

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN



BÀI TẬP LỚP NHÀ

Nhóm MTAQ

Xử lý ảnh số và video số

Giảng viên – PGS. TS. Lý Quốc Ngọc

Thành phố Hồ Chí Minh - 2022

MỤC LỤC

THÔNG TIN CÁC THÀNH VIÊN TRONG NHÓM	4
I. Tuần 1	5
II. Tuần 2	11
III. Tuần 3	15
IV. Tuần 4	20
V. Tuần 5	23
VI. Tuần 6	27
VII. Tuần 7	31
VIII. Tuần 8	41
IX. Tuần 9	45
X. SEMINAR	46
1. Nhóm MyBack - Khảo sát phân đoạn ảnh dựa vào superpixel và ứng dụng trong phân đoạn ảnh y khoa.....	46
<i>a) Động lực nghiên cứu.....</i>	<i>46</i>
<i>b) Phương pháp.....</i>	<i>46</i>
2. Nhóm VIP - Khảo sát Image Sticking và ứng dụng để ghép nối các ảnh dưới các view khác nhau	47
<i>a) Động lực nghiên cứu.....</i>	<i>47</i>
<i>b) Phương pháp.....</i>	<i>47</i>
3. Nhóm 4T - Khảo sát Seamless cloning và ứng dụng để ghép ảnh trông tự nhiên hơn. (Khảo sát giải pháp truyền thống, deep learning)	47
<i>a) Động lực nghiên cứu.....</i>	<i>48</i>
<i>b) Phương pháp.....</i>	<i>48</i>
4. Nhóm Bất Ổn - Khảo sát Inpainting và ứng dụng để hồi phục dữ liệu (tài liệu bị ô vàng, rách nát).....	48
<i>a) Động lực nghiên cứu.....</i>	<i>48</i>
<i>b) Phương pháp.....</i>	<i>49</i>

5. Nhóm Động Lực - Khảo sát Super Video Resolution và ứng dụng trong tăng chất lượng ảnh, video. (Khảo sát giải pháp truyền thống, deep learning).....	49
<i>a) Động lực nghiên cứu.....</i>	<i>49</i>
<i>b) Phương pháp.....</i>	<i>50</i>
6. Nhóm CDKH - Khảo sát saliency map và ứng dụng trong phát hiện vùng quan tâm trong ảnh (Khảo sát giải pháp truyền thống, deep learning)	50
<i>a) Động lực khoa học</i>	<i>50</i>
<i>b) Động lực ứng dụng</i>	<i>50</i>
7. Nhóm Chiến thần - Khảo sát 2D, 3D Pose Estimation và ứng dụng trong nhận dạng hành động. (Khảo sát giải pháp truyền thống, deep learning)	51
<i>a) Động lực khoa học</i>	<i>51</i>
<i>b) Động lực ứng dụng</i>	<i>51</i>
8. Nhóm MTAQ - Hand detection and Finger Pose Estimation và ứng dụng trong giao tiếp người-máy	52
<i>a) Động lực khoa học</i>	<i>52</i>
<i>b) Động lực ứng dụng</i>	<i>52</i>
9. Nhóm Cù Lao Sweep - Khảo sát Optical flow và ứng dụng trong phát hiện camera motion, object motion trong Visual Slam.....	52
<i>a) Động lực nghiên cứu.....</i>	<i>52</i>
<i>b) Phương pháp.....</i>	<i>53</i>

THÔNG TIN CÁC THÀNH VIÊN TRONG NHÓM

MSSV	Họ Tên	Email	Ghi chú
20120201	Phạm Gia Thông	20120201@student.hcmus.edu.vn	
20120240	Dương Thị An	20120240@student.hcmus.edu.vn	
20120635	Phan Đình Anh Quân	20120635@student.hcmus.edu.vn	
20120328	Hoàng Đức Nhật Minh	20120328@student.hcmus.edu.vn	Nhóm trưởng

I. Tuần 1

1/ Thế nào là chuyển đổi số?

Chuyển đổi số là quá trình thay đổi tổng thể và toàn diện của cá nhân, tổ chức về cách sống, cách làm việc và phương thức sản xuất dựa trên công nghệ số.

2/ Lấy một số ví dụ về chuyển đổi số

Học online thông qua ứng dụng zoom, học sinh và giáo viên không cần nhất thiết phải đến trường để giảng dạy cho việc học (giảng dạy số).

Sàn thương mại điện tử như Shopee, Lazada, Tiki đã thay đổi việc kinh doanh truyền thống thay vào đó bằng việc kinh doanh bán hàng trên Internet mà không cần phải có mặt bằng.

Be, Gojeck, Grab, Uber đã thay đổi hình thức kinh doanh vận tải truyền thống, giúp kết nối các khách hàng có nhu cầu sử dụng vận tải (di chuyển hành khách, giao hàng) và các tài xế có xe, có thời gian rảnh và muốn kiếm thêm thu nhập một cách dễ dàng.

3/ Trong các cuộc cách mạng công nghiệp, những cốt lõi của từng cuộc cách mạng là gì?

Cuộc cách mạng thứ nhất dựa trên cơ khí máy móc (chạy bằng sức nước và hơi nước).

Cuộc cách mạng thứ hai dựa trên nền tảng năng lượng điện và các dây truyền sản xuất lớn ra đời.

Cuộc cách mạng thứ ba dựa trên nền tảng sự phát triển của chất bán dẫn, siêu máy tính, máy tính cá nhân và Internet.

Cuộc cách mạng thứ tư ra đời dựa trên nền tảng là dữ liệu. Yếu tố cốt lõi của kỹ thuật số trong cách mạng công nghiệp lần thứ tư là: Trí tuệ nhân tạo, IoT và dữ liệu lớn.

4/ Trong thiết kế sản phẩm, KHDL Thị giác nằm ở đâu?

Đồ họa cho phép một nhà thiết kế tương tác tổng hợp các hình dạng sản phẩm khác nhau, trực quan hóa chúng trong các cài đặt khác nhau và phân tích hiệu suất chức năng của chúng.

Ví dụ: người thiết kế thực hiện trong việc vẽ nhà, vẽ xe, mô phỏng hoạt họa 3D cho sản phẩm,...

5/ Trong kiểm tra sản phẩm, KHDL Thị giác nằm ở đâu?

Những lỗi sản phẩm xảy ra trong quá trình sản xuất là điều không thể tránh khỏi. Vì vậy, việc phát hiện sớm và loại bỏ sản phẩm khỏi dây chuyền là một điều cần thiết. Việc áp dụng kỹ thuật của khoa học dữ liệu thị giác vào phát hiện các khuyết điểm của sản phẩm trở nên rất quan trọng. Khoa học dữ liệu thị giác đã được áp dụng vào việc phát hiện khuyết điểm trên các bề mặt kim loại, phát hiện trái cây và rau củ bị hư, phát hiện các lỗi trong việc sản xuất các sản phẩm điện, ...

6/ Dữ liệu thị giác nằm ở đâu trong quy trình sản xuất và quản lý xã hội?

Quản lý trật tự và đảm bảo an ninh xã hội, giao thông:

- + Camera thông minh giám sát diện rộng tại khu dân cư, giám sát khuôn mặt tại cổng ra vào, nhận diện kẻ tình nghi lẫn trốn
- + Hệ thống điều phối giao thông thông minh (phát hiện xe vi phạm, phân luồng xe di chuyển hiệu quả tại các điểm kẹt xe,..)

Cải thiện đời sống của con người và quy trình trong công nghiệp:

- + Bãi giữ xe thông minh kiểm soát xe ra vào
- + Robot lau dọn nhà cửa, đưa đồ
- + Humanoid robot (sắp xếp hàng hóa), autonomous mobile robot (tự hành trong xí nghiệp) trong sản xuất công nghiệp
- + Hệ thống kiểm tra, thiết kế sản phẩm, phân loại hàng hóa
- + Hệ thống phát hiện sớm các mầm bệnh ung thư, mô phỏng 3D khuôn mặt người trong thuật thẩm mỹ

7/ Hãy nêu sự khác biệt của người và kỹ thuật số trong thu nhận ảnh.

<i>Mắt người</i>	<i>Máy ảnh kỹ thuật số</i>
- Cho ra được hình ảnh 3D	- Cho ra hình ảnh 2D

<ul style="list-style-type: none"> - Không thể thu được hình ảnh trong môi trường khắc nghiệt, không thể nhìn rõ vật ở phía xa (ánh sáng quá chói, quá nóng, quá lạnh, quá tối, nguy hiểm) - Thu nhận hình ảnh và xử lý thông tin của ảnh ngay tức thì 	<ul style="list-style-type: none"> - Thu nhận được hình ảnh rõ nét, tăng cường, thu nhận hình ảnh trong ánh sáng chói, trong bóng đêm tối, trong môi trường khắc nghiệt, thu được hình ảnh rõ nét ở xa - Chỉ thu nhận hình ảnh, video mà không xử lý
---	---

8/ KHDL Thị giác đã khai thác dữ liệu 3D từ dữ liệu 2D thu nhận được như thế nào?

Thu nhận dữ liệu 2D, dãy ảnh 2D từ các thiết bị máy quay, máy ảnh kỹ thuật số, camera của robot, thu nhận các điểm ảnh, sau đó qua các khâu xử lý ảnh, đồ họa máy tính, KHDL Thị giác thu nhận được một hình ảnh mô phỏng 3D được dựng trên các hình ảnh 2D đó.

9/ Ngữ nghĩa của hình

- Hình bên trái --> Hình cô gái và bà già
- Hình bên phải --> Hình con chim bồ câu và bông hoa

10/ Nội dung của bức ảnh chứa cái gì?

Một bức ảnh gồm có 3 thành phần chính:

- + Object (vật thể)
 - . Specific (tính cụ thể, riêng biệt): vd vật thể là xe hơi hiệu toyota, màu trắng, 7 chỗ
 - . General (tính chung chung): vd vật thể là một chiếc xe hơi
 - . Abstract (tính trừu tượng): vd vật thể đó là một phương tiện giao thông
- + Scene (khung cảnh, bức tranh tổng quát gồm nhiều vật thể)
- + Event (sự kiện diễn ra trong bức ảnh)

11/ Dữ liệu thị giác ở lấy ở đâu và dùng làm gì?

Dữ liệu lấy thông qua tín hiệu từ các camera, cảm biến, sensor được trang bị trên các Robot dùng nó để robot có thể xử lý, định hướng vị trí trong không gian và đi đến đích một

cách chính xác nhất. Ngoài ra có thể giúp robot phân loại hàng hóa, sắp xếp hàng hóa hiệu quả nhất.

Dữ liệu lấy từ hình ảnh, dãy ảnh của các camera giám sát để có thể nhận diện và lấy được thông tin đối tượng, truy vấn dữ liệu đã được lưu để chứng thực, truy vết đối tượng theo dãy ảnh nếu đối tượng đó đang bị tình nghi.

Dữ liệu còn được lấy thông qua quá trình học máy của máy tính và từ đó dữ liệu thị giác sẽ được tạo sinh nhiều hơn từ một tập dữ liệu ban đầu. Có được tập dữ liệu tạo sinh, máy tính sẽ học được thêm nhiều trạng thái của vật thể trong bức ảnh giúp nhận diện được vật thể đó dễ dàng hơn.

12/ Xử lý gì ở Humanoid Robot để thực hiện được các nhiệm vụ?

Có ba quy trình cơ bản cho một robot hình người để hoàn thành nhiệm vụ.

- Humanoid Robot phải hiểu môi trường xung quanh nó.
- Robot cần lập kế hoạch cho chuyển động phù hợp với thông tin về môi trường và nhiệm vụ nhất định.
- Robot cần phải tự điều khiển để thực hiện vật lý chuyển động và nhiệm vụ mong muốn.

13/ Phân biệt 3 môn Xử lý ảnh và video số, Thị giác máy tính, Đồ họa máy tính

Tên môn học	INPUT	OUTPUT
Đồ họa máy tính	- Tập tọa độ điểm 2D, 3D - Loại đối tượng	Ảnh 2D hiện trên thiết bị ngoại vi.
Xử lý ảnh số và video	Ảnh 2D	Ảnh 2D đã tăng cường chất lượng.
Thị giác máy tính	- Ảnh 2D - Tập tọa độ điểm 2D, 3D	- Semantics of Objects - 3D objects - Scene

14/ Phân biệt nhóm tác vụ Visual Data Understanding: Detection; Recognition; Classification; Tracking; Retrieval; Counting; Generating

Tác vụ	INPUT	OUTPUT
Detection	2D Image	- <i>Boundary box</i>

		- Object ID (general, abstract)
Recognition	2D Image	Object ID (specific)
Classification	2D Image	Class ID
Tracking	Dãy ảnh (image sequence)	Location of tracked object in the sequence (xe vượt đèn đỏ, truy vết).
Retrieval	Query Image	Retrieval Image ~ Query Image (trả về nhiều ảnh giống nhau. Ví dụ: trả về ảnh của nhiều chiếc xe có màu sắc, hình dáng gần giống với xe vượt đèn đỏ).
Counting	2D Image	Number of Objects (làm nhiệm vụ thống kê,...)
Generating	2D Image	New 2D Image

15/ Cho biết những tập dữ liệu cần được tạo sinh do rất khó thu thập dữ liệu truyền thống.

Tập dữ liệu về cử động của con người: Cùng một hành động (ví dụ hắc xì), là khác nhau về động tác, góc quay,...

Tập dữ liệu về khuôn mặt người. Con người mang nhiều sắc thái, cảm xúc khác nhau và trong một số trường hợp, khuôn mặt con người có thể bị che khuất hay thay đổi vì một yếu tố bên ngoài như đeo khẩu trang, lão hóa, đeo kính râm, đội nón, thay đổi màu tóc, thay đổi màu mắt,...

Tập dữ liệu về hóa thạch của các sinh vật sống thời cổ xưa. Vì các sinh vật sống đó nay không tồn tại nữa.

16/ Hiện nay, với dữ liệu khuôn mặt nhiều nhưng có cần phải tạo sinh để tạo thêm nhiều dữ liệu nữa hay không?

