



HA NOI UNIVERSITY OF SCIENCE AND TECHNOLOGY
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Computer Vision

Chapter 4: Feature detection and Image matching

1

Chapter 4 - Content

- Edge
 - Edge Detection:
 - Role of edges & general approach of edge detection
 - Image gradient
 - Edge detectors: Sobel, Prewitt, Canny detector
 - Edge Linking
- Feature extraction and matching
 - Global features
 - Local features
 - Matching and Applications



VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

2

Local features vs Global features

- Two types of features are extracted from the image:
 - local and global features (descriptors)
- **Global features**
 - Describe the **image as a whole** to generalize the entire object
 - Include contour representations, shape descriptors, and texture features
 - Examples: Invariant Moments (Hu, Zernike), Histogram Oriented Gradients (HOG), PHOG, and Co-HOG,...
- **Local feature:**
 - the local features **describe the image patches** (key points in the image) of an object
 - represents the texture/color in an image patch
 - Examples: SIFT, SURF, LBP, BRISK, MSER and FREAK, ...



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

3

Feature extraction

- **Global features**
 - Color / Shape / Texture
- Local features
- Matching and Applications



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

4

Global features

How to distinguish these objects?






SOICT 25th Anniversary SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

5

Types of global features

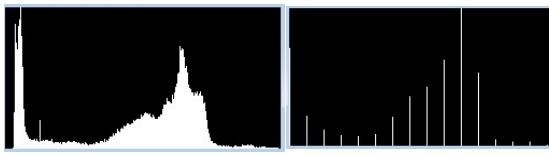
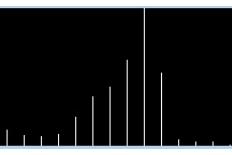
- Color features, descriptors
- Texture features
- Contour representation, Shape features

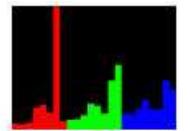
SOICT 25th Anniversary SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

6

Color features

Histogram

SOICT 25th Anniversary SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

7

Distance / Similarity

- L1 ou L2 (euclidian) distances are often used

$$d_{L1}(H, G) = \sum_{i=1}^N |h_i - g_i|$$

- Histogram intersection

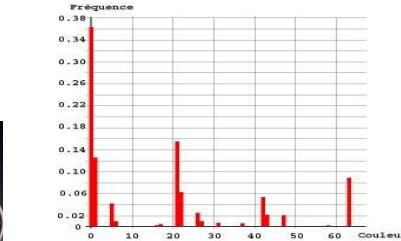
$$\cap(H, G) = \frac{\sum_i \min(h_i, g_i)}{\sum_i g_i}$$

SOICT 25th Anniversary SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

8

Advantages of histogram

- Invariant to basic geometric transformations:
 - Rotation
 - Translation
 - Zoom (Scaling)

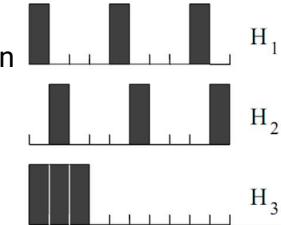


SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

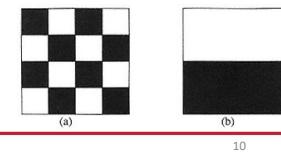
9

Some inconveniences

- The **similarity between colors in adjacent colors (bin)** is not taken into account



- The **spatial distribution** of pixel values is **not considered**: 2 different images may have the same histogram

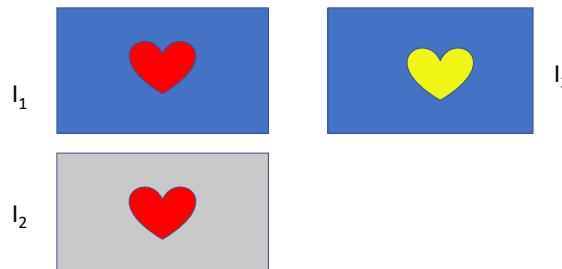


SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

10

Some inconveniences

- **Background effect:** $d(I_1, I_2) \neq d(I_1, I_3)$



- **Color representation dependency** (color space), device dependency, ...

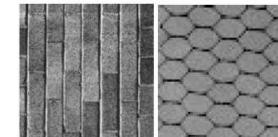
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

11

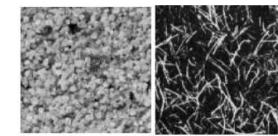
Texture features

- A **texture** can be defined as
 - a region with variation of intensity
 - as a spatial organization of pixels

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



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



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

12

Texture features

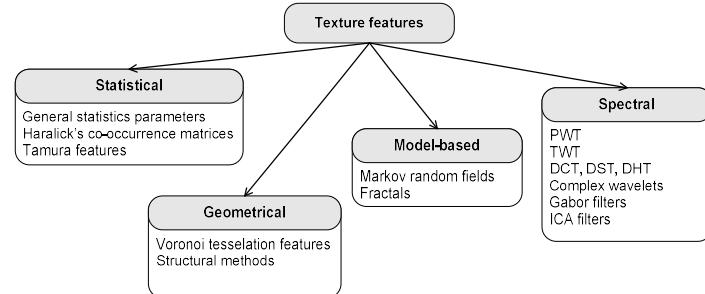
- There are several methods for analyzing textures:
 - First order statistics
 - Statistics on histogram
 - Co-occurrence matrices
 - Searching patterns
 - Frequential analysis
 - Gabor filter
 - ...
- The most difficult is to find a good representation (good parameters) for each texture



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

13

Texture features



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

14

GLCM (Grey Level Co-occurrence Matrices)

- The idea here is to identify **gray level that repeat themselves** given a **distance** and a **direction**
 - Co-occurrence matrices (Haralick)
- **Matrix of size $Ng \times Ng$**
 - Ng is the number of gray level in the image (256x256)
 - We often reduce that number to 8x8, 16x16 or 32x32
- Many matrices, one for each distance and direction
 - **Distance** : 1, 2, 3 (,4, ...)
 - **Direction** : $0^\circ, 45^\circ, 90^\circ, 135^\circ, \dots$
- Processing time can be very long



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

15

GLCM

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

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



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

16

16

GLCM: Example (1)

- Example on how to compute these matrices:

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

Image

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

Matrix for distance=1
and direction=0°

We loop over the image and for each pair of pixels following the given distance and orientation, we increment the co-occurrence matrix



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

17

GLCM: Example (2)

- Example on how to compute these matrices:

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

Image

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

Matrix for distance=1 and
direction=0°

Pair of neighbor pixels (1,4)



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

18

GLCM: Example (3)

- Example on how to compute these matrices:

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

Image

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

Matrix for distance=1 and
direction=0°

Pair of neighbor pixels (4,4)



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

19

GLCM: Example (4)

- Example on how to compute these matrices:

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

Image

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

Matrix for distance=1 and
direction=0°

...and so on (after 2 lines)



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

20

GLCM: Example (5)

- Example on how to compute these matrices (final):

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°

	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°

...and so on for each matrix (several matrices at the end)



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

21

GLCM

- Most important/popular parameters computed from GLCM:

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

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

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

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



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

is.hust.edu.vn

22

Haralick features using GLCM

- Haralick features:

- For each GLCM, we can compute up to **14 (13)** parameters characterizing the texture, of which the most important : mean, variance, energy, inertia, entropy, inverse differential moment
- Ref: <http://haralick.org/journals/TexturalFeatures.pdf>



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

is.hust.edu.vn

23

Invariances

- Rotation?

- Average on all directions

- Scaling?

- Multi-resolutions



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

24

Texture features comparision

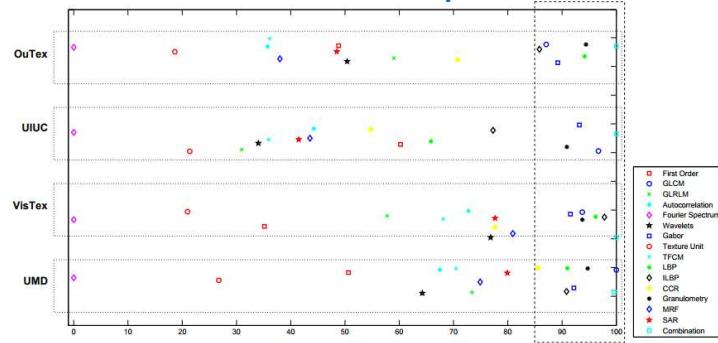


Figure 10. Comparison of results for all considered features on each data set. Each horizontal dotted box displays results for one data set. The x -axis shows normalized results (between 0 and 100) according to the minimum and maximum classification rates achieved for each data set. The vertical slashed box on the right-hand side selects the feature extraction methods that achieved the best results throughout all data sets.

