

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG
KHOA CÔNG NGHỆ THÔNG TIN**



TS. Nguyễn Thị Thanh Nhân
TS. Phùng Thế Huân
Ths. Phạm Thị Liên

**BÀI GIẢNG
XỬ LÝ ẢNH**

Tài liệu lưu hành nội bộ

TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG
KHOA CÔNG NGHỆ THÔNG TIN

TS. Nguyễn Thị Thanh Nhân
TS. Phùng Thế Huân
Ths. Phạm Thị Liên

BÀI GIẢNG
XỬ LÝ ẢNH

Thái Nguyên, tháng 6 năm 2024

MỤC LỤC

CÁC TỪ VIẾT TẮT	6
MỞ ĐẦU	7
Chương 1: TỔNG QUAN VỀ XỬ LÝ ẢNH	8
Bài 1: TỔNG QUAN VỀ XỬ LÝ ẢNH (Số tiết: 5 tiết).....	8
1.1 Các khái niệm	8
1.2- Cấu trúc của hệ thống xử lý ảnh	10
1.3 Mô hình màu và loại ảnh	12
1.3.1 Các hệ màu thông dụng	12
1.3.2 Một số loại ảnh thông dụng	25
1.4. Cấu trúc dữ liệu ảnh	27
1.4.1. Cấu trúc ảnh Raster.....	27
1.4.2 Cấu trúc ảnh vector	27
1.5. Một số định dạng ảnh phổ biến.....	28
1.5.1 Định dạng ảnh TIFF.....	28
1.5.2 Định dạng ảnh GIF	29
1.5.3 Định dạng hình ảnh JPG	29
1.5.4 Định dạng ảnh BMP	30
1.5.5 Định dạng ảnh PNG.....	30
1.6. Ứng dụng của xử lý ảnh.....	31
CHƯƠNG 2. CÁC PHÉP BIẾN ĐỔI KHÔNG PHỤ THUỘC KHÔNG GIAN VÀ CÁC PHÉP TOÁN TRÊN ĐA ẢNH	34
Bài 2: Các phép biến đổi không phụ thuộc không gian (Số tiết: 5 tiết)	34
2.1. Giới thiệu.....	34
2.2- Các phép toán không phụ thuộc không gian.....	35
Bài 3: THAO TÁC TRÊN ĐA ẢNH, PHÉP TÍCH CHẬP (Số tiết: 5 tiết).....	47
2.3-Thao tác trên đa ảnh.....	47
Chương 3: CÁC PHÉP BIẾN ĐỔI PHỤ THUỘC KHÔNG GIAN	52
3.1- Phương pháp tích chập.....	52
BÀI 4: CÁC KỸ THUẬT LỌC THÔNG DỤNG (Số tiết: 5 tiết).....	59
3.2- Các kỹ thuật lọc thông dụng	59
CHƯƠNG 4. PHƯƠNG PHÁP PHÂN ĐOẠN ẢNH	67

Bài 5: PHƯƠNG PHÁP PHÂN ĐOẠN ẢNH (Số tiết: 5 tiết)	67
4.1-Tổng quan về phân đoạn ảnh	67
4.2. Phân đoạn bằng ngưỡng	68
4.2.1 Phân ngưỡng thủ công	68
4.2.2. Phân đoạn ngưỡng tự động	69
4.3 Phân đoạn ảnh bằng k-means	71
4.4 Phân đoạn ảnh bằng MeanShift	73
4.5 Phân đoạn ảnh bằng Watershed	75
CHƯƠNG 5. PHƯƠNG PHÁP PHÁT HIỆN BIÊN	79
Bài 6: PHƯƠNG PHÁP PHÁT HIỆN BIÊN GRADIENT (Số tiết: 5 tiết)	79
5.1 Các khái niệm	79
5.2 Phương pháp phát hiện biên Gradient	81
BÀI 7: PHƯƠNG PHÁP PHÁT HIỆN BIÊN LAPLACE (Số tiết: 2 tiết)	87
5.3- Phương pháp phát hiện biên Laplace	87
5.4. Phương pháp Canny	89
Chương 6: BIẾN ĐỔI HÌNH THÁI HỌC	94
Bài 8: BIẾN ĐỔI HÌNH THÁI HỌC (Số tiết: 5 tiết)	94
6.1. Phần tử cấu trúc	94
6.2 Phép giãn ảnh - Dilation	96
6.3 Phép co ảnh- EROSION	98
6.4.Một số tính chất	99
6.4.1 Tính gia tăng	99
6.4.2- Tính phân phối với phép \cup	100
6.4.3. Tính phân phối với phép \cap	101
6.4.4 Tính kết hợp	102
6.4.5 Tính đối ngẫu	103
6.4.6 Một số tính chất khác	103
6.5- Các phép toán đóng/ mở	103
6.6- Phép biến đổi “Hit or miss”	107
6.7- Ứng dụng của kỹ thuật hình thái học	108
6.7.1 Trích chọn biên ảnh	108
6.7.2 Tô vùng – Region fill	109
6.7.3 Gán nhãn cho các thành phần liên thông	110

6.7.4. Làm mỏng đối tượng	111
6.7.5 Làm dày đối tượng	113
CHƯƠNG 7. KỸ THUẬT NHẬN DẠNG	115
Bài 9: KỸ THUẬT NHẬN DẠNG (Số tiết: 5 tiết)	115
7.2. Phương pháp máy véc tơ hỗ trợ	119
7.2.1 Giới thiệu	119
7.2.2 Phân loại tuyến tính	119
7.2.3. Phân loại tuyến tính lẻ mềm	123
7.2.4 Hàm nhân	123
7.2.5 Tuyến tính hóa phân loại phi tuyến	124
Tài liệu tham khảo	126
Các câu hỏi thường gặp	127

CÁC TỪ VIẾT TẮT

TT	Từ viết tắt	Ý nghĩa của từ
1	RGB	Đỏ_xanh lơ_xanh lục
2	CMY	Lục lơ_Đỏ nâu_Vàng
3	HSV	Sắc màu_Độ bão hòa_giá trị
4	SVM	Máy vector hỗ trợ

MỞ ĐẦU

Cùng với sự phát triển nhanh chóng của khoa học công nghệ, các kỹ thuật dựa trên trí tuệ nhân tạo và thị giác máy tính ứng dụng trong các hệ thống thông minh đạt được những kết quả vượt bậc, có nhiều triển vọng. Trong thời gian qua, các kỹ thuật mới trong lĩnh vực thị giác máy tính, mà cụ thể hơn là xử lý ảnh và nhận dạng đã được nhiều nhà khoa học, tập đoàn công nghệ chú trọng nghiên cứu, phát triển. Vì thế, các kiến thức về xử lý ảnh, nhận dạng mẫu, trí tuệ nhân tạo, học máy đã trở thành môn học quan trọng đối với sinh viên chuyên ngành liên quan đến khoa học máy tính, công nghệ thông tin, tự động hóa. Để đáp ứng nhu cầu học tập của sinh viên Khoa Công nghệ Thông tin, bài giảng “Xử lý ảnh” được biên soạn theo đề cương môn học. Bài giảng này với mục tiêu cung cấp các kiến thức cơ bản về xử lý ảnh.

Nội dung bài giảng gồm 7 chương:

Chương 1. Tổng quan về xử lý ảnh.

Chương 2. Các phép biến đổi không phụ thuộc không gian và thao tác trên đa ảnh.

Chương 3. Các phép biến đổi phụ thuộc không gian.

Chương 4. Phương pháp phân đoạn ảnh.

Chương 5. Phương pháp phát hiện biên.

Chương 6. Biến đổi hình thái học.

Chương 7. Kỹ thuật nhận dạng.

Bài giảng được biên soạn trên cơ sở kế thừa những nội dung của các giáo trình về xử lý ảnh đã được xuất bản trong và ngoài nước nhưng được cấu trúc lại, bổ sung và mở rộng cho phù hợp với sinh viên Khoa Công nghệ thông tin. Hy vọng đây sẽ là tài liệu có ích phục vụ trong học tập, tham khảo cho sinh viên ngành công nghệ thông tin cũng như độc giả quan tâm đến lĩnh vực xử lý ảnh và nhận dạng.

Mặc dù bài giảng đã được biên soạn nghiêm túc, cẩn thận, nhưng chắc chắn không thể tránh khỏi những thiếu sót trong biên soạn. Tác giả rất mong nhận được sự góp ý của các nhà khoa học, các đồng nghiệp và các em sinh viên đóng góp ý kiến để cuốn bài giảng được hoàn thiện hơn. Xin trân trọng cảm ơn.

Tác giả

Chương 1: TỔNG QUAN VỀ XỬ LÝ ẢNH

Nội dung chính của chương

Chương 1 giới thiệu kiến thức tổng quan về xử lý ảnh như: Những khái niệm liên quan đến ảnh kỹ thuật số, các thành phần hệ thống xử lý ảnh, các loại hệ màu cơ bản, các loại hệ ảnh thông dụng, các cấu trúc dữ liệu ảnh và kiểu định dạng phổ biến trong xử lý ảnh

Mục tiêu cần đạt được của chương

Sinh viên hiểu được các kiến thức cơ bản về: các khái niệm cơ bản; cấu trúc của hệ thống xử lý ảnh; mô hình màu và loại ảnh; cấu trúc dữ liệu ảnh; một số định dạng phổ biến và các ứng dụng.

Bài 1: TỔNG QUAN VỀ XỬ LÝ ẢNH (Số tiết: 5 tiết)

1.1 Các khái niệm

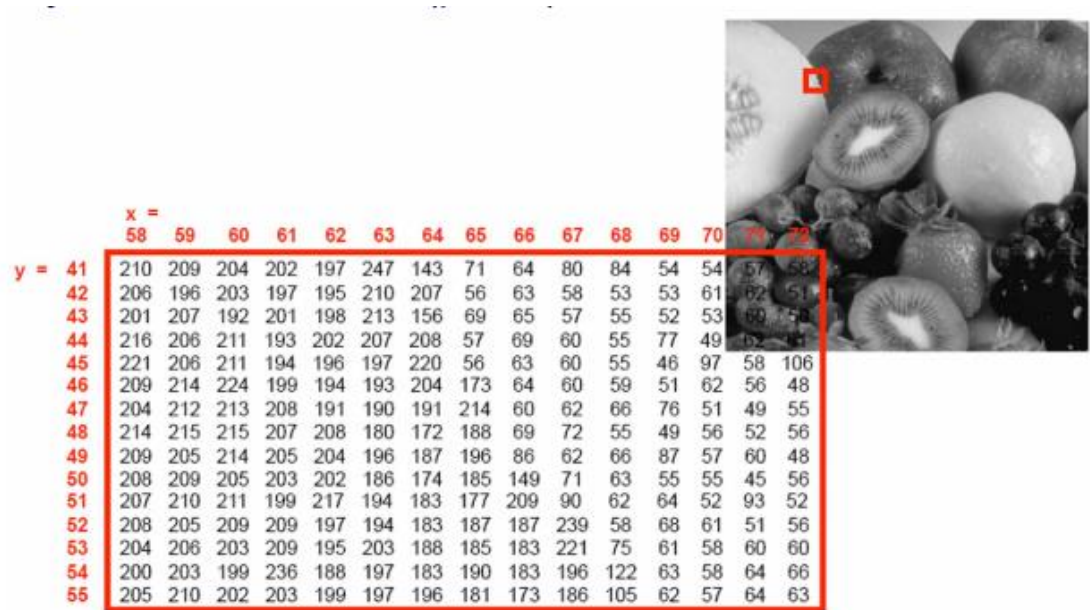
Con người thu nhận thông tin qua các giác quan, trong đó thị giác đóng vai trò quan trọng nhất. Hình ảnh chiếm 80% thông tin con người thu nhận được. Những năm trở lại đây với sự phát triển của phần cứng máy tính, xử lý ảnh và đồ họa đã phát triển một cách mạnh mẽ và có nhiều ứng dụng trong cuộc sống.

Điểm ảnh hay còn gọi là pixel (picture element, pels, image element) được xem như dấu hiệu hay cường độ sáng tại một tọa độ trong không gian của đối tượng.

Kích thước và khoảng cách giữa các điểm ảnh được biểu diễn thích hợp sao cho mắt người cảm nhận sự liên tục về không gian và mức xám (màu) của ảnh số gần với như với hình ảnh của nó trong không gian thật. Số điểm ảnh trên mỗi diện tích biểu diễn xác định độ phân giải của ảnh số. Ảnh có độ phân giải càng cao thì càng thể hiện rõ nét các đặc điểm của hình ảnh, càng làm cho hình ảnh gần trở nên gần với thực tế thực và sắc nét hơn.

Ảnh số

Ảnh số (digital image) có thể được xem là một biểu diễn dữ liệu rời rạc thể hiện thông tin về không gian và cường độ màu. Ảnh số gồm một tập hợp hữu hạn các phần tử được biểu diễn bởi giá trị số. Ảnh số có thể được biểu diễn dưới dạng ma trận 2 chiều, mỗi phần tử của ảnh số gọi là điểm ảnh. Ảnh số được xác định theo mảng 2 chiều biểu diễn cường độ sáng của điểm ảnh với giá trị cố định, cũng có thể xác định theo hàm hai chiều $f(x,y)$, trong đó x và y là các tọa độ không gian và độ lớn của hàm f được gọi là độ sáng (intensity) hay độ xám (gray level) của ảnh tại điểm đó.



Hình 1.1 Số hóa của một ảnh liên tục.

Mức xám (gray level)

Giá trị mức xám là kết quả của ánh xạ của giá trị độ sáng của một điểm ảnh màu trong không gian thực với một giá trị số nguyên dương thể hiện mức độ sáng tối của điểm ảnh đó. Các thang giá trị mức xám thường dùng là 2, 16, 32, 64, 128, 256. Ảnh đa mức xám thường dùng là 256, như vậy mức xám thường được xác định trong khoảng $[0, 255]$ tùy thuộc vào giá trị mà mỗi điểm ảnh được biểu diễn.

Các dạng ảnh

Ảnh	Số bit/pixel	Số màu
Ảnh nhị phân	1	2
Ảnh đa cấp xám	8	256
Ảnh RGB	24	256^3
Ảnh 32 bit	32	256^4

Bảng 1.1 Một số dạng ảnh

Ảnh màu: Ảnh màu thường là các ảnh chứa thông tin về đối tượng được biểu diễn dưới dạng màu sắc mà mắt thường có thể quan sát được. Mỗi điểm ảnh có cấu trúc gồm nhiều kênh màu khác nhau, thông thường trong máy tính, nó biểu diễn ba lớp màu cơ bản RGB, gồm màu đỏ, xanh lá cây và xanh lam.

Ảnh đa mức xám: Ảnh đa mức xám thường biểu diễn thông tin liên quan đến cường độ đa mức xám của đối tượng trong không gian mà không được thể hiện bởi màu sắc thực của nó.

Ảnh nhị phân: ảnh biểu diễn đối tượng bởi hai mức 0 hoặc 1, thường được dùng để biểu diễn, phân biệt sự xuất hiện đối tượng và nền trong mỗi bức ảnh.

Ảnh hồng ngoại: Biểu diễn trực quan quang phổ, liên quan đến phổ điện tử. Ảnh hồng ngoại cung cấp thông tin ảnh dựa trên phản xạ ánh sáng hồng ngoại hoặc bức xạ hồng ngoại mà các đối tượng trong khung nhìn phát ra. Dựa vào khả năng thu nhận phản xạ hoặc bức xạ hồng ngoại mà các loại camera hồng ngoại thu được hình ảnh trong điều kiện không có ánh sáng nhìn thấy hoặc ánh sáng kém.

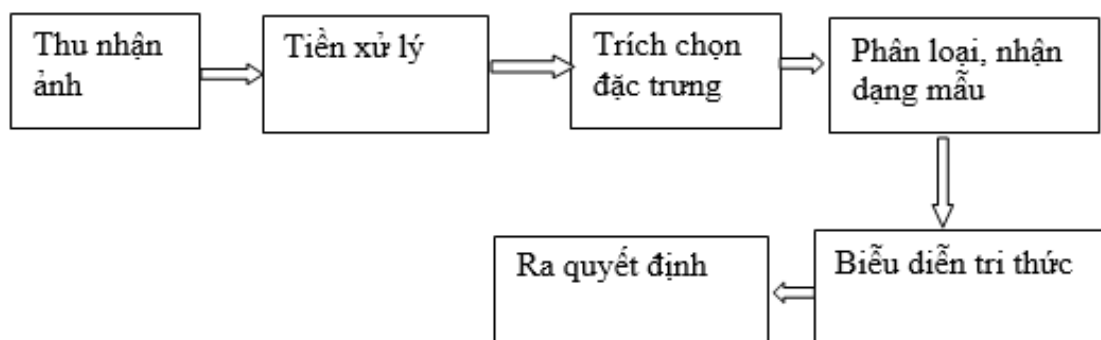
Độ phân giải của ảnh

Độ phân giải không gian của ảnh là mật độ pixel được xác định trên một ảnh số. Một số độ phân giải thông thường được sử dụng trong các thiết bị hiển thị và lĩnh vực xử lý ảnh như 640x480, 800x600, 1024x768 (HD), 1920x1080(full HD), 3840x2160 (UHD),...

Kỹ thuật xử lý ảnh: là quá trình biến đổi một hình ảnh thành một hình ảnh khác bằng máy tính điện tử một cách tự động phụ thuộc vào mục đích của người sử dụng.

1.2- Cấu trúc của hệ thống xử lý ảnh

Xử lý ảnh là quá trình biến đổi từ một ảnh ban đầu được thu nhận từ một thiết bị sang một không gian mới sao cho làm nổi bật đặc tính dữ liệu, thuận lợi cho quá trình xử lý thông tin và nâng cao độ chính xác. Một hệ thống xử lý ảnh thường gồm một số thành phần chính sau:



Hình 1.2 Sơ đồ tổng quát hệ thống xử lý ảnh

-Thu nhận ảnh: là việc hình ảnh về thế giới thực được thu nhận và chuyển qua tín hiệu ảnh rời rạc thông qua máy ảnh kỹ thuật số hoặc các thiết bị thu hình ảnh khác.

- **Tiền xử lý (preprocessing):** là bước xử lý trên ảnh đầu vào nhằm khử nhiễu, làm nổi bật một số tính chất của ảnh nhằm nâng cao chất lượng các bước xử lý sau.

Do những nguyên nhân khác nhau: có thể do chất lượng thiết bị thu nhận, do nguồn sáng hay do nhiễu ảnh có thể bị suy biến, do vậy cần phải tăng cường và khôi phục lại ảnh để làm nổi bật một số đặc tính chính của ảnh, hay làm cho ảnh gần giống nhất với trạng thái gốc- trạng thái trước khi ảnh bị biến dạng.

Có 2 loại nhiễu cơ bản trong quá trình thu nhận:

+ Nhiễu hệ thống: Đặc trưng của nhiễu hệ thống là tính tuần hoàn. Do vậy, có thể tách được loại nhiễu này bằng việc sử dụng phép biến đổi Fourier và loại bỏ các điểm đỉnh (peaks).

+ Nhiễu ngẫu nhiên: là nhiễu không rõ nguyên nhân, khắc phục bằng các giải pháp lọc, phương pháp nội suy.

Chỉnh mức xám: Đây là các kỹ thuật nhằm chỉnh sửa tính không đồng đều của thiết bị thu nhận hoặc tăng độ tương phản giữa các vùng ảnh.

Chỉnh tán xạ: Ảnh nhận được từ các thiết bị điện tử hay quang học có thể bị nhoè. Phương pháp biến đổi dựa trên tích chập của ảnh với hàm tán xạ cho phép giải quyết được bài toán hiệu chỉnh này.

-**Trích chọn đặc trưng (Feature Extracting):** là quá trình biến đổi dữ liệu ảnh đầu vào thành tập các đặc trưng. Các đặc trưng thường có đặc tính phân biệt cao của mẫu đầu vào giúp cho việc phân biệt mẫu dữ liệu ảnh dễ dàng hơn nhằm nâng cao chất lượng phân loại mẫu so với xử lý dữ liệu thô trên giá trị pixel ảnh. Việc trích chọn đặc trưng cũng có thể làm giảm kích thước thể hiện thông tin trong ảnh trong khi dữ liệu về đặc trưng ảnh có tính phân biệt cao.

Phân loại, nhận dạng mẫu là quá trình xử lý dữ liệu bằng các kỹ thuật, phương pháp phân tích đặc trưng để phân loại mẫu về các nhóm có một số tính chất chung. Các phương pháp phân loại, nhận dạng mẫu thường liên quan đến kỹ thuật học máy, bao gồm cả học có giám sát và học không có giám sát.

Biểu diễn tri thức là bước thể hiện mức cao của biểu diễn dữ liệu, các mẫu dữ liệu sau khi phân loại, nhận dạng được biểu diễn dưới dạng tri thức giúp hệ thống có khả năng “hiểu biết” ngữ nghĩa của nó theo từng kiểu ứng dụng khác nhau trong hệ thống trí tuệ nhân tạo và hệ thống thông minh.

Ra quyết định là bước cuối cùng của một hệ thống trong lĩnh vực hệ thống thông minh. Các mẫu được biểu diễn dưới dạng tri thức và được suy luận ngữ nghĩa để đưa ra các quyết định thực hiện một nhiệm vụ nào đó. Ví dụ trong hệ thống robot di chuyển tự động, khi phát hiện chướng ngại vật, robot sẽ tự động ra quyết định tìm kiếm đường đi mới và di chuyển theo đường đi khả thi.

Các mức độ của hệ thống xử lý ảnh

-Mức độ thấp: chỉ biết sử dụng các kỹ thuật xử lý ảnh đơn giản, thuần túy, không có tri thức như các thao tác tiền xử lý ảnh.

-Trung bình: có một chút về tri thức (trí tuệ nhân tạo) có nhiệm vụ tách và đặc trưng hóa các thành phần trong một ảnh nhận được từ quá trình xử lý mức độ thấp như các thao tác phân đoạn ảnh.

-Cao: nhận dạng, phân tích ảnh, ra quyết định. Cần kiến thức và sự hiểu biết.

Thảo luận: Cho biết phần mềm Photoshop đạt được ở mức độ nào của hệ thống xử lý ảnh

1.3 Mô hình màu và loại ảnh

1.3.1 Các hệ màu thông dụng

1.3.1.1- Mô hình màu

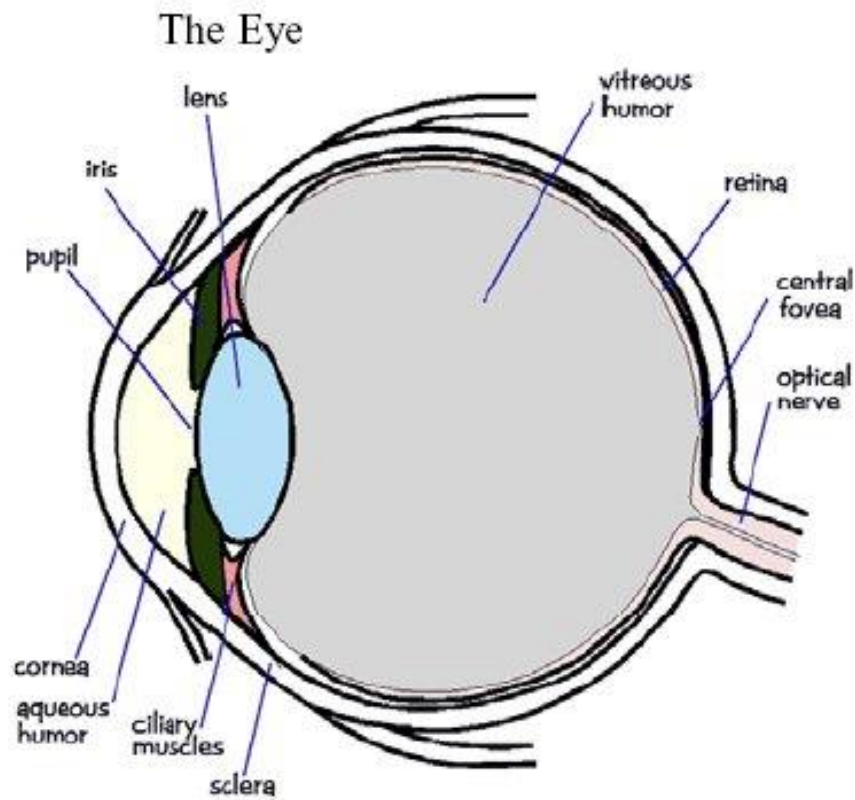
***Cảm nhận màu**

Cấu tạo của mắt:

Mắt có dạng gần hình cầu với đường kính trung bình khoảng gần 20mm. Mắt bao gồm 3 lớp: giác mạc(cornea), võng mạc(scleria) và mô mạc (choroid) và thấu kính. Giác mạc trong suốt, bao bọc phía ngoài mắt. Võng mạc bao bọc phía trong mắt. Mô mạc bao gồm một mạng mạch máu nhỏ li ti, nó là nguồn nuôi mắt chính. Mô mạc giúp là giảm số lượng ánh sáng đi vào mắt. Tròng mắt co và giãn để điều tiết lượng ánh sáng đi vào mắt. Độ mở của tròng mắt thay đổi với đường kính 2-8mm. Mặt trước của tròng mắt bao gồm sắc tố nhìn của mắt. Thấu kính mắt tập trung lượng ánh sáng vào các tế bào thần kinh mắt (hoàng điểm)

Màng hoàng điểm (retina) được tạo ảnh nếu như mắt nhận ánh sáng hội tụ từ vật thể bên ngoài mắt. Có 2 loại tế bào nhận ánh sáng là dạng hình nón (cones) và que (rod). Tế bào hình nón ở mỗi mắt gồm 6-7 triệu. Nó nằm chủ yếu ở trung tâm hoàng điểm và nhạy với màu sắc. Mắt người nhận biết các chi tiết nhỏ nhờ tế bào hình nón gọi là nhìn thấy ánh sáng mạnh. Số lượng tế bào hình que 75-150 triệu, phân bố trên bề mặt hoàng điểm. Các tế bào hình que cho một hình ảnh chung về trường chiếu sáng, nó không nhạy với màu sắc mà nhạy với ánh sáng có mức thấp gọi là nhìn thấy ánh sáng tối.

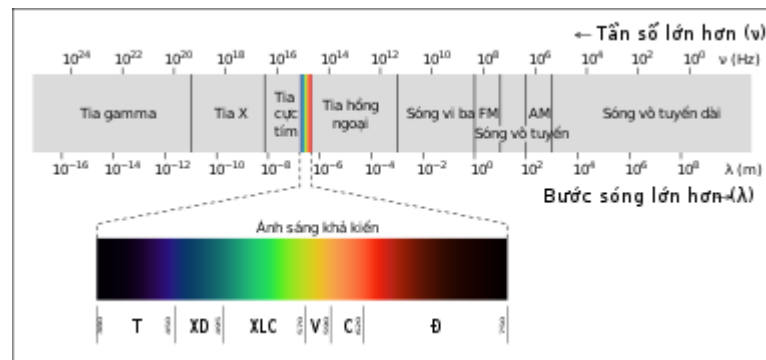
Ảnh trên võng mạc được ánh xạ sơ cấp trong vùng hoàng điểm. Việc thu nhận ảnh này do sự kích thích của các phần tử thu nhận ánh sáng mà nó biến đổi năng lượng bức xạ thành các xung điện và được não giải mã.



Hình 1.3 Cấu tạo của mắt người

Ánh sáng (Light) là bức xạ điện từ kích thích sự nhìn của chúng ta, được biểu diễn bởi phân bố năng lượng $L(\lambda)$, λ là bước sóng có giới hạn từ 350nm đến 780nm.

Màu sắc của một tia ánh sáng là cảm giác màu mà tia sáng đó gây nên ở mắt người. Màu sắc của các vật thể là màu sắc của ánh sáng do chúng phát xạ ra (nếu là vật nóng sáng) hay phản xạ từ chúng từ một nguồn chiếu sáng (nếu coi vật là không nóng sáng). Ánh sáng mà con người nhận biết (hay màu khác nhau) là dải tần hẹp trong quang phổ điện từ



Hình 1.4 Bước sóng của ánh sáng

Trong xử lý ảnh và đồ họa, mô hình màu là một chỉ số kỹ thuật của một hệ tọa độ màu 3 chiều với tập các màu nhỏ thành phần có thể trông thấy được trong hệ thống tọa độ màu thuộc một gam màu đặc trưng. Ví dụ như mô hình màu RGB (Red, Green, Blue): là một đơn vị tập các màu thành phần sắp xếp theo hình lập phương của hệ trục tọa độ Đề các.

Mô hình màu là một mô hình toán học trừu tượng gồm 3 hoặc 4 thành phần màu cơ sở để sinh ra nhiều màu khác nhau.

Hai hay 3 màu được sử dụng để mô tả các màu khác được gọi là primary colors. Thực tế là số primary colors là không giới hạn. Tuy nhiên chỉ 3 màu cơ sở đã đủ cho phần lớn các ứng dụng.