Cần phải tạo sinh để tạo thêm dữ liệu vì không bao giờ là đủ. Ứng với mỗi khuôn mặt sẽ có nhiều cảm xúc, trạng thái khác nhau, Face Agency khác nhau (khi về già, khi bị tai nạn có vết thương, vết sẹo, khi đeo khẩu trang, đeo mắt kính chỉ thấy một phần của khuôn mặt),...

17/ Vì sao lại dùng xe tự hành?

Để tận dụng thời gian chết của xe (thời gian mà xe không được con người sử dụng để lái và nằm trong bãi đỗ gara). Giảm tải được số lượng phương tiện giao thông lưu thông, tạo điều kiện cho phương tiện công cộng và xe tự hành, bảo vệ môi trường,...

Giúp tiết kiệm chi phí, mỗi người không nhất thiết phải sở hữu xe, xe vẫn tự đến và chở chúng ta đi khi được yêu cầu. Bên cạnh còn tiết kiệm thời gian, thời gian được phân bổ một cách hợp lý cho mỗi người, chỉ sử dụng xe khi cần di chuyển, không phát sinh thời gian chết, chi phí bãi đỗ gara cùng nhiều loại chi phí khác khi sở hữu xe.

18/ Hiện nay tại sao chưa có xe tự hành xuống đường?

Lý do lớn nhất: Chưa có một công ty bảo hiểm nào đứng ra bảo đảm khi có tai nạn xảy ra, vì người ta chưa tin vào xe tự hành; người kỹ sư chưa thể giải thích được tại sao xe có thể chạy được, đảm bảo an toàn.

Thời tiết khắc nghiệt.

Đường đi và lề đường sẽ tùy thuộc vào mỗi thành phố, mỗi quốc gia. Điều này sẽ xảy ra vấn đề nếu một chiếc xe tự lái chưa học lái xe, biển báo và đặc điểm đường ở địa điểm mà nó cần đi, dễ hiểu là hệ thống giao thông ở mỗi quốc gia không được đồng nhất giống nhau.

Chưa đảm bảo được sự an toàn.

II. Tuần 2

1/ Tại sao hệ thống truy vết phải theo dõi nhiều thuộc tính của con người?

- Hệ thống truy vết phải theo dõi nhiều thuộc tính của con người vì những đối tượng khác nhau có thể có nhiều đặc điểm giống nhau nên phải theo dõi nhiều thuộc tính để phân biệt các đối tượng với nhau.

2/ VR và AR khác nhau gì? Trong AR, VR thì ĐHMT có đóng góp gì?

Sự khác nhau của VR và AR

AR	VR
Xây dựng mô hình, thành phần 3D và đưa vào thế giới thực	Nhập vai hoàn toàn vào thế giới ảo
Người dùng luôn hiện diện trong thế giới thực	Người dùng sẽ được kiểm soát và quản lý bởi hệ thống
75% thực và 25% ảo	25% thực và 75% ảo
Hệ thống giúp người dùng đắm chìm một phần vào trải nghiệm	Hệ thống giúp người dùng đắm chìm hoàn toàn vào trải nghiệm
Yêu cầu băng thông từ 100 MBPS	Yêu cầu băng thông từ 50 MBPS
Người dùng vẫn liên hệ với thế giới thực khi tương tác với các thực thể ảo gần họ.	Người dùng bị cô lập với thế giới thực và đắm chìm bản thân trong một thế giới viễn tưởng.
Giao thoa giữa thế giới thực và thế giới ảo	Người dùng được đặt chân vào thực tế viễn tưởng như trong thế giới game
Ví dụ: Pokemon GO, Snapchat filters, ...	Ví dụ: Google VR, Google CardBoard, ...

Trong AR, VR thì đồ họa máy tính đã có những đóng góp là:

- Giúp xây dựng đối tượng trong AR, VR
- Hỗ trợ thiết kế đồ họa

3/ Cho ví dụ về hệ thống thị giác nào có khả năng hồi phục, sáng tác, ghi nhớ, tính toán, suy luận, dự đoán, phân biệt, tư hành, tìm kiếm, giao tiếp.

- Hồi phục: Hệ thống phục dựng hình ảnh, hệ thống khôi phục ảnh từ ảnh nhiễu, ...

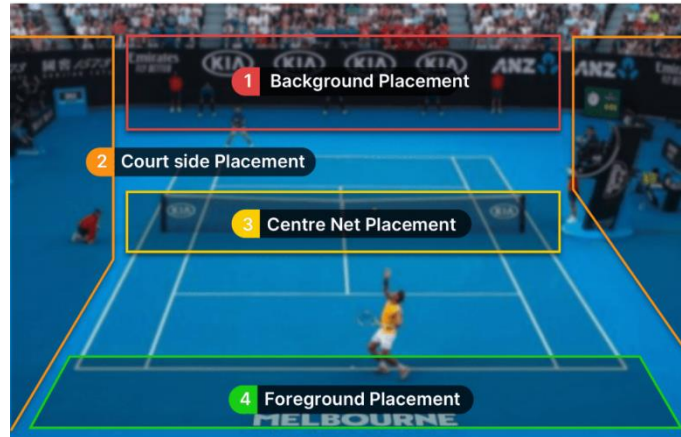
- Sáng tác: Hệ thống vẽ tranh từ mô tả, ...
- Ghi nhớ: Các hệ thống nhận diện, ...
- Tính toán: Tất cả các hệ thống
- Suy luận: Hệ thống nội suy từ ảnh 2D sang ảnh 3D, Hệ thống tạo sinh
- Dự đoán: Hệ thống chuẩn đoán bệnh ung thư phổi dựa trên ảnh X-Ray, hệ thống chuẩn đoán bệnh da liễu,...
- Phân biệt: Hệ thống nhận dạng khuôn mặt, hệ thống nhận diện biển báo giao thông,...
- Tự hành: Xe giao hàng không người lái, xe tự lái (Tesla, VinFast,...), máy bay không người lái, ...
- Tìm kiếm: Google Lens, Pinterest Lens, ...
- Giao tiếp: Hệ thống chuyển kí hiệu và biểu tượng thành lời nói, văn bản và ngược lại, Hệ thống chuyển hình ảnh thành mô tả và ngược lại

4/ Mô hình áp dụng AI, giúp tạo trí khôn cho Visual System ở chỗ nào?

Áp dụng AI vào Visual System, hay [Visual Artificial Intelligence](#) giúp máy tính không chỉ hiểu về những *đặc trưng thị giác* của ảnh như con người, mà còn rút trích được *đặc trưng ngữ nghĩa* của bức ảnh.



Phân loại từng object trong một bức ảnh, gán nhãn đúng



Real-time intelligence

5/ Không cần DHMT, XLA, TGMT mà chỉ cần mô hình và dữ liệu đã được đánh nhãn thì có đúng hay không?

Không. nhu cầu có thêm dữ liệu ảnh cho mô hình là rất cần thiết. Đồ họa máy tính, xử lý ảnh sẽ dùng các thuật toán, các phương trình toán học để biến đổi, rút trích các đặc trưng trọng điểm của ảnh hay tạo thêm dữ liệu ảnh input cho mô hình. Giúp cho mô hình học sâu có performance tốt hơn.

6/ Cho ví dụ về mức độ ngữ nghĩa cho đầu ra của hệ thống thị giác

- Tên ứng dụng: Hệ thống quét sinh trắc học vân tay người
- Input: Dữ liệu ảnh từ máy quét vân tay
- Output:
 - + Wisdom: Tự dự đoán, tạo sinh được vân tay (vân tay bị đặt méo chỉ thấy một phần, vân tay bị vết thương), xác định đó có phải là vân tay người hay không, đó có phải là một bức ảnh vân tay người hay không (lấy hình ảnh vân tay để giả), nhận biết vân tay có nhiệt độ của cơ thể (có thể dùng mô hình tay giả có vân tay để quét)
 - + Intelligence: Xác định được tính hợp lệ của vân tay (người lạ hay người quen), xác định được tính cách của người đó, xác định rõ căn bệnh
 - + Knowledge: dự đoán được độ tuổi (dấu vân tay to/nhỏ), ID của người đó từ vân tay nhận diện được, dự đoán được bệnh
 - + Information: Thông tin, đặc trưng của vân tay
 - + Data: Ảnh vân tay đã được cắt ra, có cân chỉnh độ sáng, độ rõ nét nếu cần thiết

7/ Phân biệt tác vụ Detection, Recognition, Classification, Retrieval

Detection	Recognition	Classification	Retrieval
Object ID + Location (miền bao đối tượng và phát hiện vị trí của đối tượng)	Specific ID (nhận biết đối tượng với thuộc tính đặc trưng riêng biệt)	Class ID (nhóm đối tượng cùng thuộc tính với nhau)	Truy vấn đối tượng theo đặc trưng (màu sắc, kích thước, hình dáng) và ngữ nghĩa

8/ Cho các ví dụ về Detection, Recognition, Classification, Retrieval sẽ dùng ở đâu trong tầng 3?

- Detection: phát hiện vị trí lỗi trên bề mặt sản phẩm như vết xước, vết nứt, lỗi, lõm,... phát hiện thực phẩm bị hư hại, phát hiện vật cản trong hệ thống xe tự hành.
- Recognition: nhận dạng người thông qua các thuộc tính đặc trưng để ứng dụng trong giám sát an ninh, truy vết tội phạm, nhận biết các hoạt động bất thường trong kiểm tra thi cử.
- Classification: phân loại các đối tượng tham gia giao thông (ô tô, xe gắn máy, người đi bộ,...), phân loại thực phẩm,...
- Retrieval: truy vấn đối tượng theo đặc trưng, ứng dụng trong điểm danh lớp học, tìm kiếm thông tin liên quan đến về bức ảnh, truy hồi thông tin ngôn ngữ trên tập ảnh sang dạng văn bản.

9/ Tìm một ví dụ thực hiện nhiều tác vụ nhất có thể

Hệ thống phân tích hình ảnh chụp từ X-Quang, Siêu âm, MRI,...

- Detection: Phát hiện vị trí đối tượng cần theo dõi trong ảnh, bounding box theo từng bộ phận cơ thể hoặc các chi tiết trên 1 bộ phận cần theo dõi.
- Recognition: Nhận biết các chi tiết bất thường ở các bộ phận cơ thể (có khối u, phát hiện bóng trong mô phổi,...)
- Classification: phân loại hình ảnh dựa vào các đối tượng, hình dạng, màu sắc.
- Retrieval: truy xuất hình ảnh của bệnh nhân trong các lần chụp, đưa ra các điểm bất tương đồng trong tập dữ liệu của bệnh nhân.
- Reconstruction: tái tạo hình ảnh 3D từ 2D của các chi tiết như khối u để đưa ra hình dáng chính xác của chúng.
- Generation: tạo ra báo cáo tự động sau khi phân tích hình ảnh

III. Tuần 3

1/ Công nghệ nguồn là gì?

- Công nghệ nguồn là công nghệ cốt lõi nền tảng để phát triển ứng dụng trong sản xuất, vật lý và xã hội.

2/ Tầm quan trọng của tầng 3.

- Tầng 3 là tầng ứng dụng tổng hợp những công nghệ đã được nghiên cứu ở tầng 1 và tầng 2, là nơi mà các hệ thống tự động dần được ra đời, hoạt động một cách hiệu quả và thậm chí là an toàn hơn nhất là trong thời đại kỹ thuật số như ngày nay. Thiếu tầng 3 thì những nghiên cứu, phương pháp ở tầng 1, tầng 2 chỉ còn là những lý thuyết tổng hợp, không có được nơi để ứng dụng vào.

- Ứng dụng được nhiều trong xã hội, sản xuất và đời sống của con người, là nơi kết nối giữa thành tựu AI với cuộc sống hiện đại.

- Tầng 3 còn là một trong những mỏ vàng, hái ra tiền cho ngành Business IT và được nhận nhiều đầu tư công nghệ, nguồn vốn lớn từ các doanh nghiệp. Từ đó thì kinh tế ngày càng phát triển, cuộc sống con người ngày càng đơn giản và hiện đại hơn.

3/ Tầng 1 trong xử lý ảnh thì làm gì?

- Tiền xử lý; Biến đổi; Phát hiện điểm trọng yếu, điểm neo; Rút trích đặc trưng; Xử lý dữ liệu trong miền không gian, miền không gian - thời gian, miền tần số, xử lý dữ liệu đa phương thức, các mô hình học, xử lý màu, ...

4/ Hành động (Action) khác đối tượng (Object) chỗ nào? (Action Recognition vs Object Recognition)

	INPUT	OUTPUT
Object Recognition	Ảnh, điểm ảnh	Object, Scene
Action Recognition	Dãy ảnh	Event

5/ Ngoài Object, Action thì còn nhận dạng cái gì nữa?

- Human object interaction - Nhận biết sự tương tác giữa người với đối tượng thông qua ảnh đơn
- Nhận biết hành động đúng hay sai thông qua nhận diện hành vi và ngữ cảnh

- Thinking&Action: nhận biết suy nghĩ và hành vi trong suy nghĩ thông qua sóng não

6/ Nếu nghiên cứu nhận dạng hành động hành vi thì ứng dụng gì ở tầng 3

- Ứng dụng được vào hệ thống camera giám sát ở các sân bay, hội nghị quan trọng cấp cao, giám sát ở khu vực dân cư, những điểm giao thông phức tạp, những nhà tù nguy hiểm,... Bên cạnh đó còn có thể nhận dạng hành vi thông qua các cảm xúc của khuôn mặt hay theo dõi thú cưng tại nhà.
- Có thể ứng dụng trong hệ thống chuyển hành động thành văn bản, kí hiệu, lời nói cho những người bị câm, điếc và ngược lại.

7/ Tại sao nhận dạng biển số xe lại làm trong 40 năm ở Mỹ?

- Tại Mỹ, biển số xe có thể mang tên chủ nhân, những biểu tượng hoặc cả một câu khẩu hiệu của một tiểu bang, đặc biệt, có thể vẽ bất cứ thứ gì lên bảng số xe... Vì vậy, việc nhận dạng biển số thay đổi liên tục, cần thu thập dữ liệu mới thường xuyên và không đơn thuần là nhận dạng chữ số như ở Việt Nam.

