



TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN

CHUNG KIM KHÁNH (19127644), LA TRƯỜNG PHI (19127506)

BÁO CÁO

MỞ RỘNG KHẢ NĂNG TRUY VẤN ĐỐI TƯỢNG TRONG BỘ DỮ LIỆU ẢNH KHÔNG LỒ

Truy vấn thông tin thị giác

Mục Lục

Tuần 1 (10/05/2022)	3
Tuần 2 (17/05/2022)	5
Tuần 3 (24/05/2022)	10
Tuần 4 (31/05/2022)	16
Tuần 5 (7/06/2022)	23
Tuần 6 (14/06/2022)	26
Tuần 7 (28/06/2022)	30
Seminar	31

Tuần 1 (10/05/2022)

1. Phát biểu bài toán VIR bằng ngôn từ toán học (input: Tập ảnh - output: Tập ảnh được xếp hạng dựa trên đặc trưng thị giác và đặc trưng ngữ nghĩa)

Phát biểu hình thức:

Cho trước tập dữ liệu ảnh như sau:

$$SI = \{I, i = 1 \dots nSI\}$$

Trong đó:

SI là tập dữ liệu ảnh

li là ảnh thứ i của tập SI, nSI là số phần tử của tập SI

Cho trước ảnh truy vấn QI

Cho trước sai số e (độ dị biệt giữa ảnh truy vấn và ảnh kết quả)

Cần xác định tập ảnh kết quả được sắp hạng tương ứng với e

Gọi SRle là tập ảnh kết quả ứng với sai số e, có thể biểu diễn SRle như sau:

$$SRli = \{Rli \mid d(Rli, QI) \leq d(Rli+1, QI) < e, Rli \in SI, i = 1 \dots n(e) - 1\}$$

Trong đó:

Rli là ảnh kết quả thứ i (sắp xếp theo thứ tự tăng dần dựa vào khoảng cách dị biệt d)

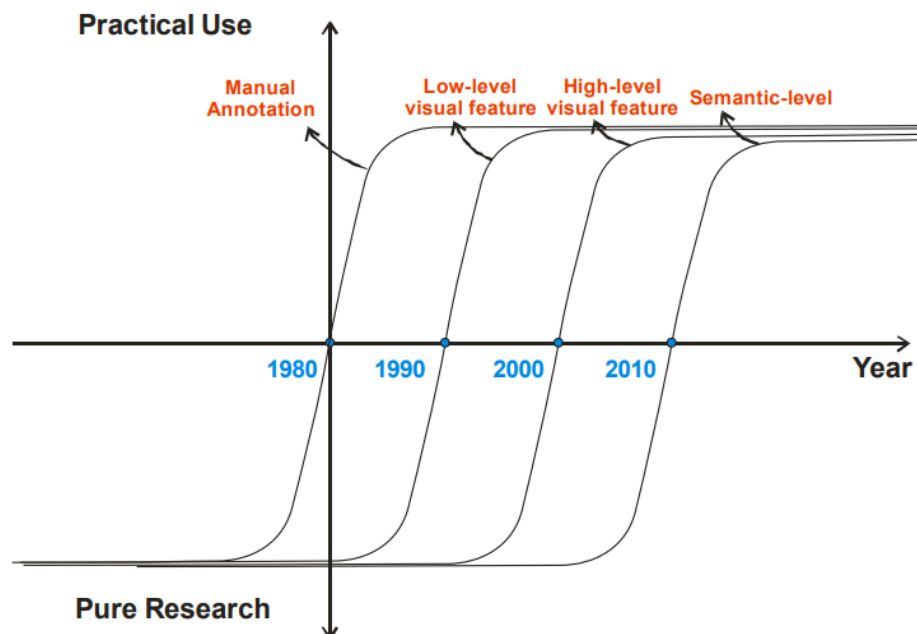
d(Rli, QI) là khoảng cách dị biệt giữa ảnh Rli và ảnh truy vấn QI

n(e) là số ảnh thuộc tập SI có khoảng cách dị biệt so với ảnh truy vấn nhỏ hơn e

2. Cho biết diễn tiến của ba công đoạn sau đến nay đã thay đổi như thế nào?

- Đánh chỉ mục (indexing)
- Rút trích đặc trưng (Feature Extraction)
- Độ đo đối sánh

3. Cho hình sau, bổ sung những đóng góp/nghiên cứu trong giai đoạn 2010 đến nay? (tách từ 2010 - 2012 - 2015 - 2020 - hiện nay)



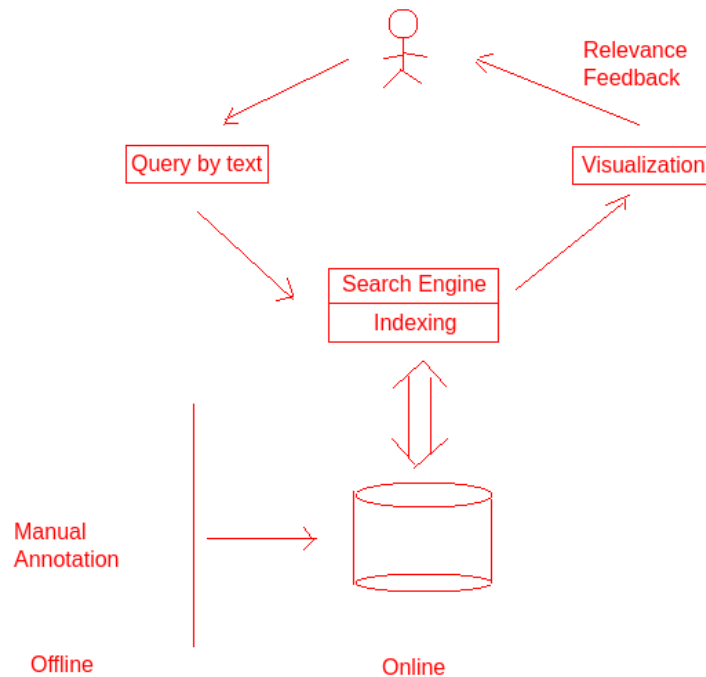
Trong sơ đồ The main development stages of VIR, từ 2010 là sự phát triển của kỹ thuật nào (Deep learning và các kỹ thuật liên quan). Cụ thể:

- Từ 2010 - 2012: Nghiên cứu Deep Learning
- Từ 2012 - 2015: Nghiên cứu các kiến trúc Deep Learning
- Từ 2015 - 2020: Các kiến trúc đã phát triển vượt bậc. Các công trình của Zisserman (Rút trích đặc trưng, Tổ chức Deep Learning, Truy vấn ảnh)

4. Liệt kê 1 số hệ thống truy vấn ảnh đặc thù (specific VIR)

- Truy vấn ảnh y phục
- Truy vấn thuộc tính mặt người
- Truy vấn ảnh landmark
- Truy vấn người
- Truy vấn logo
- Truy vấn thực phẩm
- Keyword:CBIR
- Google Image Search

5. Giải thích về hệ thống Relevance Feedback trong quy trình truy vấn ảnh



Relevance Feedback ứng dụng làm truy vấn thông minh, có sự đóng góp của người dùng thì hệ thống càng tốt hơn.