Mục đích của mô hình màu là cho phép các chỉ số kỹ thuật quy ước của một số loại màu sắc thích hợp với các màu sắc của một số gam màu khác. Chúng ta có thể nhìn thấy trong mô hình màu này, không gian màu là một tập hợp nhỏ hơn của không gian các màu có thể nhìn thấy được, vì vậy một mô hình màu không thể được sử dụng để định rõ tất cả có thể nhìn thấy. Sau đây, ta xem xét một số mô hình hay được sử dụng nhất.

1.3.1.2 . Mô hình màu RGB (Red, Green, Bule)

3 màu cơ sở là: đỏ (Red), xanh lá cây (Green), xanh da trời (Blue), sử dụng mô hình màu cộng tính

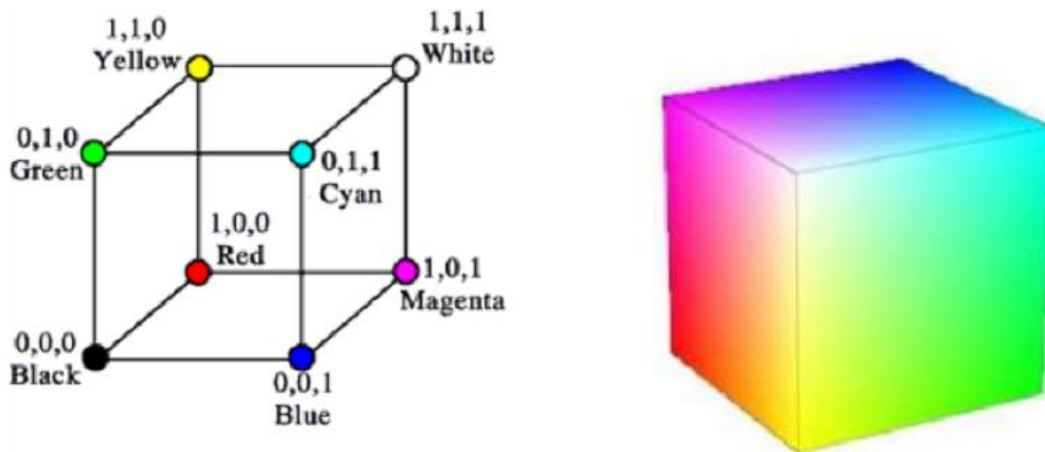
Mô hình màu RGB được biểu diễn bởi lập phương có cạnh là 1 với các trục R, G, B

Góc biểu diễn màu đen

Tọa độ (1, 1, 1) biểu diễn màu trắng.

Tọa độ trên các cạnh trục biểu diễn các màu cơ sở.

Các cạnh còn lại biểu diễn màu bù cho mỗi màu cơ sở



Hình 1.5 Mô hình màu RGB

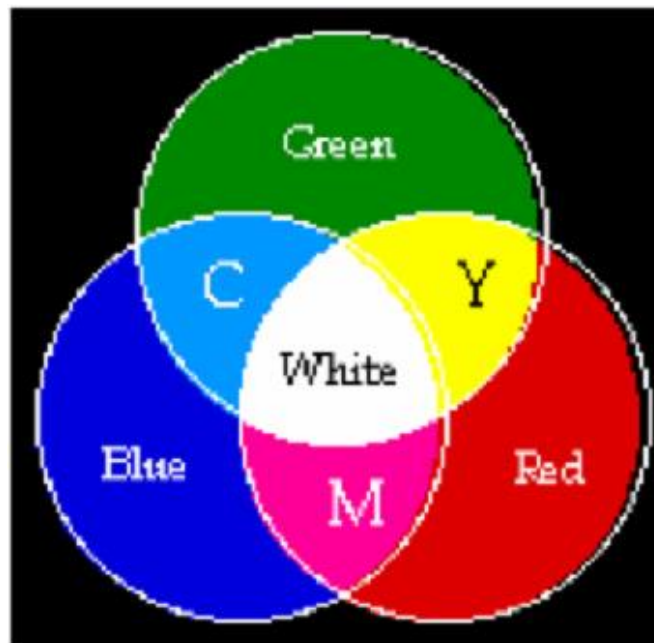
Đường chéo từ (0, 0, 0) đến (1, 1, 1) là biểu diễn màu xám.

Biểu đồ RGB thuộc mô hình cộng:

Phát sinh màu mới bằng cách cộng cường độ màu cơ sở

Gán giá trị từ 0 đến 1 cho R, G, B

Red+Blue \rightarrow Magenta (1, 0, 1)



Hình 1.6 Phối trộn màu trong mô hình RGB

Nhận xét

+ Mô hình này không thể biểu diễn mọi màu trong phổ nhìn thấy

+ Màn hình máy tính và TV sử dụng mô hình này

+Phụ thuộc vào thiết bị

Chuyển đổi ảnh RGB thành ảnh đa mức xám

$$\text{Grey}=0.2989R+0.5870G+0.1140B$$

1.3.1.3. Mô hình màu CMY, CMYK

Mô hình màu CMY

Các màu cơ sở:

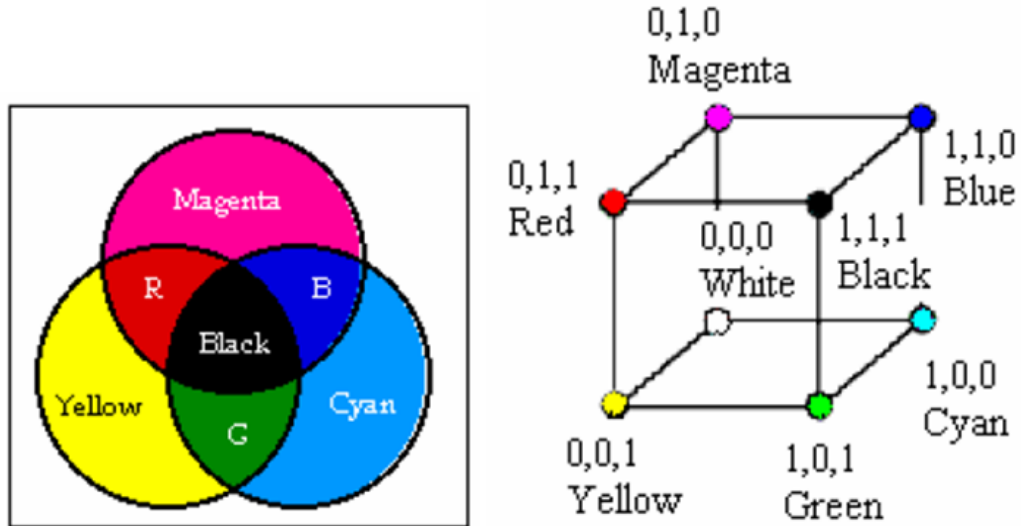
C (Cyan): xanh lơ, M(magenta): hồng sẫm, Y(yellow); vàng.

Mô hình CMY là bù của mô hình RGB.

Là phần bù tương ứng cho các màu đỏ, lục, lam và chúng được sử dụng như những bộ lọc loại trừ các màu này từ ánh sáng trắng. Vì vậy CMY còn được gọi là các phần bù loại trừ của màu gốc. Tập hợp màu thành phần biểu diễn trong hệ tọa độ Đề-các cho mô hình màu CMY cũng giống như cho mô hình màu RGB ngoại trừ màu trắng (ánh sáng trắng), được thay thế màu đen (không có ánh sáng) ở tại nguồn sáng. Các màu thường được tạo thành bằng cách loại bỏ hoặc được bù từ ánh sáng trắng hơn là được thêm vào những màu tối

Biểu đồ CMY thuộc loại mô hình trừ.

Hỗn hợp của các màu CMY **lý tưởng** là loại trừ (các màu này khi in cùng một chỗ trên nền trắng sẽ tạo ra màu đen).



Hình 1.7 Mô hình màu CMY

Khi bề mặt được bao phủ bởi lớp mực màu xanh tím, sẽ không có tia màu đỏ phản chiếu từ bề mặt đó. Màu xanh tím đã loại bỏ phần màu đỏ phản xạ khi có tia sáng trắng, mà bản chất là tổng của 3 màu đỏ, lục, lam. Vì thế ta có thể coi màu Cyan là màu trắng trừ đi màu đỏ và đó cũng là màu lam cộng màu lục. Tương tự như vậy ta có màu đỏ thẫm (magenta) hấp thụ màu lục, vì thế nó tương đương với màu đỏ cộng màu lam. Và cuối cùng màu vàng (yellow) hấp thụ màu lam, nó sẽ bằng màu đỏ cộng với lục.

Khi bề mặt của thực thể được bao phủ bởi xanh tím và vàng, chúng sẽ hấp thụ hết các phần màu đỏ và xanh lam của bề mặt. Khi đó chỉ tồn tại duy nhất màu lục bị phản xạ từ sự chiếu sáng của ánh sáng trắng. Trong trường hợp khi bề mặt được bao phủ bởi cả 3 màu xanh tím, vàng, đỏ thẫm, hiện tượng hấp thụ xảy ra trên cả 3 màu đỏ, lục và lam. Do đó, màu đen sẽ màu của bề mặt. Chuyển đổi từ mô hình màu RGB sang CMY.

$$\begin{bmatrix} C \\ M \\ Y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

Mô hình màu CMYK

CMYK để chỉ mô hình màu loại trừ.

K=Key (trong tiếng Anh nên hiểu theo nghĩa là cái gì đó *then chốt* hay *chủ yếu* để ám chỉ màu đen mặc dù màu này có tên tiếng Anh là *black* do chữ **B** đã được sử dụng để chỉ màu xanh lam (blue) trong mô hình màu RGB.

Nguyên lý làm việc của CMYK là trên cơ sở hấp thụ ánh sáng. Màu mà người ta nhìn thấy là từ phần của ánh sáng không bị hấp thụ.

Thực hiện chuyển đổi từ CMY sang CMYK

$$K = \min(C, M, Y) \quad \begin{cases} C = C - K \\ M = M - K \\ Y = Y - K \end{cases}$$

Mục đích: dùng trong in ấn

1.3.1.4 Mô hình màu HSV (Hue, Saturation, Value)

Các mô hình màu RGB, CMY được định hướng cho phần cứng trái ngược với mô hình màu HSV của Smith đề xuất năm 1978 hay còn được gọi là mẫu HSB với B là Brightness (độ sáng). Mô hình màu này dựa trên các đặc tính màu trực quan như sắc (tint), bóng (shade) và tông màu (tone) hoặc họ màu, độ thuần khiết và độ sáng của hình ảnh trong thực tế để biểu diễn

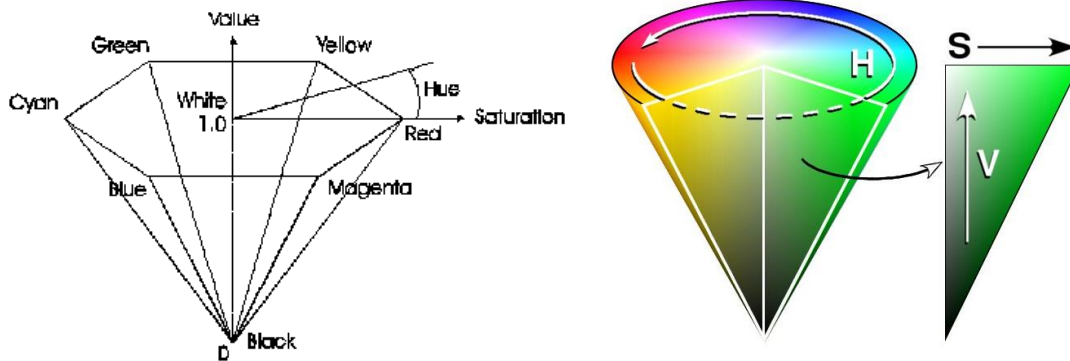
Hệ thống tọa độ có dạng hình trụ và tập màu thành phần của không gian bên trong mô hình màu được xác định là hình nón hoặc hình chóp sáu cạnh như trong hình 1.8. Đỉnh hình chóp là sáu cạnh khi $V = 1$ chứa đựng mối quan hệ giữa các màu sáng và những màu trên mặt phẳng với $V = 1$ đều có màu sáng.

H (Hue) là giá trị thể hiện sắc màu của hình ảnh (hay còn gọi là vùng màu), sắc màu thông thường được dùng để chỉ tên gọi của màu như đỏ, lục, lam, chàm, tím,... Các sắc màu khác nhau được biểu diễn trên một vòng tròn chỉ màu, giá trị từ 0^0 đến 360^0

S (Saturation) là giá trị thể hiện mức độ bão hòa màu. Giá trị của nó để chỉ mức độ thuần khiết của màu. Nói cách khác, khi ảnh có độ bão hòa cao, màu sẽ trong và rực rỡ hơn giá trị bão hòa thấp. Giá trị của S nằm trong đoạn $[0, 1]$, trong đó S đạt giá trị cao nhất (giá trị 1) là màu tinh khiết nhất, hoàn toàn không pha trắng, nghĩa là S càng lớn thì màu càng tinh khiết, nguyên chất.

V (value) là giá trị đo độ sáng của màu. Thành phần V có giá trị trong đoạn $[0,1]$ với giá trị đặc biệt $V=0$ thì ảnh hoàn toàn tối (đen), ngược lại $V=1$ là hoàn toàn sáng. Giá trị V càng lớn thì màu càng sáng.

Mô hình HSV suy diễn từ mô hình RGB. Hãy quan sát hình hộp RGB theo đường chéo từ White đến Black (gốc) -> ta có hình lục giác. Sử dụng làm đỉnh hình nón HSV



Hình 1.8 Mô hình màu HSV

$$\begin{cases} 0^{\circ} \leq H \leq 360^{\circ} \\ 0 \leq S, V \leq 1 \end{cases}$$

$V=0$, màu đen. Đỉnh lục giác có màu cực đại

Red= $(0^{\circ}, 1, 1)$

Green= $(120^{\circ}, 1, 1)$

Blue= $(240^{\circ}, 1, 1)$

Mô hình HSV trực giác hơn mô hình RGB

Chuyển đổi từ RGB sang HSV

Việc chuyển đổi ảnh từ hệ màu RGB sang hệ màu HSV tương ứng với việc kết hợp các giá trị màu thành phần R, G, B để tính các thành phần H, S, V. Trước hết, thực hiện chuyển đổi cường độ sáng từ miền giá trị $[0 \dots 255]$ thành $[0..1]$ bằng cách chia các giá trị thành phần màu R, G, B cho 255:

$$\begin{cases} R' = R/255 \\ G' = G/255 \\ B' = B/255 \end{cases}$$

Tiếp theo, thực hiện tính toán các giá trị cực đại, cực tiểu và độ chênh lệch của giá trị màu thực tế của ảnh trong các giá trị R' , G' , B' theo công thức.

$$\begin{cases} M = \max(R', G', B') \\ m = \min(R', G', B') \\ C = M - m \end{cases}$$

Tiếp theo, giá trị thành phần Hue được tính theo công thức:

$$\begin{cases} \text{undefined}, & \text{if } C = 0 \\ \frac{G' - B'}{C} \bmod 6, & \text{if } M = R' \\ \frac{B' - R'}{C} + 2, & \text{if } M = G' \\ \frac{R' - G'}{C} + 4, & \text{if } M = B' \end{cases}$$

$$H = 60^\circ \times H'$$

Tính giá trị thành phần độ sáng Value được xác định theo giá trị cực đại $V=M$.

Cuối cùng, giá trị thành phần Saturation được tính bằng công thức:

$$S_{HSV} = \begin{cases} 0, & \text{if } V = 0 \\ \frac{C}{V}, & \text{otherwise} \end{cases}$$

Chuyển đổi từ HSV sang RGB

Đầu tiên tính giá trị màu C (Chroma) trong không gian màu HSV theo công thức $C=V \times S_{HSV}$

Tiếp theo, thực hiện tính các giá trị màu (R_1, G_1, B_1) trong không gian màu RGB tương ứng với giá trị Hue và Chroma bằng cách sử dụng giá trị trung gian X cho thành phần lớn nhất của màu này với các công thức:

$$H' = \frac{H}{60^\circ}$$

$$X = C \times (1 - |H' \bmod 2 - 1|)$$

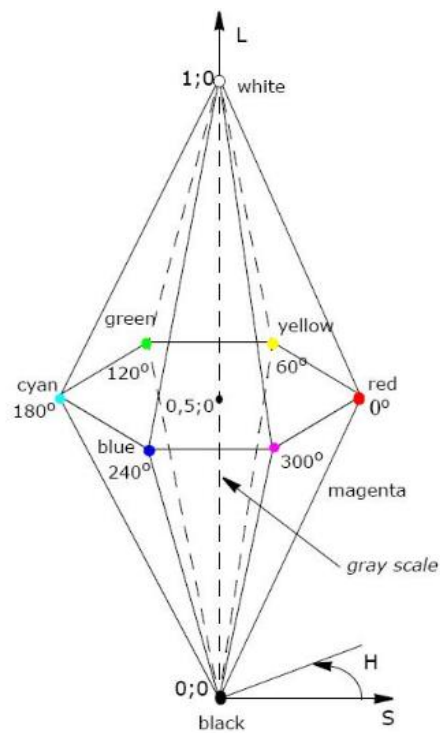
$$(R_1, G_1, B_1) = \begin{cases} (0, 0, 0) & \text{if } H \text{ is undefined} \\ (C, X, 0) & \text{if } 0 \leq H' \leq 1 \\ (X, C, 0) & \text{if } 1 \leq H' \leq 2 \\ (0, C, X) & \text{if } 2 \leq H' \leq 3 \\ (0, X, C) & \text{if } 3 \leq H' \leq 4 \\ (X, 0, C) & \text{if } 4 \leq H' \leq 5 \\ (C, 0, X) & \text{if } 5 \leq H' < 6 \end{cases}$$

Cuối cùng, các thành phần màu R, G, B được xác định bằng cách thêm vào mỗi thành phần màu một giá trị m, với $m=V-C$ là mức độ chênh lệch giữa độ sáng V và chroma là mức chênh lệch giá trị màu thực tế giữa hai không gian màu.

$$(R, G, B) = (R_1 + m, G_1 + m, B_1 + m)$$

1.3.1.5 Mô hình màu HLS

Mô hình màu HLS được xác định bởi tập hợp hình chóp sáu cạnh đôi của không gian hình trụ. Sắc màu là góc quanh trục đứng của hình chóp sáu cạnh đôi với màu đỏ tại góc 0° . Các màu sẽ xác định theo thứ tự giống như trong biểu đồ CIE khi ranh giới của nó bị xoay ngược chiều kim đồng hồ: Màu đỏ, màu vàng, màu lục, màu xanh tím, màu lam và đỏ thẫm. Điều này cũng giống như thứ tự sắc xếp trong mẫu hình chóp sáu cạnh đơn HSV.



Hình 1.9. Mô hình màu HLS

Chúng ta có thể xem mẫu HLS như một sự biến dạng của mẫu HSV mà trong đó mẫu này màu trắng được kéo hướng lên hình chóp sáu cạnh phía trên từ mặt $V=1$. Như với mẫu hình chóp sáu cạnh đơn, phần bổ sung của một màu sắc được đặt ở vị trí 180° hơn là xung quanh hình chóp sáu cạnh đôi, sự bão hòa được đo xung quanh trục đứng, từ 0 trên trục tới 1 trên bề mặt. Độ sáng bằng không cho màu đen và bằng một cho màu trắng.

Chuyển đổi từ RGB sang HLS

Hàm RGB_HLS_Conversion()

H: Sắc độ màu [0-360] với màu đỏ tại điểm 0

S: Độ bão hòa [0-1]

V: Giá trị cường độ sáng [0-1]

Max: Hàm lấy giá trị cực đại

Min: Hàm lấy giá trị nhỏ nhất

```
{
    //Xác định độ sáng
    M1= Max(R,G,B)
    M2= Min(R,G,B)
    L= (M1+M2)

    //Xác định độ bão hòa
    If M1=M2 //Trường hợp không màu
        S= 0
        H= Undefined
    Else
        If L <= 0.5 then //Trường hợp màu
            S= (M1-M2)/(M1+M2)
        Else
            S= (M1-M2)/(2-M1-M2)
        Endif

        //Xác định sắc độ
        Cr= (M1-R)/(M1-M2)
        Cg= (M1-G)/(M1-M2)
        Cb= (M1-B)/(M1-M2)

        if R=M1 then
```

$$H = C_b - C_g$$

If $G = M1$ then

$$H = 2 + C_r - C_b$$

If $B = M1$ then

$$H = 4 + C_g - C_r$$

$$H = H * 60$$

if $H < 0$ then

$$H = H + 360$$

endif

}

Chuyển đổi từ HLS sang RGB

Hàm HLS_RGB_Conversion()

H: Sắc độ màu [0-360] với màu đỏ tại điểm 0

S: Độ bão hòa [0-1]

V: Giá trị cường độ sáng [0-1]

{

If $L \leq 0.5$ then

$$M2 = L * (1 + S)$$

Else

$$M2 = L + S - L * S$$

Endif

$$M1 = 2 * L - M2$$

//Kiểm tra độ bão hòa = 0

If $S = 0$ then

If $H = \text{Undefined}$

$$R = L$$

$$G = L$$

$$B = L$$

```

    Else //Error: Dữ liệu nhập sai
    Endif

Else //Xác định giá trị của RGB
    RGB(H+120, M1,M2,Value)
    R= Value
    RGB(H, M1,M2,Value)
    G= Value
    RGB(H-120, M1,M2,Value)
    B= Value
Endif
}

//Hàm điều chỉnh giá trị của H cho phù hợp khoảng xác định
Hàm RGB(H, M1, M2, Value)
{
    If H < 0 then
        H= H+360
    If H < 60 then
        Value= M1+(M2-M1)*H/60
    If H >=60 and H < 180 then
        Value= M2
    If H>= 180 and H < 240 then
        Value = M1+(M2-M1)*(240-H)/60
    If H > 240 and H <= 360 then
        Value= M1
    Return
}

```

Thảo luận:

Tìm hiểu một số hàm chuyển đổi các mô hình màu trong các ngôn ngữ lập trình

Tìm hiểu các mô hình màu khác

1.3.2 Một số loại ảnh thông dụng

1.3.2.1 Ảnh màu

Ảnh màu được hiểu chung là ảnh thể hiện các đối tượng theo màu sắc của nó mà mắt thường cảm nhận được. Nói cách khác, ảnh màu được hiểu là một ảnh chứa một hoặc nhiều kênh màu xác định màu sắc tại các vị trí cụ thể của ảnh. Theo Thomas Young thì ảnh màu được tổ hợp từ 3 màu cơ bản là đỏ, xanh lục, xanh lam và thường thu nhận trên các dải băng tần khác nhau. Không gian màu RGB thường được dùng nhiều trong các thiết bị hiển thị điện tử so với các không gian màu khác như HSV, YcbCr. Một ảnh màu thường được lưu trữ trong bộ nhớ như là một bản đồ raster hay bản đồ pixel (bitmap) bởi một mảng hai chiều với mỗi phần tử là bộ ba giá trị màu tương ứng với mỗi kênh màu. Trong ảnh RGB, thông thường mỗi kênh màu dùng 8 bit để biểu diễn cho mỗi điểm ảnh, vậy một điểm ảnh màu cần 24 bit tương ứng với 3 kênh màu thành phần. Giá trị mỗi kênh màu được chia thành n mức màu khác nhau tương ứng từ 0 đến $n-1$, nếu 8 bit thì có 256 từ 0 đến 255. Mỗi giá trị thể hiện cường độ sáng của kênh màu tương ứng. Trong hệ màu RGB, việc lưu trữ ảnh màu theo từng kênh màu riêng biệt rất dễ dàng.

1.3.2.2 Ảnh đa mức xám

Trong kỹ thuật xử lý ảnh, ảnh đa mức xám (grey image) còn được gọi là ảnh đơn sắc (monochromatic). Ảnh đa mức xám dùng một kênh để thể hiện cường độ sáng của điểm ảnh. Mỗi điểm ảnh có một giá trị mức xám độ sáng từ màu đen (giá trị nhỏ nhất 0) đến màu trắng (giá trị lớn nhất). Như vậy khác với ảnh màu RGB, ảnh đa mức xám chỉ dùng một thành phần giá trị nên nếu dùng cùng độ bit để mã hóa cho mỗi thành phần thì ảnh đa mức xám có kích thước bằng 1/3 so với ảnh màu RGB.

Một số loại ảnh đa mức xám phổ biến phân theo kích thước bit dùng cho mỗi pixel:

- Ảnh 256 mức xám: loại ảnh này cần dùng 8 bit cho mỗi điểm ảnh, có giá trị nằm trong khoảng $[0, 255]$ tương ứng với biến đổi cường độ sáng từ đen qua trắng.

- Ảnh 8 mức xám: Loại ảnh này cần dùng 3 bit cho mỗi điểm ảnh, giá trị nằm trong khoảng $[0, 7]$. Như vậy, loại ảnh này có độ phân giải màu thấp hơn so với loại 8 bit. Giá trị điểm ảnh bằng 0 nghĩa là điểm ảnh đó tối (đen), giá trị điểm ảnh lớn nhất nghĩa là điểm ảnh đó trắng. Giá trị điểm ảnh càng lớn thì điểm ảnh đó càng sáng.

- Ảnh 2 mức xám: Mỗi pixel chỉ biểu diễn 2 mức cường độ sáng tương ứng với đen (0) và trắng (1). Để tránh nhầm lẫn ảnh đen trắng với ảnh đa mức xám, người ta thường gọi là ảnh nhị phân hay ảnh đen trắng nhị phân. Như vậy mỗi pixel chỉ cần dùng 1 bit để biểu diễn. Ảnh đen trắng thường được dùng để biểu diễn đối tượng và nền trong ảnh.

Trong một số trường hợp, màu sắc của hình ảnh không quan trọng và màu có thể nhạy cảm với các nguồn sáng mà chỉ cần biểu diễn hình dáng theo mức độ sáng tối của đối tượng trong hình ảnh thì ảnh đa mức xám được sử dụng để giảm dung lượng lưu trữ và giảm thiểu ảnh hưởng của nguồn sáng.

1.3.2.3 Ảnh đen trắng

Ảnh đen trắng là trường hợp đặc biệt của ảnh đa mức xám chỉ gồm 2 màu là đen và trắng, thường gọi là ảnh nhị phân. Ảnh nhị phân khá đơn giản, các phần tử ảnh có thể coi như các phần tử nhị phân. Ảnh nhị phân thường được dùng để biểu diễn trạng thái đối tượng, phân biệt đối tượng trong ảnh với nền, hoặc để biểu diễn các đường biên đối tượng, vùng ảnh.

1.3.2.4 Ảnh quang phổ

Ảnh quang phổ là trường hợp tổng quát của hình ảnh với các loại tín hiệu tương ứng với các loại bước sóng đặc biệt nào đó, bao gồm cả các loại bước sóng nhìn thấy (đối với ảnh số thông thường). Ảnh quang phổ có thể biểu diễn dưới dạng ảnh màu hoặc ảnh đa mức xám, ảnh nhị phân. Có nhiều ảnh loại quang phổ khác nhau như ảnh quang phổ X-ray, ảnh quang phổ hồng ngoại,...Ảnh phổ hồng ngoại là hình ảnh thu được từ tia hồng ngoại do bức xạ điện từ có bước sóng dài hơn ánh sáng nhìn thấy nhưng ngắn hơn tia bức xạ vi ba. Hồng ngoại có nghĩa là bước sóng ngoài mức đỏ, là bước sóng dài nhất trong ánh sáng nhìn thấy được.

Hiện nay loại camera hồng ngoại trở nên phổ biến, đặc biệt là các loại camera giám sát an ninh, camera trong các hệ thống tự động hóa. Camera hồng ngoại có khả năng thu bức xạ hồng ngoại để chuyển đổi qua ảnh hiển thị trong ngưỡng nhìn thấy bằng mắt thường. Do vậy, camera hồng ngoại được sử dụng nhiều trong các ứng dụng ban đêm và các điều kiện thiếu ánh sáng.

Thảo luận:

Tìm hiểu một số loại ảnh trong y tế

Tìm hiểu một số hàm chuyển đổi từ ảnh màu về ảnh đen trắng, ảnh đa cấp xám trong các ngôn ngữ lập trình

1.4. Cấu trúc dữ liệu ảnh

1.4.1. Cấu trúc ảnh Raster

Đây là cách biểu diễn ảnh thông dụng nhất hiện nay, ảnh được biểu diễn dưới dạng ma trận các điểm (điểm ảnh). Thường thu nhận qua các thiết bị như camera, scanner. Tuỳ theo yêu cầu thực tế mà mỗi điểm ảnh được biểu diễn qua 1 hay nhiều bit

Mô hình Raster thuận lợi cho hiển thị và in ấn. Ngày nay công nghệ phần cứng cung cấp những thiết bị thu nhận ảnh Raster phù hợp với tốc độ nhanh và chất lượng cao cho cả đầu vào và đầu ra.

Hầu hết các ảnh được lưu trữ theo các định dạng thông thường đều theo dạng cấu trúc raster và các biến thể nén như GIF, JPEG và PNG. Ảnh dạng raster thường có kích thước lớn hơn ảnh véc tơ.