8/ Vì sao con drone phát hiện bệnh chết nhanh trên cây hồ tiêu lại không được duyệt?

Vì drone phát hiện bệnh dựa vào lá cây hồ tiêu nhưng mầm bệnh lại tấn công từ gốc rễ, biểu hiện bệnh ở lá có thể ví như giai đoạn cuối của bệnh ung thư. Vì thế, việc phát hiện bệnh khi này không có quá nhiều tác dụng để cứu sống cây. Hơn nữa, việc duy trì hệ thống drone trên diện tích lớn rất là tốn kém và nếu vùng hồ tiêu đó không có bệnh thì việc làm này vô cùng lãng phí.

9/ Nhận diện khuôn mặt, y phục, hành vi bất thường, mật độ người ở tầng 1, 2 phải làm gì?

Ở tầng 1, ta phải nghiên cứu các công việc như: tiền xử lý, biến đổi, xử lý dữ liệu,... dựa trên nền tảng các công thức, tri thức có sẵn để tạo ra các mô hình toán mới, mô hình xử lý dữ liệu thị giác mới để hỗ trợ xây dựng các hệ thống thông minh.

Ở tầng 2, ta sẽ đi cài đặt và thử nghiệm trên các công đoạn của hệ thống thị giác: phát hiện, nhận dạng, phân đoạn, phân lớp, theo vết, truy vấn, hồi phục, tái tạo ba chiều, tự định vị,...

10/ Nhận dạng thuộc tính mặt người để làm gì? Tài liệu tham khảo: [Applications of Facial Recognition Technology in Different Industries \(anyconnect.com\)](https://anyconnect.com)

Công nghệ nhận diện khuôn mặt hiện này làm được nhiều thứ hơn là xác minh một người. Có thể kể đến như sau:

- An ninh, bảo mật: Các công ty an ninh dùng công nghệ này để bảo vệ cơ sở, vật chất của họ
- Các trạm kiểm soát xuất nhập cảnh sử dụng nhận dạng khuôn mặt để thực thi việc kiểm soát biên giới thông minh hơn.
- Các công ty ô tô có thể ứng dụng, để đảm bảo đúng tài xế của chiếc xe, đưa đón đúng hành khách của mình, tránh bị đột nhập,...
- Các công ty, hãng phim, khu trung tâm thương mại, siêu thị có thể theo dõi cảm xúc của khách hàng để đánh giá sự hài lòng về sản phẩm, cũng như thái độ của nhân viên, từ đó đưa ra các chiến lược thay đổi thích hợp hơn.

11/ Đếm số người trong đám đông để làm gì?

Có rất nhiều ứng dụng từ việc đếm số người trong đám đông, có thể kể ra như sau:

- Đánh giá tình trạng kẹt xe ở một khu vực.
- Kiểm soát số lượng người tham gia một sự kiện nào đó (Festival, Biểu tình, Đại nhạc hội,...)
- Nhận biết số người cần di tản trong 1 khu vực đang gặp hiểm họa thiên tai.

12/ Đếm số đầu người trên xe máy để làm gì?

- Để nhận biết số người đang có trên xe từ đó kiểm tra xe có chở quá số người quy định hay không và những người trên xe có đội mũ bảo hiểm hay không.

13/ Truy vấn y phục đưa vào thuộc tính thông minh ở chỗ nào?

- Biết được đâu là y phục, áo quần trên cơ thể người
- Biết được loại y phục
- Biết được thuộc tính y phục: có tay, màu sắc,...

14/ Phân biệt truy vấn thuộc tính y phục với truy vấn thuộc tính mặt người.

Truy vấn y phục	Truy vấn mặt người
Chỉ có thông tin về y phục, không thể phân biệt nếu các Object có y phục giống nhau	Có thêm ID của Object để nhận diện khuôn mặt, phân biệt riêng biệt với từng Object khác nhau

15/ Tái tạo di sản văn hóa làm gì? So sánh điều này với quay phim video.

- Để có thể tạo ra các mô hình di sản văn hóa một cách đầy đủ nhất để trưng bày ở các bảo tàng, đảm bảo việc giữ nguyên vẹn cho di sản văn hóa gốc (nếu đặt di sản văn hóa gốc thì có thể bị đánh cắp, bị hư hại do khách tham quan và nhiều yếu tố ảnh hưởng khác).

Tiêu chí	Quay phim video 2D	Tái tạo 3D
Đặc điểm	<ul style="list-style-type: none"> - 2 chiều: Chiều dài, rộng - Không nhìn thấy được các kích thước vật lý về độ sâu của đối tượng 	<ul style="list-style-type: none"> - 3 chiều: Dài – rộng – sâu - Thể hiện đầy đủ các kích thước vật lý của đối tượng (chiều cao, chiều ngang, chiều sâu)
Ưu điểm	<ul style="list-style-type: none"> - Là dạng video phổ biến, phù hợp với nhiều lĩnh vực. - Công đoạn sản xuất đơn giản. - Có thể làm thủ công hoặc trên những phần mềm làm hoạt hình 2D miễn phí. 	<ul style="list-style-type: none"> - Hình ảnh chân thực, sống động. - Hình ảnh được thể hiện với nhiều góc quay khác nhau. - Có thể dễ dàng tái sử dụng nhân vật trong những lần sau bằng cách thay đổi kiểu dáng, hành động. - Khơi gợi được sự hứng thú, mới lạ cho người xem
Nhược điểm	<ul style="list-style-type: none"> - Chỉ được chụp ảnh mà không lựa được góc quay. - Thường tốn nhiều thời gian và công sức vì phải vẽ nhiều hình và sắp xếp rất tỉ mỉ. - Video 2D không thu hút bằng video 3D 	Cần nhiều chi phí đầu tư và thời gian sản xuất lâu hơn.
Chi phí sản xuất	Thấp	Chi phí sản xuất cao hơn
Thời gian sản xuất	Nhanh chóng	Thời gian lâu hơn

16/ Tái tạo mặt 3D để làm gì? Đem lại lợi ích gì?

- Để có thể tạo ra bộ dữ liệu khuôn mặt với nhiều góc nhìn khác nhau (từ trên xuống, bên trái sang, bên phải sang, từ dưới lên chứ không nhìn thẳng trực tiếp từ 1 hướng như ở ảnh 2D)
- Tạo ra được bộ dữ liệu khuôn mặt với nhiều phương pháp chiếu sáng khác nhau (một nửa mặt sáng, nửa bên kia tối thì sẽ như thế nào)
- Tạo ra bộ dữ liệu khuôn mặt với nhiều thuộc tính (trẻ thành già, già thành trẻ, thay đổi cảm xúc của khuôn mặt), đặc điểm (khuôn mặt đeo kính, đeo khẩu trang, các chi tiết khác có thể xuất hiện từ khuôn mặt ban đầu)
 - Ưu điểm:
 - + Nhận diện được khuôn mặt với tốc độ nhanh (ngay cả khi bạn đeo kính, trang điểm hay thay đổi theo tuổi tác)
 - + Có khả năng nhận diện chính xác các đặc điểm, thuộc tính khuôn mặt
 - + Tạo ra bộ dữ liệu đa dạng làm nguồn cung cấp giàu có cho các hệ thống học máy.
 - + Ứng dụng trong việc truy vết, truy vấn đối tượng của các hệ thống nhận diện

IV. Tuần 4

1/ Liệt kê vài ứng dụng của xử lý ảnh số và video số?

- Trong công nghiệp, sản xuất:
 - + Hệ thống phân loại sản phẩm, kiểm tra lỗi, tính hoàn thiện của sản phẩm: làm rõ nét hình ảnh sản phẩm nhận được, cân chỉnh màu, tăng độ sáng
- Thông tin ảnh, truyền thông ảnh: chỉnh sửa chi tiết, thay đổi, di chuyển các đối tượng trong bức ảnh, nén video, truyền thông video, truyền hình số
- Thiên văn, nghiên cứu không gian, vũ trụ: tập trung độ sáng ở các hình ảnh ngôi sao, scale hình ảnh vẫn giữ được độ chi tiết và rõ nét, xử lý ảnh vệ tinh, viễn thám
- Người máy, tự động hóa: chống rung cho dãy ảnh khi người máy di chuyển
- Sinh học, y học: chụp cắt lớp vị trí các khối u, cân chỉnh độ rõ nét
- Điều khiển lưu lượng chuyển động: phân biệt màu của các đối tượng trong đám đông, giám sát theo dõi phát hiện chuyển động

2/ Với số lượng màu (16,8 triệu màu) như vậy thì đã đủ chưa?

Với số lượng màu (16,8 triệu màu) như vậy là vừa thiếu mà cũng vừa dư. Có một số màu trong tự nhiên mà máy không thể biểu diễn được và cũng có một số màu do thế giới số tạo ra, nhưng mà không thể thể hiện ra bên ngoài một cách toàn vẹn.

3/ Tìm và tính xem độ dị biệt giữa 2 điểm màu

Độ dị biệt của hai điểm màu = $\sqrt{(R_2 - R_1)^2 + (G_2 - G_1)^2 + (B_2 - B_1)^2}$

Source: https://en.wikipedia.org/wiki/Color_difference

4/ Xét ưu và khuyết điểm của hệ màu RGB. Gợi ý: Sự tương đồng về đặc trưng thị giác có dẫn đến sự tương đồng về đặc trưng ngữ nghĩa không?

Sự tương đồng về đặc trưng thị giác thì không dẫn đến sự tương đồng về đặc trưng ngữ nghĩa, ví dụ người này nhìn thấy hình ảnh như vậy nhưng người khác nhìn theo hướng khác sẽ ra được ý nghĩa của bức ảnh đó khác

5/ Cho vd để cho thấy hệ màu RGB gây cản trở như thế nào cho việc ứng dụng.

Số lượng các màu của hệ màu RGB quá nhiều, có thể gây ra dư thừa, không dùng đến các màu ảnh đó trong việc ứng dụng

6/ Cho ví dụ để thấy mô hình màu HSV tiện lợi hơn RGB

1. **Biểu diễn trực quan:** Mô hình màu HSV biểu thị các màu theo sắc thái (màu sắc), độ bão hòa (cường độ) và giá trị (độ sáng) của chúng. Điều này có thể trực quan hơn mô hình RGB, mô hình này biểu thị màu sắc theo các thành phần đỏ, lục và lam của chúng.
2. **Điều chỉnh các tham số riêng lẻ:** Trong mô hình màu HSV, có thể dễ dàng điều chỉnh màu sắc, độ bão hòa hoặc giá trị của màu một cách độc lập mà không ảnh hưởng đến các tham số khác. Trong mô hình màu RGB, có thể khó thực hiện các loại điều chỉnh này mà không ảnh hưởng đến các kênh màu khác.
3. **Lựa chọn màu sắc:** Mô hình màu HSV có thể hữu ích cho việc lựa chọn màu sắc vì nó cho phép bạn chỉ định sắc độ, sau đó điều chỉnh độ bão hòa và giá trị để tinh chỉnh màu sắc. Điều này có thể trực quan hơn so với việc chọn màu trong mô hình màu RGB, trong đó bạn phải chỉ định các giá trị chính xác cho các kênh màu đỏ, lục và lam.

7/ Tìm độ dị biệt giữa 2 điểm màu trong mô hình HSV?

- Có 1 vài công thức được dùng để tính độ dị biệt giữa 2 điểm màu tùy theo nhu cầu sử dụng:
- Đặt:
 - Color 1: hue = h_1 , saturation = s_1 , value = v_1
 - Color 2: hue = h_2 , saturation = s_2 , value = v_2

1. Công thức 1:

$$\text{difference} = |h_1 - h_2| + |s_1 - s_2| + |v_1 - v_2|$$

2. Công thức 2:

$$\text{difference} = \sqrt{(h_2 - h_1)^2 + (s_2 - s_1)^2 + (v_2 - v_1)^2}$$

8/ Trong 3 đặc trưng màu, vân, dáng cái nào quan trọng nhất?

Trong 3 đặc trưng màu, vân dáng thì dáng là quan trọng nhất.

9/ Có khi nào có nhiều ảnh khác nhau mà có cùng một histogram hay không?

Có, do histogram chỉ biểu thị chỉ số màu theo biểu đồ cột chứ không biểu thị sự phân bố của màu. Do vậy, nhiều ảnh khác nhau có thể có cùng một histogram nhưng khác phân bố màu.

10/ Hãy nêu những thách thức của 3D Information, Scale (PSR), Viewpoint, Occlusion, Clutter Background , Semanticgap, Lighting?

- 3D Information: thiếu dữ liệu thông tin về đối tượng trong bức ảnh, bức ảnh bị chụp mờ, chụp mất 1 phần đối tượng, màu sắc không rõ, bức ảnh đã bị scale
- Lighting: bức ảnh chụp có độ histogram bị thay đổi, một điểm ảnh có giá trị màu bị tăng cao đột ngột so với mức thông thường

11/ Viết một giải thuật tính lược đồ màu trong không gian màu HSV

RGB → HSV

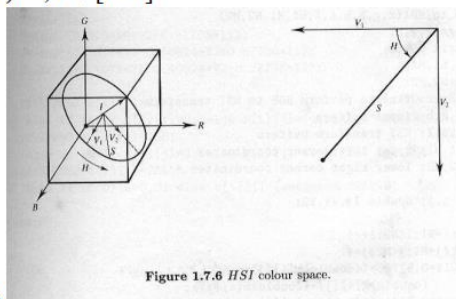
$$\begin{bmatrix} V \\ V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} \sqrt{3}/3 & \sqrt{3}/3 & \sqrt{3}/3 \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} \\ 2/\sqrt{6} & -1/\sqrt{6} & -1/\sqrt{6} \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

$$H = \tan^{-1}(V_2/V_1)$$

$$S = \sqrt{V_1^2 + V_2^2}$$

Giá trị màu tại điểm ảnh gồm bộ ba giá trị (H,S,V).

$H \in [0..360)$, $S, V \in [0..1]$.



12/ Viết giải thuật tính lược đồ hệ số góc?