6. Vẽ sơ đồ biểu diễn thể hệ thứ 3 của hệ thống truy vấn ảnh (Image Retrieval System)

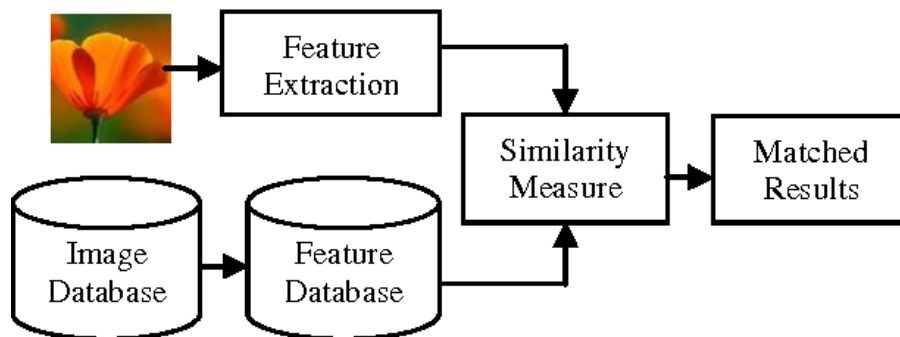
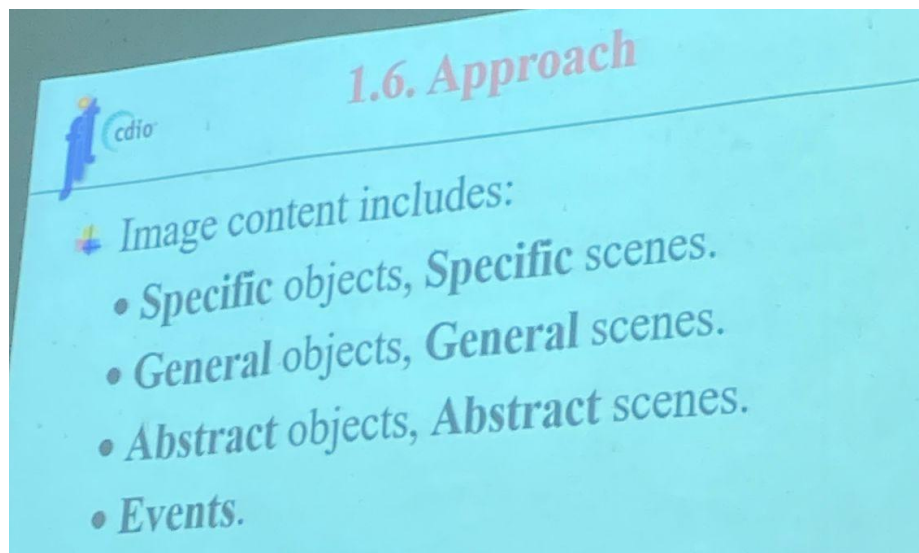


Fig 1. Block diagram of content-based image retrieval system

Tuần 2 (17/05/2022)

1. Cho ví dụ của đặc trưng ảnh (Image content):



Object: Xe Mercedes

- Specific: xe Mercedes
- General: xe
- Abstract: phương tiện giao thông

Scenes: Vịnh Hạ Long

- Specific: Vịnh Hạ Long
- General: Vịnh
- Abstract: phong cảnh

Events: Cảnh dựng xe (Có sự tương tác giữa các Object trong đó). Cảnh ghi bàn (Người sút với người chụp có liên hệ với nhau, cầu môn với quả bóng có sự tương tác với nhau)

2. Việc mất thông tin khi chuyển từ 3D sang 2D gây khó khăn như thế nào trong các công đoạn và một số ví dụ của các khó khăn khác.

Khó khăn khi chuyển từ 3D -> 2D

- Mất 1 chiều (mất thông tin) - Information

- Mất góc nhìn (viewpoint): Cùng 1 đối tượng nhưng nhìn ở góc độ khác nhau thì lại ra những hình ảnh khác nhau.

Ví dụ của một số thách thức khác trong truy vấn ảnh

- Deformation: Cùng một đối tượng nhưng lại bị biến dạng trong mỗi hoạt động.
- Light condition: Một con mèo ban ngày lại có màu khác, ban đêm lại có màu khác
- Scale: Khi máy ảnh zoom. Vị trí khác nhau thì to nhỏ khác nhau.
- Occlusion: Bị che (Ví dụ trọng tài bị che bởi các cầu thủ khác)
- Clutter background: Nền thay đổi phức tạp (Ví dụ: Tìm khung thành nhưng phía sau lưới khung thành có rất nhiều khán giả)
- Fluge storage size: Mỗi giây có hàng triệu video được đăng tải lên.
- Semantic Gap: Khó nhất (chung quy lại tất cả)

3. Các ảnh hưởng này nó như thế nào đối với Truy vấn offline và online.

Nó làm mất nhiều thời gian truy vấn hơn và kết quả truy vấn không được tốt.

4. Tính đoạn video bóng đá 90 phút quay 25 frames 1 giây. Với khung hình là 512x768. 1 pixel có 3 bytes màu (RGB). Thì đoạn video có dung lượng là bao nhiêu?

Thời gian của cả đoạn video:

$$90 \text{ phút} = 90 \times 60 = 5.400 \text{ (giây)}$$

Tổng số frame của video:

$$5400 \times 25 = 112.500 \text{ (frames)}$$

Tổng số pixel của video:

$$112.500 \times 512 \times 768 \text{ (pixel)}$$

Dung lượng của video:

$$(112.500 \times 512 \times 768) \times 3 \text{ (byte)}$$

5. So sánh VQA (Visual Question Answering) với VIR (Visual Information Retrieval)

Trả lời câu hỏi trực quan (VQA) là một tập dữ liệu chứa các câu hỏi mở về hình ảnh. Những câu hỏi này đòi hỏi sự hiểu biết về tầm nhìn, ngôn ngữ và kiến thức thông thường để trả lời. Phiên bản đầu tiên của tập dữ liệu đã được phát hành vào tháng 10 năm 2015.

Truy xuất thông tin trực quan (VIR) là một lĩnh vực nghiên cứu liên ngành có các kỹ thuật nhằm giải quyết vấn đề tìm kiếm hình ảnh và video có liên quan (tài liệu chứa) dựa trên một truy vấn.

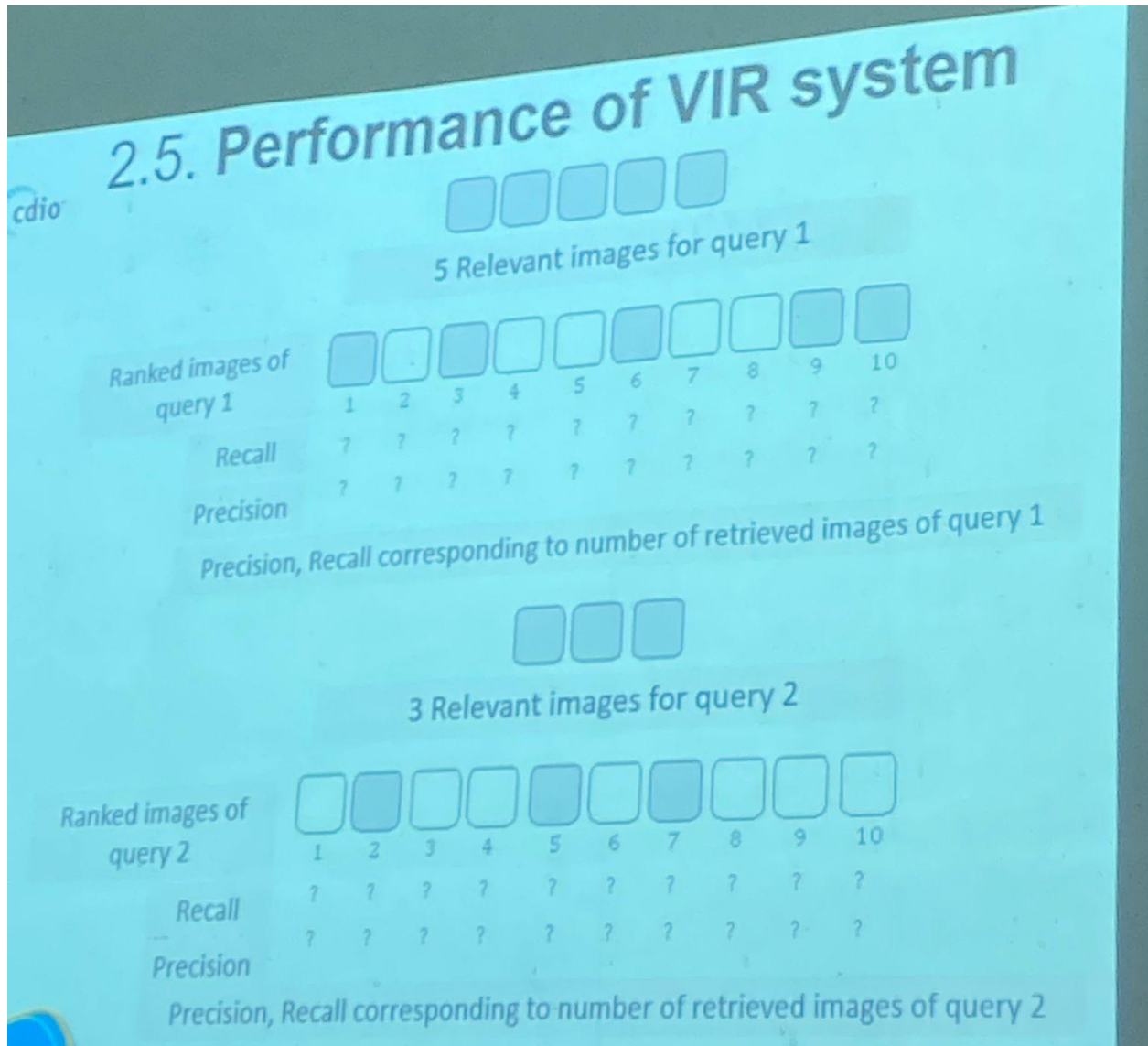
6. Cho ví dụ về Precision (độ chính xác / tìm sai) và Recall (tìm sót)

