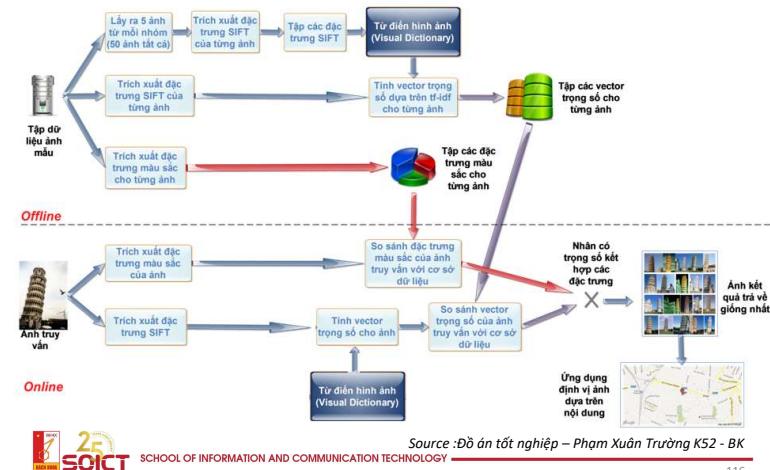
Source. <http://www-rocq.inria.fr/media>

114

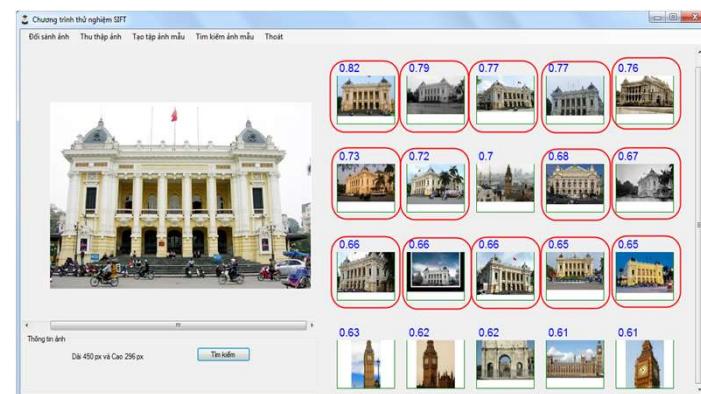
CBIR: BoW with SIFT + histogram



CBIR: BoW with SIFT + histogram



CBIR: BoW with SIFT + histogram



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](http://www.ubc.ca))
- SURF:** <https://people.ee.ethz.ch/~surf/eccv06.pdf>
- Harris corner detector: https://home.cis.rit.edu/~cnspci/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>