- Gồm có 73 phần tử.
- 72 phần tử đầu tỉ lệ các điểm ảnh tương tự nhau.
- Phần tử cuối là số lượng điểm ảnh không nằm trên biên.

$$H_i = \frac{m(I_0)}{n_{I_0}}$$

với $m(I)$ là số lượng điểm ảnh có hệ số góc α .

- Có những trường hợp các ảnh khác nhau nhưng lại có chung lược đồ màu hoặc lược đồ hệ số góc. Điều này là có thể tại vì các ảnh đó có tỉ lệ xuất hiện của các màu tương tự nhau nhưng vị trí xuất hiện trong ảnh khác nhau dẫn đến ảnh khác nhau. Tương tự như lược đồ hệ số góc.

V. Tuần 5

1/ Trong các phép biến đổi màu, có phép thay đổi độ tương phản(contrast). Vậy khi nào ta làm tăng độ tương phản, khi nào làm giảm độ tương phản? Vì sao $g(x,y) = a.f(x,y)$ giúp thay đổi độ tương phản?

Khi ta muốn làm tăng độ chênh lệch sáng tối của 2 điểm dữ liệu trên bức ảnh.

Cụ thể:

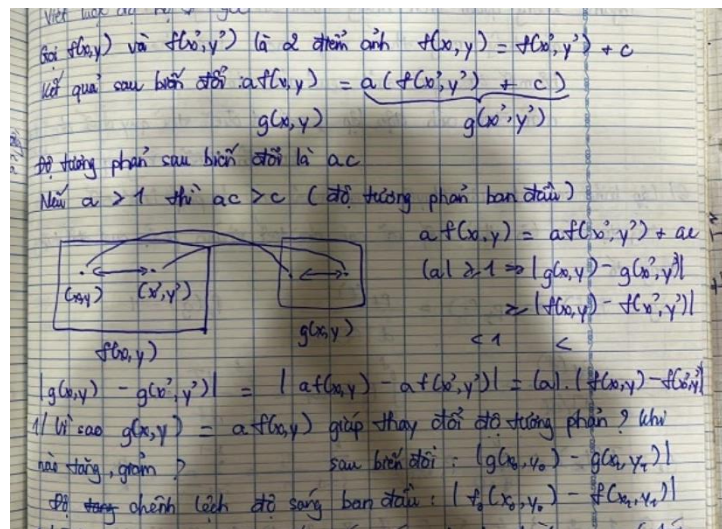
- Chọn 2 pixel $f(x,y)$ và $f(x1, y1)$ trên ảnh gốc, tương ứng với 2 pixel $g(x,y)$, $g(x1,y1)$ trên ảnh đã được tăng độ tương phản. Áp dụng công thức thay đổi độ sáng: $g(x, y) = a.f(x, y)$, ta có:

$$\begin{aligned} & \text{ABS}(g(x, y) - g(x1, y1)) \\ &= \text{ABS}(a.f(x, y) - a.f(x1, y1)) \\ &= \text{ABS}(a) \cdot \text{ABS}(f(x, y) - f(x1, y1)) \end{aligned}$$

- Nếu $a \geq 1$:

$$\text{ABS}(g(x, y) - g(x1, y1)) \geq \text{ABS}(f(x, y) - f(x1, y1)) \Rightarrow \text{Tăng độ tương phản.}$$

- Nếu $a < 1$: $\text{ABS}(g(x, y) - g(x1, y1)) < \text{ABS}(f(x, y) - f(x1, y1)) \Rightarrow$ Giảm độ tương phản.



2/ Tập ảnh đang có mức xám $[f1, f2]$. Làm sao chuẩn hóa tập ảnh về mức xám $[g1, g2]$

Phát biểu bài toán: $g(x) = af(x) + b$

$$\begin{cases} g_1 = af_1 + b & (1) \\ g_2 = af_2 + b & (2) \end{cases}$$

$$(2) - (1) \Rightarrow \begin{cases} g_1 = af_1 + b \\ a = \frac{g_2 - g_1}{f_2 - f_1} \end{cases}$$

$$\Leftrightarrow \begin{cases} b = \frac{g_1 f_2 - f_1 g_2}{f_2 - f_1} \\ a = \frac{g_2 - g_1}{f_2 - f_1} \end{cases}$$

$$\text{Vậy, } g(x) = \frac{g_2 - g_1}{f_2 - f_1} f(x) + \frac{g_1 f_2 - f_1 g_2}{f_2 - f_1}$$

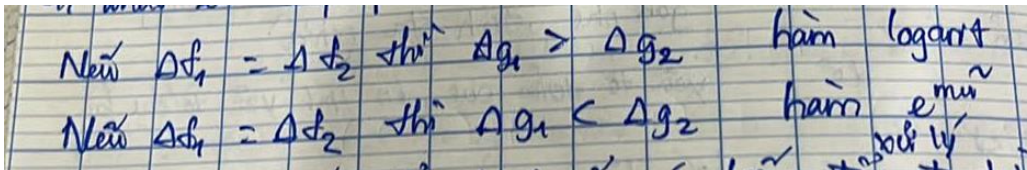
3/ Tại sao dùng phép biến đổi phi tuyến. Dùng phép biến đổi tuyến tính có được không?

Vì phép biến đổi phi tuyến vì sự thay đổi giá trị của phép biến đổi này không đồng đều. Nếu như trong phép biến đổi tuyến tính, kết quả của phép biến đổi của khoảng (a_1, b_1) và (a_2, b_2) (giả sử 2 khoảng này có độ lớn như nhau) là tương tự nhau, còn trong phép biến đổi phi tuyến, kết quả của phép biến đổi của 2 khoảng này là khác nhau

Các điểm màu được thay đổi khác nhau, không thể như nhau được

4/ Phép biến đổi phi tuyến khác gì so với phép biến đổi tuyến tính?

Sự khác nhau cơ bản là độ lớn của sự thay đổi của kết quả sau khi áp dụng phép biến đổi. Trong phép biến đổi tuyến tính, độ lớn này là không thay đổi (xét trường hợp các khoảng cách của input là như nhau), còn trong phi tuyến, độ lớn thay đổi tùy theo phép biến đổi đang áp dụng.



5/ Hình dưới đây, các cặp ảnh, với ảnh bên trái có thể dùng phép biến đổi tuyến tính hay phép biến đổi phi tuyến để tạo thành ảnh tương ứng bên phải không?

Không. Vì các bức ảnh bên trái có nhiều chi tiết, đường nét chưa rõ, thậm chí không thể thấy được, rất mờ, nhưng các ảnh bên phải thì lại được biến đổi rất rõ ràng đến từng chi tiết. Vì thế ta không thể dùng phép biến đổi tuyến tính hay phi tuyến được, ta phải dùng cân bằng lược đồ xám (histogram qualization)

3.1.3. Phép biến đổi dựa trên phân bố xác suất



6/ Lập bảng so sánh: nguyên lý, phương pháp, độ phức tạp tính toán, hiệu quả, ưu điểm và nhược điểm của các phương pháp biến đổi màu: tuyến tính, phi tuyến, dựa trên phân bố xác suất(Histogram Specification và Histogram Equalization).

Tiêu chí	Tuyến tính	Phi tuyến	Phân bố xác suất	Đặc tả hình học
Nguyên lý	Biến đổi tập giá trị màu	Biến đổi màu từng giá trị màu, điểm ảnh	Biến đổi màu dựa trên phân bố mức xám toàn bộ bức ảnh	Biến đổi màu dựa trên phân bố mức xám đã cho trước mà chưa biết ảnh g
Phương pháp tính	$g = a.f + b$	$g = \log f$ $g = e^f$	$g = T(f)$ $T[R] = T[p-1] + H[p]$	$g = G(T[f])$ $T[p] = T[p-1] + H_f[p]$ $G[p] = G[p-1] + H_g[p]$
Độ phức tạp thuật toán	$O(N)$	$O(N)$	$O(N)$ $O(GL)$	$O(N)$ $O(GL-f)$ $O(GL-g)$

Hiệu quả	Thay đổi độ sáng, độ tương phản với cùng 1 công thức tuyến tính	Tăng độ sáng của những vùng sáng trong hình ảnh và giảm độ sáng của những vùng tối, cải thiện độ tương phản.	Có xét ngữ nghĩa của bức ảnh và chỉnh theo	Có xét ngữ nghĩa của bức ảnh và chỉnh theo
Khuyết điểm	Chỉnh cả ảnh, không thể điều chỉnh từng nơi	Chỉ đổi giá trị màu của một điểm ảnh theo giá trị màu của điểm ảnh đó theo công thức	Đôi khi tạo ra bức ảnh không tốt hơn ảnh gốc (vì cân bằng ở ảnh cũ đã tốt sẵn)	Đôi khi tạo ra bức ảnh không tốt hơn ảnh gốc (vì cân bằng ở ảnh cũ đã tốt sẵn)

VI. Tuần 6

1/ Trong đồ họa máy tính, ảnh lưu dưới dạng vector. Nếu ảnh lưu thành các tọa độ nguyên, thì khi áp dụng phép biến đổi hình học lên tọa độ nguyên như vậy thì gặp khó khăn gì?

Có rất nhiều khó khăn như: xoay ảnh, tăng kích thước ảnh phóng to thu nhỏ,...Cụ thể, khi ta muốn scale ảnh lớn hơn, nếu ảnh lưu thành các tọa độ nguyên(dạng ô lưới), mapping từng điểm của ảnh nguồn sang từng điểm của ảnh đích, sẽ xuất hiện những lỗ trống, làm ảnh đích bị nhòe, mờ,...Nguyên nhân xuất hiện lỗ trống là vì: cơ chế mapping từng điểm trên ảnh là 1:1, ảnh được scale sẽ có kích thước lớn hơn ảnh nguồn, do vậy, ta sẽ không đủ ô màu pixel ở ảnh nguồn để mapping cho ảnh đích.

2/ Vì sao khi biến đổi vị trí điểm ảnh, ta phải thực hiện phép biến đổi ngược? Gồm mấy công đoạn biến ?

Thông thường, thì biến đổi thuận, một điểm ở ảnh đích sẽ mapping với 1 điểm ở ảnh gốc, có thể dư số pixel, không phủ kín.

Tọa độ ảnh nguồn với ảnh đích đều là tọa độ nguyên. Chẳng hạn, scale ảnh lớn hơn 2 lần, biến đổi theo chiều thuận sẽ xuất hiện những lỗ trống. Cách khắc phục, ta biến đổi theo chiều ngược lại(ảnh đích --> ảnh nguồn). Một điểm (x,y) trong ảnh nguồn tương ứng với một điểm (x',y') trong ảnh đích, hay $(x,y) = T^{-1}(x', y')$. Những điểm kết quả mapping từ ảnh đích sang ảnh nguồn, không rơi vào tọa độ nguyên, thì ta sẽ thực hiện phép nội suy giá trị màu.

Gồm có 2 công đoạn đó là xử lý theo vị trí và nội suy giá trị màu

3/ Phép biến đổi hình học lên đồ họa máy tính, thì có 2 phép biến đổi: Phép biến đổi vị trí điểm ảnh và phép nội suy giá trị màu của điểm ảnh, cho ví dụ. Ưu điểm và nhược điểm của 2 phép nội suy giá trị màu: Người láng giềng gần và trung bình cộng.

Trung bình cộng:

- Ưu: có tính trọng số của 4 điểm biên cạnh kề, nhanh, dễ.
- Nhược: điểm ảnh được mapping từ ảnh đích sang ảnh nguồn không rơi vào trọng tâm của ô pixel, mà sẽ lệch về một bên nào đó. Tính trung bình trong trường hợp này không hợp lý lắm.

Láng giềng gần:

- Ưu: tính được giá trị cho tọa độ điểm ở ảnh đích tốt hơn. Do có đánh trọng số của điểm biên cạnh gần với điểm được mapping từ ảnh đích sang ảnh nguồn.
- Nhược: tính toán khó khăn.

4/ Tìm hiểu phép nội suy giá trị màu theo phương pháp BICUBIC

Nội suy xoắn bậc ba là phương pháp tính toán giá trị của điểm ảnh mới dựa trên giá trị trung bình của 16 điểm ảnh gốc gần nhất (4 x 4).

Hàm toán học của phương pháp nội suy xoắn bậc ba :

$$u(x) = \begin{cases} \frac{3}{2}|x|^3 - \frac{5}{2}|x|^2 + 1 & \text{khi } 0 \leq |x| < 1 \\ -\frac{1}{2}|x|^3 + \frac{5}{2}|x|^2 - 4|x| + 2 & \text{khi } 1 \leq |x| < 2 \\ 0 & \text{khi } 2 \leq |x| \end{cases}$$

Trong đó, x là khoảng cách giữa điểm nội suy và điểm lưới.

Để tính toán điểm nội suy P' từ 8 điểm ban đầu P(1,1), P(1,2),...P(4,4). Theo chiều ngang của lưới ta:

Tính điểm P'(1) từ 4 điểm P(1,1), P(1,2), P(1,3), P(1,4)

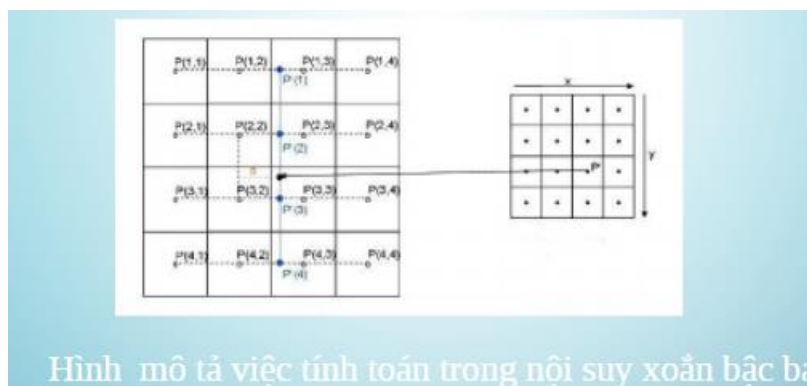
Tính điểm P'(2) từ 4 điểm P(2,1), P(2,2), P(2,3), P(2,4)

Tính điểm P'(3) từ 4 điểm P(3,1), P(3,2), P(3,3), P(3,4)

Tính điểm P'(4) từ 4 điểm P(4,1), P(4,2), P(4,3), P(4,4)

Theo chiều dọc của lưới, điểm P' được tính nội suy từ 4 điểm P'(1), P'(2), P'(3), P'(4)[15].