Khác với ảnh vector, ảnh raster khi phóng to thường bị hiện tượng răng cưa, đối tượng không sắc nét. Nếu ảnh gốc kích thước nhỏ, khi phóng quá to so với ban đầu đối tượng thường không giữ lại được diện mạo, bị mờ.

1.4.2 Cấu trúc ảnh vector

Xét về mặt cấu trúc tổ chức, ảnh vector được tạo nên từ những yếu tố chính của hình học như điểm rời rạc, các đường thẳng, đường cong, đa giác và các vùng tương ứng với các đối tượng. Trên cơ sở đó vector được tạo thành dựa trên những biểu thức toán học (hoặc xấp xỉ), các vector này đi qua các điểm chính với mỗi điểm có một tọa độ x, y nhất định trên hệ trục tọa độ. Nhờ vậy, các điểm ảnh chi tiết trên đối tượng khi phóng sẽ được nội suy dựa vào những điểm chính và biểu thức toán học để tính giá trị điểm ảnh giữa các điểm chính.

Ưu điểm của ảnh vector là khi phóng to hoặc thu nhỏ ảnh không bị vỡ, đường biên giữa các vùng không bị hiện tượng răng cưa. Kích thước ảnh vector thường nhỏ, chứa đối tượng đơn giản. Ảnh vector được dùng nhiều trong các trường hợp thiết kế logo, icon avatar, ảnh nghệ thuật vector.

Hạn chế của ảnh vector là hình ảnh hiển thị thường không “tự nhiên”, có sự chuyển màu và không sắc nét với ảnh ngoại cảnh, phân phối màu phức tạp.

Thông tin về ảnh

- Chiều rộng: số pixel/ dòng (số byte/dòng)
- Chiều cao: số dòng

- độ phân giải: số pixel/inch
- số bit/ pixel

số bit/ pixel ≤ 8 phải có pallete màu (bảng màu)

> 8 dùng màu trực tiếp

Pal:palette

Pal[index].R, Pal[index].G, Pal[index].B,

Nếu số bit/pixel > 1 ta có ảnh cấp xám (gray level image), nếu $= 1$ ta có ảnh nhị phân với 2 cấp xám ví dụ như đen và trắng.

16 bit/pixel

5 6 5

R	G	B
---	---	---

24 bit/pixel, mỗi phần màu là một byte

8	8	8
R	G	B

Ảnh có 3 giá trị $R=G=B$ là ảnh đa cấp xám không cần lưu giá trị palatte vì bản thân nó là giá trị điểm ảnh (pallete đúng bằng chỉ số giá trị điểm ảnh).

Thảo luận: Cho biết các thuộc tính của một bức ảnh trên máy tính

1.5. Một số định dạng ảnh phổ biến

1.5.1 Định dạng ảnh TIFF

Định dạng ảnh TIFF (tagged image format file) được nghiên cứu và giới thiệu vào năm 1986, là một định dạng file ảnh chất lượng cao và được sử dụng nhiều trong các ứng dụng thu nhận ảnh từ máy scan. Chuẩn định dạng TIFF là một trong những tiêu chuẩn quan trọng, được sử dụng nhiều trong ngành công nghiệp in ấn và xuất bản. File ảnh dạng TIFF thường có kích thước lớn hơn nhiều so với các file ảnh nén theo chuẩn JPEG. Định dạng TIFF lưu trữ dữ liệu hình ảnh dạng nén hoặc không nén và có thể sử dụng các kỹ thuật nén không mất dữ liệu.

Đặc điểm của ảnh theo định dạng TIFF là thường không bị mất dữ liệu hình ảnh khi lưu trữ ra thiết bị nhớ và đọc lại để xử lý trong máy tính, thường được sử dụng để

biểu diễn hình ảnh có màu sắc phức tạp. Ảnh định dạng TIFF sử dụng trong các trường hợp đòi hỏi chất lượng cao như hình ảnh in ấn, phân tích mẫu.

1.5.2 Định dạng ảnh GIF

Định dạng ảnh GIF (graphics interchange format) được phát triển từ năm 1987, thường được dùng trong biểu diễn và truyền hình ảnh trong môi trường Web. Ảnh định dạng GIF thường biểu diễn hình ảnh thành các frame để tạo ảnh chuyển động. Với mục đích tạo ra định dạng trao đổi hình ảnh nên các file ảnh theo định dạng GIF thường có kích thước nhỏ, chất lượng hình ảnh vừa phải, đáp ứng được trong môi trường mạng. Khác với JPEG, GIF sử dụng thuật toán nén ít mất thông tin mà không làm giảm chất lượng hình ảnh sau khi nén. Trong kỹ thuật nén ảnh theo GIF, dữ liệu lưu bằng cách sử dụng màu chỉ mục (index), mỗi hình ảnh có thể bao gồm 256 màu.

Một trong những ưu điểm của GIF là nén theo chuẩn Lossless nên ảnh thường không bị mất dữ liệu khi nén, hình ảnh dạng GIF được tự động nhận biết trên hầu hết các trình duyệt web. Vì chuẩn GIF lưu trữ dữ liệu theo bảng chỉ mục nên nó thường được dùng để tạo các khung nhìn khác nhau tạo nên hiệu ứng chuyển động, vì hình ảnh giữa các frame có mức độ tương tự cao nên sẽ tiết kiệm được không gian nhớ so với video thông thường. Ảnh GIF sử dụng tốt đối với các trường hợp biểu diễn hình ảnh đơn giản như những bản vẽ chỉ có nét, bảng màu sắc và những minh họa đơn giản, tạo những hình ảnh động, hình ảnh Web không có quá nhiều màu sắc, những ảnh avatar có kích thước nhỏ.

1.5.3 Định dạng hình ảnh JPG

Định dạng JPG được đề xuất năm 1992 trong công bố của tác giả Haines. Định dạng JPG được gắn liền với chuẩn nén ảnh JPEG (Joint photographic expert group) và lưu trong máy tính theo file JPG. Định dạng JPG là một trong những phương pháp sử dụng phổ biến nhất hiện nay cho các file ảnh kỹ thuật số và xử lý tính toán trong máy tính. Định dạng JPG gắn liền với thuật toán nén mất thông tin, tức là khi nén dữ liệu để lưu trữ, thông tin sẽ bị mất trong quá trình nén và giải nén. Do đó, chất lượng hình ảnh sẽ bị giảm so với ảnh ban đầu. Tuy nhiên, với phương pháp nén mất thông tin thì kích thước file lưu trữ của ảnh cũng giảm đáng kể. Phương pháp nén JPEG thường được dùng để nén ảnh số có mất mát thông tin. Các file ảnh dùng nén theo chuẩn JPEG thường có tên file mở rộng là *.jpg, *.jpeg, *.jfif hay *.jpe.

Các ảnh sử dụng phương pháp nén JPEG tương thích với hầu hết các trình duyệt web hiện nay. Ảnh JPG sử dụng tốt và hiệu quả đối với các loại ảnh tĩnh, ảnh có màu sắc phức tạp, ảnh đa mức xám, ảnh ngoại cảnh và ảnh chân dung.

1.5.4 Định dạng ảnh BMP

BMP là loại định dạng bitmap, được phát triển vào năm 1994. BMP là loại định dạng và lưu trữ file ảnh đồ họa dạng lưới (raster) được sử dụng để lưu trữ ảnh số dạng thô. File ảnh dạng BMP thường có kích thước lớn và dữ liệu không nén do vậy cũng không mất thông tin trong quá trình lưu file và đọc ảnh từ file. Dữ liệu hình ảnh BMP độc lập với các thiết bị hiển thị như Graphics adapter, đặc biệt trên các ứng dụng chạy trong môi trường Microsoft Windows và hệ điều hành OS/2.

Định dạng BMP có ưu điểm là không làm mất thông tin của ảnh đang xử lý, nên nó phù hợp cho việc in ấn, chỉnh sửa hình ảnh. Mặt khác, vì ảnh không nén nên file ảnh BMP được đọc dễ dàng bằng các chương trình phần mềm dùng chung với những thuật toán đơn giản. Tuy nhiên, ảnh không hỗ trợ nén cũng ảnh hưởng không tốt cho việc lưu trữ vì dung lượng file thường lớn hơn các loại định dạng khác.

1.5.5 Định dạng ảnh PNG

PNG (Portable Network Graphics) được đề xuất năm 1996 là một định dạng file đồ họa dạng raster. PNG hỗ trợ nén dữ liệu không bị mất thông tin. Định dạng PNG được xem là một dạng cải tiến và thay thế cho GIF trong môi trường ảnh vector và được sử dụng nhiều trên internet. Chuẩn định dạng PNG thường sử dụng hai dạng khác nhau là PNG-8 và PNG-24. Trong trường hợp ảnh có màu sắc phức tạp, không phân bố theo dạng vector thì PNG có dung lượng lớn hơn JPEG.

Ưu điểm của định dạng PNG là hình ảnh các đối tượng không bị cạnh răng cưa khi phóng to ảnh, điểm ảnh được biểu diễn dạng vector. Ảnh định dạng PNG được nén theo chuẩn không mất thông tin do vậy khi giải nén ảnh vẫn giữ nguyên được chất lượng ban đầu trước khi nén.

Ảnh dạng PNG thích hợp với các loại hình ảnh chứa đối tượng phân phối màu đơn giản, tuân theo quy luật như văn bản, các loại hình vẽ. Với các loại hình ảnh mà nền trong suốt hoặc có thể được thiết lập giữa mờ đục lưu trữ theo định dạng PNG cho ảnh chất lượng cao với kích thước file nhỏ. Bên cạnh đó, nó cũng được dùng trong quá trình chỉnh sửa hình ảnh nhằm không làm mất thông tin của ảnh đang xử lý. Ngoài ra định dạng PNG sử dụng tốt cho các hình ảnh web/blog, những mảng màu phẳng, thiết kế logo, hình ảnh có nền trong suốt hoặc bán trong suốt.

Thảo luận: Tìm hiểu định dạng ảnh khác ngoài 5 định dạng trình bày ở trên

1.6. Ứng dụng của xử lý ảnh

Ngày nay, với sự hỗ trợ của các hệ thống tính toán lớn, các thuật toán tiên tiến ra đời cho phép máy tính có thể hiểu biết và quyết định tốt hơn con người trong một số lĩnh vực nhất định. Ví dụ hệ thống xử lý ảnh bằng mạng nơ ron nhân tạo học sâu có thể nhận dạng, phân loại các kiểu đối tượng khác nhau tốt hơn và nhanh hơn con người. Xử lý ảnh có rất nhiều ứng dụng trong hầu hết các lĩnh vực của đời sống xã hội dân sự, an ninh quốc phòng, hàng không vũ trụ như:

Lĩnh vực quân sự, an ninh, quốc phòng: Tự động nhận dạng, phát hiện tội phạm, theo vết và truy tìm thủ phạm thông qua hình ảnh hiện trường phạm tội và các vấn đề hỗ trợ do tìm tội phạm qua hệ thống giám sát an ninh toàn cầu, quốc gia.

Trong lĩnh vực y tế: Phân tích hình ảnh, chuẩn đoán bệnh qua các loại hình ảnh tia Gamma, X-quang, scan PET/CT (cắt lớp phát xạ), ảnh cực tím và đặc biệt với sự thành công của kỹ thuật học sâu đã giúp cho các chuẩn đoán hình ảnh y học đạt kết quả cao.

Trong lĩnh vực viễn thám, vũ trụ: thám hiểm vũ trụ, do thám, phân tích và phát hiện vật thể trong vũ trụ.

Trong lĩnh vực giao thông, dân sự: Các hệ thống khôi phục ảnh, chỉnh sửa, điều chỉnh độ phân giải, xử lý màu sắc, mã hóa và truyền tin, nhận dạng và phân loại hành động trong các hệ thống giám sát an ninh; hệ thống xe không người lái, giám sát sản phẩm sản xuất công nghiệp, robot phục vụ dân sự, giám sát bãi xe thông minh, kiểm soát-điều khiển giao thông thông minh.

Hệ thống nhận dạng chữ, form

Image→Table + Text + Image : nhận dạng được chữ, ảnh, bảng

Image→form

Nhận dạng chữ OCR (Optical character Recognition)

Nhận dạng nhãn OMR (Optical Mark Recognition)

-Nhận dạng tiếng Anh

Omnipage 12.0 của hãng ScanSoft

Fine Reader 6.0 của hãng ABBRY.

-Nhận dạng tiếng việt

VN Docr 4.0 của Viện CNTT.

Image Scan của công ty Tcapro

Hoạt hình: biến đổi hình học, bóp méo hình học (morph)

Dự báo thời tiết, cháy rừng, lũ lụt, sâu bệnh, khoáng sản.

Nhận dạng vân tay, Water Marking

Dấu thông tin trong ảnh, Nhận dạng mã vạch

Các robot tự động phục vụ.

Thảo luận: Trình bày các ứng dụng xử lý ảnh mà em gặp trong thực tế

Câu hỏi

1. Hãy phân biệt ảnh màu, ảnh đa mức xám và ảnh nhị phân?
2. Trình bày các độ phân giải thường dùng?
3. Trình bày cấu trúc của hệ thống xử lý ảnh?
4. Trình bày các thiết bị thu nhận ảnh hiện nay?
5. Trình bày các loại nhiễu thường gặp trên ảnh?
6. Đọc vào một ảnh định dạng RGB, chuyển ảnh về ảnh đa mức xám
7. Đọc vào một ảnh định dạng RGB, chuyển ảnh về ảnh CMY
8. Đọc vào một ảnh định dạng RGB, chuyển ảnh về ảnh HSV
9. Đọc vào một ảnh có định dạng JPG, thay đổi ảnh về kích thước 256×256 pixels, sau đó lưu ra các file ảnh theo các định dạng GIF, PNG, BMP và TIFF. Dùng phần mềm Paint thay đổi kích thước ảnh đã lưu thành 1.000×1.000 pixels. Đọc ảnh và hiển thị các ảnh theo các định dạng đã tạo và cho nhận xét về chất lượng ảnh.
10. Dùng phần mềm Paint để tạo ảnh kích thước 480×480 pixels có chứa một số hình đơn giản có sẵn trong Paint, tô đầy một màu cho mỗi đối tượng. Lưu ảnh thành các định dạng JPG, PNG, GIF, BMP và TIFF. Hãy nhận xét về dung lượng lưu trữ theo mỗi định dạng. Đọc các ảnh và hiển thị, đánh giá chất lượng hình ảnh tương ứng.
11. Dùng phần mềm Paint để thay đổi kích thước các ảnh ở câu 1.8 thành 1.024×1.024 pixels. Cho biết dung lượng lưu trữ và chất lượng các ảnh theo các định dạng tương ứng.

12. So sánh giữa các định dạng ảnh phổ biến

13. Đọc vào một ảnh có định dạng JPG, thay đổi ảnh về kích thước 256×256 pixels, sau đó lưu ra các file ảnh theo các định dạng GIF, PNG, BMP và TIFF. Dùng phần mềm Paint thay đổi kích thước ảnh đã lưu thành 1.000×1.000 pixels. Đọc ảnh và hiển thị các ảnh theo các định dạng đã tạo và cho nhận xét về chất lượng ảnh.

14. Dùng phần mềm Paint để tạo ảnh kích thước 480×480 pixels có chứa một số hình đơn giản có sẵn trong Paint, tô đầy một màu cho mỗi đối tượng. Lưu ảnh thành các định dạng JPG, PNG, GIF, BMP và TIFF. Hãy nhận xét về dung lượng lưu trữ theo mỗi định dạng. Đọc các ảnh và hiển thị, đánh giá chất lượng hình ảnh tương ứng.

15. Dùng phần mềm Paint để thay đổi kích thước các ảnh ở câu 1.8 thành 1.024×1.024 pixels. Cho biết dung lượng lưu trữ và chất lượng các ảnh theo các định dạng tương ứng

CHƯƠNG 2. CÁC PHÉP BIẾN ĐỔI KHÔNG PHỤ THUỘC KHÔNG GIAN VÀ CÁC PHÉP TOÁN TRÊN ĐA ẢNH

Nội dung chính của chương:

Chương 2 trình bày các phép biến đổi không phụ thuộc không gian và các phép toán trên đa ảnh

Mục tiêu cần đạt được của chương:

Sinh viên hiểu được khái niệm cơ bản về histogram, các phép biến đổi phụ thuộc không gian và các phép toán trên đa ảnh, từ đó có thể lập trình được các thuật toán.

Bài 2: Các phép biến đổi không phụ thuộc không gian (Số tiết: 5 tiết)

2.1. Giới thiệu

Các phép toán không phụ thuộc không gian là các phép toán toàn cục không phụ thuộc vào vị trí của điểm ảnh. Hay còn gọi là toán tử xử lý điểm ảnh dùng các ánh xạ nhằm biến đổi giá trị của một điểm ảnh chỉ dựa vào giá trị của chính nó mà không quan tâm đến giá trị của điểm ảnh khác.

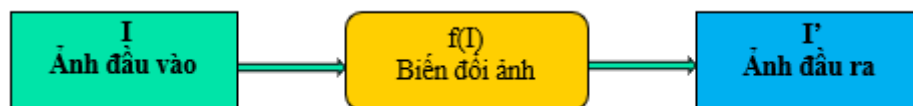
Một cách toán học ánh xạ đó được định nghĩa:

$$I'(m,n) = f(I(m,n))$$

$I(m,n)$ là giá trị cường độ sáng tại tọa độ (m,n)

$I'(m,n)$ là giá trị cường độ sáng thu được sau phép biến đổi

f là hàm biến đổi.



Hình 2.1 Biến đổi ảnh

Một số phép toán

- + Tăng giảm độ sáng
- + Hiệu chỉnh độ tương phản
- + Phép toán thống kê
- + Tách theo ngưỡng và chồng nhiều ảnh

...

2.2- Các phép toán không phụ thuộc không gian

Định nghĩa: (Histogram) Lược đồ mức xám

Lược đồ mức xám là một biểu đồ cung cấp tần suất xuất hiện của mỗi mức xám trong ảnh

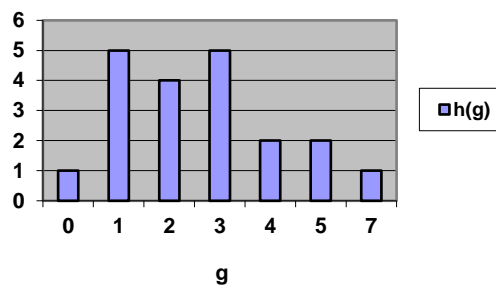
Tần suất của cấp xám g là số điểm ảnh có mức xám g trong ảnh

$h(g)$: số điểm ảnh có giá trị bằng g

$$h(g) = \# \{P \mid I(P) = g\}$$

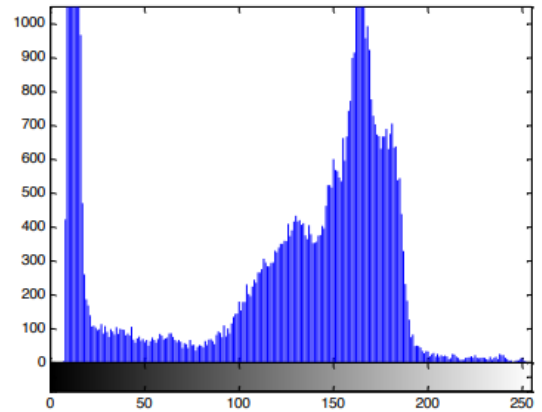
giả sử ta có ảnh $I = \begin{pmatrix} 1 & 3 & 2 & 7 & 5 \\ 1 & 4 & 3 & 2 & 5 \\ 1 & 1 & 2 & 3 & 3 \\ 0 & 1 & 2 & 3 & 4 \end{pmatrix}$

g	0	1	2	3	4	5	7
$h(g)$	1	5	4	5	2	2	1



Hình 2.2 Histogram của ảnh I

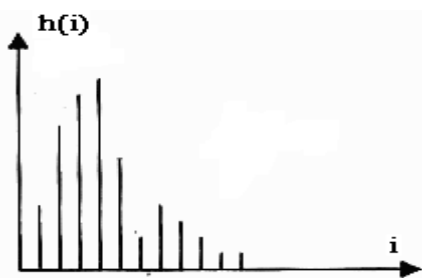
Nếu phát triển đến 256 màu các điểm sẽ liên kề nhau, tạo thành đường cong.



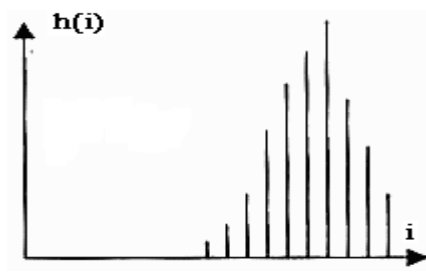
Hình 2.3 Histogram của một ảnh thực tế

Đối với ảnh màu, thì sẽ có 3 lược đồ xám tương ứng với 3 màu Red, Green, Blue.

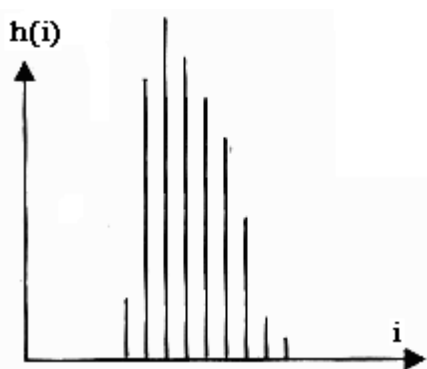
Lược đồ xám cung cấp rất nhiều thông tin về phân bố mức xám của ảnh.



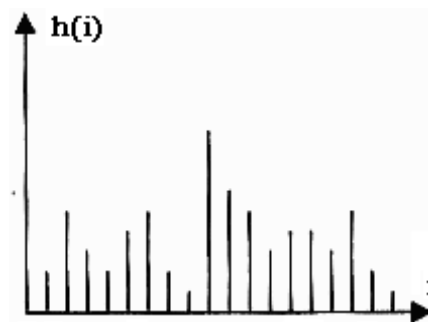
ảnh tối



ảnh sáng



ảnh có độ tương phản thấp



ảnh có độ tương phản cao

Hình 2.4 Các histogram của các ảnh có tính chất khác nhau

Thuật toán xây dựng lược đồ xám:

H là bảng chứa lược đồ xám (là vecto có N phần tử)

Khởi tạo: đặt tất cả các phần tử của bảng là 0

Tạo bảng: với mỗi điểm ảnh $I(x,y)$ tính $H[I(x,y)]=H[I(x,y)]+1$

- **Kỹ thuật 1:** Tăng giảm độ sáng của ảnh

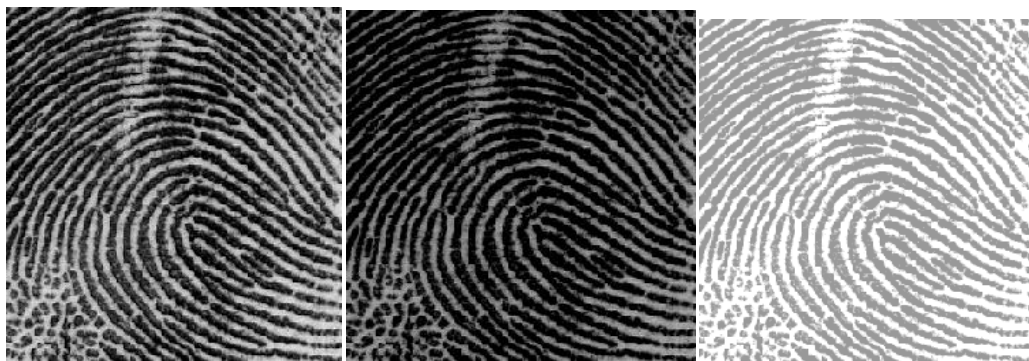
Cộng vào mọi điểm ảnh một giá trị C

$$I'(x, y) = I(x, y) + C$$

$C > 0$ ảnh sáng lên

$C < 0$ ảnh tối đi.

Ví dụ: Ảnh đã bị tối đi và sáng lên



Ảnh gốc

Ảnh tối

Ảnh sáng

Hình 2.5 Tăng giảm độ sáng của ảnh

Chú ý: khi cộng vượt qua 255 thì quay lại từ 0

Đối với ảnh màu trong hệ không gian màu RGB, việc thay đổi mức sáng có thể thực hiện thay đổi trên từng kênh màu R-G-B. Đối với hệ không gian màu HSV thì V là thành phần biểu diễn mức sáng, việc thay đổi mức sáng tương ứng với thay đổi thành phần V tương ứng.

***Kỹ thuật 2:** Tăng giảm độ tương phản

Độ tương phản được hiểu là mức độ chênh lệch về độ sáng giữa các đối tượng lân cận nhau, hoặc với các vùng nền trong ảnh. Nếu mức độ chênh lệch về độ sáng của đối tượng với vùng nền càng thấp thì ta nói ảnh đó có độ tương phản thấp và ngược lại độ chênh lệch càng cao có nghĩa là độ tương phản ảnh càng cao. Một ảnh có độ tương phản thấp thì khi quan sát thấy ảnh ít có sự sắc nét giữa các thành phần sáng-tối với nhau

$$I'(x,y) = \alpha I(x,y) + \beta$$

α, β là các hằng số

$\alpha > 1$ độ tương phản tăng

$\alpha < 1$ độ tương phản giảm



Ảnh gốc

Tăng độ tương phản

Giảm độ tương phản

Hình 2.6 Tăng giảm độ tương phản

* Kỹ thuật 3: Kéo dẫn độ tương phản

Kéo dẫn histogram từ [0, 255]

$$output(x) = \begin{cases} 0 & x < low \\ \frac{255 * (x - low)}{high - low} & low \leq x \leq high \\ 255 & x > high \end{cases}$$

low và high là giá trị cấp xám cao nhất và thấp nhất trong ảnh

* Kỹ thuật 4: Hiệu chỉnh gamma

$$I'(x, y) = (I(x, y))^\gamma = e^{\gamma \ln(x, y)}$$

* Kỹ thuật 5: Tách ngưỡng

$\theta > 0$: ngưỡng

$$I'(x, y) = \begin{cases} \text{Max} & \text{nếu } I(x, y) \geq \theta \\ \text{Min} & \text{nếu } I(x, y) < \theta \end{cases}$$

chuyển ảnh đa cấp xám về ảnh nhị phân. Chia ảnh thành 2 phần nền và vùng. Nếu $\text{Max}=1$, $\text{Min}=0$ thì ta có ảnh đen trắng

tồn tại sai số

sai số: $g \in \text{nền}$, xếp g vào vùng ảnh (dính, bắt nhầm)

$g \notin \text{nền}$, xếp g vào nền (đứt nét, bỏ sót)

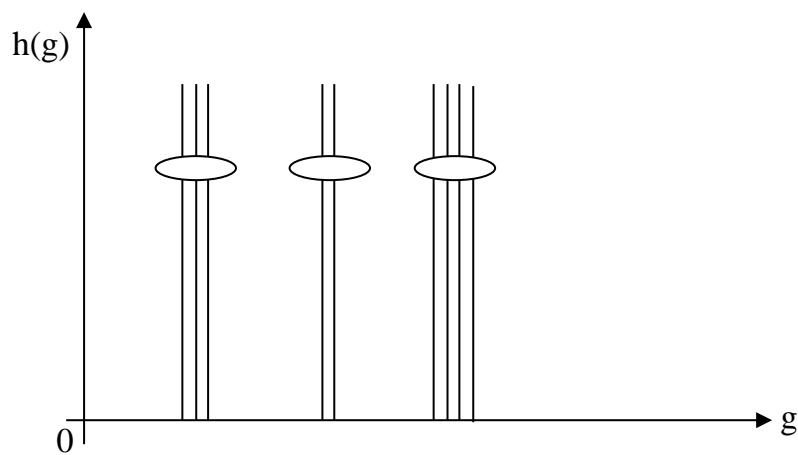
Nhược điểm: Có thể xảy ra các sai sót nền thành ảnh hoặc ảnh thành nền làm cho ảnh bị đứt nét.

* **Kỹ thuật 6:** Bó cụm -bunching (tách theo nhiều ngưỡng)

Mục đích: giảm bớt cấp xám của ảnh bằng cách nhóm lại các cấp xám gần nhau thành cấp xám mới. Trong trường hợp hai nhóm thì chính là tách ngưỡng

Cho một tập các ngưỡng theo thứ tự tăng dần $\theta_1, \theta_2, \dots, \theta_k$

$$I'(x, y) = \begin{cases} \theta_i & \text{nếu } \theta_{i-1} \leq I(x, y) \leq \theta_i \\ 0 & \text{nếu } I(x, y) < \theta_1 \\ \text{Max} & \text{nếu } I(x, y) > \theta_k \end{cases}$$



Hình 2.7 Bó cụm

Thông thường có nhiều nhóm với kích thước khác nhau, để tổng quát khi biến đổi người ta sẽ lấy cùng kích thước.