Khi recall = 100% thì show hết tập kết quả trong dữ liệu ra.

7. Khi nào thì ảnh trả về đúng với ảnh trong query.

Muốn xác định được thì phải tạo một bản xác định <Ground Truth>, sau đó đánh nhãn tập dữ liệu trong query, khi 1 ảnh nằm đúng trong nhãn của nó thì được xác định là đúng với query

8. Tính (ô màu xám là đúng)



5 Relevant images for query 1

Ranked images of query 1	1	2	3	4	5	6	7	8	9	10
Recall (lấy số ô màu xám chia cho 5)	1/5	1/5	2/5	2/5	2/5	3/5	3/5	3/5	4/5	1
Precision	100 %		50 %			33,3			25 %	20 %

						%				
--	--	--	--	--	--	---	--	--	--	--

Q1	Q2
1	2
3	5
6	7
9	
10	

$$\text{MAP} = \frac{1}{2} \left[\frac{1}{5} \left(\frac{1}{1} + \frac{2}{3} + \frac{3}{6} + \frac{4}{9} + \frac{5}{10} \right) + \frac{1}{3} \left(\frac{1}{2} + \frac{2}{5} + \frac{3}{7} \right) \right] \approx 0.5325$$

9. Phân biệt đặc trưng toàn cục và đặc trưng cục bộ khác nhau thế nào?

Đặc trưng toàn cục:

- Màu: vàng, đen, xanh,...
- Vân: tròn, xoắn,...
- Dáng: vuông, tròn, tam giác,...

=> Rút ra 1 vector đại diện cho đặc trưng chung của cả bức ảnh. Thường không phân biệt được sự thay đổi của vị trí trong ảnh. -> Kết quả Histogram không thay đổi.

Đặc trưng cục bộ:

=> Để ý tới sự thay đổi vị trí trong ảnh. Xét trên lân cận điểm ảnh chứ không xét toàn ảnh.

Đặc trưng HOG

(Có chia lưới và chia shield là cục bộ)

10. Rút đặc trưng cục bộ đem lại ích lợi gì trong truy vấn?

Để chỉ ra chính xác đối tượng cần tìm (Ví dụ: tìm xe Mercedes Benz thì nó sẽ đưa ra xe Mercedes Benz)

11. Tìm hiểu về Deep Feature?

Deep feature là phản hồi nhất quán của một node hoặc layer trong mô hình phân cấp đối với đầu vào đưa ra phản hồi có liên quan đến đầu ra cuối cùng của mô hình. Một feature được coi là “sâu hơn” so với tính năng khác tùy thuộc vào mức độ kích hoạt sớm trong cây quyết định hoặc khung công tác khác.

Đưa mạng nào đó vào tập huấn luyện và trả về kết quả cuối cùng với rất nhiều đặc trưng quan trọng.

12. Local Feature dựa trên lưới lợi gì cho truy vấn và dựa trên vùng lợi gì cho truy vấn?

Local feature đề cập đến một mẫu hoặc cấu trúc riêng biệt được tìm thấy trong một hình ảnh, chẳng hạn như một điểm, cạnh hoặc bản vá hình ảnh nhỏ. Chúng thường được liên kết với một bản vá hình ảnh khác với môi trường xung quanh ngay lập tức bởi kết cấu, màu sắc hoặc cường độ.

Tuần 3 (24/05/2022)

1. Trong thiết kế web, làm sao ta có thể biết được ta có thiết kế đúng màu sắc so với yêu cầu hay không một cách tự động.

Sử dụng so sánh qua các đặc trưng

2. So sánh 3 vector đặc trưng trong local features (grid + descriptor, region + descriptor, interest point + descriptor)

- Sử dụng region ta có thể loại bỏ được background
- Sử dụng grid thì ngoài màu ta còn xác định được cả vị trí
- Interest point bất biến khi thay đổi view, scale

Từ những đặc trưng tạo ra 1 chuỗi codebook. Từ codebook tạo ra một cái đại diện gọi là visual word.

3. Tiêu chí dưới để đánh giá color histogram có giúp tăng độ chính xác, độ phủ trong truy vấn ko.

Precision

Recall

MAP (Mean Average Precision)

Precision Recall là một thước đo hữu ích của sự thành công của dự đoán khi các lớp rất mất cân bằng. Trong truy xuất thông tin, độ chính xác là thước đo kết quả liên quan, trong khi thu hồi là thước đo có bao nhiêu kết quả thực sự có liên quan được trả về.

MAP (Mean Average Precision) Đây là **độ đo tổng hợp kết quả của nhiều query, được sử dụng rất phổ biến**. MAP cũng chứa thông tin của precision và recall, có xét đến độ quan trọng của thứ hạng kết quả. Average Precision: trung bình của các precision tại các điểm mà mỗi kết quả đúng trả về.

4. Kiểm tra đặc trưng histogram có những đặc trưng này không (Đồng nghĩa hay đa nghĩa). Nó có làm ảnh hưởng đến độ đo Precision (độ chính xác) hay recall (độ phủ) không?

- **Tính đồng nghĩa** (Nhiều đặc trưng thị giác khác nhau nhưng cùng một ngữ nghĩa): Ví dụ hình ảnh con mèo có màu đen và hình mèo màu trắng. Khác màu nhưng cùng là mèo.
- **Tính đa nghĩa** (Cùng đặc trưng thị giác giống nhau nhưng có ngữ nghĩa khác nhau): Ví dụ một hình cầu màu đỏ có thể là 1 viên bi hoặc 1 trái banh.

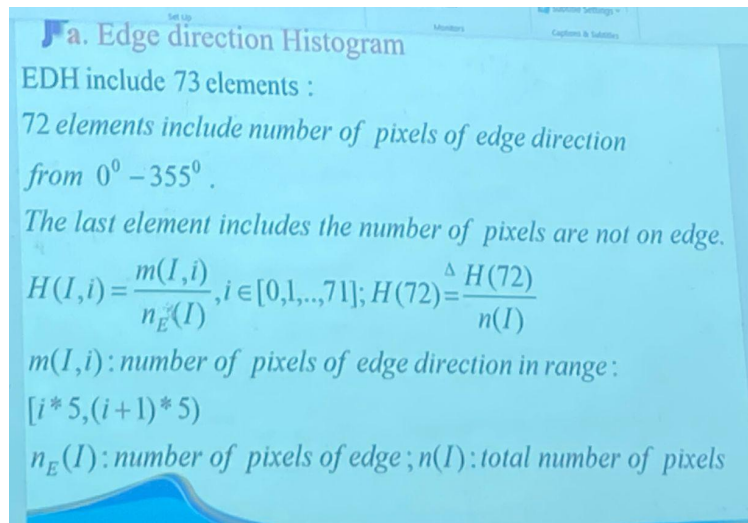
5. Nếu dùng đặc trưng pixel liên kết màu thì có trả lời được câu hỏi 1 hay không?