Việc tính toán được mô tả như hình dưới đây:



Phương pháp nội suy xoắn bậc ba tốn nhiều thời gian và bộ nhớ trong xử lý, được sử dụng trong các trường hợp không cần xem xét đến vấn đề thời gian. Phương pháp nội suy này thường được sử dụng phổ biến trong các phần mềm xử lý ảnh thông dụng như Photoshop, After Effects,...

5/ Viết giải thuật quay ảnh quanh tâm với 1 góc theta

điểm C với góc tọa độ rồi quay góc θ rồi tính trên trục (x, y)

$$T(\vec{CO}) = \begin{bmatrix} 1 & 0 & -x_c \\ 0 & 1 & -y_c \\ 0 & 0 & 1 \end{bmatrix}$$

$$L = \begin{bmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$T(\vec{CO}), R(\theta), T(\vec{CO}) = \begin{bmatrix} \cos \theta & -\sin \theta & -x_c \cos \theta + y_c \sin \theta + x_c \\ \sin \theta & \cos \theta & -x_c \sin \theta - y_c \cos \theta + y_c \\ 0 & 0 & 1 \end{bmatrix}$$

$$S(C, s_x, s_y) = T(\vec{x}) \cdot S(C, s_x, s_y) \cdot T(\vec{CO})$$

6/ Vì sao các median khử nhiễu tốt?

Để khử nhiễu --> lấy phần tử trung dung.

Trong lân cận O, sắp xếp tăng dần của độ sáng của các pixel trong lân cận đó. $I_1 < I_2 \dots \text{med}(I_i) = I_{v+1}$ VD: lân cận 3x3: $I_1 < I_2 < I_3 < I_4 < \dots < I_9 \rightarrow$ Lấy I_5

Khi ở thể trung dung như thế này, nếu dữ liệu ở 2 cực thay đổi ntn, thì phần tử ở giữa ko thay đổi. VD: trong lân cận đang xét, những phần tử bất ảo cũng tồn tại, tránh được việc lấy nhầm điểm bất ảo vào mô hình.

Tổng độ dị biệt của giá trị median với giá trị độ xám của những pixel còn lại trong lân cận là bé nhất so với những vị trí pixel khác. Vị trí median, các điểm lân cận đem lại tổng độ dị biệt đó là bé nhất.

7/ Tìm phép làm trơn ảnh mà bảo toàn được biên cạnh

Các phương pháp làm nhiễu hiện có sử dụng giá trị gốc của hình ảnh và tối thiểu hóa hàm năng lượng E để tính toán ảnh đã khử nhiễu \hat{x} . Đầu tiên, ta nhận được một hàm E từ một hình ảnh nhiễu y, và sau đó một số thấp tương ứng với hình ảnh không bị nhiễu thông qua quy trình ánh xạ. sẽ được tính toán. Sau đó, ta có thể xác định một ảnh đã được khử nhiễu \hat{x} bằng cách tối thiểu hóa hàm E:

$$\hat{x} \in \arg \min_x E(x)$$

Từ quan điểm Bayes, ước tính xác suất MAP của x là

$$\hat{x} = \arg \max_x P(x|y) = \arg \max_x \frac{P(y|x)P(x)}{P(y)}$$

Và có thể được xây dựng tương đương như:

$$\hat{x} = \arg \max_x (\log P(y|x) + \log P(x))$$

trong đó số hạng đầu tiên $P(y|x)$ là một hàm likelihood của x, và số hạng thứ hai $P(x)$ đại diện cho hình ảnh trước đó. Trong trường hợp của AWGN, hàm mục tiêu thường có thể được xây dựng dưới dạng như sau:

$$\hat{x} = \arg \min_x \frac{1}{2} \|y - x\|_2^2 + \lambda R(x)$$

Trong đó $\|y - x\|_2^2$ là thuật ngữ biểu thị sự khác biệt giữa hình ảnh gốc và hình ảnh bị nhiễu. $R(x) = -\log P(x)$ biểu thị thuật ngữ chính quy và λ là tham số chính quy. Đối với các phương pháp làm giảm biến phân, điểm quan trọng để giải quyết là tìm một hình ảnh phù hợp trước ($R(x)$). Các mô hình thành công trước đây bao gồm priors

8/ Làm trơn ảnh bằng toán tử trung bình thì ưu khuyết điểm là gì?

Ưu điểm: Tính toán nhanh.

Khuyết điểm

- Mờ biên cạnh.
- Khử nhiễu không hiệu quả khi các giá trị xung quanh đang xét là giá trị nhiễu
- Kết quả ảnh trả về có thể bị độ nhiễu cao (bị hột)

9/ Ứng dụng công thức toán để làm trơn ảnh

7) Làm trơn ảnh = toán tử trung bình
khuyết điểm là gì?
Toán tử gaussian

$$h = \frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

$$g(x, y) = \sum_i \sum_j f(x-i, y-j) \cdot h(i, j), (i, j) \in \mathbb{Z}$$

$$h(i, j) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{i^2 + j^2}{2\sigma^2}}$$

8) ứng dụng công thức để làm trơn ảnh

VII. Tuần 7

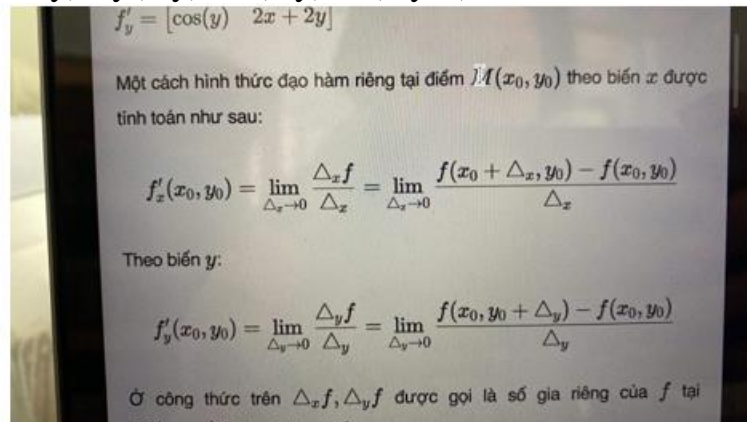
1/Khi lọc nhiễu bằng phương pháp làm trơn ảnh với kỹ thuật “lọc trung bình”, thì ta có ưu điểm, khuyết điểm gì?

Ưu	Nhược
<p>Phương pháp Averaging lọc tốt đối với những ảnh bị nhiễu random như nhiễu Gaussian hoặc nhiễu Uniform.</p> <p>Tính toán dễ</p>	<p>Khi khử nhiễu, luôn phải cần sự tham gia trợ giúp của tất cả pixel trong khu vực đang duyệt.</p> <p>Quá trình làm mờ nhiễu cũng đồng thời làm mờ ảnh vì bộ lọc trung bình có xu hướng cân bằng cường độ sáng các điểm ảnh.</p> <p>Khi ảnh có quá nhiều thành phần bất ảo, thì phương pháp này không có tác dụng</p>

2/ Làm sao để lấy đạo hàm, đạo hàm chỉ trong miền liên tục. Trong miền rời rạc, thì tính đạo hàm như thế nào? Chuyển công thức trong miền liên tục thành rời rạc.

Ta biết rằng, (x, y) là tọa độ nguyên là tọa độ của điểm ảnh khi nói đến f(x,y). f(x, y) là giá trị độ sáng cũng lấy giá trị nguyên 0 à 255

$$f_x(x,y) = f(x,y) - f(x+1,y), f_y(x,y) = f(x,y) - f(x,y-1)$$



Cho delta = 1 (hoặc h = 1):

$$\bullet \quad \partial f / \partial x(x, y) = \lim_{h \rightarrow 0} ([f(x + h) - f(x, y)] / h)$$

3/ Làm sao khắc phục xấp xỉ differencing

Lấy cặp điểm 2 bên để định vị tâm.

4/ Robert khác differencing như thế nào?

Lấy cạnh đường chéo

5/ Trình bày nguyên lý biên cạnh dựa vào đạo hàm bậc nhất và đạo hàm bậc 2 của đồ thị ảnh?

Đạo hàm bậc nhất lấy điểm theo ngưỡng cho phép dùng cho độ chênh lệch cao

Đạo hàm bậc 2 để biết điểm biến thiên xác định cực đại cực tiểu của đồ thị -> dễ định mức hơn

6/ Vì sao cường độ gradient bậc 2 nó bé?

Vì khi cường độ gradient bé, rất khó để xác định ngưỡng biên cạnh. Khi đó, đạo hàm bậc 2 sẽ tạo ra 1 điểm uốn phân rõ ngưỡng âm và ngưỡng dương

7/ Nếu lấy trung bình cộng trọng số thì khác phức nhiễu như thế nào?

Vì khi làm vậy là ta đang làm trơn ảnh, trung bình cộng giảm bớt nhiễu, nếu lấy 1 lần thì dính ngay vào chỗ nhiễu nên là nó sẽ làm cùng lúc 2 công việc là phát hiện biên cạnh với khử nhiễu cùng lúc luôn

8/ Khi $k = 1$, $k = 2$, $k = 2^{1/2}$ thì f_x , f_y là gì?

- Với $k = 1$:

$$G_x = \frac{1}{3} \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix}$$

$$G_y = \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}$$

- Với $k = 2$:

$$G_x = \frac{1}{4} \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$

$$G_y = \frac{1}{4} \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

- Với $k = 2^{\frac{1}{2}}$:

$$G_x = \frac{1}{2\sqrt{2}} \begin{bmatrix} -1 & 0 & 1 \\ -\sqrt{2} & 0 & \sqrt{2} \\ -1 & 0 & 1 \end{bmatrix}$$

$$G_y = \frac{1}{2\sqrt{2}} \begin{bmatrix} 1 & \sqrt{2} & 1 \\ 0 & 0 & 0 \\ -1 & -\sqrt{2} & -1 \end{bmatrix}$$

9/ Nhận xét gì về cách thức của 3 phương pháp

- Với cách tính của Prewitt (k=1): sử dụng các lân cận với các trọng số, mọi điểm đều như nhau để tính gradient tại một điểm. Chính vì thế, phương pháp tính này rất nhạy với nhiễu. Ngược lại, phương pháp Robert chỉ dùng 1 cặp điểm
- Với cách tính của Sobel (k=2): sử dụng các lân cận với tổng số khác nhau (các điểm liền kề ở 4 hướng có trọng số lớn hơn). Phương pháp này rất nhạy với các đường biên chéo, Đông Tây Nam Bắc nhưng rất hiệu quả trong hầu hết các trường hợp.
- Với cách tính của Frei-chen (k=sqrt(2)): cải tiến từ phương pháp của Sobel và sử dụng hệ số k tốt hơn. Với hệ số k mới này, việc phát hiện biên cạnh ở các đường chéo tốt hơn. Chính vì vậy, phương pháp Frei-chen dễ dàng phát hiện ra được các biên nhỏ và tạo ra được các đường biên nhỏ.

Đồng đều với biến thiên độ xám (nhưng lâu)

10/ Khi phát hiện biên cạnh bằng gradient, thì phản ứng phụ gì?

- Tại những điểm nhiễu cũng có biến thiên lớn về độ sáng, biên cạnh của vậy, thì khi lấy ra những điểm biên cạnh, ta sẽ vô tình lấy ra các điểm nhiễu.

11/ Vì sao Prewitt lại sử dụng nhiều cặp?

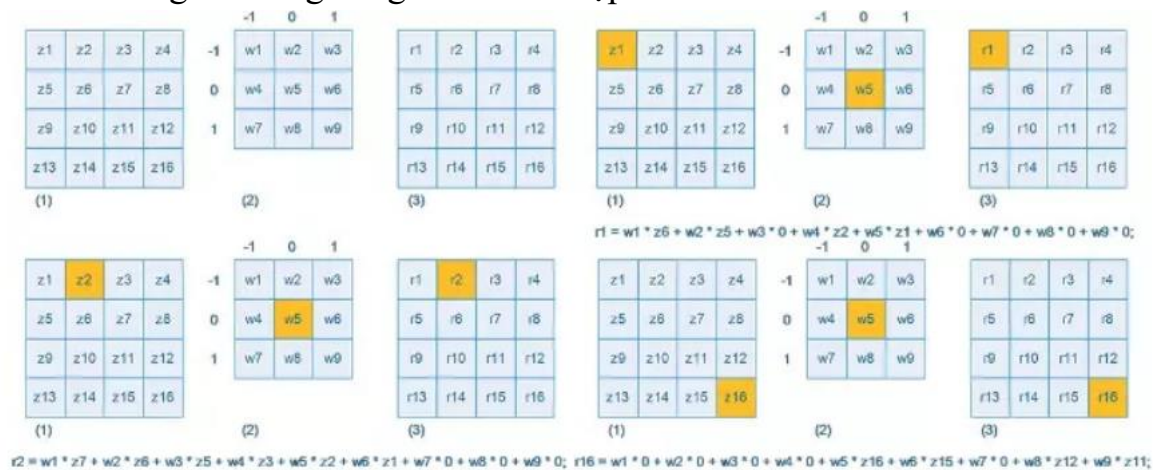
Phương pháp Prewitt khi xấp xỉ đạo hàm, thì lấy lân cận nhiều hơn. Đạo hàm theo phương ngang lấy 3 cặp để tính (A0-A2, A7-A3, A6-A5), chiều dọc (A0-A6),.. lấy 3 cặp để xấp xỉ.

12/ Trong toán tử gradient, tại sao lại sử dụng phép tích chập?