Cách làm:

Chia các nhóm đều nhau có cùng kích thước là bunch size

$$I'(x, y) = \frac{I(x, y)}{bunch_size} * bunch_size$$

Đây là phép chia lấy phần nguyên

Ví dụ:

$$I = \begin{pmatrix} 1 & 2 & 4 & 6 & 7 \\ 2 & 1 & 3 & 4 & 5 \\ 7 & 2 & 6 & 9 & 1 \\ 4 & 1 & 2 & 1 & 2 \end{pmatrix}$$

bunch_size=3, khi đó ta thu được

$$I'(1,1) = \frac{1}{3} * 3 = 0$$

$$I' = \begin{pmatrix} 0 & 0 & 3 & 6 & 6 \\ 0 & 0 & 3 & 3 & 3 \\ 6 & 0 & 6 & 9 & 0 \\ 3 & 0 & 0 & 0 & 0 \end{pmatrix}$$

[0, 1, 2] [3, 4, 5] [6, 7, 8] [9, 10, 11]

0 3 6 9

*Kỹ thuật 7: Tách ngưỡng tự động

Chia dãy cấp xám thành 2 phần: tìm θ chia ảnh thành 2 phần.

Kỹ thuật nhằm tìm ra ngưỡng θ dựa vào Histogram một cách tự động theo nguyên lý vật lý. Vật thể tách làm 2 phần nếu tổng độ lệch trong từng phần là tối thiểu.

- *Vật lý: Khi tách ra 2 phần thì xuất hiện lực hấp dẫn.*
- *Cho $\theta = [1, 255]$ thì cái nào có lực hút yếu nhất bị tách ra, do đó tìm vị trí có lực hút yếu nhất.*
- *Mômen quán tính: vị trí làm cho momen quay nhỏ nhất.*
- *Hàm: giá trị cực trị địa phương*
- *Thống kê: vị trí mà momen bậc 2, bậc 3 histogram bằng nhau.*
- *Thông tin: entropy cực đại, $\sum p_i \log p_i \rightarrow \text{Max}$ ($E \text{ nền} + E \text{ vùng} \rightarrow \text{max}$, số các điểm là cực đại cho 2 phần)*
- *Độ lệch: $\sigma_1 + \sigma_2 \rightarrow \text{min}$*

σ_1 độ lệch các pixel nền

σ_2 độ lệch các pixel vùng

- Đường cong

Điểm chia là điểm lồi hoặc lõm của đường cong.

Gọi $t(g)$ số điểm ảnh có giá trị $\leq g$

G là số mức xám được xét kể cả khuyết thiếu.

P là số điểm ảnh được xét.

$m(g)$ giá trị trung bình các điểm ảnh $\leq g$

$$m(g) = \left(\sum_{i=0}^g i h(i) \right) / t(g)$$

$i.h(i)$ là momen quán tính của điểm i

$m(g)$ là moment quán tính trung bình của các điểm có mức xám $\leq g$.

$$A(g) = t(g) / (P - t(g))$$

$$B(g) = [m(g) - m(G-1)]^2$$

$$F(g) = A(g) * B(g) - 1.$$

Tìm θ sao cho $f(\theta) = \max \{f(g), \text{ với } g \leq G-1\}$

Ví dụ: $G=9$ hãy tách ngưỡng tự động ảnh này

1	4	2	8	5	7
4	2	8	5	7	1
0	8	5	7	1	4
0	0	7	1	4	2
0	0	0	4	2	8
0	0	0	0	8	5
0	0	0	0	0	7

$$P=42$$

g	h(g)	t(g)	g*h(g)	sum(i*h(i))	m(g)	A(g)	B(g)	f(g)
0	15	15	0	0	0	0.56	9.14	4.08
1	4	19	4	4	0.21	0.83	7.91	5.54
2	4	23	8	12	0.52	1.21	6.26	6.58
4	5	28	20	32	1.14	2	3.53	6.08
5	4	32	20	52	1.63	3.2	1.96	5.26
7	5	37	35	87	2.35	7.4	0.45	2.35
8	5	42	40	127	3.02	####	0	####

f(2) cho giá trị lớn nhất vì vậy ngưỡng tự động là 2

Thảo luận: Cho ảnh I, G=5, P=20 hãy tách ngưỡng tự động ảnh sau

$$I = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ 0 & 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

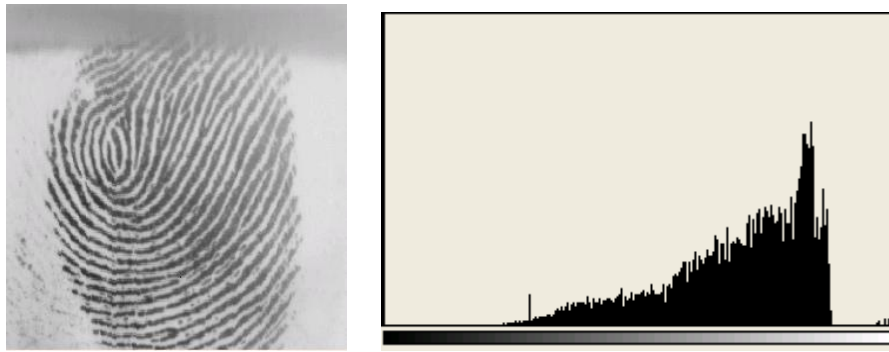
*** Kỹ thuật 8: Cân bằng histogram**

Kỹ thuật cân bằng lược đồ ảnh thường được sử dụng để tăng cường độ tương phản ảnh bằng cách giãn đều giá trị cường độ sáng của các điểm ảnh phân bố trên vùng rộng hơn mà vẫn giữ được nội dung chính của bức ảnh.

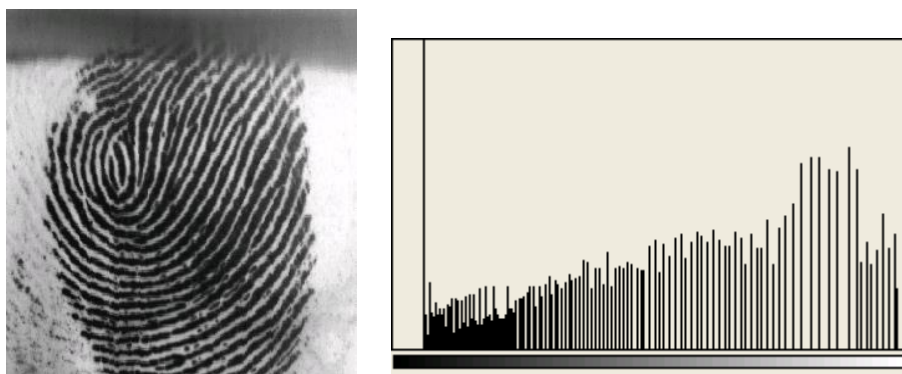
Phương pháp cân bằng lược đồ ảnh có ưu điểm là đơn giản, tính toán dễ dàng, đồng thời cho phép phục hồi lại trạng thái ảnh ban đầu khi cần thiết. Tuy nhiên, hạn chế quan trọng của nó là dễ dàng làm tăng cường nhiều trong ảnh, đồng thời làm giảm các chi tiết quan trọng của hình ảnh.

Cân bằng lý tưởng: với mọi g, g': g ≠ g' thì h(g)=h(g').

Trong trường hợp không cân bằng, chia Histogram thành các đoạn và các vùng chia là xấp xỉ bằng nhau (về số điểm trong Histogram)



Hình 2.8 Ảnh gốc và histogram



Hình 2.9 Ảnh sau khi cân bằng và histogram

+ Phương pháp cân bằng toàn cục là công việc điều chỉnh, làm cân bằng lại sự phân bố các giá trị độ sáng điểm ảnh. Kỹ thuật cân bằng toàn cục được sử dụng trong việc xử lý ảnh từ vệ tinh, chụp X-quang, ảnh đo nhiệt bức xạ....

Thuật toán

$T(g)$: là số điểm ảnh $\leq g$

$T(g) = \text{Sum}(h(i); 0 \leq i \leq g)$

M, N là kích thước của ảnh I

New level: là số mức xám cần cân bằng.

$F(g) = \max(0, \text{round}((t(g)/(M*N))*\text{new_level}-1))$

đặt $TB = (M*N)/\text{new_level}$

Khi đó $f(g) = \max(0, \text{round}(t(g)/TB-1))$

Ví dụ 1: Cho I và new_level=6, hãy cân bằng histogram

1	4	2	8	5	7	g	h(g)	t(g)	f(g)
4	2	8	5	7	1	0	15	15	1
0	8	5	7	1	4	1	4	19	2
0	0	7	1	4	2	2	4	23	2
0	0	0	4	2	8	4	5	28	3
0	0	0	0	8	5	5	4	32	4
0	0	0	0	0	7	7	5	37	4
						8	5	42	5

Thảo luận: Cho I và new_level=4, hãy cân bằng histogram

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

+ Phương pháp cân bằng cục bộ là quan tâm đến từng phần của ảnh thay vì thực hiện cân bằng trên toàn bộ bức ảnh. Cách tiếp cận cân bằng cục bộ giúp giải quyết vấn đề trong ảnh có các vùng quá tối hoặc quá sáng được cân bằng tốt hơn. Một trong những giới hạn chính của phương pháp cân bằng cục bộ là phải thực hiện lặp đi lặp lại việc tính toán trên các phân vùng nhỏ nên dẫn đến thời gian tính toán rất lớn, không phù hợp xử lý thời gian thực.

Kỹ thuật 9: Ảnh âm bản

$$I'(x,y) = L - I(x,y) \quad \forall (x,y)$$

$$I(x,y) \in [0, L]$$

Ảnh âm bản có màu sắc là các màu bù của ảnh gốc.

Kỹ thuật 10: Sử dụng bảng tra

Look up table (LUT)

LUT đơn giản là một mảng sử dụng các giá trị điểm ảnh hiện tại làm chỉ mục tới các giá trị mới được lưu trong bảng. Khi đó tránh được các tính toán lặp cần thiết. Ví dụ:

Ảnh đầu vào					LUT		Ảnh đầu ra				
7	7	5	4	2	0	0	5	5	3	2	1
7	6	4	3	0	1	0	5	4	2	1	0
7	6	4	3	1	2	1	5	4	2	1	0
6	6	4	2	0	3	1	4	4	2	1	0
5	5	3	1	0	4	2	3	3	1	0	0
					5	3					
					6	4					
					7	5					

Câu hỏi

Cho ảnh I, sau

	1	3	5	7	8	9
	6	5	2	3	5	2
I=	4	8	9	6	2	7
	5	3	0	2	1	7
	7	8	4	9	0	6
	8	4	9	3	5	2

1. Vẽ histogram của ảnh I
2. Tăng độ sáng của ảnh với $C=2$
3. Giảm độ của ảnh $C=-2$
4. Cho $G=10$, hãy tách ngưỡng tự động ảnh I
5. Thực hiện kỹ thuật bảng tra ảnh I, cho biết LUT

LUT

0	0
1	0

2	1
3	1
4	2
5	2
6	2
7	3
8	4
9	4

6. Cho biết ảnh âm bản của ảnh I, cho biết $L=10$
7. Cho $\text{new_level}=4$, hãy thực hiện cân bằng ảnh I trên
8. Cho $\text{bunch_size}=3$, hãy thực hiện phương pháp bó cụm ảnh I
9. Cho ngưỡng bằng 5, hãy thực hiện tách ngưỡng đưa về ảnh đen trắng
10. Tìm hiểu các thủ tục thực hiện các phép biến đổi không phụ thuộc vào không gian.

	0	0	0	0	0	0		8	8	8	8	8	8
	0	0	0	0	0	0		8	8	8	8	8	8
I1-I2	0	0	0	0	0	0	8- I1-I2 =	8	8	8	8	8	8
	0	0	0	0	1	1		8	8	8	8	7	7
	0	0	0	1	1	1		8	8	8	7	7	7
	0	0	1	1	1	1		8	8	7	7	7	7

*Kỹ thuật 2: Kỹ thuật cộng ảnh

Giả sử ta có 2 ảnh I_1 , I_2 có cùng kích thước và thuộc tính. Kỹ thuật cộng ảnh nhằm tạo ra ảnh mới lai ghép giữa I_1 và I_2 cũng có kích thước và thuộc tính giống 2 ảnh ban đầu.

$$I_\alpha [i, j] = (1-\alpha)I_1 [i, j] + \alpha I_2 [i, j]$$

$$\alpha \in [0, 1]$$

Ứng dụng: trong hoạt hình, tái tạo hình ảnh.

*Kỹ thuật 3: Lọc trung vị đa ảnh

Định nghĩa 1:

Cho dãy $x_1; x_2, \dots; x_n$ đơn điệu tăng (giảm). Nếu n lẻ thì trung vị của dãy là $x_{[(n+1)/2]}$

Nếu n chẵn thì trung vị của dãy là $(x_{n/2} + x_{[n/2]+1})/2$

Hay chúng ta có thể có định nghĩa khác

Định nghĩa 2:

X là trung vị của dãy x_1, x_2, \dots, x_n

$$\#\{k | x \geq x_k\} = \#\{k | x \leq x_k\}$$

Ví dụ: Trung vị của dãy 1, 4, 5, 2, 6, 8, 7 là 5.

Mệnh đề 1

Cho dãy x_1, x_2, \dots, x_n

$$\sum_{k=1}^n |x - x_k| \rightarrow \min \text{ tại trung vị}$$

Chứng minh

+ Xét trường hợp n chẵn

$$\text{Đặt } M = \frac{n}{2}$$

Ta có:

$$\begin{aligned} \sum_{i=1}^n |x - x_i| &= \sum_{i=1}^M |x - x_i| + \sum_{i=1}^M |x - x_{M+i}| \\ &= \sum_{i=1}^M (|x - x_i| + |x_{M+i} - x|) \geq \sum_{i=1}^M |x_{M+i} - x_i| \\ &= \sum_{i=1}^M [(x_{M+1} - x_M) + (x_M - x_i)] \\ &= \sum_{i=1}^M |x_{M+i} - \text{Med}(\{x_i\})| + \sum_{i=1}^M |x_i - \text{Med}(\{x_i\})| \\ &= \sum_{i=1}^n |x_i - \text{Med}(\{x_i\})| \end{aligned}$$

+ Nếu n lẻ:

Bổ sung thêm phần tử $\text{Med}(\{x_i\})$ vào dãy. Theo trường hợp n chẵn ta có:

$$\begin{aligned} \sum_{i=1}^n |x - x_i| + |\text{Med}(\{x_i\}) - \text{Med}(\{x_i\})| &\rightarrow \min \text{ tại } \text{Med}(\{x_n\}) \\ \sum_{i=1}^n |x - x_i| &\rightarrow \min \text{ tại } \text{Med}(\{x_n\}) \end{aligned}$$

Cho trước dãy ảnh I_1, I_2, \dots, I_n . Ảnh trung vị của dãy ảnh là I_{Tv} trong đó mỗi $I_{Tv}(P)$ là trung vị của dãy $I_1(P), I_2(P), \dots, I_n(P)$.

Ví

dụ

$$\begin{array}{ccc}
 1 & 2 & 3 \\
 I_1 = & 4 & 5 & 6 \\
 7 & 8 & 9
 \end{array}
 \quad
 \begin{array}{ccc}
 4 & 5 & 6 \\
 I_2 = & 1 & 2 & 3 \\
 7 & 8 & 9
 \end{array}
 \quad
 \begin{array}{ccc}
 1 & 2 & 3 \\
 I_3 = & 7 & 8 & 9 \\
 4 & 5 & 6
 \end{array}$$

$$\begin{array}{ccc}
 1 & 2 & 3 \\
 I_{TV} = & 4 & 5 & 6 \\
 7 & 8 & 9
 \end{array}$$

• **Kỹ thuật 4: Lọc trung bình đa ảnh**

Cho dãy x_1, x_2, \dots, x_n khi đó trung bình của dãy là: $x = \frac{1}{n} \sum_{i=1}^n x_i$

Cho trước dãy ảnh I_1, I_2, \dots, I_n có cùng kích thước và thuộc tính. Ảnh trung bình của dãy là ảnh $I_{TB}(P)$ là trung bình của dãy $I_1(P), I_2(P), \dots, I_n(P)$. P là tọa độ của ảnh.

$$I_{TB} = \frac{1}{n} \sum_{k=1}^n I_k$$

Ứng dụng: để lấy ảnh đặc trưng trong dãy ảnh.

Mệnh đề 2

Cho dãy x_1, x_2, \dots, x_n .

$$\sum_{k=1}^n (x - x_k)^2 \rightarrow \min \text{ tại giá trị trung bình}$$

chứng minh

$$\phi(x) = \sum_{k=1}^n (x - x_k)^2$$

$$\phi'(x) = 2 \sum_{k=1}^n (x - x_k) = 0$$

$$nx = \sum_{k=1}^n x_k \Rightarrow x = \frac{1}{n} \sum_{k=1}^n x_k$$

$$\Rightarrow \phi \rightarrow \min \text{ tại } x = TB(\{x_i\})$$

Ví dụ:

$$I_1 = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \quad I_2 = \begin{pmatrix} 4 & 5 & 6 \\ 1 & 2 & 3 \\ 7 & 8 & 9 \end{pmatrix} \quad I_3 = \begin{pmatrix} 1 & 2 & 3 \\ 7 & 8 & 9 \\ 4 & 5 & 6 \end{pmatrix} \quad I_{TB} = \begin{pmatrix} 2 & 3 & 4 \\ 4 & 5 & 6 \\ 6 & 7 & 8 \end{pmatrix}$$

Thảo luận

1. Cho 2 ảnh

$$I1 = \begin{pmatrix} 9 & 8 & 6 \\ 5 & 4 & 3 \\ 4 & 2 & 5 \end{pmatrix} \quad I2 = \begin{pmatrix} 6 & 4 & 2 \\ 0 & 1 & 1 \\ 3 & 1 & 2 \end{pmatrix}$$

Thực hiện trừ 2 ảnh

2. Cho 2 ảnh

$$I1 = \begin{pmatrix} 9 & 8 & 6 \\ 5 & 4 & 3 \\ 4 & 2 & 5 \end{pmatrix} \quad I2 = \begin{pmatrix} 6 & 4 & 2 \\ 0 & 1 & 1 \\ 3 & 1 & 2 \end{pmatrix}$$

Thực hiện cộng 2 ảnh, với giá trị α tự cho.

3. Cho 3 ảnh

$$I1 = \begin{pmatrix} 9 & 8 & 6 \\ 5 & 4 & 3 \\ 4 & 2 & 5 \end{pmatrix} \quad I2 = \begin{pmatrix} 6 & 4 & 2 \\ 0 & 1 & 1 \\ 3 & 1 & 2 \end{pmatrix} \quad I3 = \begin{pmatrix} 4 & 8 & 3 \\ 5 & 4 & 7 \\ 2 & 4 & 6 \end{pmatrix}$$

Thực hiện lọc trung vị 3 ảnh trên

4. Cho 3 ảnh

$$I1 = \begin{pmatrix} 9 & 8 & 6 \\ 5 & 4 & 3 \\ 4 & 2 & 5 \end{pmatrix} \quad I2 = \begin{pmatrix} 6 & 4 & 2 \\ 0 & 1 & 1 \\ 3 & 1 & 2 \end{pmatrix} \quad I3 = \begin{pmatrix} 4 & 8 & 3 \\ 5 & 4 & 7 \\ 2 & 4 & 6 \end{pmatrix}$$

Thực hiện lọc trung bình 3 ảnh trên

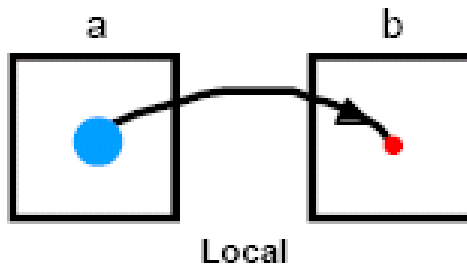
Chương 3: CÁC PHÉP BIẾN ĐỔI PHỤ THUỘC KHÔNG GIAN

Nội dung chính của chương sẽ trình bày các phép biến đổi phụ thuộc không gian, cụ thể là phương pháp tích chập, các mẫu thường dùng, và các kỹ thuật lọc thông dụng.

Mục tiêu cần đạt được của chương

Sinh viên hiểu các khái niệm, phương pháp và biết cách thao tác các phương pháp đã học.

Các phép toán phụ thuộc không gian là các phép toán cục bộ phụ thuộc vào vị trí của điểm ảnh. Thường phụ thuộc vào các lân cận của điểm ảnh



Các phép toán cuộn trên cửa sổ, các thao tác tách cạnh

Các phép biến đổi hình học hai chiều: quay tịnh tiến (co dãn), xê dịch, biến đổi affine

3.1- Phương pháp tích chập

Tích chập (Convolution) là kỹ thuật cơ bản và quan trọng, có nhiều ứng dụng trong xử lý ảnh. Tích chập được sử dụng nhiều trong các phép toán trên ảnh như đạo hàm ảnh, làm trơn ảnh, trích xuất cạnh và gần đây tích chập còn được sử dụng trong các mạng nơ ron học sâu tích chập.

Trong toán học, tích chập là phép toán tuyến tính. Phép tích chập thường được ký hiệu phép nhân tròn là \otimes

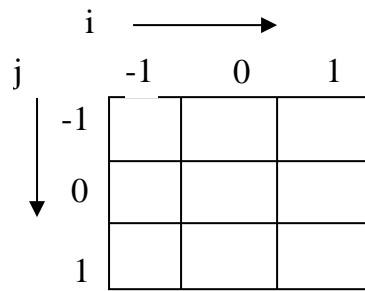
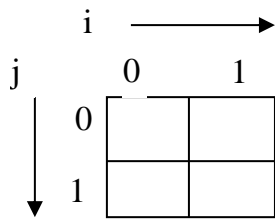
Cho ảnh có kích thước $M \times N$

Nhân lọc (kernel) hay mặt nạ (mask), mẫu có kích thước $m \times n$ (kích thước này thường nhỏ hơn nhiều kích thước ảnh). Thông thường mẫu có kích thước 3×3 , 5×5 , 7×7

Phép toán tích chập được xác định bởi công thức:

$$I \otimes T(x, y) = \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} I(x+i, y+j) * T(i, j) \quad (2.1)$$

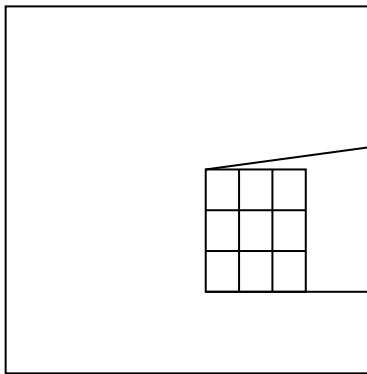
$$\text{Hoặc} \quad I \otimes T(x, y) = \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} I(x-i, y-j) * T(i, j) \quad (2.2)$$



Input image

cửa sổ cuộn

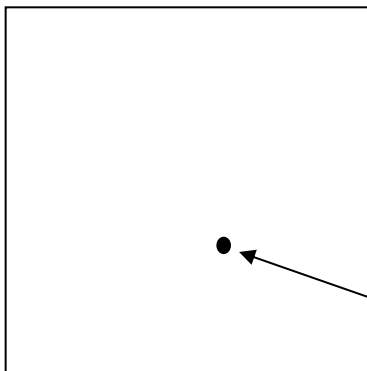
mặt nạ cuộn



I_0	I_1	I_2
I_3	I_4	I_5
I_6	I_7	I_8

M_0	M_1	M_2
M_3	M_4	M_5
M_6	M_7	M_8

output image



$$I_0 * M_0$$

$$I_1 * M_1$$

$$I_2 * M_2$$

$$I_3 * M_3$$

$$I_4 * M_4$$

$$I_5 * M_5$$

$$I_6 * M_6$$

$$I_7 * M_7$$

$$I_8 * M_8$$

$$+ \quad \underline{\hspace{2cm}}$$

New pixel

Hình 3.1 Minh họa phép tích chập

Ma trận mặt nạ T trong tích chập ảnh được dịch chuyển theo từng điểm ảnh và áp lên ảnh để thực hiện tích chập cho từng vị trí trên ảnh mới. Vị trí điểm gốc của mặt nạ được lấy làm chuẩn khi tích chập được gọi là điểm neo (anchor point) của mặt nạ tích chập. Điểm neo sẽ xác định vị trí khớp giữa mặt nạ tích chập với vị trí trên ảnh đầu vào để tích chập. Thông thường điểm neo được chọn là tâm của mặt nạ lọc. Giá trị mỗi phần tử trên mặt nạ được xem như hệ số tổ hợp với lần lượt giá trị độ xám của từng điểm ảnh trong vùng tương ứng với mặt nạ.

Phép tích chập được hình dung là việc thực hiện dịch chuyển mặt nạ lần lượt qua tất cả các vị trí trên ảnh, bắt đầu từ góc trên trái đến dưới phải của ảnh. Quá trình dịch chuyển điểm neo được đặt tương ứng tại điểm ảnh đang xét và tiến hành tính tích chập. Ở mỗi lần dịch chuyển, thực hiện tính toán kết quả tích chập mới cho điểm ảnh đang xét bằng công thức tính chập như trên.

Ví dụ 1:

$$T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad I = \begin{pmatrix} 1 & 3 & 5 & 7 \\ 2 & 4 & 6 & 8 \\ 3 & 4 & 6 & 7 \\ 1 & 1 & 2 & 2 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 3 & 5 & 7 \\ 2 & 4 & 6 & 8 \\ 3 & 4 & 6 & 7 \\ 1 & 1 & 2 & 2 \end{pmatrix} \otimes \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 8 & 14 & 20 & 7 \\ 10 & 16 & 21 & 8 \\ 8 & 12 & 15 & 7 \\ 2 & 3 & 4 & 2 \end{pmatrix}$$

VD2:

$$I = \begin{pmatrix} 1 & 2 & 4 & 5 & 8 & 7 \\ 2 & 1 & 1 & 4 & 2 & 2 \\ 4 & 5 & 5 & 8 & 8 & 2 \\ 1 & 2 & 1 & 1 & 4 & 4 \\ 7 & 2 & 2 & 1 & 5 & 2 \end{pmatrix}$$

$$T = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$I \otimes T = \begin{pmatrix} 2 & 3 & 8 & 7 & 10 & * \\ 7 & 6 & 9 & 12 & 4 & * \\ 6 & 6 & 6 & 12 & 12 & * \\ 3 & 4 & 2 & 6 & 6 & * \\ * & * & * & * & * & * \end{pmatrix} \quad \text{Tính theo (2.1)}$$

Tính theo công thức 2.2

$$I \otimes T = \begin{pmatrix} * & * & * & * & * & * \\ * & 2 & 3 & 8 & 7 & 10 \\ * & 7 & 6 & 9 & 12 & 4 \\ * & 6 & 6 & 6 & 12 & 12 \\ * & 3 & 4 & 2 & 6 & 6 \end{pmatrix}$$

*** Nhân xét:**

- Trong quá trình thực hiện phép nhân chập có một số thao tác ra ngoài ảnh, ảnh không được xác định tại những vị trí đó dẫn đến ảnh thu được có kích thước nhỏ hơn. Hoặc trong trường hợp có tính thì những điểm ảnh coi là nhận giá trị 0

- Ảnh thực hiện theo công thức 2.1 và 2.2 chỉ sai khác nhau 1 phép dịch chuyển để đơn giản ta sẽ hiểu phép nhân chập là theo công thức 2.1

Đa số các phép toán trong xử lý ảnh đều dùng phép nhân chập (hay phép cuộn)

Các mẫu thường dùng

$$T_1 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

$$T_2 = \begin{pmatrix} 1 & 3 & 1 \\ 3 & 16 & 3 \\ 1 & 3 & 1 \end{pmatrix}$$

$$T_3 = \begin{pmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{pmatrix}$$

T_1 dùng để lọc nhiễu

T_2 dùng để lọc trơn bề mặt

T_3 dùng để tách cạnh

(lọc thông thấp)

(lọc thông trung bình)

-phát hiện ra các điểm

có giá trị thay đổi hẳn

so với các giá trị bên cạnh

I				$I \otimes T_1$			
1	1	1	1	5	8	8	5
1	2	2	1	10	17	17	10
1	4	4	1	10	17	17	10
1	1	1	1	7	12	12	7

Thảo luận: Nhân chập ảnh I với mẫu T2, T3

Để giảm nhiễu, vết mờ ta dùng lọc thông thấp(chỉ cho tần số thấp đi qua): lọc trung vị, lọc trung bình.

Để làm ảnh sắc nét dùng lọc thông cao (chỉ cho tần số cao đi qua).