Khi ta dùng đặc trưng pixel liên kết màu thì trong lược đồ liên kết màu sẽ chia thành 2 biên, 1 biên là những pixel liên kết màu và 1 biên là những pixel không liên kết màu. Khi tách ra thì chúng ta có thể dễ dàng phân biệt được. Nhưng khi hoán vị vị trí của các pixel trên trang web thì his vẫn không thay đổi dẫn đến ta không phân biệt được sự khác nhau về màu sắc giữa 2 trang web 1 cách tự động.

6. Sử dụng color vector có phân biệt được màu khác nhau (2 màu trang web được hoán vị vị trí các đối tượng) hay không?

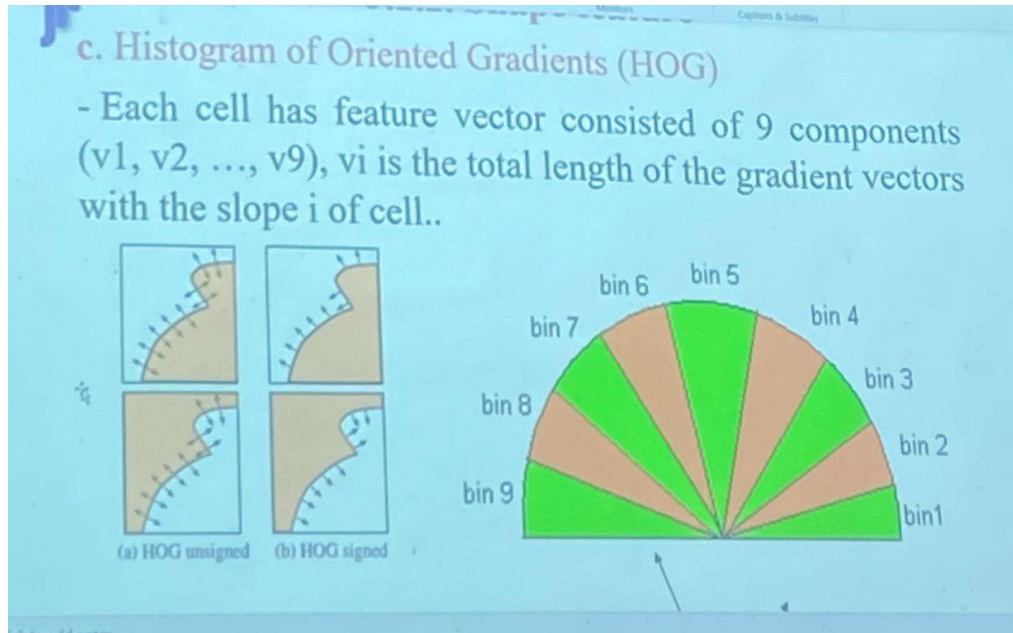
Không phân biệt được (do hoán vị)

7. Tại sao lại có the last element (pin cuối), bỏ H(72) được không? Tại sao tại cần pixel biên cạnh.



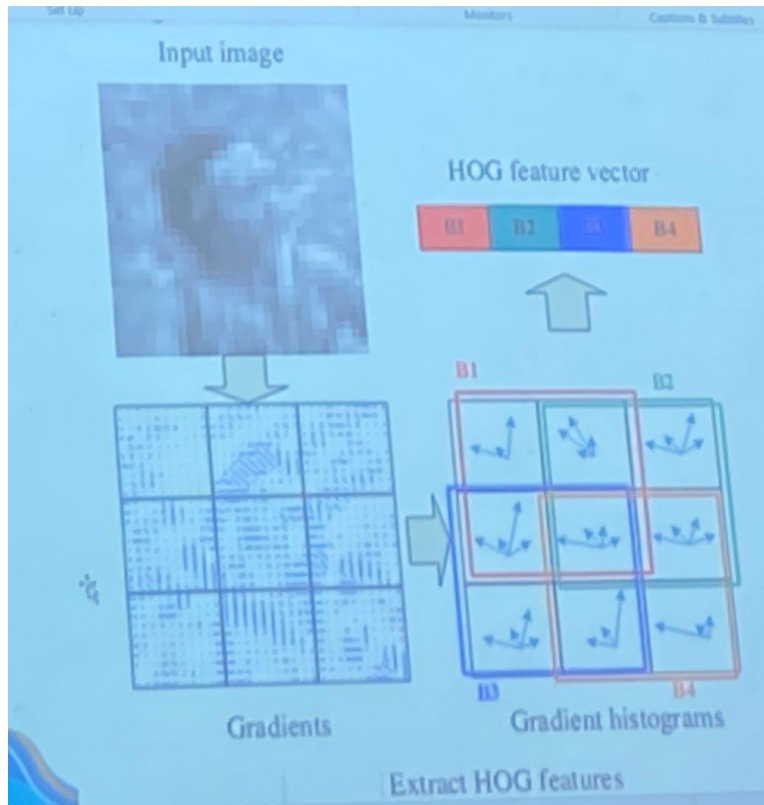
Vì các pixel càng nằm gần biên thì càng làm lộ rõ bản chất của đối tượng nên chúng ta không thể bỏ H(72)

8. Khi nào dùng cái HOG unsigned khi nào dùng HOG signed



- **HOG unsigned** khi mà muốn xác định đối tượng không quan trọng đối tượng sáng hay tối.
- **HOG signed** khi mà đối tượng luôn sáng hơn hay tối hơn background.

9. Sau khi nó tính Gradients Histogram xong thì nó lấy 4 cells thành 1 góc mà không xét liên tục với nhau (2 blocks có đè lên nhau).



Để tính Feature vector (vector đặc trưng) có số chiều lớn hơn.

Ảnh là 2 chiều nên bị khống chế theo 2 phương x và y. Khi trải ra thì nó sẽ ít quan tâm chiều dọc. Vậy nên khi chia nhóm vậy nó sẽ khống chế quan tâm đều 2 phương.

Overlap (đề lên nhau) để nhấn nhá -> hiện rõ viền ảnh hơn.

10. Ảnh (320x200) lấy đặc trưng HOG. Tính số block, số chiều vector đặc trưng.

Thay W thành 320, H thành 200

Số block = $(W/8 - 1) * (H/8 - 1) = (320/8 - 1) * (200/8 - 1) = 33\ 696$ (blocks)

Số vector = Số block * 36 = (components)

11. Tại sao tính compact lại là độ đầy đặn độ no tròn (compact = p_c/p). Giá trị biến thiên từ đâu đến đâu?

Tiến về 1 là đầy đặn, no tròn

Tiến về 0 là thon dài

12. Tính compact của 2 miếng đất D_1 và D_2

Miếng đất D_1 : 2mx50m ($A = 100m^2$, $P = 104$)

Miếng đất D_2 : 10mx10m ($A = 100m^2$, $P = 40m$)

$A_c = \pi R^2 = 100 \Rightarrow R?$

$P_c = 2\pi R = 35,45$

$Compact(D_1) = P_c/104 = 0.34$

$Compact(D_2) = P_c/40 = 0.89$

Tuần 4 (31/05/2022)

1. Viết chương trình tìm bao lồi của một vùng. Bao lồi là một tập lồi có diện tích bé nhất chứa đối tượng - Hình câu 1.

```
import cv2
import numpy as np
import sys

if __name__ == "__main__":
    if(len(sys.argv)) < 2:
        file_path = "sample.jpg"
    else:
        file_path = sys.argv[1]

    # read image
    src = cv2.imread(file_path, 1)

    # show source image
    cv2.imshow("Source", src)

    # convert image to gray scale
    gray = cv2.cvtColor(src, cv2.COLOR_BGR2GRAY)

    # blur the image
    blur = cv2.blur(gray, (3, 3))

    # binary thresholding of the image
    ret, thresh = cv2.threshold(blur, 200, 255, cv2.THRESH_BINARY)
```