Source : William Robson Schwartz et al. Evaluation of Feature Descriptors for Texture Classification – 2012 JEL
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY is.hust.edu.vn 25

25

Shape features

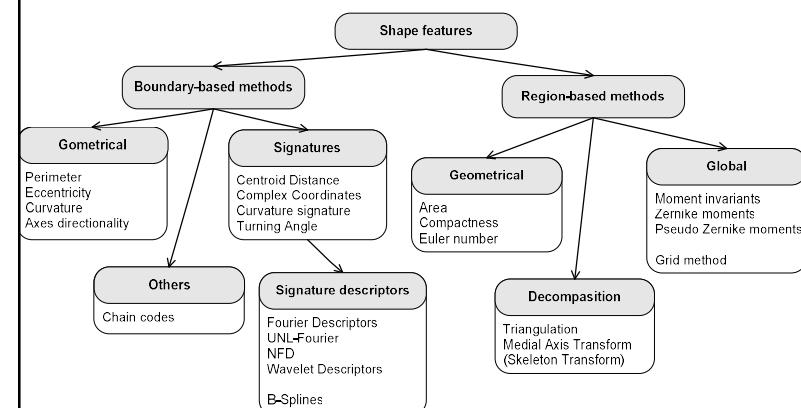
- Contour-based features
 - Chain coding, polygon approximation, geometric parameters, angular profile, surface, perimeter, ...
 - Region based:
 - Invariant moments, ...



26

25

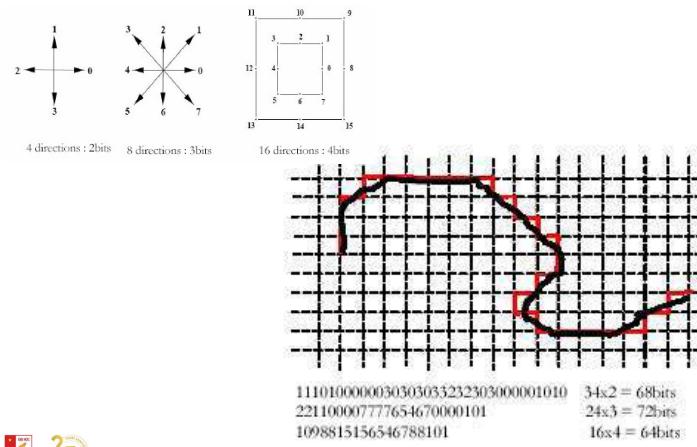
Shape features



• SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

2

Examples: Freeman chain coding

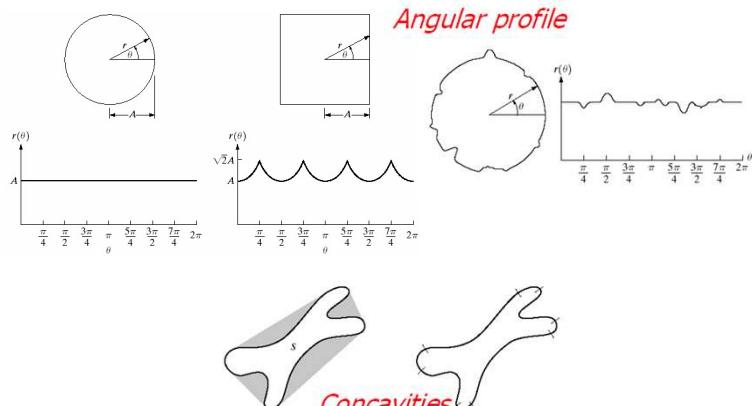


28

27

28

Examples: angular profile, ...



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

29

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

30

Invariant moments (Hu's moments)

invariant to
translation,
scale, and
rotation, and
reflection

Change for
image
reflection

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

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

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

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



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

31

Examples : 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	2.66884	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	z	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>

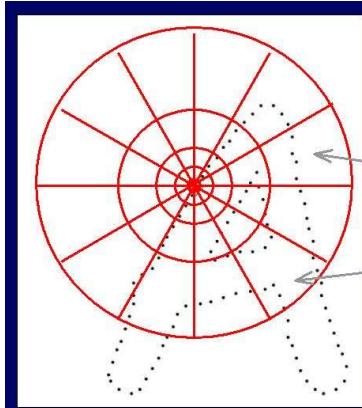
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

32

31

32

Shape Context



Count the number of points inside each bin, e.g.:

Count = 4

⋮

Count = 10

Compact representation of distribution of points relative to each point



https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/shape/sc_digits.html

SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

is.hust.edu.vn

33

Feature extraction

- Global features
- **Local features and matching**
 - Local features
 - Interest point detector
 - Local descriptors
 - Matching and applications



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

34

Local features

- Why local features?
- Image matching: A challenging problem



Source: CS131 - Juan Carlos Niebles and Ranjay Krishna

35

Image matching

A challenging problem



by Diva Sian



by swashford

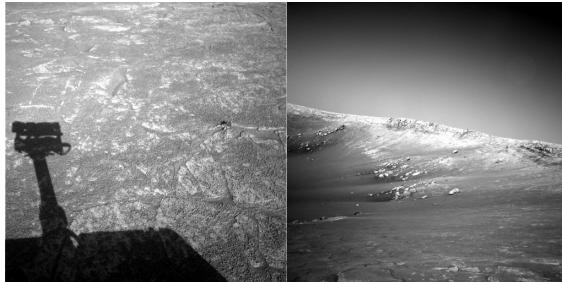


by scgbt

Slide credit: Steve Seitz

36

Harder Still?



NASA Mars Rover images

Slide credit: Steve Seitz

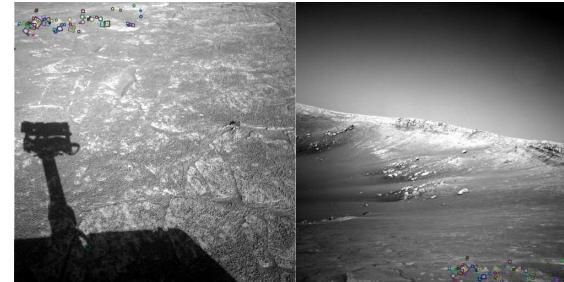


SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

37

37

Answer Below (Look for tiny colored squares)

NASA Mars Rover images with SIFT feature matches
(Figure by Noah Snavely)

Slide credit: Steve Seitz



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

38

38

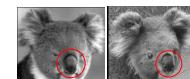
Motivation for using local features

- Global representations have major limitations
- Instead, describe and match only local regions
- Increased robustness to

– Occlusions



– Articulation



– Intra-category variations



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

39

39

Local features and alignment

- We need to match (align) images
- Global methods sensitive to occlusion, lighting, parallax effects. So look for local features that match well.
- How would you do it by eye?



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

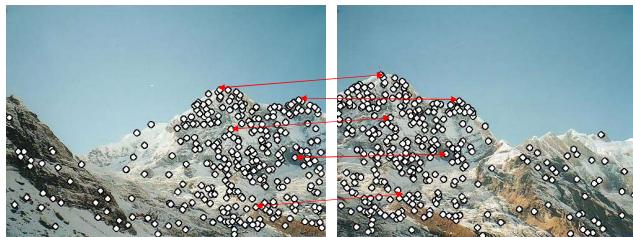
[Darya Frolova and Denis Simakov]

40

40

Local features and alignment

- Detect feature points in both images
- Find corresponding pairs



[Darya Frolova and Denis Simakov]

41

Local features and alignment

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

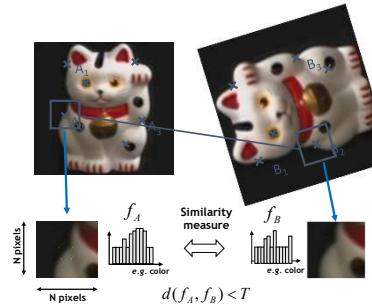


[Darya Frolova and Denis Simakov]

[Darya Frolova and Denis Simakov]

42

Local features for image matching



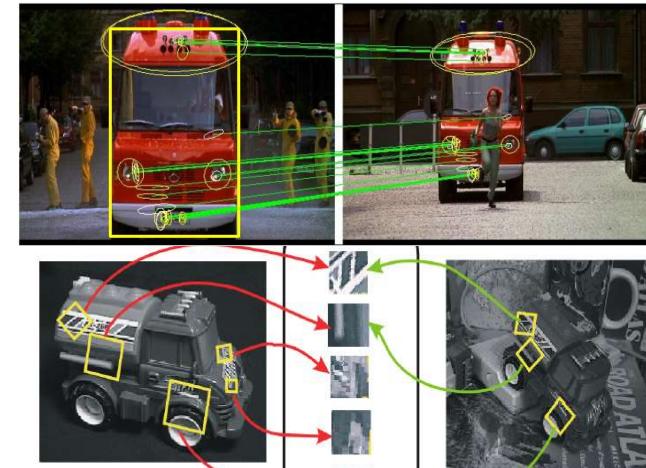
1. Find a set of distinctive key-points
2. Define a region around each keypoint
3. Extract and normalize the region content
4. Compute a local descriptor from the normalized region
5. Match local descriptors



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Slide credit: Bastian Leibe

43



Source : Jim Little, Lowe: features, UBC.