Một số mặt nạ làm ảnh sắc nét.

$$\begin{pmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{pmatrix} \begin{pmatrix} -1 & -1 & -1 \\ -1 & 9 & -1 \\ -1 & -1 & -1 \end{pmatrix} \begin{pmatrix} 1 & -2 & 1 \\ -2 & 5 & -2 \\ 1 & -2 & 1 \end{pmatrix}$$

Mẫu tách cạnh

$$A = \begin{pmatrix} -1 & 1 \\ -1 & 1 \end{pmatrix} \quad B = \begin{pmatrix} -1 & -1 \\ 1 & 1 \end{pmatrix}$$

Theo chiều dọc

theo chiều ngang

$$I = \begin{pmatrix} 3 & 3 & 3 & 0 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 3 & 3 \\ 0 & 0 & 0 & 3 & 3 & 3 \\ 0 & 0 & 0 & 3 & 3 & 3 \end{pmatrix} \quad I \otimes A = \begin{pmatrix} 0 & 0 & -6 & 0 & 0 \\ 0 & 0 & -6 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 6 & 0 & 0 \\ 0 & 0 & 6 & 0 & 0 \end{pmatrix} \quad I \otimes B = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -6 & -6 & 0 & 6 & 6 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad I \otimes A + I \otimes B =$$

$$\begin{pmatrix} 0 & 0 & -6 & 0 & 0 \\ 0 & 0 & -6 & 0 & 0 \\ -6 & -6 & 0 & 6 & 6 \\ 0 & 0 & 6 & 0 & 0 \\ 0 & 0 & 6 & 0 & 0 \end{pmatrix}$$

Câu hỏi

1. Cho ảnh I, mẫu T_1 . Thực hiện nhân chập ảnh I với mẫu T_1

$$I = \begin{pmatrix} 1 & 3 & 5 & 7 \\ 2 & 4 & 6 & 8 \\ 3 & 4 & 6 & 7 \\ 1 & 1 & 2 & 2 \end{pmatrix} \quad T_1 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

2. Cho ảnh I, mẫu T_2 . Thực hiện nhân chập ảnh I với mẫu T_2

$$I = \begin{pmatrix} 1 & 3 & 5 & 7 \\ 2 & 4 & 6 & 8 \\ 3 & 4 & 6 & 7 \\ 1 & 1 & 2 & 2 \end{pmatrix} \quad T_2 = \begin{pmatrix} 1 & 3 & 1 \\ 3 & 16 & 3 \\ 1 & 3 & 1 \end{pmatrix}$$

3. Cho ảnh I, mẫu T_3 . Thực hiện nhân chập ảnh I với mẫu T_3

$$I = \begin{pmatrix} 1 & 3 & 5 & 7 \\ 2 & 4 & 6 & 8 \\ 3 & 4 & 6 & 7 \\ 1 & 1 & 2 & 2 \end{pmatrix} \quad T_3 = \begin{pmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{pmatrix}$$

4. Cho ảnh I, mẫu A. Thực hiện nhân chập ảnh I với mẫu A

$$I = \begin{pmatrix} 1 & 3 & 5 & 7 \\ 2 & 4 & 6 & 8 \\ 3 & 4 & 6 & 7 \\ 1 & 1 & 2 & 2 \end{pmatrix} \quad A = \begin{pmatrix} -1 & 1 \\ -1 & 1 \end{pmatrix}$$

5. Cho ảnh I, mẫu B. Thực hiện nhân chập ảnh I với mẫu B

$$\mathbf{I} = \begin{pmatrix} 1 & 3 & 5 & 7 \\ 2 & 4 & 6 & 8 \\ 3 & 4 & 6 & 7 \\ 1 & 1 & 2 & 2 \end{pmatrix} \quad \mathbf{B} = \begin{pmatrix} -1 & -1 \\ 1 & 1 \end{pmatrix}$$

6. Tìm hiệu lớp tích chập trong mạng nơ ron tích chập

BÀI 4: CÁC KỸ THUẬT LỌC THÔNG DỤNG (Số tiết: 5 tiết)

3.2- Các kỹ thuật lọc thông dụng

* Kỹ thuật 1: Lọc trung vị

Lọc trung vị (median filter) là kỹ thuật lọc phi tuyến được dùng để khử nhiễu khá hiệu quả đối với các loại nhiễu đốm (speckle noise) và nhiễu muối tiêu (salt-pepper noise).

Với mỗi P, lấy cửa sổ W(P)

Tính trung vị TV(P)= trung vị W(P)

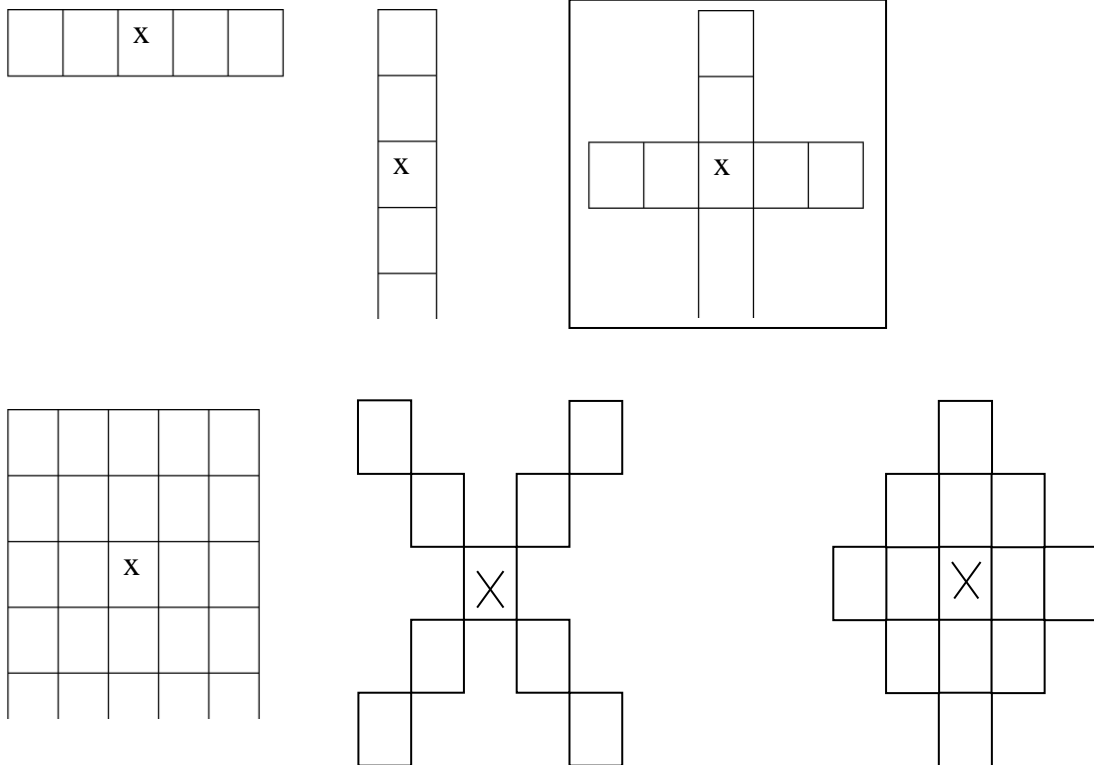
$$I'(P) = \begin{cases} I(P) & \text{nếu } |I(P) - TV(P)| \leq \theta \\ TV(P) & \text{nếu ngược lại} \end{cases}$$

$\theta > 0$ là một ngưỡng cho trước

Lọc trung vị là lọc phi tuyến. Cho 2 dãy $x(m)$ và $y(m)$.

Trung vị $\{x(m)+y(m)\} \neq \text{Trung vị}\{x(m)\} + \text{Trung vị}\{y(m)\}$

Các cửa sổ lọc trung vị có rất nhiều dạng khác nhau



Hình 3.2 Ví dụ một số dạng cửa sổ



a



b

Hình 3.3 a) ảnh bị nhiễu muối tiêu, b) lọc trung vị cửa sổ 3x3

Ví dụ:

$$I = \begin{pmatrix} 1 & 2 & 3 & 2 \\ 4 & 16 & 2 & 1 \\ 4 & 2 & 1 & 1 \\ 2 & 1 & 2 & 1 \end{pmatrix}$$

$W(3 \times 3); \theta = 2$

$$TV = \begin{pmatrix} 0 & 2 & 2 & 0 \\ 2 & 2 & 2 & 1 \\ 2 & 2 & 1 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$

$$|I-TV| = \begin{pmatrix} 1 & 0 & 1 & 2 \\ 2 & 14 & 0 & 0 \\ 2 & 0 & 0 & 0 \\ 2 & 0 & 1 & 1 \end{pmatrix}$$

$$I_{kq} = \begin{pmatrix} 1 & 2 & 3 & 2 \\ 4 & \textcircled{2} & 2 & 1 \\ 4 & 2 & 1 & 1 \\ 2 & 1 & 2 & 1 \end{pmatrix}$$

Giá trị 16, sau phép lọc có giá trị 2, các giá trị còn lại không thay đổi giá trị.

* Kỹ thuật 2: Lọc trung bình

Lọc trung bình là kỹ thuật lọc tuyến tính. Lọc trung bình hoạt động như một bộ lọc thông thấp.

Với mỗi P, lấy cửa sổ W(P)

Tính trung bình TB(P) trong cửa sổ W(P)

$$I(P) = \begin{cases} I(P) & \text{nếu } |I(P) - TB(P)| \leq \theta \\ TB(P) & \text{nếu ngược lại} \end{cases}$$

$\theta > 0$ là một ngưỡng cho trước

Việc tính ảnh trung bình chính là thực hiện cuộn với mặt nạ với các phần tử có trọng số bằng nhau và bằng $1/m \times n$

Ví dụ với mặt nạ 3×3 trọng số mỗi phần tử bằng $1/9$

$$\frac{1}{9} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$



a



b

Hình 3.4 a) Ảnh bị nhiễu muối tiêu, b) ảnh lọc trung bình

Ví dụ:

$$I = \begin{pmatrix} 1 & 2 & 3 & 2 \\ 4 & 16 & 2 & 1 \\ 4 & 2 & 1 & 1 \\ 2 & 1 & 2 & 1 \end{pmatrix}$$

$$W_{3 \times 3}, \theta = 2$$

$$TB = \begin{pmatrix} 3 & 3 & 3 & 1 \\ 3 & 4 & 3 & 1 \\ 3 & 4 & 3 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

$$|I-TB| = \begin{pmatrix} 2 & 1 & 0 & 1 \\ 1 & 12 & 1 & 0 \\ 1 & 2 & 2 & 0 \\ 1 & 0 & 1 & 0 \end{pmatrix}$$

$$I_{kq} = \begin{pmatrix} 1 & 2 & 3 & 2 \\ 4 & \textcircled{4} & 2 & 1 \\ 4 & 2 & 1 & 1 \\ 2 & 1 & 2 & 1 \end{pmatrix}$$

Giá trị 16 sau phép lọc trung bình có giá trị 3, các giá trị còn lại giữ nguyên sau phép lọc.

Thảo luận: Cho ảnh I, thực hiện lọc trung bình và trung vị cho biết cửa sổ $W_{3 \times 3}$, $\theta=3$

$$I = \begin{pmatrix} 1 & 4 & 2 & 8 & 5 & 7 \\ 4 & 2 & 8 & 5 & 7 & 1 \\ 2 & 8 & 5 & 7 & 1 & 4 \\ 8 & 5 & 7 & 1 & 4 & 2 \\ 5 & 7 & 1 & 4 & 2 & 8 \\ 7 & 1 & 4 & 2 & 8 & 5 \end{pmatrix}$$

Tác dụng của lọc trung vị và lọc trung bình: làm trơn ảnh, giảm nhiễu, hiệu quả với việc giảm nhiễu Gaussian.

* Kỹ thuật 3: Lọc trung bình theo k giá trị gần nhất

Giả sử ta có ảnh I, điểm ảnh P, cửa sổ $W(P)$, ngưỡng θ và số k. Khi đó lọc trung bình theo k giá trị gần nhất bao gồm các bước sau:

B1: tìm k giá trị gần nhất với $I(P)$

$$\{I(q) \mid q \in W(P)\} \rightarrow \{k \text{ giá trị gần } I(p) \text{ nhất}\}$$

B2: tính trung bình $\{k \text{ giá trị gần } I(P) \text{ nhất}\} \rightarrow TB_k(P)$

B3: tính giá trị

$$I(P) = \begin{cases} I(P) & \text{nếu } |I(P) - TB_k(P)| \leq \theta \\ TB_k(P) & \text{nếu ngược lại} \end{cases}$$

Ví dụ: $W(3, 3)$, $\theta=2$, $k=3$

	x	

	1	2	3	2
I=	4	16	2	1
	4	2	1	1
	2	1	2	1

	$TB_k(P)$				$ I(P)-TB_k(P) $				ảnh kết quả			
0	2	2	2		1	0	1	0	1	2	3	2
3	8	2	1		1	8	0	0	4	8	2	1
3	2	1	1		1	0	0	0	4	2	1	1
2	1	2	1		0	0	0	0	2	1	2	1

Nhận xét:

Nếu $k \geq$ kích thước của số W thì kỹ thuật này trở thành lọc trung bình

Nếu $k=1$ thì ảnh kết quả không thay đổi.

\Rightarrow Chất lượng của kỹ thuật phụ thuộc vào số phân tử lựa chọn k .

*** Kỹ thuật 4: Giả trung bình**

$$TB(P)=0,5 \{ \max I(q) + \min I(q) | q \in W(P) \}$$

Phương pháp này thường dùng để lọc nhiễu ngẫu nhiên như nhiễu Gauss hoặc nhiễu thuần nhất.

*** Kỹ thuật 5: Lọc cực đại, lọc cực tiểu**

$$I(P) = \max \{ I(q) | q \in W(P) \}$$

$$\text{hoặc } I(P) = \min \{ I(q) | q \in W(P) \}$$

Câu hỏi

1. Cho ảnh I, $W_{3 \times 3}$, $\theta=3$. Thực hiện lọc trung bình ảnh sau:

	1	2	3	4	5	6
	0	9	8	6	4	7
I=	9	5	4	3	5	8
	8	4	2	5	4	9
	7	0	3	1	0	0
	6	5	4	3	2	1

2. Cho ảnh I, $W_{3 \times 3}$, $\theta=3$. Thực hiện lọc trung vị ảnh sau:

	1	2	3	4	5	6
	0	9	8	6	4	7
I=	9	5	4	3	5	8
	8	4	2	5	4	9
	7	0	3	1	0	0
	6	5	4	3	2	1

3. Cho ảnh I, $W_{3 \times 3}$, $\theta=3$, $k=3$. Thực hiện lọc trung bình k láng giềng ảnh sau:

	1	2	3	4	5	6
	0	9	8	6	4	7
I=	9	5	4	3	5	8
	8	4	2	5	4	9
	7	0	3	1	0	0
	6	5	4	3	2	1

4. Cho ảnh I, $W_{3 \times 3}$. Thực hiện kỹ thuật giả trung bình ảnh sau:

	1	2	3	4	5	6
	0	9	8	6	4	7
I=	9	5	4	3	5	8
	8	4	2	5	4	9
	7	0	3	1	0	0
	6	5	4	3	2	1

5. Cho ảnh I, $W_{3 \times 3}$. Thực hiện kỹ thuật lọc cực đại ảnh sau:

	1	2	3	4	5	6
	0	9	8	6	4	7

I=	9	5	4	3	5	8
	8	4	2	5	4	9
	7	0	3	1	0	0
	6	5	4	3	2	1

6. Cho ảnh I, $W_{3 \times 3}$. Thực hiện kỹ thuật lọc cực tiểu ảnh sau:

	1	2	3	4	5	6
	0	9	8	6	4	7
I=	9	5	4	3	5	8
	8	4	2	5	4	9
	7	0	3	1	0	0
	6	5	4	3	2	1

7. Tìm hiểu các thủ tục thực hiện các bộ lọc đã học

CHƯƠNG 4. PHƯƠNG PHÁP PHÂN ĐOẠN ẢNH

Nội dung chính của chương sẽ trình bày về phân đoạn ảnh và các phương pháp phân đoạn ảnh.

Mục tiêu cần đạt được của chương

Sinh viên hiểu các khái niệm, phương pháp và biết cách thao tác các phương pháp phân đoạn ảnh đã học.

Bài 5: PHƯƠNG PHÁP PHÂN ĐOẠN ẢNH (Số tiết: 5 tiết)

4.1-Tổng quan về phân đoạn ảnh

Phân đoạn (Segmentation) được hiểu là quá trình nhóm điểm ảnh thành các vùng sao cho các điểm ảnh trong cùng một vùng có chung đặc điểm, tính chất nào đó, ví dụ như sự tương tự về màu sắc, cường độ hoặc kết cấu. Nói cách khác, phân đoạn ảnh là một quá trình chia một ảnh số thành nhiều vùng khác nhau, mỗi vùng gồm tập hợp các điểm ảnh có cùng tính chất. Mục tiêu của việc phân đoạn là để đơn giản hóa hoặc thay đổi cách biểu diễn một ảnh theo tiêu chuẩn nào đó nhằm làm cho nó ý nghĩa hơn và dễ dàng trong phân tích, trích xuất đặc trưng. Kết quả quá trình phân đoạn ảnh là tập hợp các vùng hay còn gọi là các phân đoạn có thể bao gồm toàn bộ ảnh hoặc tập hợp các vùng đối tượng được trích xuất ra từ ảnh.

Phân đoạn ảnh có ý nghĩa quan trọng trong việc phát hiện các đặc tính và cấu trúc đối tượng trong ảnh vì mỗi vùng thường miêu tả cấu trúc tương ứng với đối tượng trong vùng đó. Ví dụ số lượng các vùng trong ảnh thể hiện rằng trong ảnh đó có chứa nhiều đối tượng khác nhau và ngược lại mỗi đối tượng chứa các vùng trong nó tương ứng các phần khác nhau của cùng đối tượng. Phân đoạn ảnh có nhiều ứng dụng quan trọng trong nhiều lĩnh vực khác nhau như phân tích hình ảnh y học, phát hiện và nhận dạng đối tượng, phát hiện hành động trong các hệ thống camera giám sát thông minh, hệ thống điều khiển và giao thông thông minh, xe không người lái, xử lý hình ảnh vệ tinh,...

Phân đoạn ảnh được xem là quá trình tiền xử lý, có ý nghĩa quan trọng trong hệ thống xử lý ảnh. Phân đoạn tốt sẽ giúp cho quá trình xử lý ở các bước tiếp theo đạt hiệu quả cao hơn, nâng cao độ chính xác và cũng như tiết kiệm chi phí tính toán.

Các vùng phân tách được dùng trong nhiều ngữ cảnh khác nhau và đặc biệt có thể được sử dụng làm đại diện thay thế cho đối tượng và biểu diễn dưới dạng nhiều hình thức khác nhau

Các kỹ thuật phân đoạn:

+ Phân đoạn dựa theo miền đồng nhất

- + Phân đoạn dựa vào biên
- + Phân đoạn dựa vào ngưỡng biên độ
- + Phân đoạn theo kết cấu (texture)

4.2. Phân đoạn bằng ngưỡng

Ngưỡng được dùng để chỉ một giá trị nào đó mà ta dùng để phân tách một tập hợp các điểm ảnh thành các vùng phân biệt. Giá trị ngưỡng thường được xác định dựa vào giá trị đặc biệt như giá trị trung bình, giá trị trung vị hoặc dựa vào khảo sát thống kê theo kinh nghiệm. Có nhiều cách phân loại khác nhau, tuy nhiên nếu căn cứ vào số lượng ngưỡng dùng cho phân đoạn thì có phương pháp phân đoạn sử dụng ngưỡng đơn, phân đoạn ngưỡng kép hoặc phân đoạn đa ngưỡng. Nếu căn cứ vào tính chất biến thiên của giá trị ngưỡng thì có thể phân thành phương pháp dùng ngưỡng cố định và ngưỡng động hay ngưỡng thích ứng. Trong đó, ngưỡng động được hiểu theo nghĩa giá trị của ngưỡng sẽ thay đổi tùy thuộc vào sự biến thiên của tập dữ liệu ảnh theo không gian và thời gian. Thông thường giá trị ngưỡng được xác định bằng cách áp dụng phương pháp phân tích thống kê.

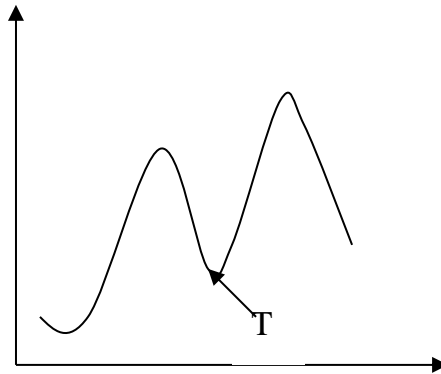
4.2.1 Phân ngưỡng thủ công

Phân ngưỡng thủ công là phương pháp đơn giản với chi phí tính toán thấp. Là quá trình phân loại điểm ảnh là điểm ảnh nền hay điểm ảnh đối tượng dựa trên một ngưỡng cho trước, kết quả cho ra ảnh nhị phân, các điểm đối tượng có giá trị là 1, các điểm nền có giá trị là 0

Xét đối tượng sáng trên nền tối, cho ngưỡng T , các điểm ảnh có giá trị cấp xám trên ngưỡng T sẽ là điểm đối tượng, còn các điểm ảnh có giá trị cấp xám nằm dưới ngưỡng T sẽ là điểm nền (làm ngược lại với ảnh có đối tượng tối trên nền sáng).

$$out(x, y) = \begin{cases} 1 & in(x, y) \geq T \\ 0 & in(x, y) < T \end{cases}$$

Một ngưỡng được chọn phụ thuộc vào dữ liệu của ảnh.



Hình 4.1 Chọn ngưỡng dựa vào histogram

Chọn ngưỡng dựa vào histogram

Với histogram có nhiều đỉnh và thung lũng thì ta chọn khe giữa 2 đỉnh làm ngưỡng.

Việc lựa chọn ngưỡng tối ưu vẫn còn là một vấn đề khó khăn.

4.2.2. Phân đoạn ngưỡng tự động

Trong trường hợp tổng quát, ảnh biến đổi phức tạp yêu cầu phân đoạn ảnh cần phải được thiết kế hiệu quả, linh động đáp ứng được với nhiều tình huống khác nhau và kết quả bền vững. Vì vậy, việc lựa chọn ngưỡng phân đoạn cần được thiết lập một cách linh động bởi hệ thống tự động hoặc theo thống kê kinh nghiệm nào đó. Thông thường, quá trình phân đoạn ảnh cần xem xét đến các yếu tố liên quan về thông tin đối tượng trong ảnh, môi trường và nền nhằm xác định các vùng liên quan đến đối tượng khác nhau hơn là chỉ sử dụng giá trị ngưỡng cố định. Các thông tin liên quan để phân đoạn thường bao gồm các yếu tố sau:

- Đặc điểm về cường độ sáng của các đối tượng
- Kích thước của đối tượng
- Tỷ lệ chiếm chỗ bởi các đối tượng trong ảnh
- Số lượng đối tượng khác nhau xuất hiện trong cùng ảnh

Như vậy, quá trình chọn ngưỡng tự động được hiểu là quá trình xác định giá trị ngưỡng để phân đoạn ảnh được thực hiện hoàn toàn bằng chương trình mà không nhờ sự can thiệp của con người. Việc xác định ngưỡng tự động được thực hiện trên cơ sở phân tích sự phân bố cường độ giá trị xám của hình ảnh. Lược đồ ảnh thường được sử dụng để phân tích phân bố giá trị xám và sử dụng kiến thức kinh nghiệm để lựa chọn ngưỡng một cách thích hợp, thích ứng với từng trường hợp, điều kiện sáng cụ thể. Khi

thông tin được xác định càng mang tính tổng quát thì miền áp dụng phân đoạn càng tăng lên.

Hầu hết phương pháp tự động thường lựa chọn ngưỡng sử dụng kích thước đối tượng, xác suất xuất hiện và ước lượng phân phối cường độ sáng bằng cách tính giá trị lược đồ cường độ sáng của ảnh.

4.2.2.1 Phương pháp lặp chọn ngưỡng

Phương pháp lặp lựa chọn ngưỡng bắt đầu với một ngưỡng xấp xỉ nào đó và thực hiện liên tục quá trình lặp để tinh chỉnh ngưỡng. Vấn đề này được kỳ vọng có thể lựa chọn ngưỡng mới từ ngưỡng đã được chọn trước đó sao cho việc phân đoạn được tốt hơn. Quá trình điều chỉnh ngưỡng đóng vai trò rất quan trọng quyết định đến thành công của cách tiếp cận này và được mô tả như sau:

Thuật toán: Phương pháp lặp chọn ngưỡng

1. Lựa chọn một ước lượng ngưỡng khởi tạo T . Ta có thể khởi tạo giá trị này bằng giá trị trung bình cường độ xám của ảnh.
2. Chia hình ảnh thành 2 nhóm R_1 và R_2 bằng cách sử dụng ngưỡng T .
3. Tính các giá trị trung bình cường độ xám của 2 vùng R_1 và R_2 tương ứng là μ_1 và μ_2 .
4. Chọn giá trị ngưỡng mới là $T = \text{mean}(\mu_1, \mu_2)$
5. Lặp lại bước 2 đến bước 4 cho đến khi giá trị trung bình μ_1 và μ_2 không thay đổi.

4.2.2.2 Ngưỡng thích nghi cục bộ

Nếu ánh sáng chiếu lên vật thể (kể cả nền) không đồng nhất thì các phương pháp chọn ngưỡng phân đoạn ở phần trước sẽ không phù hợp. Ánh sáng không đồng nhất có thể do bóng (shadow) hoặc hướng của nguồn sáng khi chiếu lên vật thể. Trong hầu hết các trường hợp, cùng một giá trị ngưỡng có thể không phù hợp để áp dụng cho toàn bộ ảnh. Khác với các phương pháp trước thực hiện phân tích lược đồ mức xám ảnh áp dụng trên toàn bộ ảnh, phương pháp thích nghi cục bộ giải quyết vấn đề ánh sáng không đồng nhất hoặc phân phối không đều giá trị xám của nền, tập trung vào các vùng nhỏ của ảnh để chia ảnh thành các ảnh con nhằm tiến hành xác định ngưỡng cho mỗi vùng con đó. Một số kỹ thuật đã được nghiên cứu để giải quyết phù hợp vấn đề này một cách hiệu quả như cách tiếp cận trực tiếp.

Cách tiếp cận trực tiếp là phân chia ảnh thành $m \times n$ vùng ảnh con nhỏ hơn (chia thành lưới) và chọn ngưỡng T_{ij} cho mỗi ảnh con thứ ij dựa vào lược đồ xám tương ứng với $a \leq i, j \leq m$. Phân đoạn tổng thể thu được là tập hợp các vùng phân đoạn ảnh nhỏ.

4.2.2.3 Phân đoạn ngưỡng kép

Thuật toán: Ngưỡng kép để phát triển vùng

1. Lựa chọn 2 ngưỡng T_1 và T_2 theo thuật toán chọn ngưỡng.
2. Phân đoạn ảnh thành ba tập gồm R_1, R_2, R_3 :
 R_1 tập gồm các điểm ảnh có giá trị cường độ xám dưới ngưỡng T_1 ;
 R_2 tập gồm các điểm ảnh có giá trị xám trong khoảng ngưỡng từ T_1 đến T_2 ;
 R_3 tập gồm các điểm ảnh còn lại có giá trị cường độ xám trên ngưỡng T_2 .
3. Duyệt qua mỗi điểm ảnh thuộc vùng R_2 , nếu điểm ảnh nào có một vùng lân cận thuộc R_1 thì ta gán lại điểm ảnh này đến vùng R_1 .
4. Lặp lại bước 3 cho đến khi không có điểm ảnh nào cần được gán lại.
5. Gán lại bất cứ điểm ảnh nào còn lại ở vùng R_2 và R_3 .

4.3 Phân đoạn ảnh bằng k-means

K-means là thuật toán phân cụm do MacQueen đề xuất trong lĩnh vực thống kê năm 1967, mục đích của thuật toán k-means là phân chia thành k cụm dữ liệu $\{C_1, C_2, \dots, C_k\}$ từ một tập dữ liệu ban đầu gồm n đối tượng trong không gian d chiều $X_i = (x_{i1}, x_{i2}, \dots, x_{id})$ ($i=1, \dots, n$), sao cho dữ liệu trong các cụm là tương đồng với nhau nhất, nghĩa là hàm tiêu chuẩn đánh giá lỗi $E = \sum_{i=1}^k \sum_{x \in C_i} D^2(x - m_i)$ đạt giá trị tối thiểu với m_i là trọng tâm của cụm C_i , D là khoảng cách giữa hai đối tượng.

Trọng tâm của mỗi cụm được biểu diễn dưới dạng một vector trong không gian d chiều, trong đó giá trị mỗi phần tử của vector là trung bình thành phần tương ứng của các đối tượng dữ liệu (các vector) trong cụm đang xét. Tham số đầu vào của thuật toán là số cụm k , tập dữ liệu gồm n phần tử và tham số đầu ra là tập trọng tâm đại diện cho các cụm dữ liệu được phân chia. Độ đo khoảng cách D giữa các đối tượng dữ liệu thường được sử dụng là khoảng cách Euclide, vì nó là mô hình khoảng cách dễ lấy đạo hàm và xác định các cực trị tối thiểu. Hàm tiêu chuẩn và độ đo khoảng cách có thể được xác định cụ thể hơn tùy vào ứng dụng cụ thể hoặc quan điểm của người dùng. Thuật toán k-means bao gồm các bước cơ bản sau:

INPUT: Một tập dữ liệu gồm n đối tượng và số các cụm k cần phân chia

OUTPUT: Tập trọng tâm (m) đại diện cho các cụm C_i ($i=1,\dots,k$) sao cho hàm tiêu chuẩn E đạt giá trị tối thiểu.

Bước 1: Khởi tạo

Chọn k đối tượng m_j ($j=1,\dots,k$) là trọng tâm ban đầu của k cụm từ tập dữ liệu (lựa chọn trọng tâm ban đầu có thể được thực hiện ngẫu nhiên hoặc theo kinh nghiệm).