Home-class exercise_Innovation

```
# find contours
im2, contours, hierarchy = cv2.findContours(thresh, cv2.RETR_TREE, \
      cv2.CHAIN_APPROX_SIMPLE)

# create hull array for convexHull points
hull = []

# calculate points for each contour
for i in range(len(contours)):
    hull.append(cv2.convexHull(contours[i], False))

# create an empty black image
drawing = np.zeros((thresh.shape[0], thresh.shape[1], 3), np.uint8)

# draw contours and hull points
for i in range(len(contours)):
    color_contours = (0, 255, 0) # color for contours
    color = (255, 255, 255) # color for convex hull
    # draw contours
    cv2.drawContours(drawing, contours, i, color_contours, 2, 8, hierarchy)
    # draw convex hull
    cv2.drawContours(drawing, hull, i, color, 2, 8)

cv2.imshow("Output", drawing)

cv2.waitKey(0)
cv2.destroyAllWindows()
```

2. Tìm trục chính và trục phụ của vùng (alpha1, alpha2) - Hình
câu 1. Tìm bằng vuông góc PCA. Tìm những trục có
phương sai biến thiên dữ liệu lớn nhất.sh

3. Viết chương trình tìm hình chữ nhật bé nhất bao lồi - Hình câu 1.

```
import cv2
import numpy as np

image = cv2.imread('shapes.png')
original = image.copy()

gray = cv2.cvtColor(image, cv2.COLOR_BGR2GRAY)
thresh = cv2.threshold(gray, 0, 255, cv2.THRESH_BINARY_INV + cv2.THRESH_OTSU)[1]

ROI_number = 0
cnts = cv2.findContours(thresh, cv2.RETR_EXTERNAL, cv2.CHAIN_APPROX_SIMPLE)
cnts = cnts[0] if len(cnts) == 2 else cnts[1]
for c in cnts:
    x,y,w,h = cv2.boundingRect(c)
    cv2.rectangle(image, (x, y), (x + w, y + h), (0,0,255), 2)

cv2.imshow('image', image)
cv2.imshow('Thresh',thresh)
cv2.waitKey()
```

Note: VCO (Visual Concept Ontology): Phả hệ tri thức, dựa vào những đặc trưng mô tả hình dung hình dạng.

Đặc trưng về SIFT

4. Cho một đặc trưng Texture Feature về màu, về dáng hơn ở ưu thế nào, ngữ cảnh nào so với Deep Feature.

VCO (Visual Concept Ontology) nằm ở mức Middle Visual Feature

Raw data -> Low Visual Feature -> Middle Visual Feature -> High Lines Visual Feature -> Semantic Memory.

- Handcrafted Feature không phụ thuộc vào tập dữ liệu hình ảnh.
- Deep Feature cần phụ thuộc vào tập dữ liệu hình ảnh.

Ví dụ: truy vấn về màu sắc.

- Nếu dùng Deep thì tập dữ liệu học cần chuẩn bị nhiều cỡ nào, có bao nhiêu màu (giả sử 16 triệu 8 màu).
- Nếu dùng Color Histogram hoặc Color Moment (Handcrafted Feature) chỉ cần rút trích đặc trưng ra.

5. Vân là gì?

Vân (texture) là **tập điểm** ảnh thể hiện sự phân bố về vị trí và màu sắc khác nhau.

Những đặc trưng cơ bản mô tả vân: Haralick, Wavelet.

6. Làm sao để đặc tả được tính chất thô ráp của vân Coarseness.

Tìm sự khác biệt lớn nhất giữa các điểm trung bình trên 2 trục tọa độ (các phần tử cấu thành).

7. Dùng biểu thức toán nào có thể tính được Coarseness

- ❖ Averages at each point over neighborhood $2^k \times 2^k$ ($k = 0 \rightarrow 5$)

$$A_k(x, y) = \sum_{x-2^{k-1}}^{x+2^{k-1}-1} \sum_{y-2^{k-1}}^{y+2^{k-1}-1} \frac{g(i, j)}{2^{2k}}$$

$g(i, j)$ is grayscale of pixel $p(i, j)$

- ❖ Take differences between pairs of average points in both horizontal & vertical orientations

$$E_{k, horizontal}(x, y) = |A_k(x + 2^{k-1}, y) - A_k(x - 2^{k-1}, y)|$$

$$E_{k, vertical}(x, y) = |A_k(x, y + 2^{k-1}) - A_k(x, y - 2^{k-1})|$$

$$E_k(x, y) = \max(E_{k, horizontal}, E_{k, vertical})$$

- ❖ At each point, pick best size which gives highest output

$$E_k = \max(E_0, E_1, \dots, E_L)$$

$$S_{best}(x, y) = 2^k$$

- ❖ Take average of S_{best}

$$F_{crs} = \frac{\sum_{p \in R} S_{best}(p_x, p_y)}{m * n}$$

m, n : width, height of image

- ❖ Normalize F_{crs} to interval [0 1] by divide to $2^{k_{max}}$

8. Tìm biểu thức toán mô tả sự hữu hướng (Directionality)

$$F_{dir} = \tan^{-1} \left(\frac{\sum_{x=1}^M \sum_{y=1}^N m(x, y) \sin d(x, y)}{\sum_{x=1}^M \sum_{y=1}^N m(x, y) \cos d(x, y)} \right),$$

Trong đó M và N là kích thước của lát giềng, và $m(x, y)$; $d(x, y)$ được tính:

$$m(x, y) = \sqrt{\delta_x(x, y)^2 + \delta_y(x, y)^2},$$

$$d(x, y) = \tan^{-1} \frac{\delta_y(x, y)}{\delta_x(x, y)} + \frac{\pi}{2},$$

9. Viết chương trình để tính Edge Direction.

```
def hysteresis(img, weak, strong=255):
    M, N = img.shape
    for i in range(1, M-1):
        for j in range(1, N-1):
            if (img[i,j] == weak):
                try:
                    if ((img[i+1, j-1] == strong) or (img[i+1, j] == strong) or
                        (img[i+1, j+1] == strong)
                        or (img[i, j-1] == strong) or (img[i, j+1] == strong)
                        or (img[i-1, j-1] == strong) or (img[i-1, j] == strong)
                        or (img[i-1, j+1] == strong)):
                        img[i, j] = strong
                except IndexError as e:
                    pass
            else:
                img[i, j] = 0
```

```
return img
```

10. Viết biểu thức toán để tìm đặc trưng vân dạng đường (Linelikeness). Giải thích vì sao mô tả dc cấu trúc vân dạng đường.

$$F_{lin}(x, y) = 1 - \frac{\sum_{i=x-\omega}^{x+\omega} \sum_{j=y-\omega}^{y+\omega} (D(i, j) - \mu_D)^2}{(2\omega + 1)^2},$$

11. Để phân biệt indoor, outdoor dùng gì?

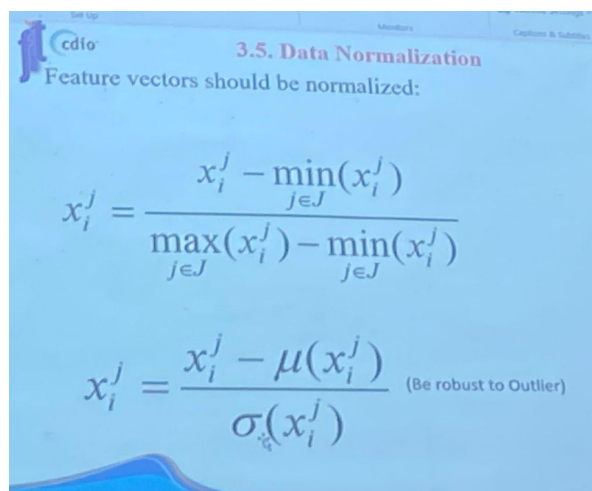
Dùng đặc trưng color moment. Màu sắc ngoài đường sẽ sặc sỡ hơn so với trong nhà.

Để phân biệt city và landscape thì thực hiện lược đồ hệ số góc. Vì thành phố có các đường ngang dọc.

Color gồm Chrominance (HS, UV, ab) và Luminance (V, L, L)

12. Giải thích dữ liệu nếu ko chuẩn hoá thì sẽ gây ảnh hưởng, tác vụ gì cho truy vấn.

Xử lý nhiễu bằng công thức thứ 2.