44

43

44

Local

- Objectifs:
 - Look for similar objects/regions
 - Partial query

Look for pictures that contain sunflowers



- Solution:
 - Describing local regions
 - Adding spatial constraints if need



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

45

Local feature extraction

- Local features: how to determine image patches / local regions

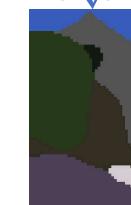
Dividing into patches with regular grid



Without knowledge about image content



Keypoint detection



Based on the content of image

SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

46

Common Requirements

- 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!

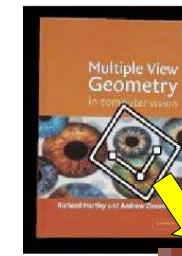


SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Slide credit: Darya Frolova, Denis Simakov

47

Invariance: Geometric Transformations



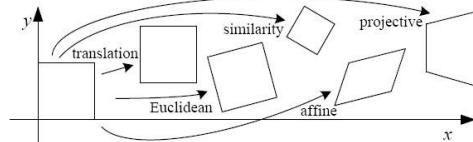
Slide credit: Steve Seitz



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

48

Invariance: Geometric Transformations



Levels of Geometric Invariance

Slide credit: Steve Seitz

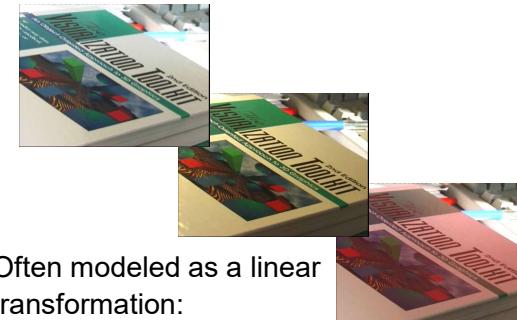
49



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

49

Invariance: Photometric Transformations



- Often modeled as a linear transformation:
 - Scaling + Offset



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Slide credit: Tinne Tuytelaars

50

50

Requirements

- 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

51

51

Main questions

- **Where** will the interest points come from?
 - What are salient features that we'll *detect* in multiple views?
- How to **describe** a local region?
- How to **establish correspondences**, i.e., compute matches?



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

52

52

Feature extraction

- Global features
- **Local features and matching**
 - Local features
 - **Interest point detector**
 - Local descriptors
 - Matching and applications

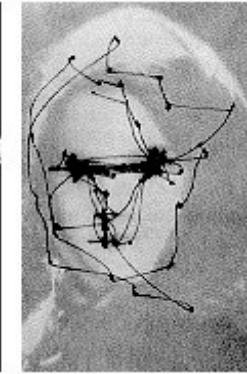


SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

53

53

Interest points: why and where?



Yarbus eye tracking



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

54

54

Interest points: why and where?

- Where will the interest points come from?

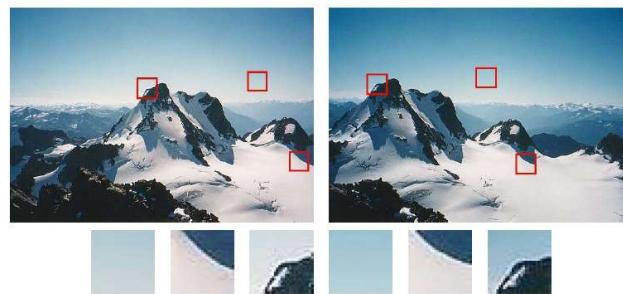


Figure 4.3: Image pairs with extracted patches below. Notice how some patches can be localized or matched with higher accuracy than others.



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

55

55

Keypoint Localization

- Goals:
 - Repeatable detection
 - Precise localization
 - Interesting content

⇒ Look for two-dimensional signal changes



Slide credit: Bastian Leibe

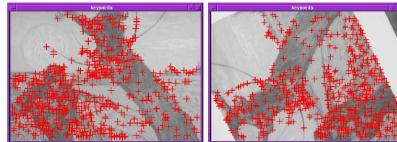


SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

56

56

Finding Corners



- Key property:
 - In the region around a corner, image gradient has two or more dominant directions
- Corners are *repeatable* and *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

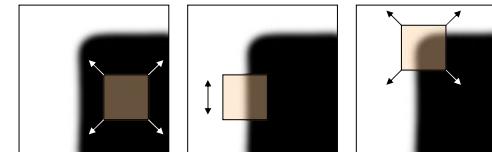
Slide credit: Svetlana Lazebnik

57

Corners as distinctive interest points

Design criteria

- We should easily recognize the point by looking through a small window (*locality*)
- Shifting the window in *any direction* should give a *large change* in intensity (*good localization*)



"flat" region:
no change in
all directions

"edge":
no change along
the edge
direction

"corner":
significant change
in all directions



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Slide credit: Alyosha Efros

58

Corners versus edges



$$\sum I_x^2 \rightarrow \text{Large} \quad \sum I_y^2 \rightarrow \text{Large}$$

Corner



$$\sum I_x^2 \rightarrow \text{Small} \quad \sum I_y^2 \rightarrow \text{Large}$$

Edge



$$\sum I_x^2 \rightarrow \text{Small} \quad \sum I_y^2 \rightarrow \text{Small}$$

Nothing



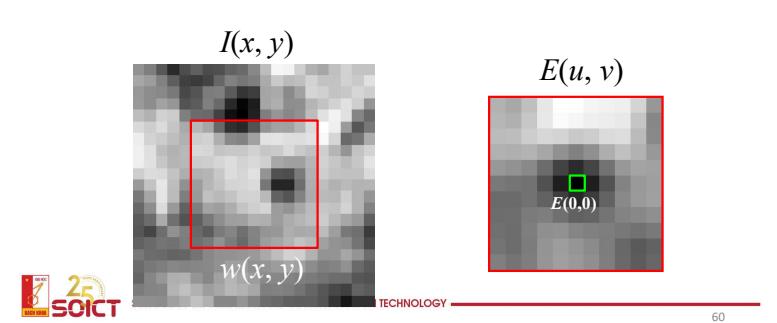
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

59

Corner Detection by Auto-correlation

Change in appearance of window $w(x,y)$ for shift $[u,v]$:

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

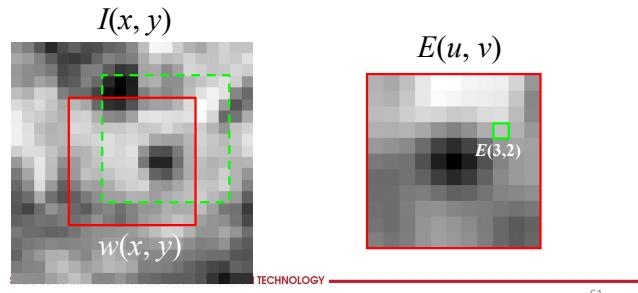


60

Corner Detection by Auto-correlation

Change in appearance of window $w(x,y)$ for shift $[u,v]$:

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



61

Corner Detection by Auto-correlation

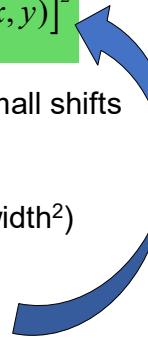
Change in appearance of window $w(x,y)$ for shift $[u,v]$:

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

We want to discover how E behaves for small shifts

But this is **very slow** to compute naively.
 $O(\text{window_width}^2 * \text{shift_range}^2 * \text{image_width}^2)$

$O(11^2 * 11^2 * 600^2) = 5.2$ billion of these



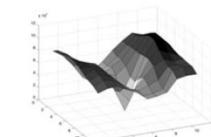
Corner Detection by Auto-correlation

Change in appearance of window $w(x,y)$ for shift $[u,v]$:

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

We want to discover how E behaves for small shifts

But we know the response in E that we are looking for – **strong peak**. → Approximation



63

Local quadratic approximation of $E(u,v)$ in the neighborhood of $(0,0)$ is given by the **second-order Taylor expansion**:

$$E(u,v) \approx E(0,0) + [u \ v] \begin{bmatrix} E_u(0,0) \\ E_v(0,0) \end{bmatrix} + \frac{1}{2} [u \ v] \begin{bmatrix} E_{uu}(0,0) & E_{uv}(0,0) \\ E_{vu}(0,0) & E_{vv}(0,0) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}$$

Notation: partial derivative

64

we are looking for **strong peak**

$$E(u, v) \approx E(0,0) + [u \ v] \begin{bmatrix} E_u(0,0) \\ E_v(0,0) \end{bmatrix} + \frac{1}{2} [u \ v] \begin{bmatrix} E_{uu}(0,0) & E_{uv}(0,0) \\ E_{vu}(0,0) & E_{vv}(0,0) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}$$

↑
Ignore function value; set to 0 ↑
Ignore first derivative, set to 0 ↑
Just look at shape of second derivative

 SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

65

Corner Detection: Mathematics

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

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

 SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

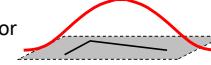
66

Harris Detector formulation

Change of intensity for the shift $[u, v]$:

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

↑
Window function ↑
Shifted intensity ↑
Intensity

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

1 in window, 0 outside Gaussian

Source: R. Szeliski

 SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

67

Harris detector formulation

- This measure of change can be approximated by:

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

where M is a 2×2 matrix computed from image derivatives:

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

Gradient with respect to x, times gradient with respect to y

Sum over image region – the area we are checking for corner

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

Slide credit: Rick Szeliski

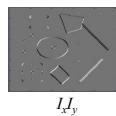
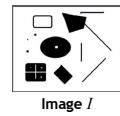
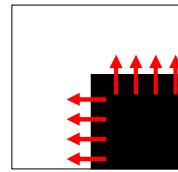
 SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

68

What does this matrix reveal?