Tích chập là 1 phép toán thực hiện đối với 2 hàm số f và g, kết quả cho ra 1 hàm số thứ 3. Phép tích chập khác với tương quan chéo ở chỗ nó cần lật kernel theo chiều ngang và dọc trước khi tính tổng của tích. Sử dụng phép tích chập trong toán tử gradient để ra được ảnh mới với bộ lọc của gradient, là một công cụ vô cùng đặc lực trong xử lý ảnh theo vùng (area processes)

13/ Tính kết quả của phép tích chập

Phép tích chập được hình dung thực hiện bằng việc dịch chuyển ma trận kernel lần lượt qua tất cả các điểm ảnh trong ảnh, bắt đầu từ góc bên trái trên của ảnh. Và đặt anchor point tương ứng tại điểm ảnh đang xét. Ở mỗi lần dịch chuyển, thực hiện tính toán kết quả mới cho điểm ảnh đang xét bằng công thức tích chập



Cứ làm lần lượt theo thứ tự từ trái sang phải từ trên xuống dưới cho đến khi kết thúc. Chú ý: Có 1 vấn đề là tại 1 số điểm của ma trận kernel sẽ nằm bên ngoài ma trận ảnh khi ma trận này quét qua, thì tại những điểm đó chúng ta sẽ lấy giá trị của nó là 0

14/ Viết giải thuật tính f_x

Code chỉ thực hiện với ảnh xám

```
void Convolution::doConvolution(Mat& sourceImage, Mat& destinationImage)
{
```

```
    int nr = sourceImage.rows;
    int nc = sourceImage.cols;
```

```
    // Tạo matrix để lưu giá trị pixel sau khi thực hiện tích chập.
    destinationImage.create(Size(nc, nr), CV_8UC1);
```

```
    // Đi lần lượt từng pixel của ảnh nguồn.
```

```
    for (int i = 0; i < nr; i++)
    {
```

```
        // Lấy địa chỉ dòng của ảnh đích, để lưu kết quả vào.
        uchar* data = destinationImage.ptr<uchar>(i);
```

```

for (int j = 0; j < nc; j++)
{
    // Lưu tổng giá trị độ xám của vùng ảnh tương ứng với kernel
    int g_val = 0;

    // Duyệt mask, giá trị pixel đích là tổ hợp tuyến tính của mask với ảnh
    nguồn.
    for (int ii = 0; ii < _kernel.size(); ii++)
    {
        // _kernelIndex: mảng chỉ số truy cập nhanh
        int index_r = i - _kernelIndex[ii].x;

        // Với pixel nằm ngoài biên, bỏ qua.
        if (index_r < 0 || index_r > nr - 1)
            continue;

        int index_c = j - _kernelIndex[ii].y;
        if (index_c < 0 || index_c > nc - 1)
            continue;

        g_val += _kernel[ii] * sourceImage.at<uchar>(index_r,
index_c);
    }

    // Gán giá trị cho matrix đích.
    data[j] = g_val;
}
}
}

```

15/ Dùng 1 bộ lọc Laplace nữa để làm gì?

Để toán tử biên ảnh đem những điểm nhiễu ra

16/ Ghi bộ lọc Gauss chập với Laplace

0	-1	0
-1	4	-1
0	-1	0

-1	-1	-1
-1	8	-1
-1	-1	-1

17/ f chập với 1 bộ lọc nào ấy thì sẽ như thế nào?

Thực hiện tính toán kết quả mới với dữ liệu trong bộ lọc cho điểm ảnh đang xét bằng công thức tích chập

18/ Còn khi lấy toán tử laplace xong thì được gì? ảnh đầu, đạo hàm bậc 2 Laplace biên cạnh đâu? Viết thêm hàm để xác định biên cạnh sau khi dùng toán tử Laplace.

Laplace xong thì ta chưa có điểm biên cạnh. Biên cạnh nằm ở chỗ chuyển tiếp từ âm sang dương, và từ dương sang âm.

Lấy gradient xong thì có ảnh biên độ của vector gradient, dùng threshold để phân phưởng à cho BEI(ảnh nhị phân của biên cạnh).

19/ Chứng minh vì sao laplace bất biến với phép quay?

Phép biến đổi Laplace là bất biến đối với phép quay. Còn toán tử gradient thì không vì hình quay à trường vector tiếp tuyến biến đổi.

Tính chất tuyến tính

Cho các hàm gốc f_k với các chỉ số tăng là α_k , biến đổi Laplace là \bar{f}_k , $k = 1, 2, \dots, n$. Khi đó biến đổi Laplace của hàm tổ hợp tuyến tính f của các hàm f_k

$$f(t) = \sum_{k=1}^n c_k f_k(t), \text{ với } c_k \text{ là hằng số}$$

là hàm \bar{f} được xác định bởi

$$\bar{f}(s) = \sum_{k=1}^n c_k \bar{f}_k(s), \quad (1.3.1)$$

với miền xác định $\text{Re } s > \max \alpha_k$.

Chứng minh. Suy ra từ định nghĩa và tính chất tuyến tính của tích phân.

Ví dụ 1.3.1

Từ kết quả của ví dụ 1.1.2 và tính chất tuyến tính ta có biến đổi Laplace của các hàm sau

a) Ta có

$$\begin{aligned} \mathcal{L}\{\sin \alpha t\} &= \mathcal{L}\left\{\frac{1}{2i}(e^{i\alpha t} - e^{-i\alpha t})\right\} \\ &= \frac{1}{2i}\left[\frac{1}{s - i\alpha} - \frac{1}{s + i\alpha}\right] \\ &= \frac{\alpha}{s^2 + \alpha^2}, \quad \text{Re } s > |\text{Im } \alpha| \end{aligned}$$

b) Tương tự ta có

$$\mathcal{L}\{\cos \alpha t\} = \mathcal{L}\left\{\frac{1}{2}(e^{i\alpha t} + e^{-i\alpha t})\right\} = \frac{s}{s^2 + \alpha^2}, \quad \text{Re } s > |\text{Im } \alpha|$$

$$\text{c) } \mathcal{L}\{\cosh \alpha t\} = \mathcal{L}\left\{\frac{1}{2}(e^{\alpha t} + e^{-\alpha t})\right\} = \frac{s}{s^2 - \alpha^2}, \quad \text{Re } s > |\text{Re } \alpha|$$

$$\text{d) } \mathcal{L}\{\sinh \alpha t\} = \mathcal{L}\left\{\frac{1}{2}(e^{\alpha t} - e^{-\alpha t})\right\} = \frac{\alpha}{s^2 - \alpha^2}, \quad \text{Re } s > |\text{Re } \alpha|.$$

Tính chất đồng dạng

Nếu $\mathcal{L}\{f(t)\} = \bar{f}(s)$, f có chỉ số tăng là α_0 thì

$$\mathcal{L}\{e^{at} f(t)\} = \bar{f}(s-a), \quad \text{Re } s > \alpha_0 + \text{Re } a \quad (1.3.3)$$

Chứng minh

Theo định nghĩa ta có

$$\mathcal{L}\{e^{at} f(t)\} = \int_0^{\infty} e^{-(s-a)t} f(t) dt = \bar{f}(s-a).$$

Cho $\mathcal{L}\{f(t)\} = \bar{f}(s)$, f là hàm gốc có chỉ số tăng α_0 và $c > 0$ là hằng số. Khi đó

Tính chất dịch chuyển ảnh

$$\mathcal{L}\{f(ct)\} = \frac{1}{c} \bar{f}\left(\frac{s}{c}\right), \quad \text{Re } s > c\alpha_0 \quad (1.3.2)$$

Chứng minh

$$\mathcal{L}\{f(ct)\} = \int_0^{\infty} e^{-st} f(ct) dt = \frac{1}{c} \int_0^{\infty} e^{-su/c} f(u) du = \frac{1}{c} \bar{f}\left(\frac{s}{c}\right).$$

20/ Lấy Laplace chập với Gauss để làm gì?

Khi lấy f chập với gauss--> làm trơn ảnh, khử nhiễu, vô hình chung ảnh sẽ không rõ nét, biến thiên độ sáng ko mạnh như ảnh cũ, không rõ nét --> cứu cánh để phát hiện biên là dùng laplace của f chập gauss = laplace của gauss chập (có được bộ lọc) với f

21/ Vì sao phương pháp Laplace of Gaussian phát hiện được biên cạnh. Và có ưu điểm gì hơn so với các phương pháp trước.

Khi làm trơn ảnh bằng trung bình có trọng số gauss. Phép trung bình có trọng số f(x,y) với hàm lọc gauss G(x,y, α) -> Bước 1, làm trơn ảnh, tức là khử nhiễu. Nhưng khử nhiễu, thì vô tình điểm biên cạnh cũng bị ảnh hưởng theo, biến thiên độ xám sẽ nhỏ hơn ảnh cũ. Lúc này, ta dùng toán tử Laplace thì sẽ lấy được biên cạnh rất tốt.

22/ Xây dựng bộ lọc của Laplace Gauss. Lưu ý đã dùng bộ lọc Gauss, thì cho lân cận lớn hơn 3x3.(tham khảo sách)

Xây dựng bộ lọc của laplace của gauss. Lưu ý chọn lân cận lớn đừng chọn 3x3 thì gauss không có gì để xét. Tìm kiếm những giải pháp làm trơn ảnh, mà giảm đi tác dụng phụ, mất mát biên cạnh

23/ Khi làm trơn ảnh, vô tình làm cho biên cạnh bị mờ đi. Tìm kiếm giải pháp làm trơn ảnh để cải thiện việc mất mát thông tin về biên cạnh này.(tham khảo sách) và tìm kiếm giải pháp làm trơn ảnh có thể chống chọi được với nhiễu.

Phương pháp canny: tham khảo thêm cuốn golalex--> phương pháp canny, phát hiện biên cạnh. Rất là tinh tế ở chỗ, vừa xét yếu tố toàn cục và local, nó cố gắng liên kết những pixel của biên cạnh có cường độ yếu vào chuỗi những pixel có cường độ mạnh hơn. Chỗ mờ vẫn liên lạc với những chỗ rõ trước đó. Và khi phát hiện biên cạnh, kết quả là bằng phương pháp canny: biên của nó rất mảnh mai, liên tục, không bị rời rạc, đứt nét.

24/ Lập bảng thông kê so sánh các phương pháp phát hiện biên cạnh: Gradient, Laplace, Laplace of Gauss, Canny.

Tên	Nguyên lý	Phương pháp	Ưu điểm	Nhược điểm	Chống chọi với nhiễu
Gradient	Dùng toán tử gradient để phát hiện sự biến thiên độ sáng	Tích chập với bộ lọc khác nhau: Các bộ lọc được lấy từ các phương pháp rời rạc hóa gradient các toán tử gradient.	Tính toán đơn giản	Nhạy cảm với nhiễu, khó xác định hướng của cạnh, độ phức tạp tính toán cao	Cùng 1 nguyên lý có nhiều phương pháp. TB có trọng số thì có chống chọi với nhiễu. Gradient lấy từng cặp 1, differencing thì ko chống chọi với nhiễu
Laplace		<p>Xác định biên dựa vào giá trị 0 của đạo hàm bậc hai của ảnh. Laplacian của một ảnh tại điểm I(x,y) được tính theo (10):</p> $L(x,y) = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2} \quad (10)$ <p>Laplacian được kết hợp với bộ lọc làm mịn ảnh để tìm biên [5]. Xét công thức sau:</p> $h(r) = -e^{-\frac{r^2}{2\sigma^2}} \quad (11)$ <p>Ở đây $r^2 = x^2 + y^2$ và σ là độ lệch chuẩn (standard deviation). Nếu thực hiện phép tích chập của hàm này với ảnh cần tìm biên, kết quả là ảnh sẽ bị mờ đi, mức độ mờ phụ thuộc vào giá trị của σ. Laplacian của h tức đạo hàm bậc hai của h theo r là:</p> $\nabla^2 h(r) = -\left[\frac{r^2 - \sigma^2}{\sigma^4}\right] e^{-\frac{r^2}{2\sigma^2}} \quad (12)$	khắc phục trong việc cho kết quả là ảnh nhận được có cấu trúc không rõ nét do tạo nên những đường biên dày, không sắc nét	Ảnh hưởng bởi nhiễu, độ phức tạp tính toán cao.	Laplacian có thể nhạy cảm với nhiễu trong ảnh, điều này có thể khiến chúng tạo ra các cạnh giả hoặc bỏ sót các cạnh thực.

Laplace of Gaussian		Để nhận được các đường biên mỏng và rõ nét, ta phải tiến hành các bước xử lý tiếp theo như loại bỏ những điểm không phải là cực trị (nonmaximum suppression) đồng thời áp dụng kỹ thuật liên kết biên (edge linking)	thích hợp cho các vùng ảnh độ tương phản thay đổi có tính nhảy bậc	cho kết quả là ảnh nhận được có cấu trúc không rõ nét do tạo nên những đường biên dày, không sắc nét	Thành phần làm mịn Gaussian của LoG có thể giúp giảm độ nhạy với nhiễu bằng cách làm mịn các biến thể nhỏ về cường độ. Điều này có thể cải thiện độ chính xác của phát hiện cạnh, đặc biệt là trong các hình ảnh nhiễu.
Canny		sử dụng hai mức ngưỡng cao và thấp. Ban đầu ta dùng mức ngưỡng cao để tìm điểm bắt đầu của biên, sau đó chúng ta xác định hướng phát triển của biên dựa vào các điểm ảnh liên tiếp có giá trị lớn hơn mức ngưỡng thấp. Ta chỉ loại bỏ các điểm có giá trị nhỏ hơn mức ngưỡng thấp. Các đường biên yếu sẽ được chọn nếu chúng được liên kết với các đường biên khỏe	Cực đại hóa tỷ số tín hiệu trên nhiễu làm cho việc phát hiện các biên thực càng chính xác. • Đạt được độ chính xác cao của đường biên thực. • Làm giảm đến mức tối thiểu số các điểm nằm trên đường biên nhằm tạo ra các đường biên mỏng, rõ	nếu chọn ngưỡng quá thấp sẽ tạo ra biên không đúng, ngược lại nếu chọn ngưỡng quá cao thì nhiều thông tin quan trọng của biên sẽ bị loại bỏ	Nếu mức ngưỡng càng thấp, số đường biên được phát hiện càng nhiều (nhưng kèm theo là nhiễu và số các đường biên giả cũng xuất hiện càng nhiều). Ngược lại nếu ta đặt mức ngưỡng càng cao, ta có thể bị mất những đường biên mờ hoặc các đường biên sẽ bị đứt đoạn

26/ Nguyên lý để phát hiện được biên cạnh là gì?