Sau khi chuẩn hoá thì dữ liệu sẽ trong khoảng (0-1). Các thành phần của vector đặc trưng đều tham gia vào, các thành phần chỉ để ý về giá trị tuyệt đối mà không để ý về giá trị tương đối.

Trong vector thì khoảng cách giữa các thành phần trong vector có trên lệch quá lớn làm cho những thành phần có giá trị thấp không có sức ảnh hưởng trong mức ý nghĩa.

Tuần 5 (7/06/2022)

1. Dựa vào biểu thức trên, biện luận các thành phần trọng yếu. (theo lamda1 và lamda2):

λ_1 và λ_2 là các giá trị riêng của M. Vì vậy, giá trị của các giá trị riêng này quyết định xem một vùng là góc, cạnh hay phẳng.

- Khi nào $|R|$ nhỏ, điều này xảy ra khi λ_1 và λ_2 nhỏ, vùng này phẳng.
- Khi $R < 0$, điều này xảy ra khi $\lambda_1 \gg \lambda_2$ hoặc ngược lại, vùng là một cạnh.
- Khi R lớn, điều này xảy ra khi λ_1 và λ_2 lớn và $\lambda_1 \sim \lambda_2$, vùng là một góc.

3.4.1. Point Detector
Harris operator (corner point detector).

Chéo hóa ma trận M ta nhận được một hệ cơ sở trực giao gồm hai vectơ riêng là $\{e_1, e_2\}$ ứng với hai giá trị riêng λ_1 và λ_2 .

Ta viết lại như sau:

$$E(u, v) \cong \begin{bmatrix} u & v \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}$$

$$E(u, v) \cong \lambda_1 u^2 + \lambda_2 v^2$$

Associate Prof. LÝ QUỐC NGỌC 12

2. Giải thích sao ra được biểu thức $R(M)$

3.4.1. Point Detector
Harris operator (corner point detector).

Giả sử λ_1 và λ_2 là hai nghiệm của phương trình đặc trưng:

$$\det(M - \lambda U) = 0, \text{ với } U \text{ là ma trận đơn vị } 2 \times 2.$$

$$\det(M - \lambda U) = \begin{vmatrix} a - \lambda & b \\ b & c - \lambda \end{vmatrix} = (a - \lambda)(c - \lambda) - b^2.$$

$$\det(M - \lambda U) = \lambda^2 - (a + c)\lambda + ac - b^2.$$

ta có :

$$R(M) = (ac - b^2) - k(a + c)^2 = \lambda_1 \lambda_2 - k(\lambda_1 + \lambda_2)$$

Với k là hệ số do Harris đề nghị $k \in [0.04, \dots, 0.15]$. Hệ số này thường được gọi là hệ số Harris.

Associate Prof. LÝ QUỐC NGỌC 15

Để tránh tính giá trị riêng λ_1, λ_2 , Harris đề nghị đại lượng xác định điểm góc :

$$R(M) = \det(M) - k \cdot \text{trace}^2(M)$$

3. Trong các phương pháp tìm điểm trọng yếu như harris và DOG thì kiến thức toán quan trọng nào ảnh hưởng nhiều nhất:

- Đạo hàm, chéo hoá ma trận, định lý vi-ét của pt bậc 2

4. Vì sao đặc trưng SIFT phải tính hướng chính của các vector đặc trưng

Để bảo toàn các vector với phép xoay

5. Đặc trưng SIFT được phân vào loại đặc trưng nào (màu, vân hay dáng)

SIFT được xếp vào đặc trưng vân

Tuần 6 (14/06/2022)

1. Vì sao trong truy vấn lại cần độ đo có tính giao hoán ($d(x,y) = d(y,x)$)?

Đưa ra ảnh x thì tìm được ảnh y và ngược lại đưa ảnh y tìm được ảnh x.

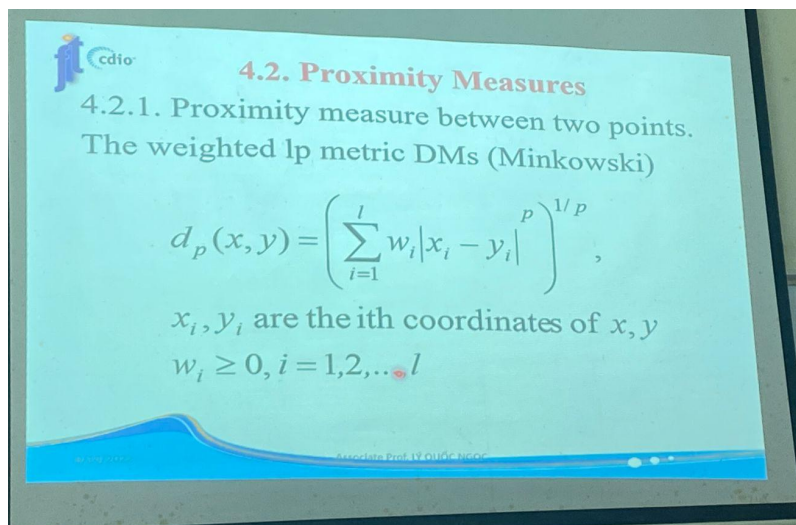
Ví dụ: tính khoảng cách từ nhà đến trường thì bằng từ nhà đến trường (tính theo đường chim bay).

2. Giải thích tiêu chí bất đẳng thức tam giác, tính bắc cầu ý nghĩa (có lợi gì) trong truy vấn ảnh là gì?

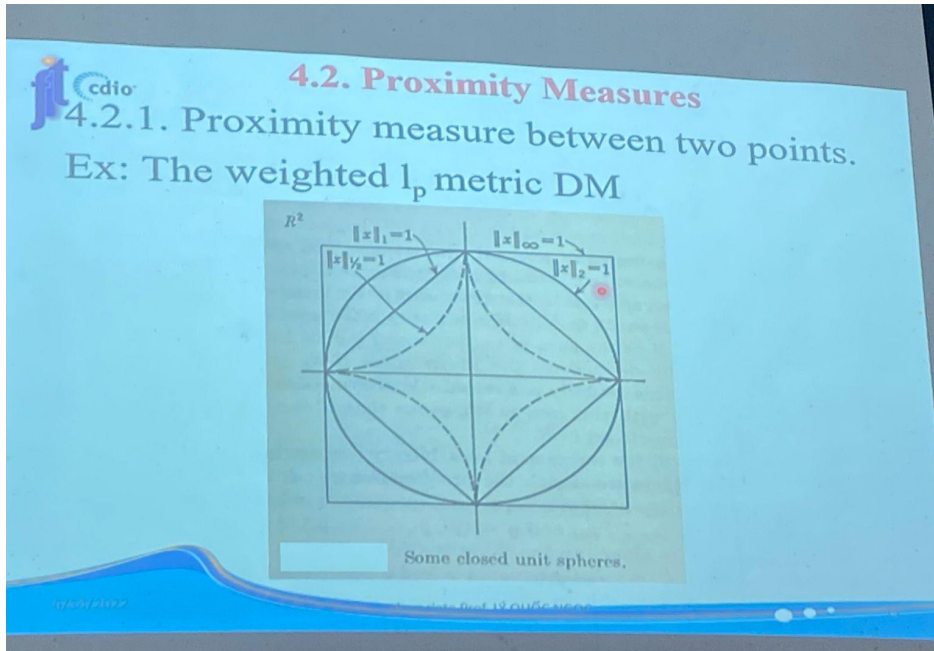
Có tính bắc cầu thì từ cái xe bò thì nó sẽ cho ra kết quả có bao gồm xe hoặc con bò.

3. Liệt kê 20 độ đo Measure và Metric

Giao hoán, phản xạ, bắc cầu



4. Dựa trên hình giải thích nếu dùng D_1 , D_2 , D vô cực thì nó ảnh hưởng đến precision và recall ntn? kết quả nào trả về lớn nhất?



Nếu chọn độ đo đối sánh khác nhau thì kết quả trả về sẽ khác nhau.