- First, let's consider an axis-aligned corner:

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



25

SOICT

SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Slide credit: David Jacobs

69

What does this matrix reveal?

- First, let's consider an axis-aligned corner:

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



- This means:

- Dominant gradient directions align with x or y axis
- If either λ is close to 0, then this is not a corner, so look for locations where both are large.

- What if we have a corner that is not aligned with the image axes?

25

SOICT

SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Slide credit: David Jacobs

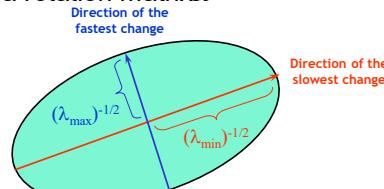
70

General case

- Since M is symmetric, we have $M = R^{-1} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} R$

(Eigenvalue decomposition)

- We can visualize M as an ellipse with axis lengths determined by the eigenvalues and orientation determined by a rotation matrix R

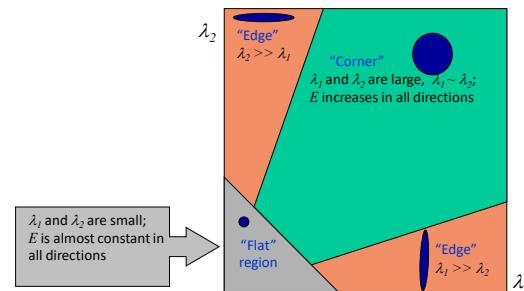


adapted from Darya Frolova, Denis Simakov

71

Interpreting the eigenvalues

- Classification of image points using eigenvalues of M :



Slide credit: Kristen Grauman

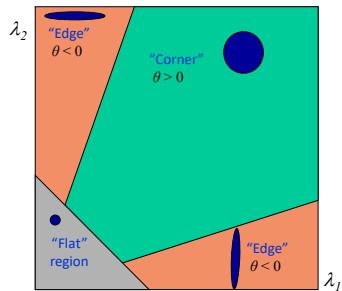
72

71

72

Corner response function

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



- Fast approximation
 - Avoid computing the eigenvalues
 - α : constant (0.04 to 0.06)



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Slide credit: Kristen Grauman

73

Window Function $w(x,y)$

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

- Option 1: uniform window

- Sum over square window



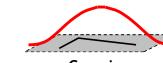
$$M = \sum_{x,y} \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}$$

1 in window, 0 outside

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



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

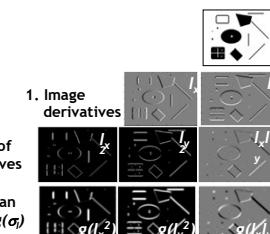
Slide credit: Bastian Leibe

74

Summary: Harris Detector [Harris88]

- Compute second moment matrix (autocorrelation matrix)

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



- Compute corner response

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



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

75

Harris Detector: Workflow



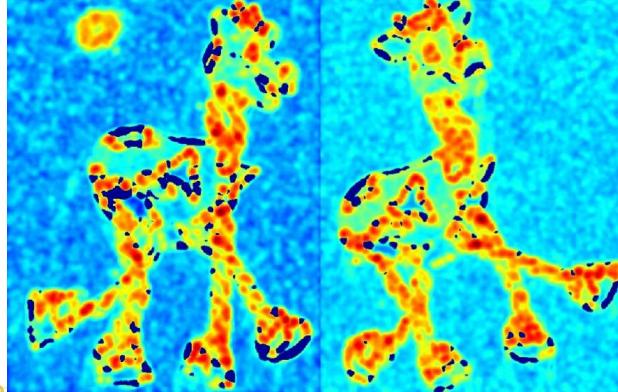
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Slide adapted from Darya Frolova, Denis Simakov

76

Harris Detector: Workflow

computer corner responses θ

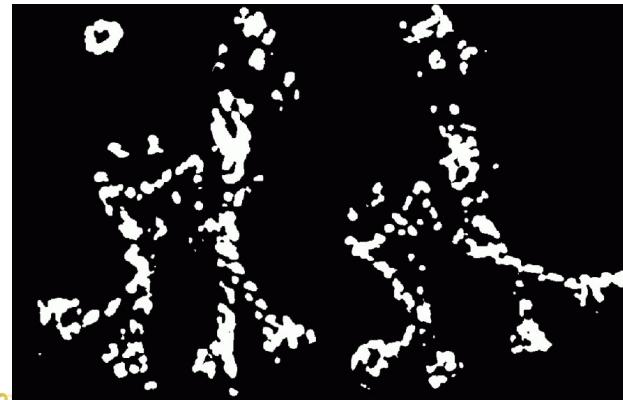


SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

77

Harris Detector: Workflow

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



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Slide adapted from Darya Frolova, Denis Simakov

78

Harris Detector: Workflow

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



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

79

Harris Detector: Workflow

Resulting Harris points



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

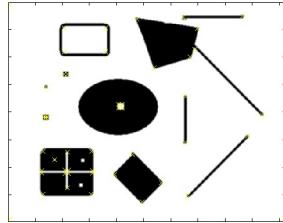
Slide adapted from Darya Frolova, Denis Simakov

80

79

80

Harris Detector – Responses [Harris88]



Effect: A very precise corner detector.



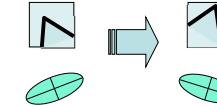
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

81

Slide credit: Krystian Mikolajczyk

Harris Detector: Properties

- Translation invariance?
- Rotation invariance?



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

Corner response θ is invariant to image rotation

Slide credit: Kristen Grauman



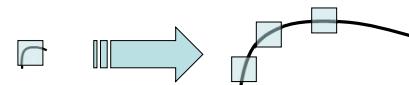
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

82

82

Harris Detector: Properties

- Translation invariance
- Rotation invariance
- Scale invariance?



Corner

All points will be classified as edges!

Not invariant to image scale!

Slide credit: Kristen Grauman



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

83

83

Detectors Scale invariance: how to?

- Exhaustive search
- Invariance
- Robustness



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

84

84

Exhaustive search

- Multi-scale approach



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

85

Invariance

- Extract patch from each image individually



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

86

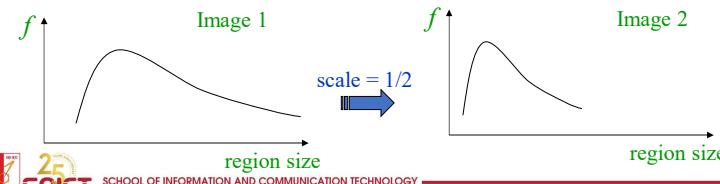
Automatic scale selection

- Solution:

— Design a function on the region, which is “scale invariant” (*the same for corresponding regions, even if they are at different scales*)

Example: average intensity. For corresponding regions (even of different sizes) it will be the same.

— For a point in one image, we can consider it as a function of region size (patch width)



87

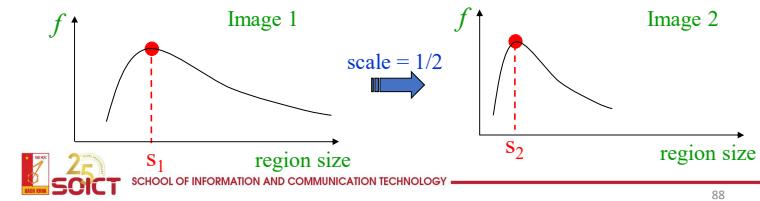
Automatic scale selection

- Common approach:

Take a local maximum of this function

Observation: region size, for which the maximum is achieved, should be *invariant* to image scale.

Important: this scale invariant region size is found in each image **independently!**



88

Automatic Scale Selection



$$f(I_{i_1 \dots i_m}(x, \sigma)) = f(I_{i_1 \dots i_m}(x', \sigma'))$$

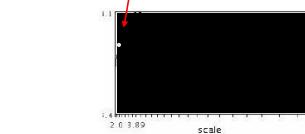
Same operator responses if the patch contains the same image up to scale factor.