Ta sẽ tận dụng khái niệm về đạo hàm sau đó áp dụng gradient -> điểm tại gradient là điểm vượt ngưỡng cho phép

VIII. Tuần 8

1/ Tính độ phức tạp của phép biến đổi Fourier?

Chuỗi Fourier là $f(x,y)$ là hàm ảnh cho trước. Nếu hàm là tuần hoàn, thì chuỗi Fourier sẽ dùng định lý chuỗi Fourier thì tạo ra phép biến đổi Fourier.

Hàm tuần hoàn f là 1 hàm tuần hoàn nào và chỉ nếu có số p sao cho $f(t + 2p) = f(t)$ với mọi t thuộc \mathbb{R} . M.X.D của $f(t)$ là $2p$ hoặc bội của p .

Chuỗi Fourier của hàm tuần hoàn: $f(t) = \frac{1}{2\pi} a_0 + \sum_{n=1}^{\infty} (a_n \cos nt + b_n \sin nt)$

BT: $f(t) = \begin{cases} 0 & t < 0 \\ e^{-at} & t \geq 0 \end{cases} \quad a > 0$, Tính $F(\omega)$

$$F(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt = \frac{1}{2\pi} \int_0^{\infty} e^{-at} e^{-i\omega t} dt = \frac{1}{2\pi} \int_0^{\infty} e^{-(a+i\omega)t} dt = \frac{1}{2\pi} \left[\frac{e^{-(a+i\omega)t}}{-(a+i\omega)} \right]_0^{\infty} = \frac{1}{2\pi} \frac{1}{a+i\omega}$$

$a > 0 \quad \lim_{t \rightarrow \infty} |e^{-(a+i\omega)t}| = \lim_{t \rightarrow \infty} |e^{-at} e^{-i\omega t}| = \lim_{t \rightarrow \infty} e^{-at} = 0$

$F(\omega) = \frac{1}{2\pi(a+i\omega)}$

$(f * h)(x,y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(u,v) h(x-u, y-v) du dv$

Để phân tích tính toán nên có chuỗi Fourier nhanh.

2/ Ẩn số trong hình là cái gì?

Ẩn số trong hình là các $H(u,v)$

3/ Cho biết ý kiến định nghĩa $H(u,v)$ tại sao nó bé và có khuyết điểm gì?

$H(u,v)$ không thể bằng 0, vì sẽ gây tiệm cận nên nó phải lên xuống từ từ \rightarrow có độ dốc, chuyển tiếp đẹp hơn

4/ Biện luận theo D_0 , càng lớn thì như thế nào, càng bé thì như thế nào và khác gì so với trước?

Càng lớn thì ảnh càng nhòe, còn (u,v) chỉ lớn một xíu sẽ mất biên cạnh, lúc trước thì lên xuống nhanh không có sự chuyển tiếp

D₀ là tần số cắt. D₀ càng lớn thì sẽ khử nhiễu kém chỉ khử về 0 với những tần số cao hơn nhưng những chỗ không bị nhiễu sẽ được bảo toàn.

D₀ càng bé thì khử nhiễu tốt nhưng những chỗ không bị nhiễu sẽ bị khử theo. Nên ảnh sẽ mờ.

5/ Vì sao ảnh trên nhiễu nhưng vẫn có thể làm nét lại trong miền tần số mà không phải trong miền không gian? 2 bộ lọc nhiễu trong miền không gian và miền tần số khác nhau gì về cơ bản?

- Trong miền không gian: $g = f * h$, nên độ phức tạp tính toán là (n^2, m^2)

$$\sum \sum f(x-i, y-j) h(i, j)$$

- Trong miền tần số:

- $fourier\{f\} = f(u, v) \rightarrow \frac{O(n^2 * n^2)}{\text{Fast fourier transform: } O(n^2 * \log n)}$

- $fourier\{h\} = h(u, v) \rightarrow O(n^2)$

- $f(u, v) * h(u, v) \rightarrow O(n^2)$

- $fourier_{-1}\{F.H\} \rightarrow \frac{O(n^2, n^2)}{\text{Fast fourier transform: } O(n^2 * \log n)}$

⇒ Trong trường hợp m, n nhỏ thì miền không gian nhanh hơn vì độ phức tạp ít hơn, nhưng khi m, n lớn thì tần số có độ phức tạp nhỏ hơn.

6/ Dùng 3 bộ lọc này để chống lại sự tấn công của nhiễu ảnh, cái nào tốt, viết hàm chống nhiễu ảnh, lập bảng so sánh?

Bộ lọc mờ trung bình (Box Filter)

- Bộ lọc làm mờ trung bình được thiết kế bằng cách thiết lập mỗi giá trị trên bộ lọc bằng: $1/(W*H)$.
- Tức là nếu Width và Height của bộ lọc bằng 3. Thì giá trị trên cửa sổ convolve sẽ là: $1/9$.

box_blur.py

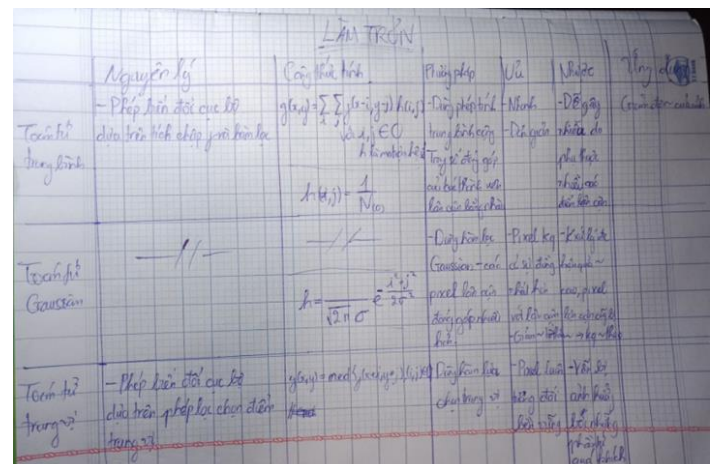
```
import os
import cv2

INPUT = 'girl_luv.jpg'
KERNEL_WIDTH = 3
KERNEL_HEIGHT = 3

if not os.path.isfile(INPUT):
    raise Exception('File not found @ %s' % INPUT)

img = cv2.imread(INPUT)

blur_img = cv2.blur(img, ksize=(KERNEL_WIDTH, KERNEL_HEIGHT)) # or cv2.boxFilter
cv2.imwrite('box_blur_%s_%d.jpg' % (os.path.splitext(os.path.basename(INPUT))[0], KERNEL_WIDTH, KERNEL_HEIGHT),
    blur_img)
```



Bộ lọc làm mờ Gaussian (Gaussian Filter)

- Công thức Gaussian:
https://docs.opencv.org/2.4/doc/tutorials/imgproc/ gaussian_median_blur_bilateral_filter/gaussian_median_blur_bilateral_filter.html# gaussian-filter
- Ngoài kích thước cửa sổ, bộ lọc Gaussian còn cho phép thiết lập thêm 2 tham số là: Sigma X và Sigma Y, tức độ lệch chuẩn theo trục x và trục y tương ứng.

gaussian_blur.py

```
import os
import cv2

INPUT = 'girl_luv.jpg'
KERNEL_WIDTH = 9
KERNEL_HEIGHT = 9
SIGMA_X = 4
SIGMA_Y = 4

if not os.path.isfile(INPUT):
    raise Exception('File not found @ %s' % INPUT)

img = cv2.imread(INPUT)

blur_img = cv2.GaussianBlur(img, ksize=(KERNEL_WIDTH, KERNEL_HEIGHT), sigmaX=SIGMA_X, sigmaY=SIGMA_Y)

cv2.imwrite('gaussian_blur_%s_%d_%d_%d.jpg' % (os.path.splitext(os.path.basename(INPUT))[0], KERNEL_WIDTH,
KERNEL_HEIGHT, SIGMA_X, SIGMA_Y), blur_img)
```

median_blur.py

```
import os
import cv2

INPUT = 'girl_luv.jpg'
KERNEL_SIZE = 9

if not os.path.isfile(INPUT):
    raise Exception('File not found @ %s' % INPUT)

img = cv2.imread(INPUT)

blur_img = cv2.medianBlur(img, KERNEL_SIZE)

cv2.imwrite('median_blur_%s_%d.jpg' % (os.path.splitext(os.path.basename(INPUT))[0], KERNEL_SIZE), blur_img)
```

7/ D_0 càng lớn thì lấy biên cạnh càng nhiều hay càng ít?

Biên cạnh phải đủ lớn thì mới lấy được các điểm do D_0 điều khiển

8/ Giải thích $H(u,v)$ tại sao lại giúp lọc nhiễu được? Vì sao có thể khử nhiễu được?

$$H(u, v) = \begin{cases} 1 & \text{if } D(u, v) \leq D_0 \\ 0 & \text{if } D(u, v) > D_0 \end{cases}$$

$$D(u, v) = \left[\left(u - \frac{M}{2} \right)^2 + \left(v - \frac{N}{2} \right)^2 \right]^{\frac{1}{2}}$$

D_0 là tần số cắt. D_0 càng lớn thì sẽ khử nhiễu kém chỉ khử về 0 với những tần số cao hơn nhưng những chỗ không bị nhiễu sẽ được bảo toàn.

D_0 càng bé thì khử nhiễu tốt nhưng những chỗ không bị nhiễu sẽ bị khử theo. Nên ảnh sẽ mờ.

9/ Phân tích độ phức tạp tính toán khi sử dụng tích chập trong miền không gian và miền tần số

Nếu trong miền không gian: $O((N * N))$

Miền không gian: $M=N$

$$f(x,y) * h(x,y) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(x-m, y-n) \cdot h(m,n)$$

Độ phức tạp: $O(n^2 * m^2)$

Miền tần số:

Bước 1: $J(f) = F(u,v) : O(n^2 \log_2 n)$ FFT

Bước 2: $J(h) = H(u,v) : O(n^2)$

Bước 3: $F(u,v) \cdot H(u,v) : O(n^2)$

Bước 4: $J^{-1}\{F \cdot H\} : O(n^2 \log_2 n)$

Độ phức tạp: $O(n^2 \log_2 n)$

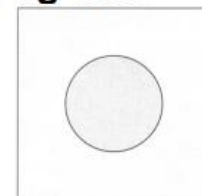
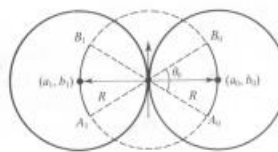
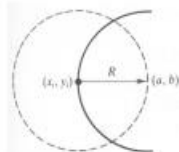
Kết luận: Phương pháp nào nhanh hơn sẽ phụ thuộc vào giá trị m, n .

10/ Cách phát hiện đường tròn

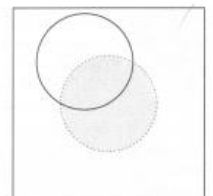
$$(x-a)^2 + (y-b)^2 = R^2$$

$$\begin{cases} x = a + R \cos \theta \\ y = b + R \sin \theta \end{cases} \Rightarrow \begin{cases} a = x - R \cos \theta \\ b = y - R \sin \theta \end{cases} \Rightarrow \begin{cases} x = a + R \cos \theta \\ y = b + R \sin \theta \end{cases}$$

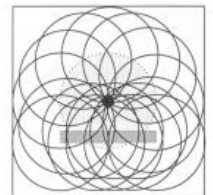
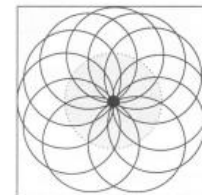
$$\begin{cases} x = a + R \cos \theta \\ y = b + R \sin \theta \end{cases} \Rightarrow \begin{cases} a = x - R \cos \theta \\ b = y - R \sin \theta \end{cases}$$



(a)



(b)



IX. Tuần 9

1/ Tóm lược lại bài toán

Tìm phép biến đổi trực giao, hệ cơ sở mới cho không gian con trong hệ

2/ Biểu thức toán ra như thế nào?