Bước 2: Tính khoảng cách

Với mỗi đối tượng X_i ($1 \leq i \leq n$) tính khoảng cách từ X_i tới mỗi trọng tâm m_j với $j=1,\dots,k$

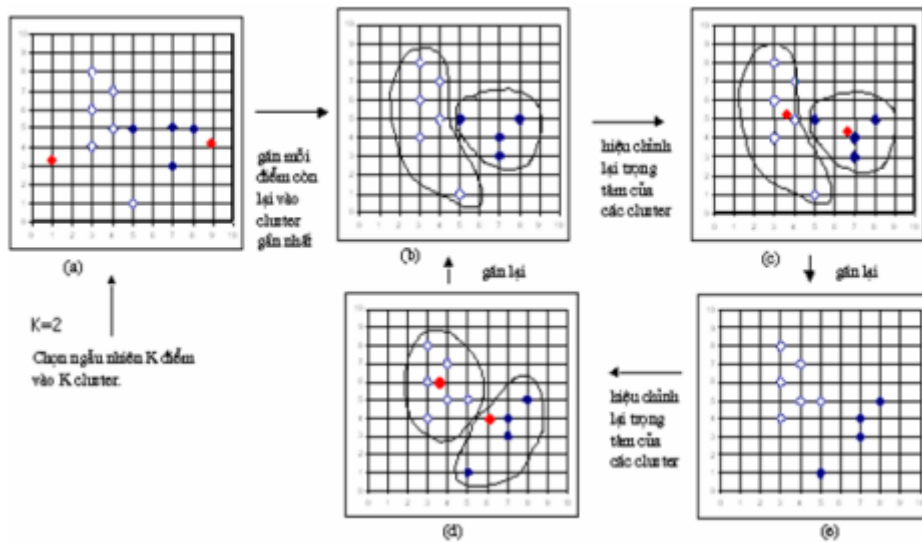
Gán lại cụm cho đối tượng sao cho khoảng cách từ nó đến cụm có trọng tâm gần nhất.

Bước 3: Cập nhật lại trọng tâm

Mỗi cụm C_j với $j=1,\dots,k$, cập nhật trọng tâm cụm m_j bằng cách tính giá trị trung bình của tất cả đối tượng dữ liệu trong cụm.

Bước 4: Điều kiện dừng

Lặp các bước 2 và 3 cho đến khi các trọng tâm của cụm không thay đổi.



Hình 4.2. Minh họa phân cụm bằng k-means

Thuật toán k-means là phương pháp phân tích phân cụm đơn giản có thể áp dụng đối với tập dữ liệu lớn. Tuy nhiên nhược điểm của k-means là chỉ áp dụng với dữ liệu có thuộc tính số và khám phá ra các cụm có dạng hình cầu, k-means rất nhạy cảm với nhiễu và các phần tử ngoại lai trong dữ liệu.

Chất lượng phân cụm dữ liệu bằng thuật toán k-means phụ thuộc nhiều vào các tham số đầu vào như số cụm k và các trọng tâm được khởi tạo ban đầu. Trong trường hợp các trọng tâm khởi tạo ban đầu quá lệch so với trọng tâm thực của cụm thì kết quả phân cụm của k-means đạt được rất thấp, nghĩa là các cụm dữ liệu được phân tách rất lệch so với cụm trong thực tế của nó và thời gian hội tụ cũng chậm. Kinh nghiệm cho thấy chưa có một giải pháp tối ưu nào để chọn tham số đầu vào chung cho nhiều bài toán khác nhau. Giải pháp thường được sử dụng nhất là thử nghiệm với các giá trị đầu vào k khác nhau sau đó chọn giải pháp tốt nhất.

Đến nay, đã có rất nhiều thuật toán kế thừa ý tưởng của thuật toán k-means áp dụng trong phân cụm dữ liệu để giải quyết tập dữ liệu có kích thước lớn đang được áp dụng rất hiệu quả và phổ biến như thuật toán k-medoid, PAM, CLARA, CLARANS, k-prototypes,...

4.4 Phân đoạn ảnh bằng MeanShift

Phương pháp MeanShift được giới thiệu năm 1975 bởi Fukunaga và Hostetle. Đây là kiểu thuật toán lặp, bắt đầu với giá trị khởi tạo x , xác định trước một hàm nhân $K(x_i - x)$ có chức năng tìm kiếm trọng số các điểm lân cận để ước lượng lại giá trị trung bình. Hàm nhân hay được sử dụng là hàm Gaussian. Hàm Gaussian xác định khoảng cách đến giá trị giá trị ước lượng hiện tại được tính theo công thức

$$K(x_i - x) = e^{-c\|x_i - x\|^2}$$

Trung bình có trọng số các giá trị trong cửa sổ được xác định bởi K như sau:

$$m(x) = \frac{\sum_{x_i \in N(x)} K(x_i - x)x_i}{\sum_{x_i \in N(x)} K(x_i - x)}$$

Với $N(x)$ là lân cận của x , tập các điểm sao cho $K(x_i - x) \neq 0$

Giá trị khoảng cách giữa x và giá trị trung bình các lân cận xác định bởi $m(x) - x$ được gọi là khoảng cách MeanShift. Giá trị ước lượng x sẽ được cập nhật bằng giá trị trung bình lân cận $x \leftarrow m(x)$ và lặp lại quá trình cập nhật x cho đến khi $m(x)$ hội tụ.

Tổng quát về thuật toán MeanShift được mô tả như sau:

Cho trước một tập hữu hạn S , với miền giá trị trong không gian Euclidean có n chiều X , K là một hàm nhân phẳng có tính chất:

$$K(x) = \begin{cases} 1 & \text{if } \|x\| \leq \lambda \\ 0 & \text{if } \|x\| > \lambda \end{cases}$$

Với mỗi bước lặp, thực hiện gán giá trị $s \leftarrow m(s)$ cho tất cả $s \in S$;

Vấn đề đầu tiên là làm thế nào để ước lượng hàm mật độ cho bởi tập thưa các mẫu. Một trong những cách tiếp cận đơn giản nhất là làm trơn dữ liệu, có thể bằng cách điều chỉnh nó với một hàm nhân cố định có độ rộng xác định như sau:

$$f(x) = \sum_i K(x - x_i) = \sum_i k\left(\frac{\|x - x_i\|^2}{h^2}\right)$$

Với x_i là các mẫu vào và $k(r)$ là hàm nhân, h là một tham số trong thuật toán và được gọi là băng thông (bandwidth).

Các tiếp cận này được gọi là ước lượng mật độ nhân hay là kỹ thuật cửa sổ Parzen. Một khi tính xong giá trị $f(x)$ theo công thức trên, cực đại cục bộ được xác định bằng cách sử dụng phương pháp độ dốc gradient hoặc bằng phương pháp tối ưu hóa nào đó. Vấn đề hạn chế của cách tiếp cận này là khi số chiều lớn hơn nó trở nên khó đánh giá $f(x)$ trên không gian tìm kiếm phức tạp. Thay vào đó, MeanShift sử dụng một biến thể khác trong tối ưu hóa như phương pháp khởi động lại nhiều lần dựa vào độ giảm gradient. Bắt đầu bằng một ước lượng cực đại địa phương y_k có thể là một giá trị ngẫu nhiên x_1 , meanshift tính giá trị gradient của mật độ ước lượng $f(x)$ tại y_k và lấy một bước dốc gradient theo hướng hiện tại.

Một số kiểu của hàm nhân:

Cho X là không gian Euclidean n chiều, \mathbb{R}^n . Ký hiệu thành phần thứ i của x là x_i . Chuẩn của x là một số không âm $\|x\|^2 = x^T x$

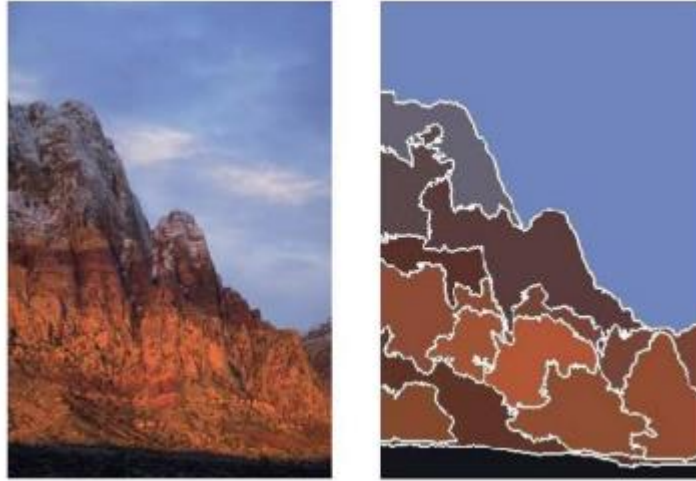
Một hàm $K: X \rightarrow \mathbb{R}$ được gọi là hàm nhân nếu tồn tại tính chất $k: [0, \infty] \rightarrow \mathbb{R}$, ví dụ như $K(x) = k(\|x\|^2)$ và k có các tính chất sau:

- Hàm không âm,
- Hàm không tăng $k(a) \geq k(b)$ nếu $a < b$
- Liên tục từng phần $\int_0^\infty k(r) dr < \infty$

Trong thực tế ứng dụng, meanshift có 2 hàm được dùng phổ biến là

- Hàm phẳng $k(x) = \begin{cases} 1 & \text{if } x \leq \lambda \\ 0 & \text{if } x > \lambda \end{cases}$
- Hàm Gaussian $k(x) = e^{-\frac{x}{2\sigma^2}}$ với độ lệch chuẩn σ như là tham số băng thông h .

Kỹ thuật meanshift có nhiều ứng dụng trong thực tế như ước lượng chuyển động, nhận dạng đối tượng, phân đoạn ảnh,....

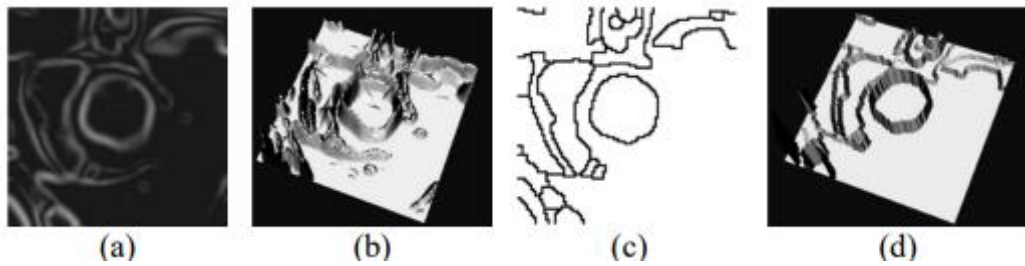


Hình 4.3. Kết quả phân đoạn ảnh bằng meanshift

4.5 Phân đoạn ảnh bằng Watershed

Kỹ thuật phân đoạn watershed được sử dụng nhiều trong xử lý ảnh nhằm tách đối tượng (foreground) khỏi nền (background) với thông tin về các đường biên đối tượng. Về cơ bản, ý tưởng của kỹ thuật này dựa vào nguyên lý nước ở lưu vực dâng lên vùng cao hơn. Phương pháp thực hiện biến đổi bắt đầu từ “lưu vực” (hay còn gọi là vũng) được gọi là vùng thấp, lưu vực tương đối bằng phẳng và xây dựng các “đường ranh giới” (được gọi là các đập chắn) được xác định bởi các cạnh (ngăn cách các vũng). Do vậy, thay vì làm trực tiếp trên ảnh đa mức xám, phương pháp Watershed hoạt động dựa trên giá trị gradient của cường độ sáng điểm ảnh. Việc xác định các đường ranh giới được thực hiện thông qua kỹ thuật phát hiện cạnh, như kỹ thuật Canny. Một ảnh được xem như bề mặt địa hình của một khu vực nào đó, với các pixel có độ sáng cao thuộc về “vùng cao” và pixel có độ sáng thấp thuộc về “vùng thấp” (hay vũng). Sau đó xuất phát từ vũng để thực hiện lan tỏa bằng cách “dâng nước lên” và gộp các vùng lại theo nguyên tắc “dâng nước”, nước sẽ dần lấp đầy các vùng thấp và đường biên ngăn cách giữa các vũng tạo thành phân đoạn ảnh. Khi dâng nước lên đạt đỉnh cục bộ theo đường biên giữa các vũng thì hình thành nên vách ngăn tại đường biên đó. Quá trình được thực hiện lặp lại cho đến khi tất cả các pixel đều “ngập nước”, các vách ngăn chính là đường biên giữa các vùng.

Ví dụ minh họa quá trình phân đoạn bằng kỹ thuật Watershed như sau: Hình (a) là ảnh thể hiện giá trị cường độ gradient mô tả độ dốc giữa các vùng và xác định các vùng cao và vùng thấp; hình (b) ảnh thể hiện kết quả dùng phương pháp phát hiện cạnh để tìm các đường ranh giới chính; hình (c) thể hiện kết quả của việc “dâng nước” gộp vùng và kết quả phân đoạn thành các vùng.



Hình 4.4. Phân đoạn ảnh bằng Watershed11: (a) ảnh gradient, (b) thể hiện độ cao gradient, (c) vách ngăn theo gradient, (d) thể hiện độ cao vách ngăn

Quá trình phân đoạn theo thuật toán Watershed thực hiện như sau:

- 1) Ảnh vào được chuyển qua ảnh đa mức xám
- 2) Dùng thuật toán tính cường độ và độ dốc biến đổi mức sáng (gọi là ảnh gradient)
- 3) Với ảnh gradient nhận được, ta liên tưởng dữ liệu hình ảnh gradient với một đồ địa hình sao cho vùng có cường độ xám thấp là vùng và ngược lại. Tại mỗi điểm ảnh, việc đánh giá dựa vào giá trị mức xám của điểm ảnh đó. Thực hiện quét các điểm ảnh theo trình tự đã sắp xếp để xây dựng các vùng. Mỗi vùng được gán một nhãn riêng biệt để phân biệt giữa các vùng.
- 4) Thực hiện quá trình “dâng nước” làm “ngập nước” các điểm ảnh. Bắt đầu tại điểm thấp nhất của vùng rồi cho nước dâng dần lên. Khi nước trong các vùng cạnh nhau có thể hòa vào nhau thì hình thành nên các vách để ngăn 2 vùng này nhập với nhau. Quá trình được thực hiện lặp lại cho đến khi mọi điểm của bề mặt địa hình đều được ngập nước.



Hình 4.5. Phân đoạn ảnh bằng thuật toán Marker-controlled watershed

Phân đoạn bằng kỹ thuật Watershed cho kết quả tốt hơn nếu có sự xác định hoặc đánh dấu các vị trí đối tượng và vị trí nền một cách hợp lý. Để giải quyết vấn đề này, phương pháp phân đoạn theo kiểm soát đánh dấu (marker controlled) Watershed được thực hiện như sau:

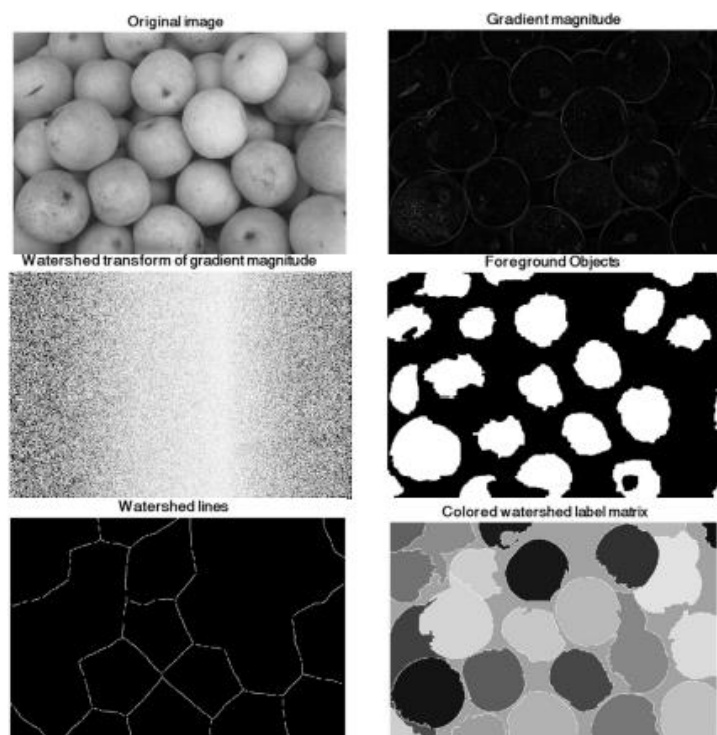
- 1) Tính hàm phân đoạn, các vùng tốt được xem là các đối tượng muốn phân loại

2) Tìm các đánh dấu vùng tiền cảnh (foreground markers) là các đốm (gồm tập điểm ảnh) liên thông trong vùng của mỗi đối tượng.

3) Tìm các đánh dấu vùng nền là các vùng điểm ảnh không thuộc về đối tượng nào cả trong ảnh.

4) Điều chỉnh phân đoạn sao cho cực tiểu hóa các vị trí đánh dấu tiền cảnh và nền.

5) Tính biến đổi Watershed theo hàm điều chỉnh phân đoạn



Hình 4.6. Minh họa các bước phân đoạn ảnh bằng Watershed

Thảo luận: Lấy demo minh họa các thuật toán phân đoạn

Câu hỏi

1. Phân đoạn ảnh là gì, lấy ví dụ minh họa.
2. Trình bày kỹ thuật phân ngưỡng thủ công, cho ví dụ.
3. Trình bày phương pháp lựa chọn ngưỡng, cho ví dụ.
4. Trình bày phương pháp ngưỡng thích nghi cục bộ, cho ví dụ.
5. Trình bày phương pháp phân đoạn ngưỡng kép, cho ví dụ.
6. Trình bày các bước của thuật toán phân đoạn bằng k-mean
7. Lấy ví dụ minh họa phân đoạn ảnh bằng thuật toán k-means

8. Trình bày các ưu nhược điểm của phân đoạn bằng k-means
9. Lập trình phương pháp phân đoạn ảnh bằng k-means
10. Trình bày các bước của thuật toán phân đoạn bằng MeanShift
11. Lấy ví dụ minh họa phân đoạn ảnh bằng thuật toán MeanShift
12. Trình bày các ưu nhược điểm của phân đoạn bằng MeanShift
13. Lập trình phương pháp phân đoạn ảnh bằng MeanShift
14. Trình bày các bước của thuật toán phân đoạn bằng Watershed
15. Lập trình phương pháp phân đoạn ảnh bằng Watershed

CHƯƠNG 5. PHƯƠNG PHÁP PHÁT HIỆN BIÊN

Nội dung chính của chương sẽ trình bày các khái niệm về biên, các phương pháp phát hiện biên đạo hàm bậc nhất, đạo hàm bậc hai.

Mục tiêu cần đạt được của chương

Sinh viên hiểu các khái niệm, phương pháp và biết cách thao tác các phương pháp đã học.

Bài 6: PHƯƠNG PHÁP PHÁT HIỆN BIÊN GRADIENT (Số tiết: 5 tiết)

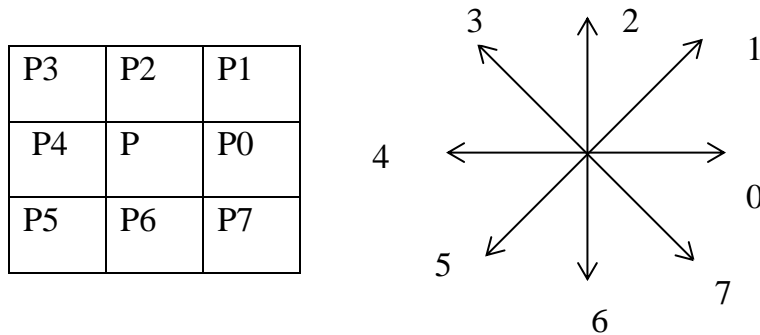
5.1 Các khái niệm

-Phát hiện biên

Phát hiện đường biên (cạnh) trong ảnh là một trong những phương pháp quan trọng trong xử lý ảnh. Có thể xem phương pháp phát hiện biên liên quan đến trích xuất đặc trưng mức thấp. Có nhiều cách tiếp cận khác nhau để biểu diễn ảnh, phân tích, ‘hiểu’ bức ảnh dựa trên đường biên từ việc phân tích và trích xuất cạnh. Cạnh thường ít nhạy cảm với sự thay đổi ánh sáng và thay đổi hướng nhìn. Phát hiện cạnh thường dựa vào sự tương phản cường độ sáng của đối tượng trong ảnh, độ tương phản cao thể hiện qua sự khác biệt lớn về cường độ sáng, sẽ làm nổi bật đặc trưng và chỉ ra được đường biên của đối tượng trong ảnh. Điều này tương tự với thị giác con người trong việc trực quan và ‘cảm nhận’ về đường biên bao quanh đối tượng từ đó xác định được đối tượng khác nhau một cách rõ ràng. Về cơ bản, ranh giới của đối tượng được xác định là thay đổi cường độ sáng đột ngột. Vị trí đường biên chính là vị trí xảy ra sự thay đổi ngược nhau về cường độ sáng. Để phát hiện vị trí cạnh, một phép biến đổi vi phân bậc nhất được thực hiện để làm nổi bật sự thay đổi ngược nhau qua giá trị đạo hàm bậc nhất, bậc hai để tìm cực trị.

Bài toán phát hiện đường biên đối tượng trong ảnh liên quan đến các bài toán phân loại và nhận dạng các đối tượng có trong ảnh. Do vậy, nhiệm vụ phát hiện đường biên được xem là một nhiệm vụ quan trọng trong xác định việc liệu đối tượng có tồn tại hay không, có bao nhiêu đối tượng hoặc có bao nhiêu thành phần nhỏ của đối tượng lớn hơn. Phát hiện đường biên đối tượng làm giảm không gian dữ liệu cần xử lý và loại bỏ các phần dữ liệu không cần thiết nhưng vẫn đảm bảo các thuộc tính quan trọng về cấu trúc của đối tượng trong bức ảnh. Có rất nhiều phương pháp phát hiện biên khác nhau, mỗi phương pháp thường hiệu quả cao đối với một số loại biên cụ thể. Các phương pháp phát hiện đường biên cơ bản liên quan đến các kỹ thuật lọc ảnh. Để lọc đường biên, người ta thường sử dụng một bộ lọc kích thước $n \times n$ với giá trị khởi tạo của bộ lọc được xác định theo một phương pháp nhất định. Một trong những bộ lọc được biết đến và sử

dụng nhiều là bộ lọc Roberts, Sobel, Prewitt kết quả của phép lọc tiếp tục được nhị phân bằng ngưỡng để phân loại.



Hình 5.1 Các lân cận của điểm P

Một số các định nghĩa

- Các điểm 4 và 8-láng giềng

Giả sử (i, j) là một điểm ảnh, các điểm 4-láng giềng là các điểm kề trên, dưới, trái, phải của (i, j) :

$$N_4(i, j) = \{(i', j') : |i - i'| + |j - j'| = 1\},$$

và những điểm 8-láng giềng gồm:

$$N_8(i, j) = \{(i', j') : \max(|i - i'|, |j - j'|) = 1\}.$$

P_0, P_1, \dots, P_7 là các 8 láng giềng của P

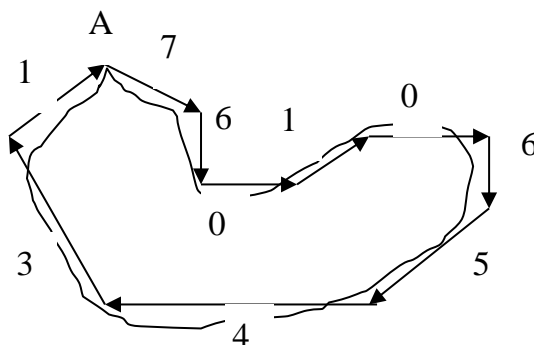
P_0, P_2, P_4, P_6 là các 4 láng giềng của P

$$N_4 = \{I(i-1, j), I(i+1, j), I(i, j-1), I(i, j+1)\}$$

$$N_8 = N_4 \cup \{I(i+1, j-1), I(i-1, j-1), I(i-1, j+1), I(i+1, j+1)\}$$

Chain Code (Mã xích)

Các vectơ giữa các điểm biên liên tiếp được mã hoá, sử dụng 8 hướng (từ 0-7), mỗi hướng được mã hoá bởi 3 bit. Mã xích chứa điểm bắt đầu theo sau bởi chuỗi các từ mã.



Hình 5.2 Ví dụ về chain code

A 111 110 000 001 000 110 101 100 011 001

5.2 Phương pháp phát hiện biên Gradient

Gradient là một vecto có các thành phần biểu thị tốc độ thay đổi giá trị của điểm ảnh theo 2 hướng x và y.

$$\text{gradient}(f(x, y)) = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right) = \frac{\partial f}{\partial x} \vec{i} + \frac{\partial f}{\partial y} \vec{j}$$

$$\frac{\partial f(x, y)}{\partial x} \approx \frac{f(x + \Delta x, y) - f(x, y)}{\Delta x}$$

$$\frac{\partial f(x, y)}{\partial y} \approx \frac{f(x, y + \Delta y) - f(x, y)}{\Delta y}$$

Trong đó $\Delta x, \Delta y$ là khoảng cách (tính bằng số điểm) theo hướng x và y.

*** Nhận xét:**

Tuy ta nói là lấy đạo hàm nhưng thực chất chỉ là mô phỏng và xấp xỉ đạo hàm bằng các kỹ thuật nhân chập vì ảnh số là tín hiệu rời rạc nên đạo hàm không tồn tại.

Lấy $\Delta x = \Delta y = 1$ hoặc $\Delta x = \Delta y = -1$

$$\frac{\partial f(x, y)}{\partial x} \approx f(x + 1, y) - f(x, y) \approx f(x, y) - f(x - 1, y)$$

$$\frac{\partial f(x, y)}{\partial y} \approx f(x, y + 1) - f(x, y) \approx f(x, y) - f(x, y - 1)$$

Do đó, mặt nạ nhân chập theo hướng x là $A = \begin{pmatrix} -1 & 1 \end{pmatrix}$

và hướng y là $B = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$

Với một ảnh liên tục $f(x, y)$ các đạo hàm riêng của nó cho phép xác định vị trí cực đại cục bộ theo hướng của biên.