K. Grauman, B. Leibe

89

Example

Function responses for increasing scale (scale signature)



25 SOICT SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

25 SOICT SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

90

Example

Function responses for increasing scale (scale signature)

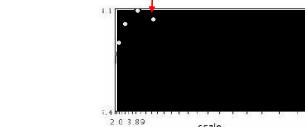


25 SOICT SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

91

Example

Function responses for increasing scale (scale signature)



25 SOICT SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

25 SOICT SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

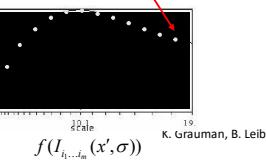
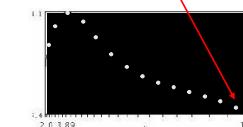
92

91

92

Example

Function responses for increasing scale (scale signature)

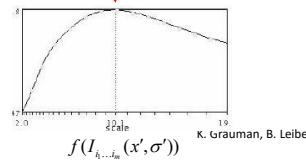
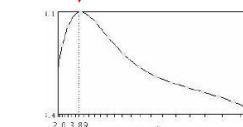


SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

93

Example

Function responses for increasing scale (scale signature)

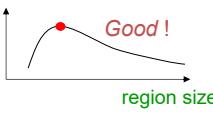
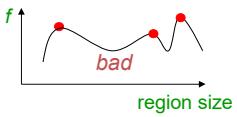
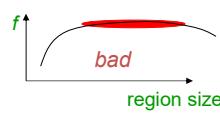


SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

94

Scale Invariant Detection

- A “good” function for scale detection:
has one stable sharp peak



- For usual images: a good function would be one which responds to contrast (sharp local intensity change)



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

95

What is a useful signature function?

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

Kernels:

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

(Laplacian)

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

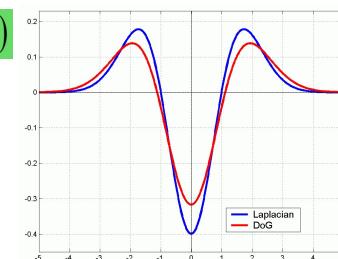
(Difference of Gaussians)

where Gaussian

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



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

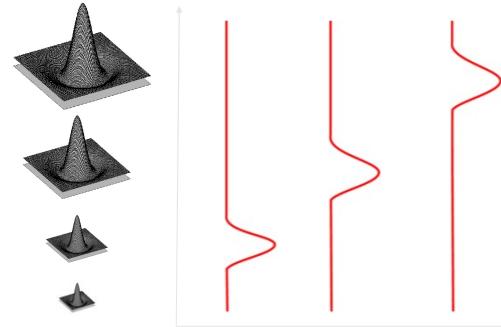


Note: both kernels are invariant to scale and rotation

96

What is a useful signature function?

- Laplacian-of-Gaussian = “blob” detector



Source: K. Grauman, B. Leibe

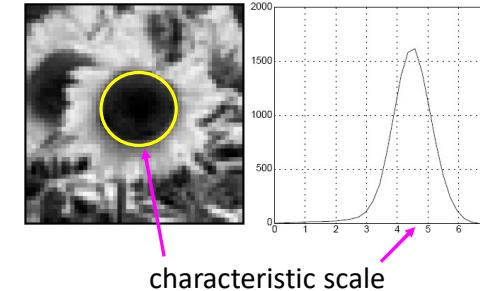


SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

97

Characteristic scale

- We define the **characteristic scale** as 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

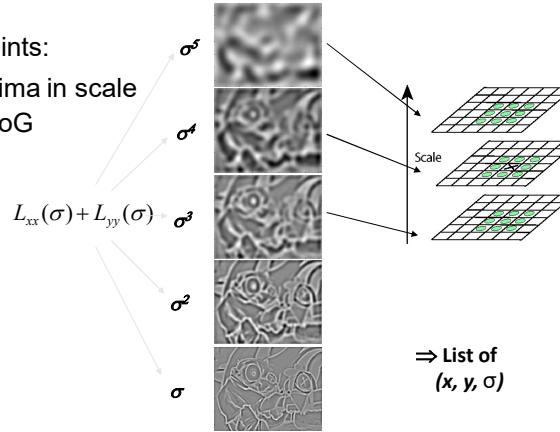
Source: Lana Lazebnik

98

Laplacian-of-Gaussian (LoG)

- Interest points:

Local maxima in scale space of LoG



Source: K. Grauman, B. Leibe



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

99

Example: Scale-space blob detector

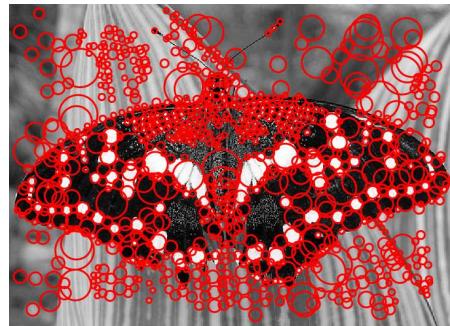


SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Source: Lana Lazebnik

100

Example: Scale-space blob detector



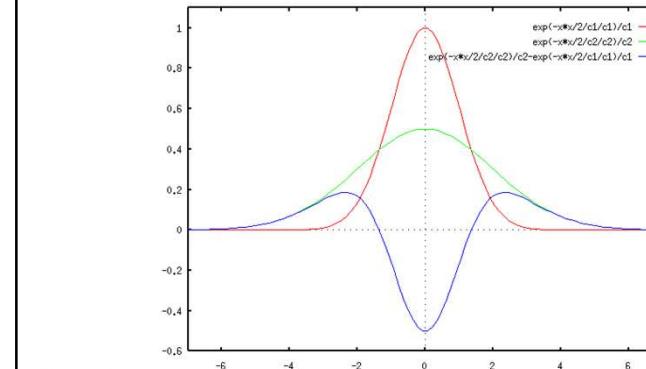
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Source: Lana Lazebnik

101

Alternative approach

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



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

102

Alternative approach

- 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, $K = \sqrt{2}$

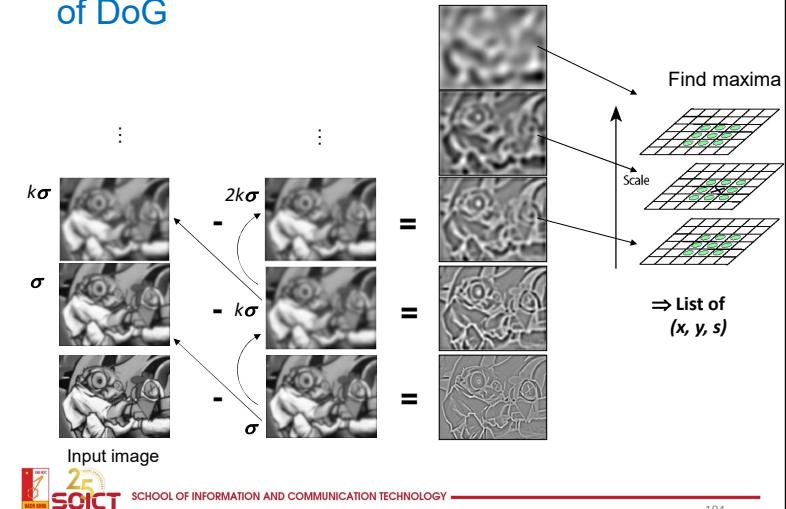


SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Source: K. Grauman, B. Leibe

103

Find local maxima in position-scale space of DoG



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

104

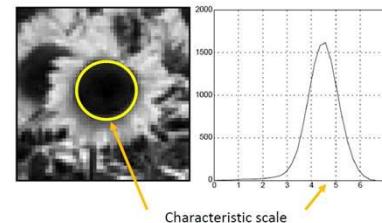
104

Harris-Laplacian

- Harris-Laplacian¹**

Find local maximum of:

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



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

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



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

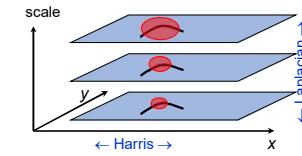
105

Scale Invariant Detectors

- Harris-Laplacian¹**

Find local maximum of:

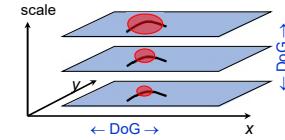
- Harris corner detector in space (image coordinates)
- Laplacian in scale



- SIFT (D.Lowe)²**

Find local maximum of:

- Difference of Gaussians in space and scale



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

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



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

106

DoG (SIFT) keypoint Detector

- DoG at multi-octaves
- Extrema detection in scale space
- Keypoint location
 - Interpolation
 - Removing instable points
- Orientation Assignment

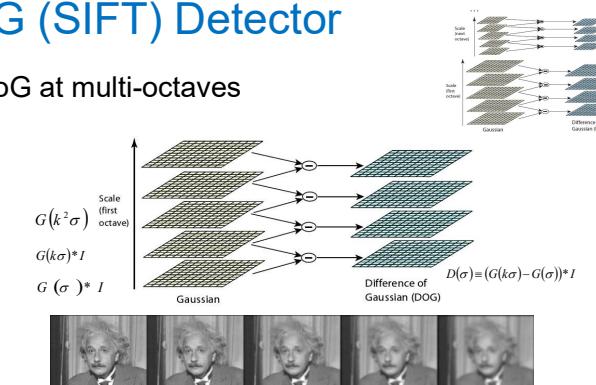


SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

107

DoG (SIFT) Detector

- DoG at multi-octaves**



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

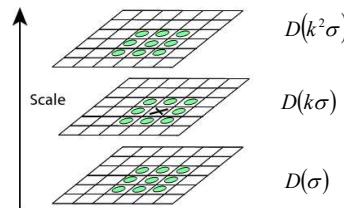
108

107

108

DoG (SIFT) Detector

- Scale-Space Extrema Choose all extrema within 3x3x3 neighborhood



X is selected if it is larger or smaller than all 26 neighbors (its 8 neighbors in the current image and 9 neighbors each in the scales above and below)