Nếu $D=1$ thì trả về kết quả hình thoi => Kết quả ít nhất => recall cao (kết quả tìm sót cao)

Nếu $D=2$ là hình tròn

Nếu D là vô cực thì sẽ trả về hình vuông => Kết quả nhiều nhất => recall thấp (kết quả tìm sót ít)

Precision phải dựa trên kết quả thực nghiệm.

5. Đo độ dị biệt giữa 2 histogram trong 2 không gian màu HSI

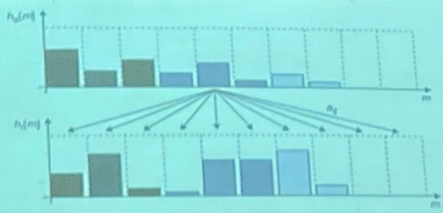
4.3. Proximity Measures on Color Features

4.3.1. Lược đồ màu

$$D_H(I_Q, I_D) = (H(I_Q) - H(I_D))^T W (H(I_Q) - H(I_D))$$

$W = [w_{ij}]$ denotes the color correlation between bin i and j

Assume that $c_i = (H_i, S_i, I_i)$, $c_j = (H_j, S_j, I_j)$,

$$w_{ij} = 1 - \frac{1}{\sqrt{5}} \left[(I_i - I_j)^2 + (S_i \cos H_i - S_j \cos H_j)^2 + (S_i \sin H_i - S_j \sin H_j)^2 \right]^{\frac{1}{2}}$$


Quadratic-form matrices compare multiple bins between the color histograms using a similarity matrix $A = [a_{ij}]$, which can take into account color similarity or color covariance.

Associate Prof. LỖ QUỐC NGỌC

Chậm hơn euclid nhưng chính xác hơn.

Không quan tâm đến độ lớn (độ scale của từng thành phần) thì dùng độ đo cosin

Còn quan tâm thì dùng độ dị biệt giữa các biến.

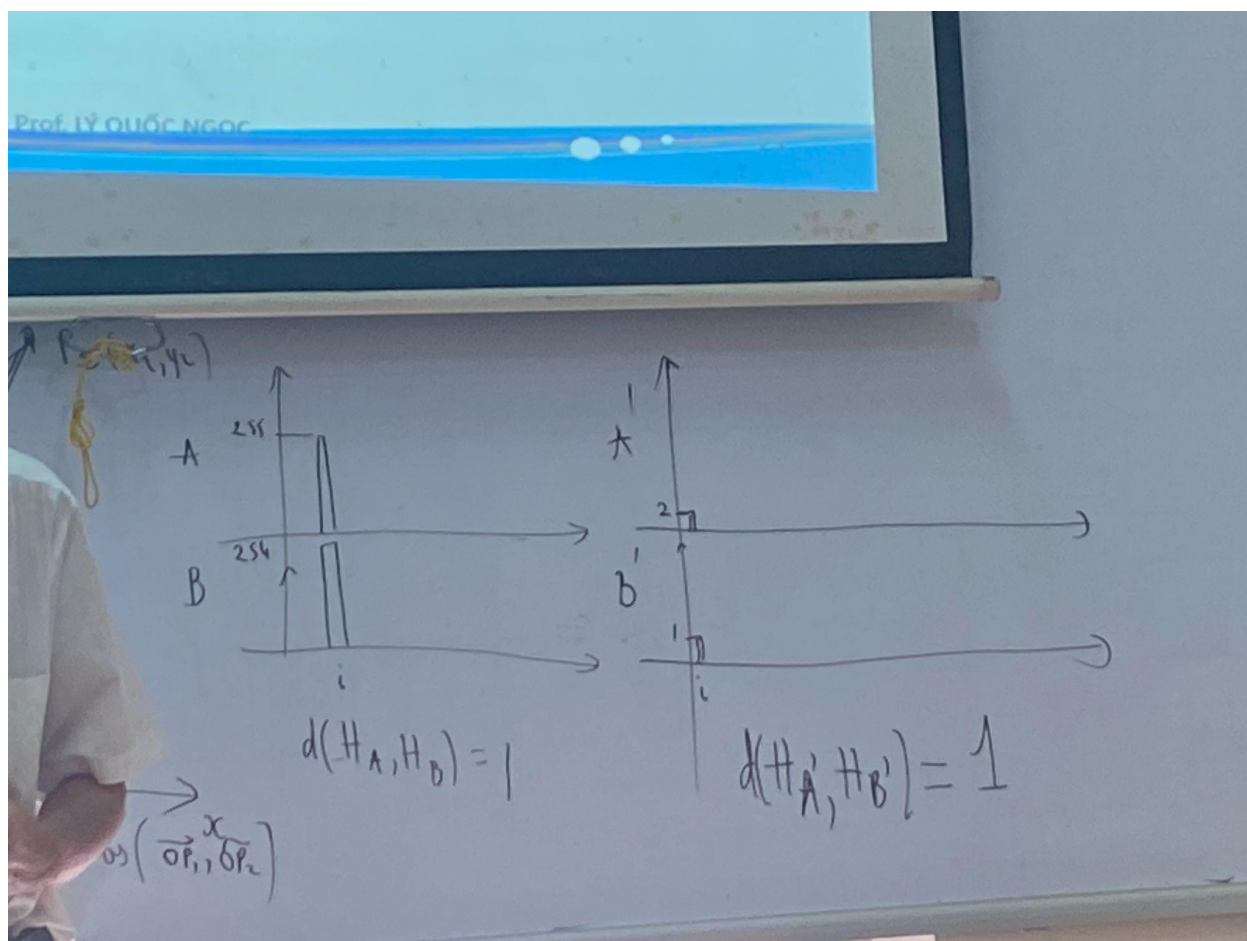
6. Nếu bỏ mẫu số thì điều gì sẽ xảy ra

4.3. Proximity Measures on Color Features

4.3.1. Lược đồ màu

$$D_H(I_Q, I_D) = \frac{\sum_{j=1}^n \min(H_{I_Q}(j), H_{I_D}(j))}{\sum_{j=1}^n H_{I_D}(j)}$$

Associate Prof. LỖ QUỐC NGỌC



Phải chia để thể hiện rõ sự khác biệt về tỉ lệ trong mỗi trường hợp. Trong trường hợp ta bỏ mẫu thì kết quả sẽ không phản ánh được tỉ lệ chính xác

Tuần 7 (28/06/2022)

1. Truy vấn theo vùng có ưu khuyết điểm gì (có chống chọi được các thách thức trong xử lý ảnh hay không)

Truy vấn ảnh theo vùng có thể khắc phục được 1 số thách thức trong xử lý ảnh như view, rotate, scale,... nhưng khi chất lượng ảnh bị thay đổi như cường độ sáng của ảnh, độ che khuất,.. Bị thay đổi thì sẽ ảnh hưởng tới kết quả truy vấn

2. Từ các vector đặc trưng thì đánh chỉ mục index như thế nào?

Visual words

- Example: each group of patches belongs to the same visual word

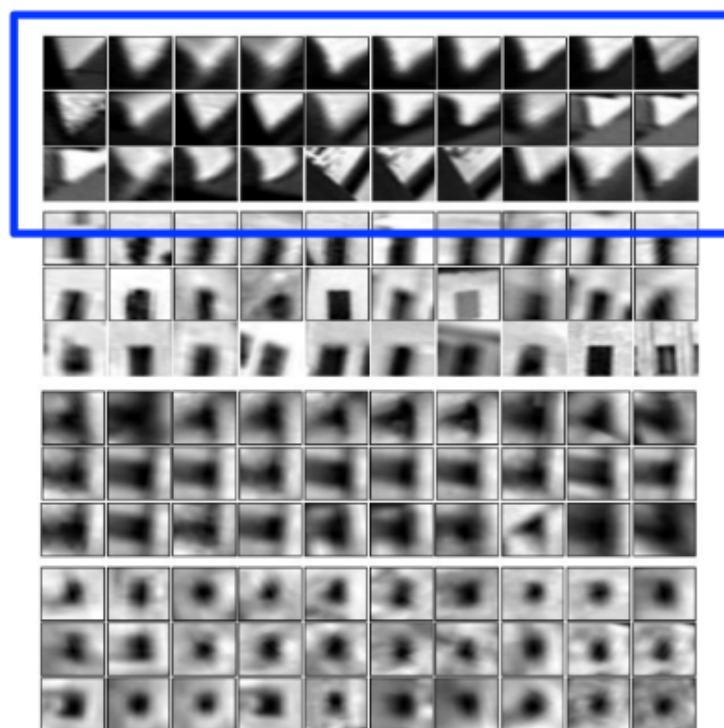
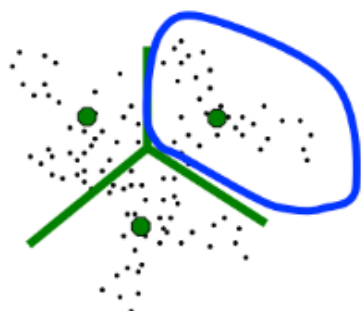


Figure from Sivic & Zisserman, ICCV 2003 Kristen Grauman

Mỗi vector đặc trưng hình là một loạt các số. Những gì sàn trượt đó làm là gán một định danh (từ trực quan) cho từng khối của các hình ảnh tương tự (đặc trưng).