3/ Phương pháp xác định K

Điều kiện dừng của vòng lặp: Until((w1m= w1m-1) & (w2m= w2m-1)..... (wkm= wkm-1))

Điều kiện kết nạp il với Cj*:

```

for l in range (1,n){
    minDistance= distance(i[l],w[i])
    minIndex=1
    for j* in range (1,k){
        if(distance(i[l],w[j*]<minDistance)){
            minDistance=distance(i[l],w[j*]);
            minIndex=j*;
        }
    }
    C[minIndex].add(i[l])

```


X. SEMINAR

1. Nhóm MyBack - Khảo sát phân đoạn ảnh dựa vào superpixel và ứng dụng trong phân đoạn ảnh y khoa

- Phân đoạn ảnh là quá trình chia một bức ảnh số thành nhiều phần khác nhau (các tập điểm ảnh) có sự đồng nhất về thuộc tính chung
- Superpixel được xem là một nhóm nhỏ các pixel với màu sắc đồng nhất và có chung ý nghĩa, màu, vân, độ sâu, thường được áp dụng cho phân đoạn ảnh hay nhận dạng vật thể
- Phân đoạn ảnh dựa trên superpixel là phương pháp chia ảnh thành các nhóm pixel hay còn gọi là superpixel theo nguyên lý chia để trị

a) Động lực nghiên cứu

Về khoa học:

- Giảm độ phức tạp tính toán vì chỉ cần quan tâm đến superpixel
- Mở ra khả năng xử lý ảnh với mức megapixel - ảnh chứa hàng triệu pixel
- Tăng độ chính xác của kết quả

Ứng dụng:

- Phát hiện vật thể
- Chẩn đoán bệnh, xác định khối u trên ảnh chụp X-Quang, ảnh nội soi
- Giúp nhận dạng đối tượng, giám sát giao thông qua hình ảnh vệ tinh và camera an ninh

b) Phương pháp

Phương pháp	Group
SLIC	Gradient-based
SEEDS	Gradient-based
FH (Felzenszwalb Huttenlocher)	Graph-based
CRS (Contour Relaxed Superpixels)	Gradient-based
CS / CIS	Gradient-based

- Phương pháp đề xuất là SLIC và FH

2. Nhóm VIP - Khảo sát Image Stitching và ứng dụng để ghép nối các ảnh dưới các view khác nhau

- Ghép ảnh là quá trình kết hợp nhiều ảnh chụp với các trường xem chồng lên nhau để tạo ra ảnh toàn cảnh được phân đoạn hoặc ảnh có độ phân giải cao.

a) Động lực nghiên cứu

Ứng dụng trong:

- Ảnh toàn cảnh trong bản đồ và hình ảnh vệ tinh với độ phân giải cao
- Hình ảnh dùng trong y tế
- UAV viễn thám
- Thực tế ảo (VR)

b) Phương pháp

Reference:

- Application of Migration Image Registration Algorithm Based on Improved SURF in Remote Sensing Image Mosaic
- Image Mosaic Algorithm Based on PCA-ORB Feature Matching
- SIFT Feature Image Stitching Based on Improved Cuckoo Algorithm
- Study and Performance Evaluation Binary Robust Invariant Scalable Keypoints (BRISK) for Underwater Image Stitching
- Automatic Stitching of Medical Images Using Feature Based Approach
- Phương pháp sử dụng: Image stitching model based on features based techniques (Mô hình ghép ảnh dựa trên kỹ thuật features based)

3. Nhóm 4T - Khảo sát Seamless cloning và ứng dụng để ghép ảnh trông tự nhiên hơn. (Khảo sát giải pháp truyền thống, deep learning)

a) Động lực nghiên cứu

Về khoa học:

- Chép một đối tượng trong ảnh sang một vị trí khác hoặc một ảnh khác.
- Giữ cho phân bố màu chuyển tiếp tự nhiên ở phần đối tượng và biên cạnh.
- Làm cho vùng biên không còn phân cách nữa, và có sự hoà trộn giữa đối tượng sao chép và hình nền.
- Ngoài ra, mục đích của việc này là nghiên cứu về “sự giả mạo” để tìm ra phương pháp nhận biết giả mạo, chống các hành vi xấu (forensics)

Ứng dụng:

- Nhân bản liên mạch
- Che dấu đối tượng
- Đổi kết cấu của đối tượng
- Ghép hình

b) Phương pháp

- Chinh sửa ảnh bằng cách giải phương trình Poisson (dựa vào lý thuyết chỉnh sửa Poisson và sự biến thiên gradient của đối tượng và ảnh nền)

4. Nhóm Bất Ổn - Khảo sát Inpainting và ứng dụng để hồi phục dữ liệu (tài liệu bị ô vàng, rách nát)

- Inpainting là một quá trình bảo tồn trong đó các phần bị hư hỏng, xuống cấp hoặc bị thiếu của các tác phẩm nghệ thuật để tạo ra một hình ảnh hoàn chỉnh.
- Quy trình inpainting có thể được áp dụng cho cả phương diện về nghệ thuật và kỹ thuật số như tranh sơn dầu hoặc acrylic, bản in ảnh hóa học, tác phẩm điêu khắc hoặc hình ảnh và video kỹ thuật số.

a) Động lực nghiên cứu

Về khoa học:

- Xử lý các pixel bị thiếu một cách hiệu quả hơn
- Xử lý các khu vực bị thiếu bất thường
- Khôi phục các vùng hư hỏng, xuống cấp (thay đổi màu sắc,...) hoặc bị thiếu pixel của ảnh trông tự nhiên hơn bằng cách ước lượng giá trị từ thông tin màu sắc được cung cấp

Ứng dụng:

- Khảo cổ: Khôi phục tài liệu bị ố vàng, rách nát
- Xử lý ảnh: Khôi phục ảnh cũ

b) Phương pháp

- Thuật toán Hole Filling SWA (Spiral weighted average algorithm)
- Giải thuật Gradient
- Thuật toán Fast Digital Image Processing
- Image Inpainting with Deep Learning

5. Nhóm Động Lực - Khảo sát Super Video Resolution và ứng dụng trong tăng chất lượng ảnh, video. (Khảo sát giải pháp truyền thống, deep learning)

a) Động lực nghiên cứu

Về khoa học:

- Là một trong những lĩnh vực nghiên cứu mới nổi ổn định nhất trong lĩnh vực xử lý hình ảnh. (Ruikar, 2020)
- Hình ảnh có độ phân giải cao cung cấp các chi tiết được tái tạo một cách cải thiện hơn của các cảnh và các đối tượng cấu thành, rất quan trọng đối với nhiều lĩnh vực nghiên cứu khác nhau như thị giác máy tính, đồ họa máy tính, v.v.
- Là một kỹ thuật phần mềm cho phép chúng ta nâng cao độ phân giải không gian của hình ảnh bằng phần cứng hiện có.

Ứng dụng:

Với sự phát triển của các ứng dụng xử lý hình ảnh, nhu cầu lớn về hình ảnh có độ phân giải cao vì hình ảnh High Resolution không chỉ cung cấp cho người xem một hình

ảnh đẹp mắt mà còn cung cấp các chi tiết bổ sung quan trọng cho việc phân tích trong nhiều ứng dụng. Các trường yêu cầu độ phân giải cao hơn, tăng cường chi tiết, khôi phục chi tiết, phóng to: hình ảnh / video y tế, hệ thống nhận dạng sinh trắc học, viễn thám, theo dõi và nhận dạng đối tượng, giám sát an ninh, truyền thông phương tiện, giám sát và phân tích video, nhận dạng mẫu trong thị giác máy tính, v.v. (Ruikar, 2020)

b) Phương pháp

Sử dụng:

- Mean Squared Error - Sai số toán phương trung bình
- Peak Signal-To-Noise Ratio - Tỉ số tín hiệu cực đại trên nhiễu
- Structural Similarity Index Measure - Chỉ mục cấu trúc tương tự

6. Nhóm CDKH - Khảo sát saliency map và ứng dụng trong phát hiện vùng quan tâm trong ảnh (Khảo sát giải pháp truyền thống, deep learning)

a) Động lực khoa học

- Nhận biết đối tượng là một trong những giai đoạn đầu tiên trước khi bước vào giai đoạn xử lý ảnh. Để có thể nhận biết được đối tượng trong ảnh thì trước tiên phải xác định được đối tượng.
- Bản chất của việc xác định đối tượng trong ảnh là chúng ta bóc tách đối tượng ra khỏi nền của ảnh trước khi thực hiện nhận biết đối tượng. Vấn đề đặt ra là làm thế nào để hệ thống thị giác máy tính có thể xác định được vùng nổi bật của ảnh (Saliency region hay Saliency map) từ một ảnh nền chưa biết.
- Như vậy nhận biết và xác định các Saliency map là bước đầu tiên nhằm mục đích nhận biết đâu là thành phần quan trọng của ảnh, điểm nổi bật của ảnh để tiện cho các xử lý phức tạp sau này. Đặc biệt, đối với những bài toán xử lý để nâng cao chất lượng ảnh thì việc sử dụng phương pháp saliency maps là điều tất yếu, hỗ trợ tìm kiếm được phần quan trọng và loại bỏ những yếu tố dư thừa của ảnh đó.
- Xác định được chủ thể chính của ảnh giúp tác vụ như detection recognition tracking được chính xác hơn

b) Động lực ứng dụng

Phát hiện và nhận dạng đối tượng: Thay vì áp dụng thuật toán phức tạp về tính toán cho toàn bộ ảnh, có thể sử dụng thuật toán chỉ các vùng nổi bật nhất của hình ảnh nơi có nhiều khả năng chứa một đối tượng nhất.

Image retargeting (làm hình ảnh có khả năng hiển thị trên các màn hình có các kích thước khác nhau): dựa trên các thông tin thị giác sẵn có của Saliency Map để ước tính chính xác tất cả các chi tiết ảnh có tính nổi bật.

Nén hình ảnh và video Đánh giá chất lượng video Cải thiện chất lượng ảnh

7. Nhóm Chiến thần - Khảo sát 2D, 3D Pose Estimation và ứng dụng trong nhận dạng hành động. (Khảo sát giải pháp truyền thống, deep learning)

a) Động lực khoa học

- HPE là một công nghệ dựa trên thị giác máy tính để xác định và phân tích để hiểu thông tin hình học và chuyển động trên các khớp nối của cơ thể con người
- HPE về cơ bản là một cách để nắm bắt một tập hợp các tọa độ cho mỗi khớp, được gọi là điểm chính, có thể mô tả tư thế của một người. Các mối kết nối giữa các điểm này là một cặp kết nối, không phải tất cả các điểm đều có thể tạo thành một cặp.
- Mục đích của HPe là tạo dựng lại bộ xương người và được xử lý cho các ứng dụng cụ thể

b) Động lực ứng dụng

Theo dõi sự thay đổi trong tư thế của một người trong một khoảng thời gian cũng có thể được sử dụng để nhận dạng hoạt động, cử chỉ và dáng đi. Có một số trường hợp sử dụng cho cùng một trường hợp, bao gồm:

- Chụp chuyển động (Motion Capture) và Thực tế tăng cường (Augmented Reality)
- Ứng dụng công nghệ mô phỏng hình ảnh bằng máy tính
- Huấn luyện Robot
- Thay vì lập trình thủ công cho rô-bốt đi theo quỹ đạo (follow trajectories), rô-bốt có thể được chế tạo để tự đi theo quỹ đạo của điểm neo của bộ xương người đang thực hiện một hành động.
- Theo dõi chuyển động (Motion Tracking) cho bảng điều khiển (consoles)
- Ứng dụng để theo dõi chuyển động của đối tượng con người để chơi trò chơi tương tác (interactive gaming).
- Phương pháp đề xuất: Hrnet

8. Nhóm MTAQ - Hand detection and Finger Pose Estimation và ứng dụng trong giao tiếp người-máy

a) Động lực khoa học

- Ý nghĩa khoa học: có thể phát hiện bàn tay con người và cùng với các điểm neo vào các vị trí thích hợp trong môi trường có nhiều cao
- qua các bài toán phát hiện đối tượng, ta thấy bàn tay không chỉ là một vật linh hoạt mà với từng tư thế khác nhau, chúng ta có thể tạo ra và tiếp nhận nhiều thông tin với nó. Và bàn tay là một vật thể xuất hiện ở khắp mọi nơi. Vì vậy việc phát hiện tốt tay, cũng như gắn các điểm neo tốt sẽ giúp chúng ta có cũng như tạo được nhiều thông tin hữu ích.

b) Động lực ứng dụng

- Ứng dụng thực tế: Nhận diện thủ ngữ, chụp ảnh bằng tay (samsung), Điều khiển drone bằng tay, điều khiển máy tính bằng tay (human-computer interaction), ... +
- Triển vọng: hiện nay mọi thứ dần trở nên ảo hóa, việc nhận diện bàn tay có thể giúp ta tạo ra bàn tay ảo trong môi trường 3D có rất ích rất nhiều trong các công nghệ thực tế ảo.
- Phương pháp đề xuất: MediaPipe
- Xử dụng hitmap và bộ dữ liệu Rendered Handpose Dataset chứa 41258 hình ảnh huấn luyện và 2728 hình ảnh ví dụ. Sử dụng segmentation masks để tách hình ảnh bàn tay ra riêng và 21 điểm neo trên bàn tay, từ những vị trí này sử dụng làm Gaussian peak để chiếu thành một hitmap

9. Nhóm Cù Lao Sweep - Khảo sát Optical flow và ứng dụng trong phát hiện camera motion, object motion trong Visual Slam

a) Động lực nghiên cứu

- Các mô hình nhận diện vật thể cần thực hiện 2 tác vụ chính đó là xác định vị trí của vật thể và dán nhãn cho vật thể đó. Gồm 2 nhóm mô hình:
- **Two stage detector (RCNN)** các mô hình này sẽ phân tách 2 tác vụ trên thành 2 giai đoạn riêng biệt. Đầu tiên là dùng một thuật toán có sẵn để xác định vật thể (vẽ bounding box cho vật thể) ví dụ như sử dụng thuật toán đề xuất khu vực Region Proposal hoặc mạng đề xuất khu vực RPN. Sau đó sử dụng một thuật toán khác để xác định nhãn của vật thể đã tìm được ở bước đầu tiên

- **One stage detector (YOLO)** các mô hình này sẽ gộp 2 tác vụ trên để xử lý đồng thời trên một mạng học sâu được huấn luyện end-to-end tức vừa xác định vật thể, vừa xác định nhãn của vật thể đó. Do 2 tác vụ được xử lý đồng thời nên tốc độ của mô hình này nhanh hơn mô hình Two stage detector

Sự phát triển của mạng học sâu CNN được áp dụng vào lĩnh vực xử lý hình ảnh. Tuy nhiên thuật toán nhóm R-CNN trước đó lại chưa áp dụng triệt để mạng CNN trong xử lý hình ảnh khi phân tách 2 tác vụ là định vị vật thể và dán nhãn vật thể thành hai tác vụ riêng biệt dẫn đến tốc độ xử lý không thể đáp ứng tốc độ của các máy quay phim hiện tại để nhận diện vật thể theo thời gian thực

Ứng dụng:

- Xây dựng hệ thống nhận diện vật thể theo thời gian thực với độ chính xác cao, ví dụ như:
- Xe tự hành: xe tự hành cần một hệ thống có độ phản hồi rất nhanh để xử lý được các tình huống khẩn cấp nhằm đảm bảo an toàn cho người dùng
- Giải trí: các trò chơi thực tế ảo tăng cường (AR) cần sự phân tích môi trường, nhận diện vật thể nhanh để có thể đưa ra những phản hồi nhanh nhất cho người dùng khi chơi game
- An ninh: cần hệ thống nhận diện vật thể nhanh và chính xác để đưa ra những cảnh báo cho người điều khiển để đảm bảo an ninh

b) Phương pháp

- Mô hình huấn luyện YOLO