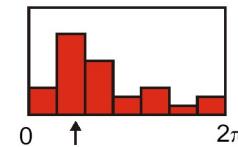
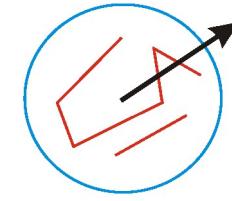


SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

109

DoG (SIFT) Detector

- Orientation assignment
 - Create histogram of local gradient directions at selected scale
 - Assign canonical orientation at peak of smoothed histogram
- Each key specifies stable 2D coordinates $(x, y, \text{scale}, \text{orientation})$



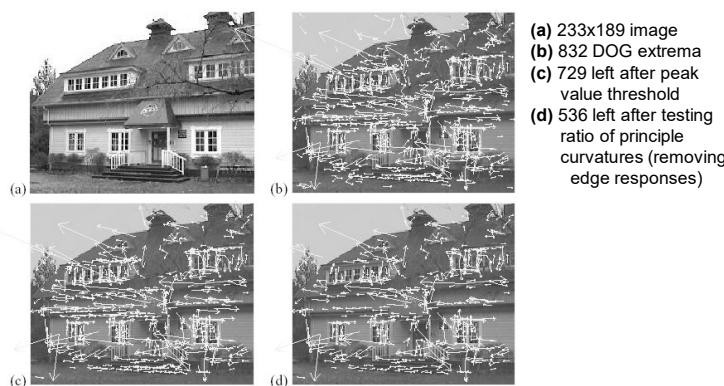
If 2 major orientations, use both.



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

110

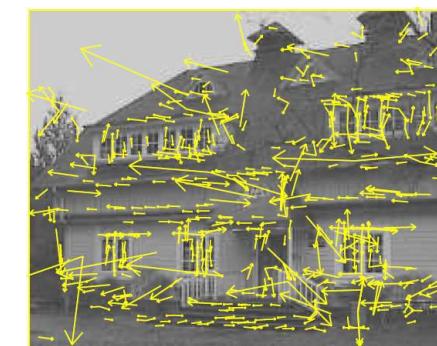
Example of keypoint detection



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

111

DoG (SIFT) Detector



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



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

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

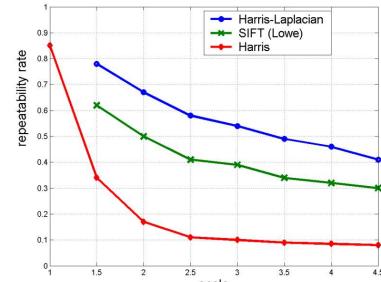
112

112

Scale Invariant Detectors

- Experimental evaluation of detectors w.r.t. scale change

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



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

113

Many existing detectors available

- 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.



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Slide credit: Bastian Leibe

114

Feature extraction

- Global features
- Local features
 - Local features
 - Interest point detectors
- Local descriptors
 - Matching



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

115

Local descriptors

- Compact, good representation for local information
- Invariant
 - Geometric transformations: rotation, translation, scaling,..
 - Camera view change
 - Illumination
- Examples
 - SIFT, SURF([Speeded Up Robust Features](#)), [PCA-SIFT](#), ...
 - LBP, BRISK, MSER and FREAK, ...



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

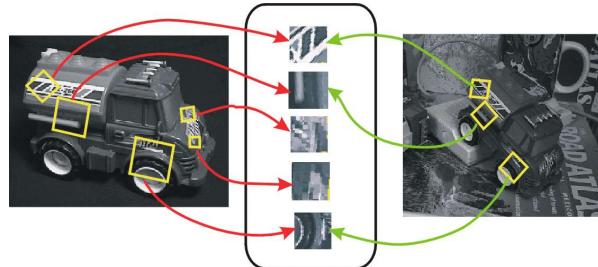
116

115

116

Invariant local features

- Image content is transformed into local feature coordinates that are invariant to translation, rotation, scale, and other imaging parameters



Following slides credit: CVPR 2003 Tutorial on Recognition and Matching Based on Local Invariant Features David Lowe
 SOICT SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

117

Advantages of invariant local features

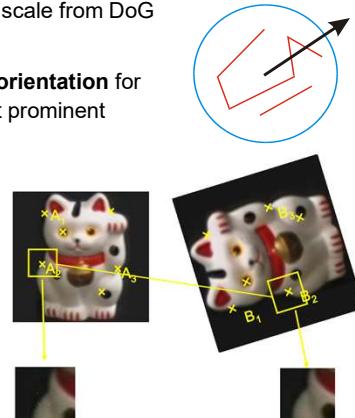
- Locality:**
 - features are local, so robust to occlusion and clutter (no prior segmentation)
- Distinctiveness:**
 - individual features can be matched to a large database of objects
- Quantity:**
 - many features can be generated for even small objects
- Efficiency:**
 - close to real-time performance
- Extensibility:**
 - can easily be extended to wide range of differing feature types, with each adding robustness

SOICT SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

118

Becoming rotation invariant

- We are given a keypoint and its scale from DoG
- We will select a **characteristic orientation** for the keypoint (based on the most prominent gradient there)
- We will describe all features **relative** to this orientation

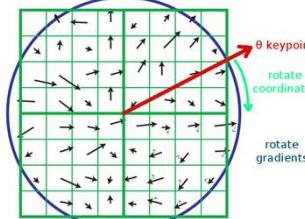


SOICT SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

119

SIFT descriptor formation

- Use the blurred image associated with the keypoint's scale
- Take image gradients over the keypoint neighborhood.
- To become rotation invariant, rotate the gradient **directions AND locations** by (-keypoint orientation)
 - Now we've cancelled out rotation and have gradients expressed at locations **relative** to keypoint orientation θ
 - We could also have just rotated the whole image by $-\theta$, but that would be slower.



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

SOICT SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

120

SIFT descriptor formation

- Using precise gradient locations is fragile. We'd like to allow some "slop" in the image, and still produce a very similar descriptor
- Using Gaussian filter : to **avoid sudden changes** in the descriptor with small changes in the position of the window, and to give **less emphasis to gradients that are far** from the center of the descriptor, as these are most affected by misregistration errors
- Create array of orientation histograms (a 4x4 array is shown)

SIFT: Distinctive Image Features from Scale-Invariant Keypoints – IJCV 2004
Image: Ashish A Gupta, PhD thesis 2013

121

SIFT descriptor formation

- Put the rotated gradients into their local orientation histograms
 - A local orientation histogram has n bins (e.g. 8 as shown).
 - To avoid all boundary affects in which the descriptor **abruptly changes** as a sample shifts smoothly from being within one histogram to another or from one orientation to another
 - interpolation is used to distribute the value of each gradient sample into adjacent histogram bins
 - Also, scale down **gradient contributions** for gradients far from the bin center: a weight of $1-d$: distance of the sample from the central value of the bin
- The SIFT authors found that best results were with 8 orientation bins per histogram and and a 4x4 histogram array → a SIFT descriptor: vector of 128 values

SIFT: Distinctive Image Features from Scale-Invariant Keypoints – IJCV 2004
Image: Ashish A Gupta, PhD thesis 2013

122

SIFT descriptor formation

- Adding robustness to illumination changes:**
 - The descriptor is made of gradients (differences between pixels)
 - **already invariant to changes in brightness** (e.g. adding 10 to all image pixels yields the exact same descriptor)
 - A higher -contrast photo will increase the magnitude of gradients linearly
 - to correct for contrast changes, **normalize the vector** (scale to length 1.0)
 - Very large image gradients are usually from unreliable 3D illumination effects (glare, etc)
 - to reduce their effect, **clamp all values** in the vector to $b \leq 0.2$ (an experimentally tuned value). Then normalize the vector again.
- Result is a vector which is fairly invariant to illumination changes

SIFT: Distinctive Image Features from Scale-Invariant Keypoints – IJCV 2004
Image: Ashish A Gupta, PhD thesis 2013

123

SIFT

- Extraordinarily robust matching technique
 - Can handle changes in viewpoint: up to about 60 degree out of plane rotation
 - Can handle significant changes in illumination
 - Sometimes even day vs. night (below)
 - Fast and efficient—can run in real time

Steve Seitz

SIFT: Distinctive Image Features from Scale-Invariant Keypoints – IJCV 2004
Image: Ashish A Gupta, PhD thesis 2013