Hướng cực đại của nó:

$$\text{tg } \varphi = \left(\frac{\partial f}{\partial y} \right) / \left(\frac{\partial f}{\partial x} \right)$$

$$\varphi = \arctg \left[\frac{\partial f}{\partial y} / \frac{\partial f}{\partial x} \right]$$

Độ lớn cực đại của vecto gradient

$$\sqrt{\left(\frac{\partial f(x, y)}{\partial x}\right)^2 + \left(\frac{\partial f(x, y)}{\partial y}\right)^2}$$

Thường để giảm thời gian tính toán, người ta còn tính gradient theo các chuẩn sau:

$$A_1 = |f_x + f_y|$$

$$\text{Hoặc } A_2 = \max(|f_x|, |f_y|)$$

Cho trước ngưỡng θ

$$\|gradient(x, y)\| \geq \theta \text{ thì } (x, y) \text{ là điểm biên}$$

Nhận thấy rằng việc lấy đạo hàm một tín hiệu có xu hướng làm tăng nhiễu trong tín hiệu đó. Độ nhạy cảm này có thể làm giảm xuống nhờ thao tác lấy trung bình cục bộ.

a) Kỹ thuật Prewitt

Kỹ thuật sử dụng 2 mặt nạ nhập chấp xấp xỉ đạo hàm theo 2 hướng x và y là:

$$H_x = \begin{pmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{pmatrix}$$

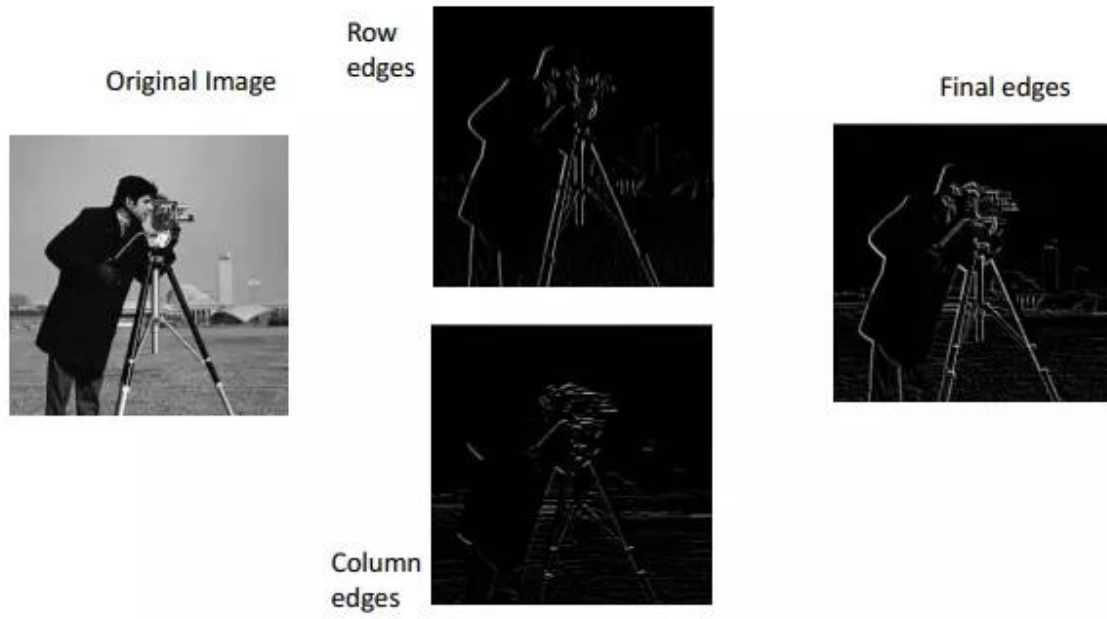
$$H_y = \begin{pmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

Các bước tính toán của kỹ thuật Prewitt

+ Bước 1: Tính $I \otimes H_x$ và $I \otimes H_y$

+ Bước 2: Tính $I \otimes H_x + I \otimes H_y$

+ Bước 3: Một ngưỡng được sử dụng để lọc trích chọn cạnh của đối tượng



Hình 5.3 Trích chọn cạnh bằng phương pháp Prewitt

b) Kỹ thuật Sobel

Đây là một trường hợp của đạo hàm bậc nhất

$$S_x = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix} \quad S_y = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{pmatrix}$$

Hai mặt nạ chỉ đơn thuần quay mặt nạ của nhau một góc $\pi/2$

Bước 1: $I \otimes S_x, I \otimes S_y$

Bước 2: $|I \otimes S_x(x, y)| + |I \otimes S_y(x, y)| \quad \forall x, y$

Bước 3: Tách ngưỡng theo θ

$$I'(x, y) = \begin{cases} 1 & \text{nếu } |I \otimes S_x(x, y)| + |I \otimes S_y(x, y)| \geq \theta \\ 0 & \text{nếu ngược lại} \end{cases}$$

$$I \otimes S_x \sim 8 \frac{\partial I(x, y)}{\partial x}$$

$$I \otimes S_y(x, y) \sim 8 \frac{\partial I(x, y)}{\partial y}$$

$$|I \otimes S_x(x,y)| + |I \otimes S_y(x,y)| \sim 8 \left(\left| \frac{\partial I}{\partial x} \right| + \left| \frac{\partial I}{\partial y} \right| \right)$$

ví dụ:

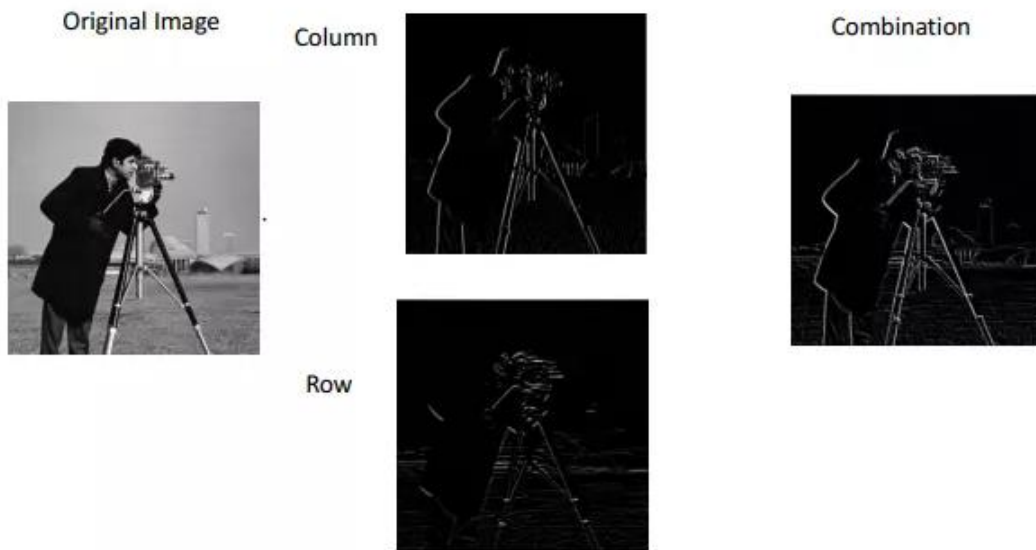
$$\begin{array}{cc} I & I \otimes S_x \\ \left(\begin{array}{cccccc} 7 & 6 & 5 & 4 & 3 & 2 & 1 \\ 0 & 7 & 6 & 5 & 4 & 3 & 2 \\ 0 & 0 & 7 & 6 & 5 & 4 & 3 \\ 0 & 0 & 0 & 7 & 6 & 5 & 4 \\ 0 & 0 & 0 & 0 & 7 & 6 & 5 \\ 0 & 0 & 0 & 0 & 0 & 7 & 6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 7 \end{array} \right) & \left(\begin{array}{cccccc} 19 & 2 & -6 & -6 & -6 & -6 & -7 \\ 20 & 17 & 0 & -8 & -8 & -8 & -12 \\ 7 & 20 & 17 & 0 & -8 & -8 & -16 \\ 0 & 7 & 20 & 17 & 0 & -8 & -20 \\ 0 & 0 & 7 & 20 & 17 & 0 & -24 \\ 0 & 0 & 0 & 7 & 20 & 17 & -20 \\ 0 & 0 & 0 & 0 & 7 & 20 & -7 \end{array} \right) \end{array}$$

$$\begin{array}{cc} I \otimes S_y & |I \otimes S_x| + |I \otimes S_y| \\ \left(\begin{array}{cccccc} -7 & -20 & -24 & -20 & -16 & -12 & -7 \\ 20 & 17 & 0 & -8 & -8 & -8 & -6 \\ 7 & 20 & 17 & 0 & -8 & -8 & -6 \\ 0 & 7 & 20 & 17 & 0 & -8 & -6 \\ 0 & 0 & 7 & 20 & 17 & 0 & -6 \\ 0 & 0 & 0 & 7 & 20 & 17 & -6 \\ 0 & 0 & 0 & 0 & 7 & 20 & 19 \end{array} \right) & \left(\begin{array}{cccccc} 26 & 22 & 30 & 26 & 22 & 18 & 14 \\ 40 & 34 & 0 & 16 & 16 & 16 & 18 \\ 14 & 40 & 34 & 0 & 16 & 16 & 22 \\ 0 & 14 & 40 & 34 & 0 & 16 & 26 \\ 0 & 0 & 14 & 40 & 34 & 0 & 30 \\ 0 & 0 & 0 & 14 & 40 & 34 & 26 \\ 0 & 0 & 0 & 0 & 14 & 40 & 26 \end{array} \right) \end{array}$$

$\theta=34$

$$\left(\begin{array}{cccccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{array} \right)$$

Toán tử Sobel được thiết kế để đáp ứng tối đa các cạnh theo chiều dọc và chiều ngang liên quan đến lưới các điểm ảnh. Nó giúp phân tích cạnh theo 2 hướng vuông góc với nhau.



Hình 5.4. Trích chọn cạnh bằng phương pháp Sobel

Thảo luận: Làm bài tập về kỹ thuật Sobel

c) -Kỹ thuật la bàn

Với mục đích nghiên cứu các mặt nạ cho kết quả tốt hơn, người ta nghĩ đến việc xem xét các lân cận theo nhiều hướng (8 hướng)

Cho T là mẫu.

Đặt $T_0 = T$, T_i nhận được từ T bằng cách quay góc $i \cdot \pi/4$:

$$A(x, y) = \max \{ | I \otimes T_i(x, y) | \mid i=0,1,2,\dots,7 \}$$

Toán tử Robinson bậc 5

$$S_0 = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}, S_1 = \begin{pmatrix} -2 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 2 \end{pmatrix}, S_2 = \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix}, S_3 = \begin{pmatrix} 0 & -1 & -2 \\ 1 & 0 & -1 \\ 2 & 1 & 0 \end{pmatrix}$$

$$S_4 = -S_0, S_5 = -S_1, S_6 = -S_2, S_7 = -S_3$$

Thảo luận: Cho Toán tử Kirsh, thực hiện xác định các mẫu theo kỹ thuật la bàn

$$H_0 = \begin{pmatrix} 5 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & -3 & -3 \end{pmatrix};$$

Các mẫu để tách cạnh, tổng các phần tử bằng 0

Phép toán làm nổi biên, tổng các phần tử bằng 1

Nếu các $|I \otimes T_i(x,y)|$ có giá trị gần như nhau thì các thông tin này không đáng tin cậy để xác định $I(x,y)$ là điểm biên.

Câu hỏi

1. Cho ảnh I sau:

$$I = \begin{pmatrix} 7 & 6 & 5 & 4 & 3 & 2 & 1 \\ 0 & 7 & 6 & 5 & 4 & 3 & 2 \\ 0 & 0 & 7 & 6 & 5 & 4 & 3 \\ 0 & 0 & 0 & 7 & 6 & 5 & 4 \\ 0 & 0 & 0 & 0 & 7 & 6 & 5 \\ 0 & 0 & 0 & 0 & 0 & 7 & 6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 7 \end{pmatrix}$$

Thực hiện kỹ thuật Prewitt

2. Cho ảnh I sau, ngưỡng $\theta = 15$

$$I = \begin{pmatrix} 3 & 4 & 5 & 6 & 7 \\ 9 & 0 & 1 & 2 & 4 \\ 3 & 1 & 5 & 7 & 9 \\ 5 & 6 & 7 & 2 & 1 \\ 2 & 3 & 4 & 8 & 2 \end{pmatrix}$$

Thực hiện kỹ thuật Sobel

3. Cho mẫu sau, thực hiện xác định các mẫu còn lại theo kỹ thuật la bàn

$$\begin{pmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{pmatrix}$$

4. Lập trình thực hiện các kỹ thuật đã học

BÀI 7: PHƯƠNG PHÁP PHÁT HIỆN BIÊN LAPLACE (Số tiết: 2 tiết)

5.3- Phương pháp phát hiện biên Laplace

Các phương pháp gradient ở trên làm việc khá tốt khi mà độ sáng thay đổi rõ nét. Khi mức xám thay đổi chậm, miền chuyển tiếp trải rộng, phương pháp cho hiệu quả hơn đó là phương pháp sử dụng đạo hàm bậc hai mà ta gọi là phương pháp Laplace. Toán tử Laplace được định nghĩa như sau:

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$$

$$\begin{aligned}\frac{\partial^2 f}{\partial x^2} &= \frac{\partial}{\partial x} \left(\frac{\partial f}{\partial x} \right) \approx \frac{\partial}{\partial x} (f(x+1, y) - f(x, y)) \\ &\approx [f(x+1, y) - f(x, y)] - [f(x, y) - f(x-1, y)] \\ &\approx f(x+1, y) - 2f(x, y) + f(x-1, y)\end{aligned}$$

Tương tự,

$$\begin{aligned}\frac{\partial^2 f}{\partial y^2} &= \frac{\partial}{\partial y} \left(\frac{\partial f}{\partial y} \right) \approx \frac{\partial}{\partial y} (f(x, y+1) - f(x, y)) \\ &\approx [f(x, y+1) - f(x, y)] - [f(x, y) - f(x, y-1)] \\ &\approx f(x, y+1) - 2f(x, y) + f(x, y-1)\end{aligned}$$

Vậy: $\nabla^2 f = f(x+1, y) + f(x, y+1) - 4f(x, y) + f(x-1, y) + f(x, y-1)$

Dẫn tới:

$$H = \begin{pmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

Kết quả nghiên cứu cho thấy phương pháp đạo hàm bậc hai rất nhạy cảm với nhiễu.

Đây là toán tử Laplace

$$L = \begin{pmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{pmatrix}$$

$$\begin{aligned}
I \otimes L &= 2I(x, y) - I(x-1, y) - I(x+1, y) \\
&\quad + 2I(x, y) - I(x, y-1) - I(x, y+1) \\
&= [I(x, y) - I(x-1, y)] - [I(x+1, y) - I(x, y)] \\
&\quad + I(x, y) - I(x, y-1) - [I(x, y+1) - I(x, y)] \\
&= \frac{\partial I(x-1, y)}{\partial x} - \frac{\partial I(x, y)}{\partial x} + \frac{\partial I(x, y-1)}{\partial y} - \frac{\partial I(x, y)}{\partial y} \\
&= -\frac{\partial^2 I(x, y)}{\partial x^2} - \frac{\partial^2 I(x, y)}{\partial y^2} \\
&= -\left(\frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2} \right) = -\nabla^2 I(x, y)
\end{aligned}$$

Phân ngưỡng: $|I \otimes L|$ theo $\theta > 0$

$$I'(x, y) = \begin{cases} 1 & \text{nếu } |I \otimes L(x, y)| \geq \theta \\ 0 & \text{nếu ngược lại} \end{cases}$$

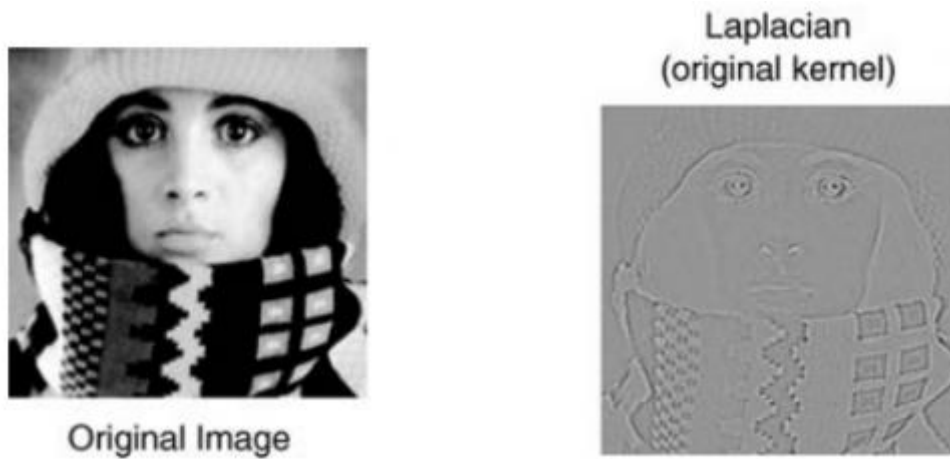
I	I ⊗ L
$ \begin{pmatrix} 6 & 5 & 4 & 3 & 2 & 1 & 0 \\ 0 & 6 & 5 & 4 & 3 & 2 & 1 \\ 0 & 0 & 6 & 5 & 4 & 3 & 2 \\ 0 & 0 & 0 & 6 & 5 & 4 & 3 \\ 0 & 0 & 0 & 0 & 6 & 5 & 4 \\ 0 & 0 & 0 & 0 & 0 & 6 & 5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 6 \end{pmatrix} $	$ \begin{pmatrix} 19 & 4 & 3 & 2 & 1 & 0 & -2 \\ -12 & 14 & 0 & 0 & 0 & 0 & 0 \\ 0 & -12 & 14 & 0 & 0 & 0 & 1 \\ 0 & 0 & -12 & 14 & 0 & 0 & 2 \\ 0 & 0 & 0 & -12 & 14 & 0 & 3 \\ 0 & 0 & 0 & 0 & -12 & 14 & 4 \\ 0 & 0 & 0 & 0 & 0 & -12 & 19 \end{pmatrix} $

$$\theta=14$$

Ảnh kết quả

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Kỹ thuật này cho đường biên mảnh, tuy nhiên rất nhạy cảm với nhiễu do đạo hàm bậc hai thường không ổn định



Hình 5.5 Trích chọn biên bằng phương pháp Laplace

Trong thực tế người ta thường dùng 1 số biến dạng khác nhau của toán tử Laplace

$$L2 = \begin{pmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{pmatrix} L3 = \begin{pmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{pmatrix}$$

Thảo luận: Thực hiện nhân chập ảnh I với mẫu L2, L3

5.4. Phương pháp Canny

Đây là một thuật toán tương đối tốt, có khả năng đưa ra đường biên mảnh, và phát hiện chính xác điểm biên với điểm nhiễu. Phương pháp Canny được đề xuất năm 1986 bởi John Canny. Ngày nay Canny được dùng khá phổ biến trong nhiều lĩnh vực khác nhau của thị giác máy tính

Thuật toán Canny gồm các bước chính được mô tả như sau

+ **Bước 1:** Làm trơn ảnh

Việc loại bỏ các yếu tố nhiễu là đặc biệt quan trọng đối với kết quả phát hiện cạnh. Sử dụng bộ lọc Gaussian để làm trơn ảnh nhằm khử nhiễu và tác động của nhiễu.

Bước này thực hiện tích chập ảnh đầu vào với mặt nạ trọng số theo phân phối Gaussian (gọi là mặt nạ Gaussian). Mặt nạ Gaussian H có thể khởi tạo theo công thức:

$$H_{ij} = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(i-(k+1))^2 + (j-(k+1))^2}{2\sigma^2}\right) \text{ ở đó } 1 \leq i, j \leq (2k+1)$$

Tính $I \otimes H$

Ví dụ mặt nạ Gaussian được xấp xỉ theo phân phối với độ lệch chuẩn $\sigma=1.4$, kích thước 5x5

$$H = \frac{1}{159} \begin{bmatrix} 2 & 4 & 5 & 4 & 2 \\ 4 & 9 & 12 & 9 & 4 \\ 5 & 12 & 15 & 12 & 5 \\ 4 & 9 & 12 & 9 & 4 \\ 2 & 4 & 5 & 4 & 2 \end{bmatrix}$$

Gọi G là kết quả lọc nhiễu: $G = I \otimes H$

+ **Bước 2:** Tính gradient ảnh dựa vào cường độ sáng điểm ảnh (thường áp dụng trên ảnh đa mức xám)

Bước này có thể được thực hiện bằng cách tích chập ảnh kết quả của bước 1 với hai toán tử Sobel theo 2 hướng dọc (trục tung) và hướng ngang (trục hoành) để xấp xỉ với đạo hàm bậc nhất theo hướng dọc G_y và ngang G_x . Sau đó tính độ lớn và hướng của gradient theo công thức.

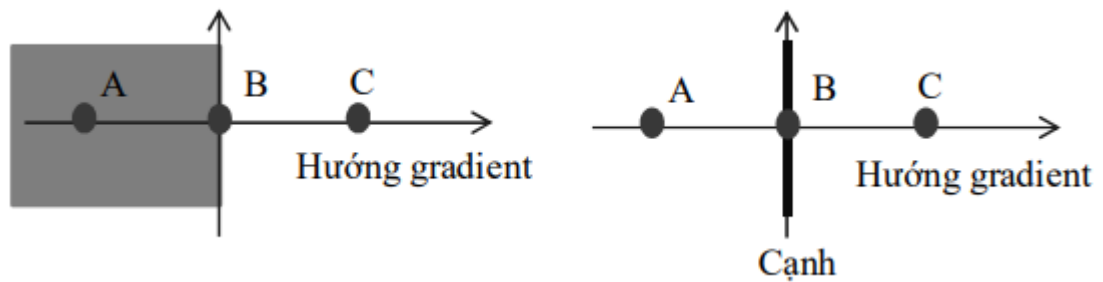
Độ lớn gradient $G_M = \sqrt{G_x^2 + G_y^2}$

Hướng gradient $G_0 = \arctan(G_y/G_x)$

+ **Bước 3:** Thực hiện loại bỏ các điểm không cực đại (non-maximum suppression) nhằm khử những giá trị thuộc cạnh giả tạo hoặc thuộc cạnh nhưng không bền vững.

Sau khi tính giá trị hướng và độ lớn của gradient, thực hiện quét qua toàn bộ ảnh để loại bỏ những điểm không mong muốn (những điểm có thể không tạo thành cạnh). Để thực hiện điều này, một điểm được đánh dấu lại nếu có giá trị cực đại

cực bộ trong vùng láng giềng của nó theo hướng gradient hiện hành, như hình minh họa.



Hình 5.6. Minh họa về tìm cực đại cực bộ theo hướng gradient

Trong minh họa này, điểm B nằm trên cạnh theo chiều dọc. Hướng gradient là vector pháp tuyến của cạnh. Điểm A và điểm C thuộc hướng gradient tại điểm B. Do đó, điểm B đánh dấu là cực đại so với điểm A và điểm C nếu nó có giá trị cực đại cực bộ (cực đại so với điểm A và điểm C). Nghĩa là điểm A và C là hai điểm ứng viên nhưng không cực đại nên bị loại bỏ. Nếu nó được đánh dấu lại sẽ được tiếp tục thực hiện bước tiếp theo, ngược lại điểm này sẽ bị xóa bỏ và không được xử lý ở các bước sau.

Kết quả của bước này cho kết quả là một ảnh (ma trận) nhị phân với các điểm ảnh thuộc cạnh đã “làm mỏng”. Tập này gồm tập các điểm được đánh dấu và các điểm nền là các điểm còn lại (không được đánh dấu).

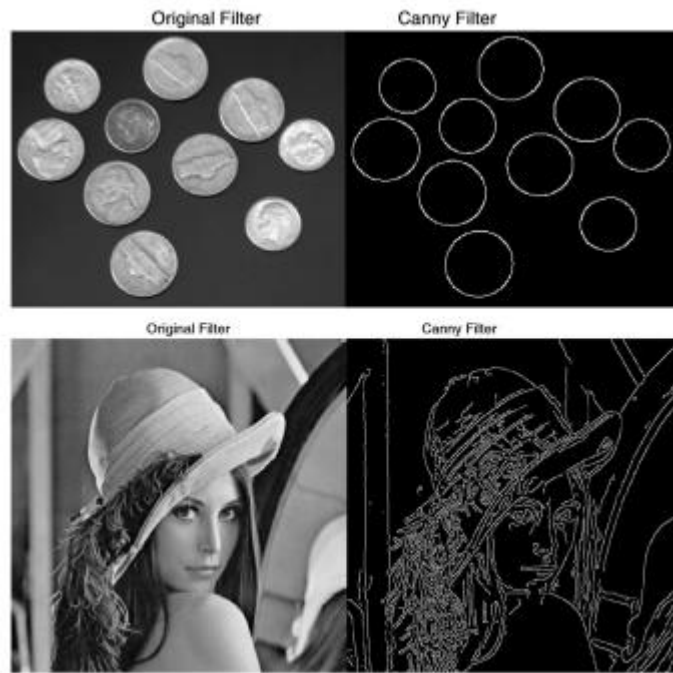
+ **Bước 4:** Sử dụng ngưỡng kép để phát hiện các đường biên tiềm năng

Bước này mới chỉ cho ra kết quả là các điểm tiềm năng (ứng cử viên) thuộc về đường biên. Thay vì sử dụng một ngưỡng, Canny dùng hai ngưỡng t_1, t_2 ($t_1 < t_2$) để lọc ra những điểm ảnh thuộc cạnh hoặc không. Tại những vị trí có giá trị gradient lớn hơn t_2 thì được xem như chắc chắn điểm thuộc cạnh. Những điểm thuộc miền giữa t_1 và t_2 thì nó không chắc chắn thuộc cạnh, có thể thuộc cạnh hoặc không, những điểm này được xem là điểm có tiềm năng thuộc cạnh tùy thuộc vào sự liên kết với cạnh. Những điểm còn lại có giá trị bé hơn t_1 thì không thuộc cạnh và được loại trực tiếp.

+ **Bước 5:** Liên kết điểm tiềm năng với các điểm thuộc cạnh.

Nếu như các điểm tiềm năng được liên kết với các điểm chắc chắn thuộc cạnh thì nó sẽ trở thành một phần của cạnh. Nghĩa là nếu điểm tiềm năng là lân

cận của điểm đã được xác định thuộc cạnh (ở bước 4) thì điểm này cũng được đánh dấu là điểm thuộc cạnh.



Hình 5.7. Kết quả trích xuất cạnh dùng thuật toán Canny

Câu hỏi

1. Cho ảnh I sau:

$$I = \begin{pmatrix} 7 & 6 & 5 & 4 & 3 & 2 & 1 \\ 0 & 7 & 6 & 5 & 4 & 3 & 2 \\ 0 & 0 & 7 & 6 & 5 & 4 & 3 \\ 0 & 0 & 0 & 7 & 6 & 5 & 4 \\ 0 & 0 & 0 & 0 & 7 & 6 & 5 \\ 0 & 0 & 0 & 0 & 0 & 7 & 6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 7 \end{pmatrix}$$

Thực hiện kỹ thuật Sobel với ngưỡng $\theta = 15$

2. Trình bày kỹ thuật Canny
3. Cho ví dụ về kỹ thuật Canny
4. Cài đặt kỹ thuật Sobel
5. Cài đặt kỹ thuật Canny

6. Viết đoạn chương trình đọc vào một ảnh ngoại cảnh sau đó sử dụng các toán tử Sobel, Prewitt, Robert để trích biên và so sánh kết quả.
7. Viết đoạn chương trình đọc vào một ảnh ngoại cảnh sau đó sử dụng phương pháp Canny để trích biên và hiển thị kết quả từng bước trong quá trình thực hiện thuật toán Canny. So sánh kết quả với các phương pháp trích biên sử dụng toán tử Sobel, Prewitt, Robert.
8. Viết chương trình để nhận dạng cạnh theo các phương pháp đã học và sau đó đánh giá, so sánh khả năng thực hiện của các phương pháp này.
9. Viết đoạn chương trình đọc vào ảnh dưới đây và thực hiện lọc cạnh bằng phương pháp Canny

Chương 6: BIẾN ĐỔI HÌNH THÁI HỌC

Nội dung chính của chương sẽ trình bày các phép biến đổi hình thái học và các ứng dụng.

Mục tiêu cần đạt được của chương

Sinh viên hiểu được định nghĩa phần tử cấu trúc, các phép biến đổi hình thái như phép giãn nở, phép co, phép đóng, phép mở, hit or miss và các ứng dụng.

Bài 8: BIẾN ĐỔI HÌNH THÁI HỌC (Số tiết: 5 tiết)

Ảnh nhị phân có thể chứa thông tin không đầy đủ, đặc biệt các vùng ảnh nhị phân là kết quả của những bộ lọc theo ngưỡng đơn giản, kết quả có thể bị méo hoặc nhiễu. Xử lý hình thái học đối với ảnh nhằm mục đích khử nhiễu và khôi phục những khuyết thiếu trong quá trình lọc phân đoạn ảnh nhằm làm cho kết quả thu được có chất lượng tốt hơn bằng cách thực hiện các phép biến đổi về hình thức và cấu trúc hình ảnh. Như vậy, biến đổi hình thái học trong xử lý ảnh là một tập hợp các phương pháp phi tuyến tính liên quan đến hình dáng và sắc thái của đối tượng. Các biến đổi hình thái chỉ thực hiện trên giá trị điểm ảnh nhị phân, tuy nhiên nó có thể mở rộng để áp dụng cho ảnh đa mức xám. Các kỹ thuật hình thái học trong xử lý dùng các mẫu hình dáng nhỏ được gọi là phần tử cấu trúc. Các phép toán (có thể gọi là toán tử) hình thái thường được xây dựng từ hai phép toán cơ bản là phép toán co (Erosion) và phép giãn (Dilation). Một vấn đề quan trọng trong sử dụng phép biến đổi hình thái là lựa chọn một phần tử cấu trúc có hình dạng thích hợp nhằm đạt được mục tiêu đặt ra và tối ưu quá trình tính toán.