Những gì lập chỉ mục sau đó là tìm thấy "giữa" của mỗi từ trực quan (nhóm các hình tương tự) để các hình ảnh khác có thể được lập chỉ mục so với các từ này. Các chấm màu xanh lá cây ở bên trái của sơ đồ là những người trung gian cho ba trong số các từ.

3. So sánh 2 cách truy vấn (point detector vs phân vùng nhóm)

Phương pháp phân đoạn dựa trên vùng tìm kiếm sự tương đồng giữa các pixel liền kề. Nghĩa là, các pixel sở hữu các thuộc tính tương tự được nhóm thành các vùng duy nhất. Như với tất cả các kỹ thuật phân đoạn, sử dụng cường độ mức xám là phương tiện phổ biến nhất để chỉ định sự giống nhau, nhưng vẫn tồn tại các khả năng khác, chẳng hạn như phương sai, màu sắc và các đặc trưng đa phương diện. Trong phương pháp phát hiện điểm, điểm được phát hiện tại một vị trí (x, y) trong một hình ảnh nơi mật độ được tập trung. Trong phương pháp phát hiện dòng, chúng ta có hai mật độ để các điểm tương ứng có nhiều khả năng được liên kết với một đường theo hướng của một mật độ là so với cái.

Seminar

- Project 1. Scalable Object Retrieval in very large image collections (Nhóm Innovation)

Chapter 4. Large scale retrieval

Các phương pháp để truy vấn với bộ dữ liệu lớn:

_HKM

_AKM -> Sử dụng k-d tree

_Sắp xếp lại không gian (Spatial re-ranking): Cải tiến phương pháp Bag-of-words ban đầu vì hệ thống truy xuất coi mọi tài liệu như một tập hợp các từ trực quan không có thứ tự -> Đề xuất phương pháp RANdom SAMple Consensus (RANSAC)

Vấn đề mất mát trong quá trình lượng tử hoá.

Chapter 5. Boosting recall

Vì các phương pháp trên giúp tăng Precision (Tìm sai) nhưng Recall không thay đổi. Vậy nên sau đây là một số phương pháp giúp tăng Recall nhưng không làm giảm Precision.

- Project 2. Scalable Object Retrieval in very large image collections (Nhóm Chaos)

Chapter 6. Object discovery and mining

Các phương pháp:

- Đồ thị so khớp (Matching graph)
- Latent Dirichlet Allocation (LDA)
- Geometric Latent Dirichlet Allocation (gLDA)
- Approximate inference
- Spatial scoring với RANdom SAMple Consensus (RANSAC)

Chapter 7. Descriptor learning

- Hồi quy lượng tử hóa (Quantization regression), so sánh với SIFT và tf-idf (term frequency - inverse document frequency)
- RANSAC

- Project 3. Advancing Large Scale Object Retrieval

Chapter 4. Improving Bag-of-Words Retrieval (Nhóm Innovation)

Ba nội dung chính trong chương này

RootSIFT	Discriminative query expansion	Database-side feature augmentation
Sử dụng khoảng cách Hellinger thay vì khoảng cách Euclidean tiêu chuẩn để đo độ tương tự giữa các mô tả SIFT dẫn đến tăng hiệu suất mạnh mẽ trong tất cả các giai đoạn của pipeline -> thay đổi rất đơn giản chỉ trong một vài dòng mã và không yêu cầu bất kỳ không gian lưu trữ bổ sung nào vì việc chuyển đổi từ SIFT sang RootSIFT có thể được thực hiện trực tuyến.	Các vector BoW các vùng được xác minh không gian được sử dụng để đưa ra các truy vấn mới, giải quyết vấn đề bỏ qua việc xác định đặc trưng ngoài lượng tử hoá và nhiễu trên bộ mô tả. Sử dụng SVM tuyến tính để học phân biệt một vector trọng số để truy vấn lại mang lại sự cải thiện đáng kể so với phương pháp Mở rộng truy vấn trung bình tiêu chuẩn (Chum et al., 2007b), trong khi duy trì tốc độ truy xuất ngay lập tức thông qua việc sử dụng hiệu quả Chỉ số đảo ngược.	Hạn chế chính của việc mở rộng truy vấn vẫn là dựa vào truy vấn để mang lại đủ số lượng kết quả có độ chính xác cao ngay từ đầu. Tăng cường tính năng bên cơ sở dữ liệu (Turcot và Lowe, 2009) là một bổ sung tự nhiên để mở rộng truy vấn -> hiệu quả nhưng không tính đến cấu trúc không gian của các đặc trưng tăng cường. => Sử dụng xác minh không gian bằng phương pháp đồng nhất.

Chapter 5. Improving Compact Image Representations (Nhóm Chaos)

- Nội chuẩn hóa (Intra-normalization)
- Vector của bộ mô tả tổng hợp cục bộ (VLAD - Vector of Locally Aggregated Descriptors)
⇒ Multi-VLAD
- Thích ứng kho từ (Vocabulary adaptation)

Chapter 6. Using Multiple Query Images to Boost Recall

- Project 4. Large-Scale Learning of Discriminative Image Representations

Chapter 3. Local Descriptor Learning

Chapter 4. Learning Descriptors from Unannotated Image Collections

- Project 5. Large-Scale Learning of Discriminative Image Representations

Chapter 5. Improving VLAD and Fisher Vector Encodings

Các mã hóa đặc trưng: với mục đích là tăng hiệu suất đối với vấn đề phân loại ảnh và truy vấn.

- Về hệ thống nhận diện hình ảnh: bằng việc sử dụng FV và bộ mô tả ảnh VLAD (có kết quả tích cực bằng việc sử dụng chuẩn hóa thích hợp và chuyển đổi tính năng cục bộ).
- Về phân loại hình ảnh: biểu diễn VLAD bằng cách sử dụng phép chiếu làm trắng không giám sát và phép chiếu các đặc điểm cục bộ đã được huấn luyện phân biệt.
- Về vấn đề rút trích dữ liệu: bằng phương pháp nội chuẩn hóa kèm theo chuẩn hóa phần dư, phép chiếu PCA cũng được ứng dụng để cải thiện mã hóa FV

Chapter 7. Learning Deep Image Representations

Deep Image Representation: Fisher Network có thể được coi là sự kết hợp giữa các biểu diễn nông | shallow representation (mà nó tổng quát) và mạng nơ-ron sâu.

- Ưu điểm: giảm độ phân biệt về chiều của các vector Fisher, điều này giảm thiểu số lượng tham số sau mỗi lần thực thi

- Project 6. AI-based VIR

Papers:

-Large-Scale Image Retrieval with Attentive Deep Local Features

Áp dụng mạng học sâu và cơ chế attention để đánh trọng số cho các vector đặc trưng quan trọng.

- Project 7. Graph-based VIR

Papers and Thesis:

- Graph-Based Image Retrieval: State of the Art
- Structured Query-Based Image Retrieval Using Scene Graphs
- Effective Graph-Based Content--Based Image Retrieval Systems for Large-Scale and Small-Scale Image Databases.