124

Sensitivity to number of histogram orientations

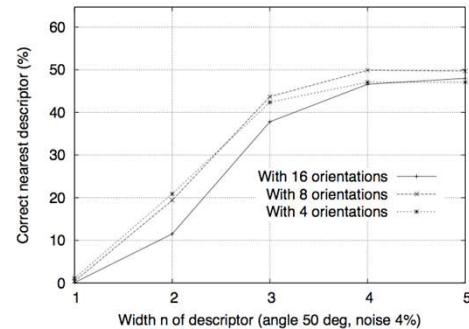


Figure 8: This graph shows the percent of keypoints giving the correct match to a database of 40,000 keypoints as a function of width of the $n \times n$ keypoint descriptor and the number of orientations in each histogram. The graph is computed for images with affine viewpoint change of 50 degrees and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

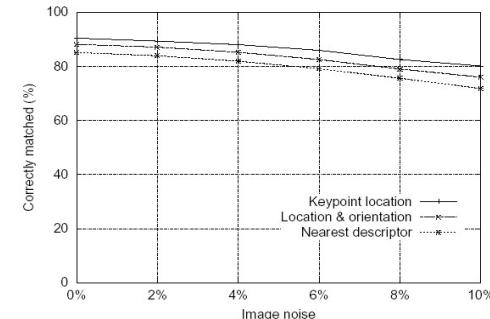


SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

125

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

126

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

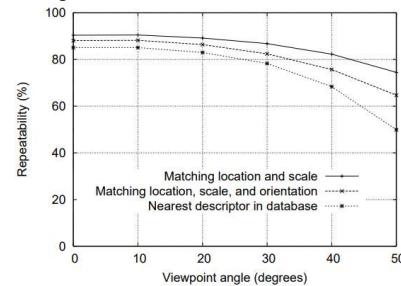


Figure 9: This graph shows the stability of detection for keypoint location, orientation, and final matching to a database as a function of affine distortion. The degree of affine distortion is expressed in terms of the equivalent viewpoint rotation in depth for a planar surface.

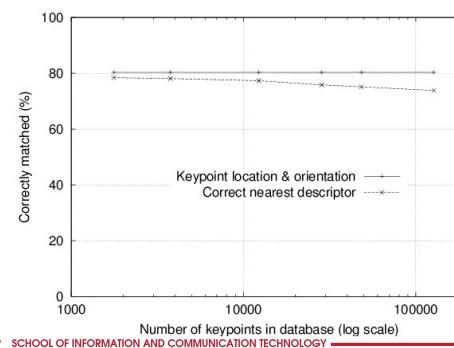


SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

127

Distinctiveness of features

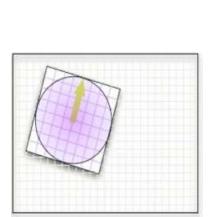
- Vary size of database of features, with 30 degree affine change, 2% image noise
- Measure % correct for single nearest neighbor match



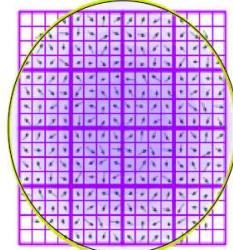
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

128

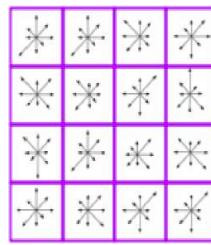
SIFT Keypoint Descriptor: summary



Blur the image
using the scale of
the keypoint
(scale invariance)



Compute gradients in
respect to the keypoint
orientation(rotation
invariance)



Compute orientation
histogram in 8
directions over 4x4
sample regions



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

129

Other detectors and descriptors

Popular features: SURF, HOG, SIFT

http://campar.in.tum.de/twiki/pub/Chair/TeachingWs13TDCV/feature_descriptors.pdf

Summary some local features:

http://www.cse.iitm.ac.in/~vplab/courses/CV_DIP/PDF/Feature_Detectors_and_Descriptors.pdf



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

130

Feature extraction

- Global features
- **Local features**
 - Local features
 - Interest point detector
 - Local descriptors
 - **Matching and applications**



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

131

Feature matching

Given a feature in I_1 , how to find the best match in I_2 ?

1. Define distance function that compares two descriptors
 - Use L1, L2, cosine, Mahalanobis,... distance
2. Test all the features in I_2 , find the one with min distance

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

132

Feature matching

- How to define the difference between two features f_1, f_2 ?
 - Simple approach: use only distance value $d(f_1, f_2)$
 - → can give good score to very ambiguous matches
 - Better approaches: add additional constraints
 - Ratio of distance
 - Spatial constraints between neighborhood pixels
 - Fitting the transformation, then refine the matches (RANSAC)



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

133

Feature matching

- Simple approach: use distance value $d(f_1, f_2)$
 - can give good score to very ambiguous matches



134

Feature matching

- Better approaches: ratio of distance = $d(f_1, f_2) / d(f_1, f_2')$
 - f_2 is best match to f_1 in I_2 ;
 - f_2' is 2nd best SSD match to f_1 in I_2
 - An ambiguous/bad match will have ratio close 1
 - Look for unique matches which have low ratio



135

Ratio of distances reliable for matching

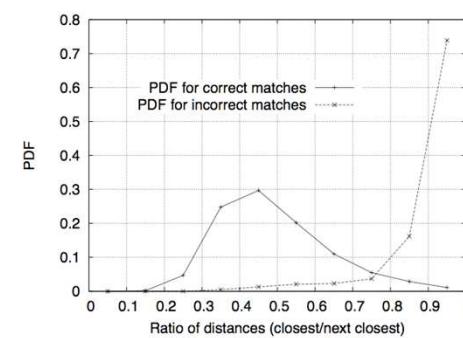


Figure 11: The probability that a match is correct can be determined by taking the ratio of distance from the closest neighbor to the distance of the second closest. Using a database of 40,000 keypoints, the solid line shows the PDF of this ratio for correct matches, while the dotted line is for matches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

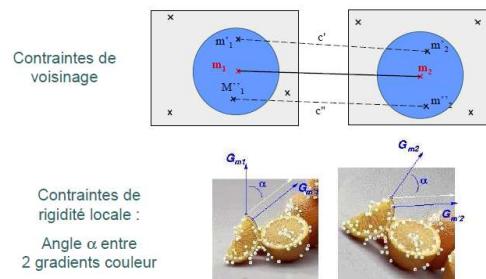


David G. Lowe, "Distinctive image features from scale-invariant keypoints," IJCV, 60, 2 (2004), pp. 91-110

136

Feature matching

- Better approaches: Spatial constraints between neighborhood pixels



Source: from slides of Valérie Gouet-Brunet



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

137

Feature matching

- Better approaches: fitting the transformation (RANSAC alg.)

- Fitting 2D transformation matrix

- Six variables

- Each point give two equations

- at least three points

- Least squares

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

- RANSAC: refinement of matches

- Compute error:

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

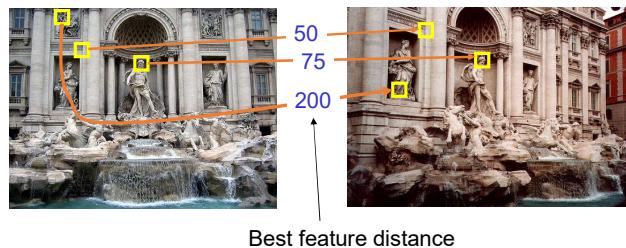


SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

138

Evaluating the results

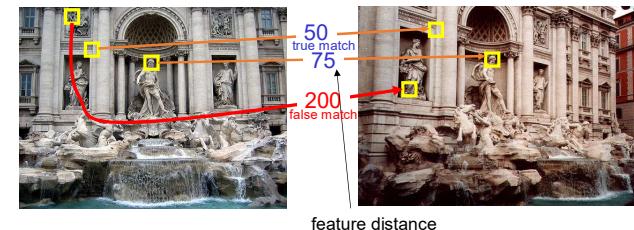
How can we measure the performance of a feature matcher?



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

139

True/false positives



The distance threshold affects performance

- True positives** = # of detected matches that are correct
 - Suppose we want to maximize these—how to choose threshold?
- False positives** = # of detected matches that are incorrect
 - Suppose we want to minimize these—how to choose threshold?



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

140

Image matching

- How to define the distance between 2 images I_1, I_2 ?
 - Using global features: easy
 $d(I_1, I_2) = d(\text{feature of } I_1, \text{feature of } I_2)$
 - Using local features:
 - Voting strategy
 - Solving an optimization problem (time consuming)
 - Building a "global" feature from local features: BoW (bag-of-words, bag-of-features), VLAD, ..

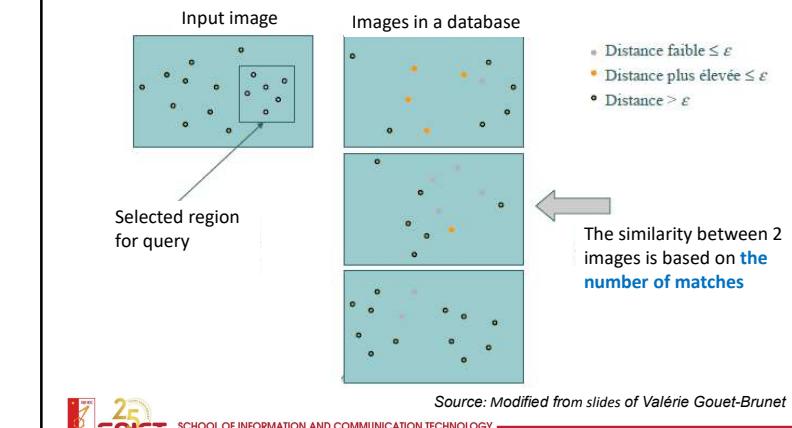


SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

141

141

Voting strategy



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

142

142

Optimization problem

- Transportation problem

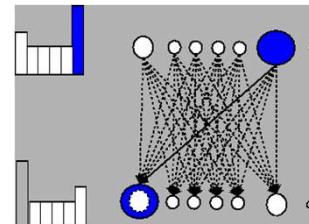
 $I_1 : \{(r_i, w_i), i=1, N\}$ Provider $I_2 : \{(r'_j, w_j), j=1, M\}$ Consommateur $d(I_1, I_2) = ???$

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

$$f_{ij} \geq 0$$

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

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



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



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

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

143

143

Bag-of-words

- Local feature \sim a word
- An image \sim a document
- Apply a technique for textual document representation: vector model



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

144

144

Visual Vocabulary ...



1. Extracting local features from a set of images



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

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

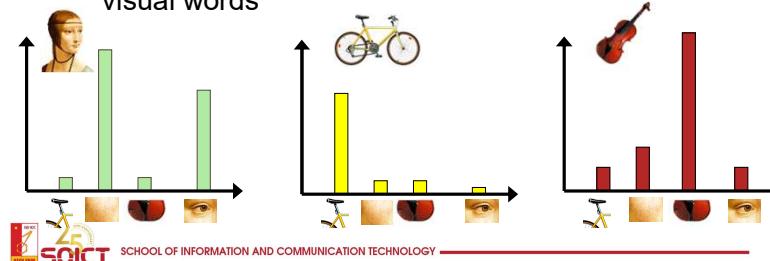
SOICT 25th Anniversary SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

145

145

Bag of words: outline

1. Extract features
2. Learn “visual vocabulary”
3. Quantize features using visual vocabulary
4. Represent images by frequencies of “visual words”



SOICT 25th Anniversary SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

146

146

Applications

Object Detection/ Recognition/ Search



Sivic and Zisserman, 2003




Lowe 2002



Rothganger et al. 2003

SOICT 25th Anniversary SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

147

Object Detection/ Recognition

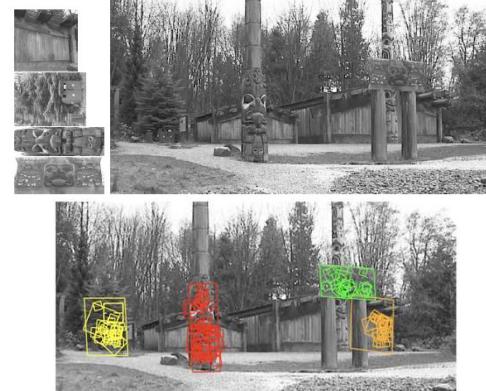


Figure 13: This example shows location recognition within a complex scene. The training images for locations are shown at the upper left and the 40x315 pixel test image taken from a different viewpoint is on the upper right. The recognized regions are shown on the lower image, with keypoints shown as squares and an outer parallelogram showing the boundaries of the training images under the affine transform used for recognition.

SOICT 25th Anniversary

148

148

Application: Image Panoramas



Slide credit: Darya Frolova, Denis Simakov

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



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

149

Automatic mosaicing



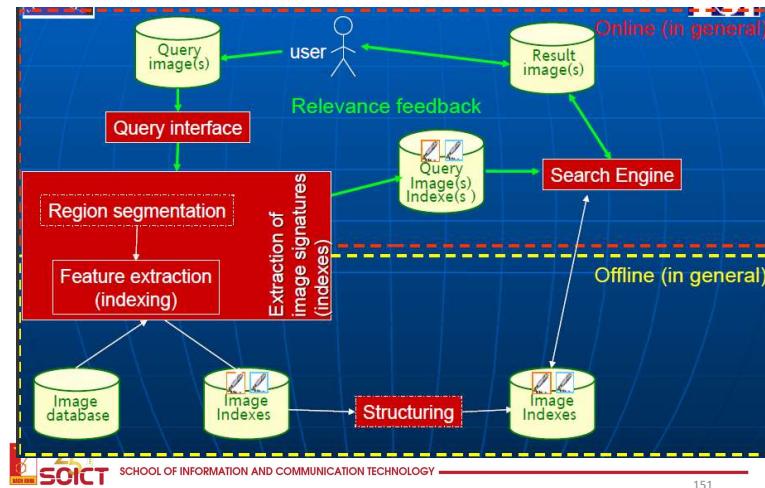
<http://www.cs.ubc.ca/~mbrown/autostitch/autostitch.html>



SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

150

CBIR (Content-based image retrieval)

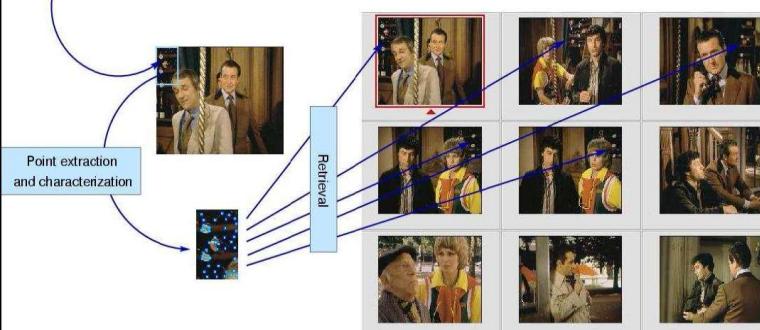


SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

151

CBIR: partial retrieval

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



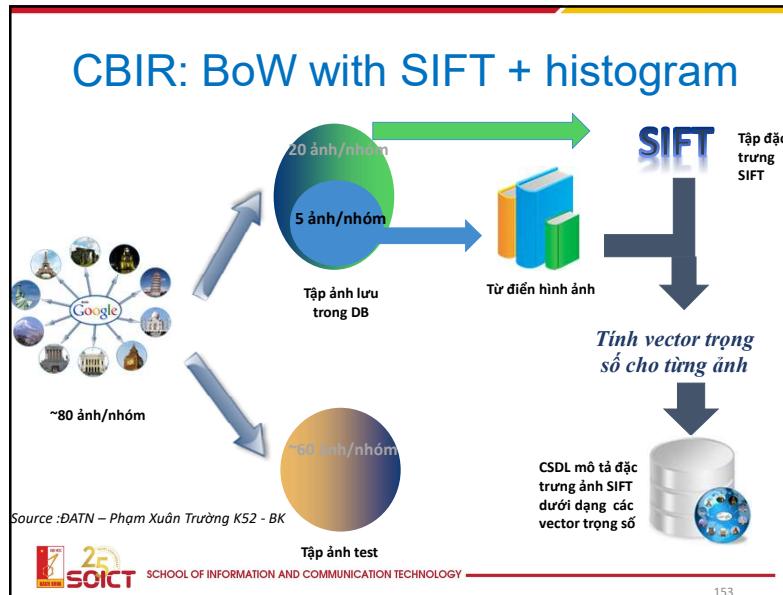
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

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

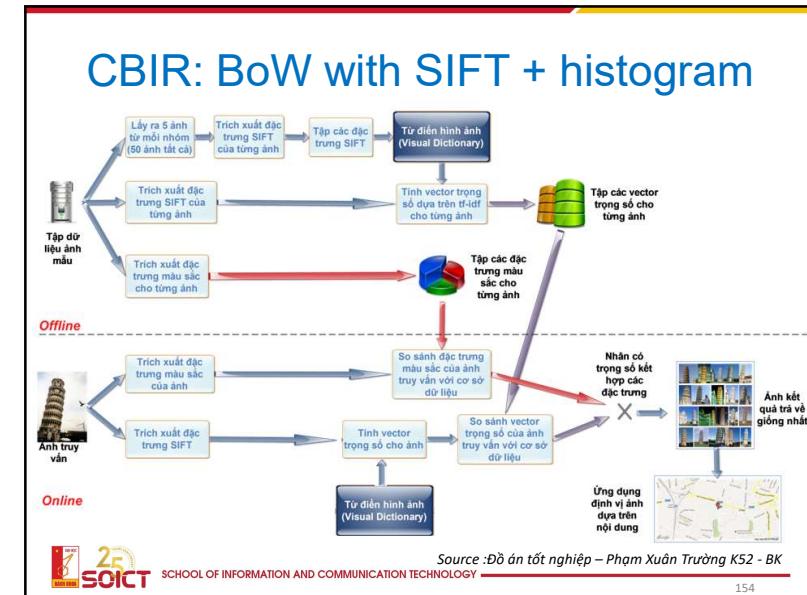
152

151

152



153



154



155



156