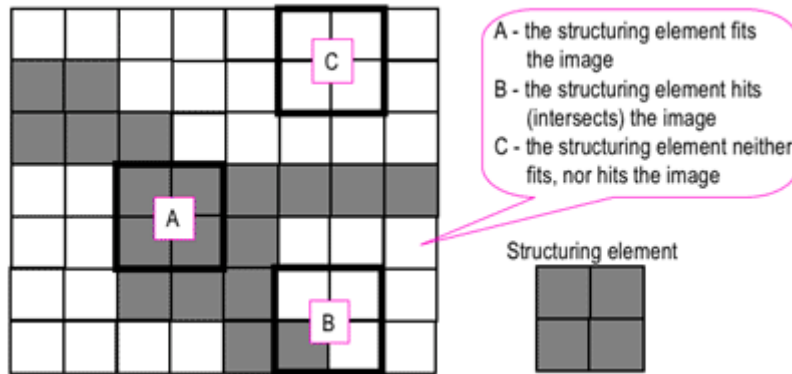
6.1. Phần tử cấu trúc

Đối với ảnh nhị phân, phần tử cấu trúc là một mẫu có kích thước nhỏ, giá trị phần tử gồm 0 và 1. Hay nói cách khác, phần tử cấu trúc là một ảnh nhị phân nhỏ, có thể là một ma trận gồm các pixel mà mỗi phần tử của nó có giá trị bằng 0 hoặc bằng 1. Trong đó, các giá trị phần tử bằng 0 được bỏ qua trong tính toán và chỉ thực hiện trên các phần tử có giá trị bằng 1. Phần tử cấu trúc ảnh nhị phân được thể hiện như sau:

$$S(i, j) \in \{0, 1\}$$

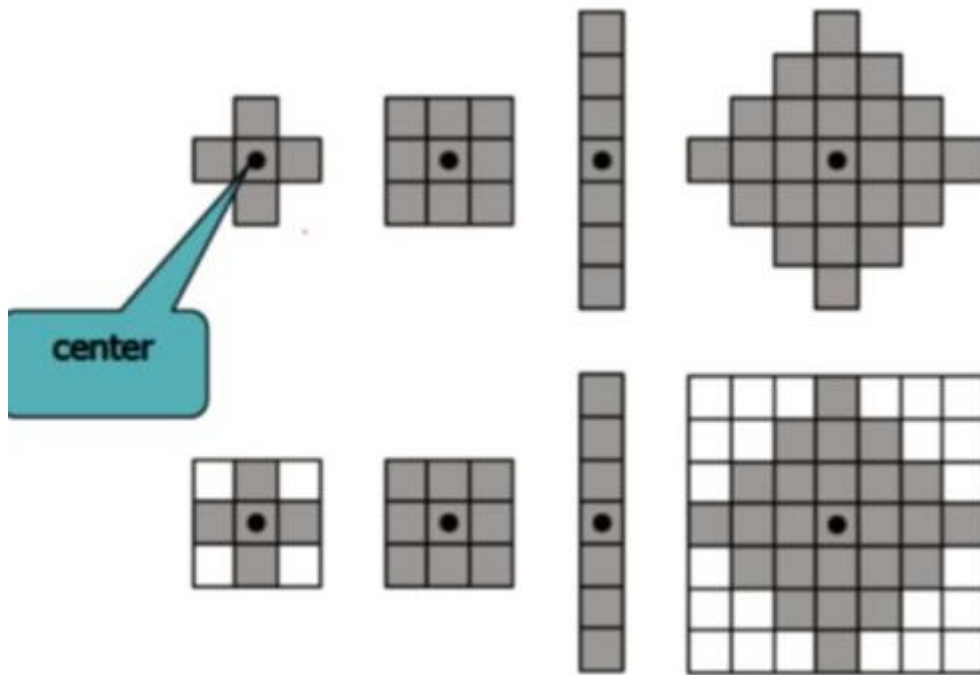
Trong quá trình xử lý hình thái học, phần tử cấu trúc được duyệt (trượt) qua tất cả các vị trí có thể trên ảnh và áp dụng (so sánh) trong vùng lân cận của điểm ảnh đang xét với kích thước vùng xác định trước theo phần tử cấu trúc. Một số toán tử kiểm tra liệu một phần tử “vừa” (fit) nằm trong vùng lân cận hay không hoặc có những toán tử

khác kiểm tra xem nó có “đạt tới” (hit) hoặc giao nhau với vùng lân cận hay không. Trong hình minh họa dưới, phần tử cấu trúc A là “vừa” nằm trong vùng lân cận, phần tử cấu trúc B là giao nhau với vùng lân cận và phần tử cấu trúc C là chưa với tới vùng lân cận của điểm ảnh đang xét



Hình 6.1. Minh họa các trường hợp của phần tử cấu trúc

Một toán tử hình thái trên ảnh nhị phân tạo ra một ảnh nhị phân mới với pixel có giá trị khác 0 nếu và chỉ nếu việc kiểm tra thành công tại vị trí đang xét trên ảnh đầu vào. Một số hình dạng của phần tử cấu trúc thường được sử dụng trên ảnh nhị phân gồm hình dạng đường theo chiều ngang và dọc, hình vuông, hình ellipse,... Các chiều của ma trận xác định kích thước của phần tử cấu trúc và đồng thời cũng xác định kích thước vùng lân cận trong quá trình xử lý hình thái học. Mẫu bố trí các số 1 và các số 0 xác định hình dạng của phần tử cấu trúc. Điểm gốc của phần tử cấu trúc thường được xác định tại một trong các phần tử của ma trận phần tử cấu trúc, mặc dù tổng quát về mặt lý thuyết, nó có thể nằm ngoài phần tử cấu trúc.



Hình 6.2. Ví dụ các phần tử cấu trúc đơn giản

6.2 Phép giãn ảnh - Dilation

$\forall p$ lân cận $v(p)$

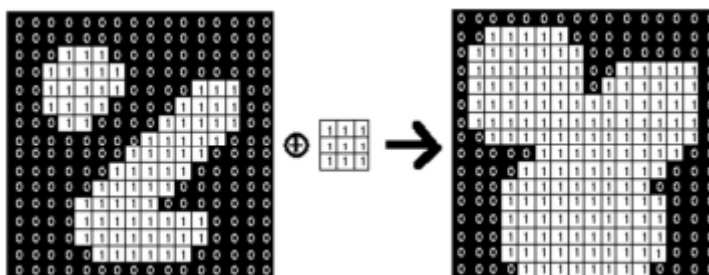
$$D(A) = \bigcup_{p \in A} v(p)$$

Giả sử có ảnh I và phần tử cấu trúc T khi đó:

Phép giãn nở của I theo mẫu T là tập hợp tất cả các x sao cho

$$I \oplus T = \bigcup_{x \in I} T_x$$

Trong đó T_x là dịch chuyển của T đến vị trí x . (gốc của T được đặt tại vị trí x)



Hình 6.3 Làm đậm vùng bằng phép toán giãn nở

Xét ảnh đen trắng

Lấy một điểm ở trong T làm gốc và dịch chuyển gốc của T đến vị trí có giá trị là 1 trong ảnh (giá trị khác 0) và thay thế phần tử 1 bằng mẫu T, làm đối với mọi tọa độ, sau đó hợp kết quả lại thì được phép giãn nở của I theo T.

$$D(I) = I \oplus T$$

$$D(I) = I'$$

$$I'(x,y) = \begin{cases} T & \text{nếu } I(x,y)=1 \\ I(x,y) & \text{nếu ngược lại} \end{cases}$$

Với phép toán D(Dilation) cho ảnh $I_{M \times N}$, mẫu $T_{m \times n}$, thì $D(I)_{M+m-1, N+n-1}$

-Phép toán “ \vee ” và “ \wedge ” trong ảnh:

\vee	1	0	*
1	1	1	1
0	1	0	*
*	1	*	*

\wedge	1	0	*
1	1	0	*
0	0	0	0
*	*	0	*

* là 1 giá trị điểm ảnh bất kỳ trong ảnh đa cấp xám

$$I_{src} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \quad I_{dst} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Ví dụ1:

$$I = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \end{pmatrix} \quad T = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

Hợp tất cả các hàng lại với nhau ta có kết quả

$$I \oplus T = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

Ví dụ 2:

$$I = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} T = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \end{pmatrix} \Rightarrow D(I) = \begin{pmatrix} 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 \end{pmatrix}$$

Phép toán giãn nở ảnh thường dùng để nối các đường đứt nét.

Thảo luận: Thực hiện giãn nở ảnh I theo phần tử cấu trúc T đã cho (x ký hiệu vị trí gốc của T).

$$I = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix} \quad T = \begin{pmatrix} 1 & x & 1 \end{pmatrix}$$

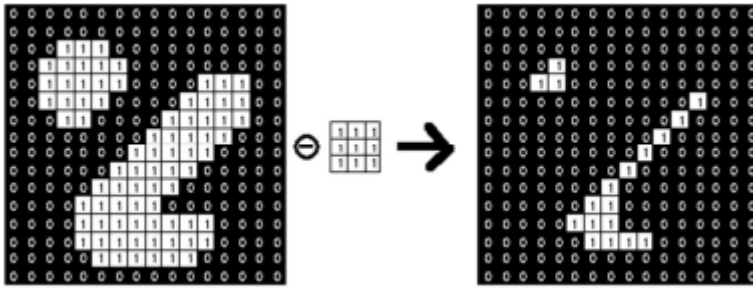
6.3 Phép co ảnh- EROSION

$$E(A) = \{p / v(p) \subset A\}$$

Mục đích: Xoá nhiễu

Toán tử Erosion trên ảnh nhị phân I với phần tử cấu trúc T là tập hợp các điểm x sao cho T_x là tập con của I.

Kết hợp với phép Dilation để làm trơn biên ảnh và lấp các lỗ hổng trên ảnh



Hình 6.3 Làm mỏng vùng bằng phép toán co ảnh

Cho ảnh $I_{M \times N}$, $T_{m \times n}$, $E(I) = I \ominus T = I'$

$$I'(x, y) = \begin{cases} 1 & \text{nếu cửa sổ đang xét có mẫu } \equiv T \\ 0 & \text{nếu ngược lại} \end{cases}$$

$$E(I) = I \ominus T = \{x \mid T_x \subseteq I\}$$

Phương pháp cụ thể: chọn một điểm trên T làm gốc(=1)

Ví dụ 1:

$$I = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \end{pmatrix} T = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$

$$E(I) = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$$

Ví dụ 2:

$$I = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} T = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix} \Rightarrow E(I) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

6.4. Một số tính chất

6.4.1 Tính gia tăng

$$X \subseteq X' \rightarrow X \oplus B \subseteq X' \oplus B \quad \forall B$$

$$X \ominus B \subseteq X' \ominus B \quad \forall B$$

$$B \subseteq B' \rightarrow X \oplus B \subseteq X \oplus B' \quad \forall X$$

$$X \ominus B \supseteq X \ominus B' \quad \forall X$$

Chứng minh

$$+ X \oplus B = \bigcup_{x \in X} B_x \subseteq \bigcup_{x \in X'} B_x = X' \oplus B$$

$$+ X \ominus B = \{x \mid B_x \subseteq X\} \subseteq \{x \mid B_x \subseteq X'\} = X' \ominus B$$

$$+ X \oplus B = \bigcup_{x \in X} B_x \subseteq \bigcup_{x \in X} B'_x = X \oplus B'$$

$$+ X \ominus B = \{x \mid B_x \subseteq X\} \supseteq \{x \mid B'_x \subseteq X\} = X \ominus B'$$

6.4.2- Tính phân phối với phép \cup

$$(i) \quad X \oplus (B \cup B') = (X \oplus B) \cup (X \oplus B')$$

$$(ii) \quad X \ominus (B \cup B') = (X \ominus B) \cap (X \ominus B')$$

Chứng minh:

$$(i) \quad X \oplus (B \cup B') = (X \oplus B) \cup (X \oplus B')$$

Ta có: $B \cup B' \supseteq B$

$$\Rightarrow X \oplus (B \cup B') \supseteq X \oplus B \quad (\text{tính gia tăng})$$

Tương tự:

$$\Rightarrow X \oplus (B \cup B') \supseteq X \oplus B'$$

$$\Rightarrow X \oplus (B \cup B') \supseteq (X \oplus B) \cup (X \oplus B') \quad (1)$$

Mặt khác,

$$\forall y \in X \oplus (B \cup B') \Rightarrow \exists x \in X \text{ sao cho } y \in (B \cup B')_x$$

$$\Rightarrow \begin{cases} y \in B_x \\ y \in B'_x \end{cases} \Rightarrow \begin{cases} y \in X \oplus B \\ y \in X \oplus B' \end{cases}$$

$$\Rightarrow y \in (X \oplus B) \cup (X \oplus B')$$

$$\Rightarrow X \oplus (B \cup B') \subseteq (X \oplus B) \cup (X \oplus B') \quad (2)$$

Từ (1) và (2) ta có: $X \oplus (B \cup B') = (X \oplus B) \cup (X \oplus B')$

$$(ii) X \ominus (B \cup B') = (X \ominus B) \cap (X \ominus B')$$

Ta có: $B \cup B' \supseteq B$

$$\Rightarrow X \ominus (B \cup B') \subseteq X \ominus B \quad (\text{tính gia tăng})$$

Tương tự: $X \ominus (B \cup B') \subseteq X \ominus B'$

$$\Rightarrow X \ominus (B \cup B') \subseteq (X \ominus B) \cap (X \ominus B') \quad (3)$$

Mặt khác,

$$\forall x \in (X \ominus B) \cap (X \ominus B')$$

$$\text{Suy ra, } \begin{cases} x \in X \ominus B \\ x \in X \ominus B' \end{cases} \Rightarrow \begin{cases} B_x \subseteq X \\ B'_x \subseteq X \end{cases}$$

$$\Rightarrow (B \cup B')_x \subseteq X$$

$$\Rightarrow x \in X \ominus (B \cup B')$$

$$\Rightarrow X \ominus (B \cup B') \supseteq (X \ominus B) \cap (X \ominus B') \quad (4)$$

Từ (3) và (4) ta có: $X \ominus (B \cup B') = (X \ominus B) \cap (X \ominus B')$.

Ý nghĩa: ta có thể phân tích các mẫu phức tạp trở thành các mẫu đơn giản, thuận tiện cho việc cài đặt.

6.4.3. Tính phân phối với phép \cap

$$(X \cap Y) \ominus B = (X \ominus B) \cap (Y \ominus B)$$

Chứng minh

Ta có, $X \cap Y \subseteq X$

$$\Rightarrow (X \cap Y) \ominus B \subseteq X \ominus B$$

Tương tự: $(X \cap Y) \ominus B \subseteq Y \ominus B$

$$\Rightarrow (X \cap Y) \ominus B \subseteq (X \ominus B) \cap (Y \ominus B) \quad (5)$$

Mặt khác,

$$\forall x \in (X \ominus B) \cap (Y \ominus B)$$

$$\text{Suy ra } \begin{cases} x \in X \ominus B \\ x \in Y \ominus B \end{cases} \Rightarrow \begin{cases} B_x \subseteq X \\ B_x \subseteq Y \end{cases}$$

$$\Rightarrow B_x \subseteq X \cap Y$$

$$\Rightarrow x \in (X \cap Y) \ominus B$$

$$\Rightarrow (X \cap Y) \ominus B \supseteq (X \ominus B) \cap (Y \ominus B) \quad (6)$$

Từ (5) và (6) ta có: $(X \cap Y) \ominus B = (X \ominus B) \cap (Y \ominus B)$.

6.4.4 Tính kết hợp

$$\text{i) } (X \oplus B) \oplus B' = X \oplus (B \oplus B')$$

$$\text{ii) } (X \ominus B) \ominus B' = X \ominus (B \oplus B')$$

Chứng minh:

$$\text{(i) } (X \oplus B) \oplus B' = X \oplus (B' \oplus B)$$

$$\begin{aligned} \text{Ta có, } (X \oplus B) \oplus B' &= \left(\bigcup_{x \in X} B_x \right) \oplus B' \\ &= \bigcup_{x \in X} (B_x \oplus B') = \bigcup_{x \in X} (B \oplus B')_x \\ &= X \oplus (B' \oplus B) \end{aligned}$$

$$\text{(i) } (X \ominus B) \ominus B' = X \ominus (B \oplus B')$$

Trước hết ta đi chứng minh: $B'_x \subseteq X \ominus B \Leftrightarrow (B' \oplus B)_x \subseteq X$

Thật vậy, do $B'_x \subseteq X \ominus B$ nên $\forall y \in B'_x \Rightarrow y \in X \ominus B$

$$\Rightarrow B_y \subseteq X$$

$$\Rightarrow \bigcup_{y \in B'_x} B_y \subseteq X$$

$$\Rightarrow (B' \oplus B)_x \subseteq X$$

Mặt khác, $(B' \oplus B)_x \subseteq X \Leftrightarrow (B'_x \oplus B) \subseteq X$

$$\Leftrightarrow \bigcup_{y \in B'_x} B_y \subseteq X$$

$$\Rightarrow \forall y \in B'_x \text{ ta có } B_y \subseteq X$$

$$\Rightarrow \text{hay } \forall y \in B'_x \text{ ta có } y \in X \ominus B$$

$$\text{Do đó, } B'_x \subseteq X \ominus B$$

$$\text{Ta có, } (X \ominus B) \ominus B' = \{x / B_x \subseteq X\} \ominus B'$$

$$= \{x / B'_x \subseteq X \ominus B\}$$

$$= \{x / (B' \oplus B)_x \subseteq X\} \text{ (do chứng minh ở trên)}$$

$$= X \ominus (B \oplus B').$$

6.4.5 Tính đối ngẫu

$$X^c \oplus B = (X \ominus B)^c$$

X là một đối tượng thì X^c là nền

6.4.6 Một số tính chất khác

$$A \oplus B = B \oplus A$$

$$A \ominus B \neq B \ominus A$$

$$(A \ominus B) \oplus B \neq A \neq (A \oplus B) \ominus B$$

Nhận xét: Trong quá trình thực hiện có thể có một số thao tác ra ngoài ảnh ta có thể mở rộng ảnh với phần mở rộng xem như là nền. Khi kết thúc thao tác thì trả lại ảnh ban đầu.

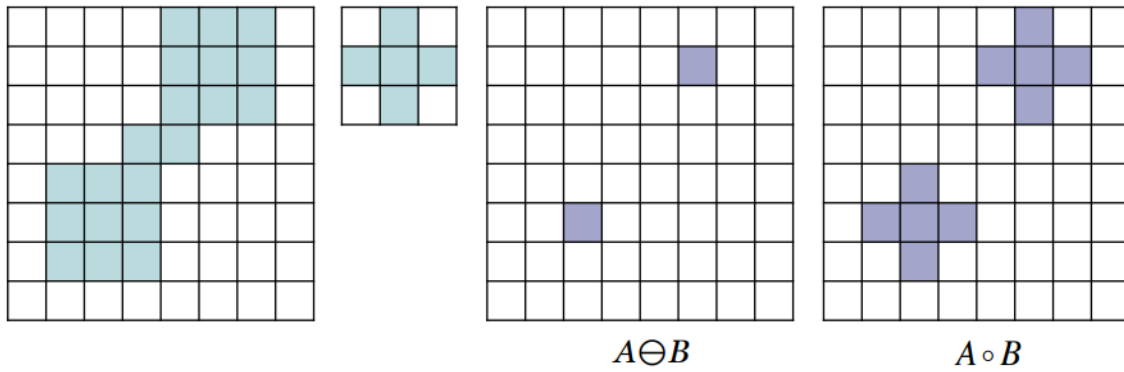
6.5- Các phép toán đóng/ mở

- Phép mở ảnh -Opening

Cho ảnh nhị phân I và T là phần tử cấu trúc, phép mở ảnh (opening) (ký hiệu là o) giữa ảnh I và phần tử cấu trúc T được xác định bởi công thức sau:

$$\text{OPEN}(I) = D(E(I)) = I \circ T = (I \ominus T) \oplus T$$

$$A \circ B = (A \ominus B) \oplus B$$



Hình 6.4 Ví dụ phép mở ảnh



Hình 6.5. Phép toán mở rộng ảnh để khử các đối tượng nhỏ

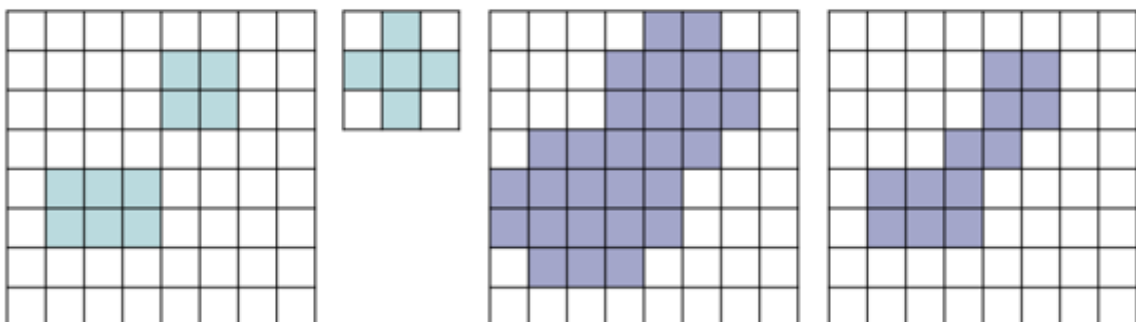
Phép mở xóa bỏ những đoạn mảnh, loại bỏ nhiễu nhưng làm tăng số đoạn đứt gãy

Phép đóng ảnh -Closing

Tập hợp I là ảnh gốc, T là phần tử cấu trúc và phép đóng ảnh ký hiệu là \bullet

$$\text{CLOSE}(I) = E(D(I)) = I \bullet T = (I \oplus T) \ominus T$$

$$A \bullet B = (A \oplus B) \ominus B$$



Hình 6.6 Ví dụ phép đóng ảnh



Hình 6.7 Phép đóng khử những lỗ nhỏ trong đối tượng

Trong xử lý ảnh, phép đóng và phép mở là những phép xử lý cơ bản trong khử nhiễu hình thái học. Phép mở giúp loại bỏ những đối tượng nhỏ trong khi phép đóng giúp loại bỏ những lỗ nhỏ trong đối tượng.

Ví dụ:

$$I = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \end{pmatrix}, T = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Tính $D(I)$, $E(I)$, $OPEN(I)$, $CLOSE(I)$

$$I \oplus T = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{pmatrix} \quad I \oplus T \ominus T = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{pmatrix}$$

$$I \ominus T = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix} \quad I \ominus T \oplus T = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$

Một số tính chất

- X bị chặn bởi các cận OPEN và CLOSE

Giả sử, X là một đối tượng ảnh, B là mẫu, khi đó, X sẽ bị chặn trên bởi tập CLOSE của X theo B và bị chặn dưới bởi tập OPEN của X theo B. Tức là:

$$(X \oplus B) \ominus B \supseteq X \supseteq (X \ominus B) \oplus B$$

Chứng minh:

Ta có: $\forall x \in X \Rightarrow B_x \subseteq X \oplus B$ (Vì $X \oplus B = \bigcup_{x \in X} B_x$)

$\Rightarrow x \in (X \oplus B) \ominus B$ (theo định nghĩa phép co)

$\Rightarrow (X \oplus B) \ominus B \supseteq X$ (7)

Mặt khác,

$\forall y \in (X \ominus B) \oplus B$, suy ra:

$\exists x \in X \ominus B$ sao cho $y \in B_x$ (Vì $(X \ominus B) \oplus B = \bigcup_{x \in X \ominus B} B_x$)

$\Rightarrow B_x \subseteq X \Rightarrow y \in X$

Suy ra: $X \supseteq (X \ominus B) \oplus B$ (8)

Từ (7) và (8) Ta có: $(X \oplus B) \ominus B \supseteq X \supseteq (X \ominus B) \oplus B$.

Tính bất biến :

(i) $((X \oplus B) \ominus B) \oplus B = X \oplus B$

(ii) $((X \ominus B) \oplus B) \ominus B = X \ominus B$

chứng minh

Thật vậy, từ t/c trên ta có $X \subseteq (X \oplus B) \ominus B$

$\Rightarrow X \oplus B \subseteq ((X \oplus B) \ominus B) \oplus B$ (do tính chất gia tăng) (9)

Mặt khác, cũng từ t/c trên ta có $(X \ominus B) \oplus B \subseteq X \forall X$

Do đó, thay X bởi $X \oplus B$ ta có, $((X \oplus B) \ominus B) \oplus B \subseteq X \oplus B$ (10)

Từ (9) và (10) Ta có: $((X \oplus B) \ominus B) \oplus B = X \oplus B$

Tương tự ta có, $((X \ominus B) \oplus B) \ominus B = X \ominus B$

một số tính chất khác

$A_1 \subseteq A_2$

$\text{OPEN}(A_1) \subseteq \text{OPEN}(A_2)$

$$\text{OPEN}(\text{OPEN}(I)) = \text{OPEN}(I)$$

$$\text{CLOSE}(A_1) \subseteq \text{CLOSE}(A_2)$$

$$\text{CLOSE}(\text{CLOSE}(I)) = \text{CLOSE}(I)$$

Thảo luận

Thực hiện phép đóng, phép mở ảnh I theo phần tử cấu trúc T đã cho (x ký hiệu vị trí gốc của T).

$$I = \begin{matrix} & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ I = & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{matrix} \quad T = \begin{matrix} & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \end{matrix} \quad \begin{matrix} & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \end{matrix}$$

6.6- Phép biến đổi “Hit or miss”

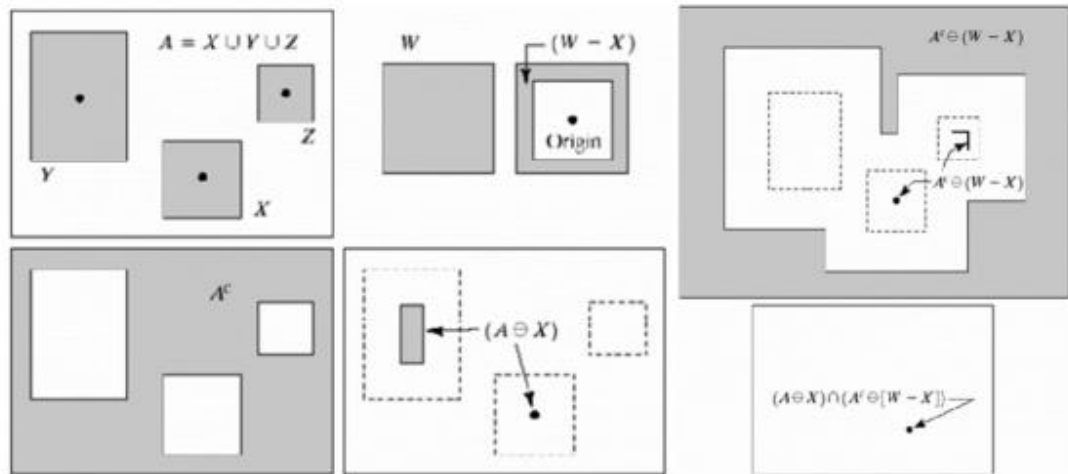
Phép toán “Hit or miss” là phép biến đổi thường được dùng để phát hiện, so khớp các mẫu đối tượng trong ảnh (template), ví dụ sử dụng phép “Hit or miss” để tìm chính xác hình dáng được cho bởi đối tượng X.

Đối tượng X có thể vừa phía trong của nhiều đối tượng, do đó ta xem nền cục bộ là W-X. Đầu tiên, thực hiện toán tử erosion A và X ($A \ominus X$) cho tất cả các điểm ảnh mà X có thể nằm vừa phía trong A.

Để khớp với nền, thực hiện tính phần bù của A là A^c . Tập tất cả các vị trí sao cho X vừa khớp là phần giao của A co với X và phép co của của A^c bởi W-X.

Giả sử ký hiệu tập B là tập hợp bao gồm X và phần nền, phép biến đổi hit (vừa) của B trong A được xác định như sau:

$$A \odot B = (A \ominus X) \cap [A^c \ominus (W-X)]$$



Hình 6.7 Biến đổi Hit or miss

Ký hiệu tổng quát $B=(B_1, B_2)$, trong đó:

B_1 : Là tập được tạo thành từ các phần tử của B liên quan đến đối tượng

B_2 : Là tập được tạo thành từ các phần tử của B liên quan đến phần nền tương ứng

$B_1=X$ và $B_2 = (W-X)$

Công thức tổng quát được xác định như sau:

$$A \odot B = (A \ominus B_1) \cap [A^c \ominus B_2]$$

$A \odot B$ chứa tất cả các điểm gốc tại đồng thời B_1 “chạm tới” A và B_2 “chạm tới” A^c .

Một số trường hợp sử dụng phép biến đổi Hit-or-Miss:

- Các điểm bị cô lập: Không có điểm lân cận là láng giềng (tức là kết nối trực tiếp)
- Điểm kết thúc: chỉ có một hoặc không điểm láng giềng với nó
- Điểm nhiều láng giềng: Điểm có nhiều hơn hai láng giềng
- Điểm biên: Điểm có ít nhất một điểm lân cận của nó thuộc nền.

6.7- Ứng dụng của kỹ thuật hình thái học

6.7.1 Trích chọn biên ảnh

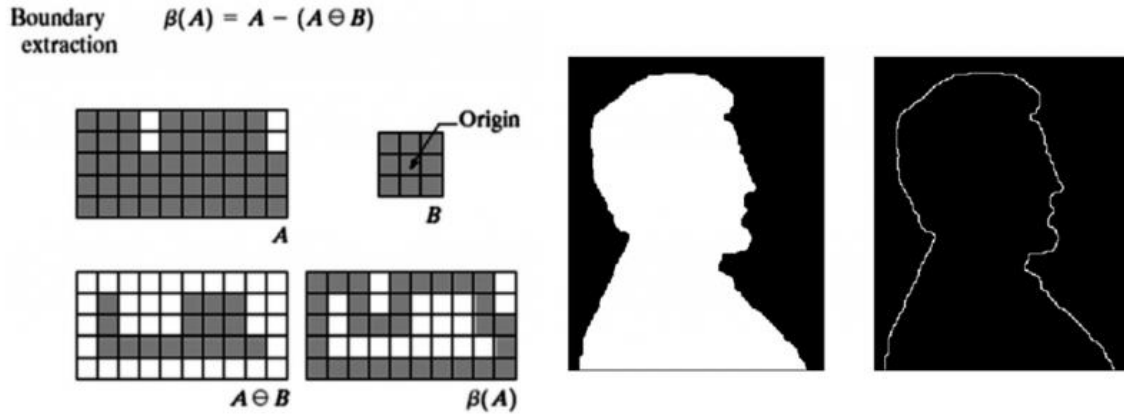
Để trích lọc biên của ảnh nhị phân A, thực hiện hai bước sau :

+ Thực hiện phép co ảnh với phần tử cấu trúc B được xác định theo kinh nghiệm tùy theo từng ứng dụng cụ thể.

+ Thực hiện khử nền ảnh bằng cách lấy ảnh gốc A trừ cho ảnh thu được sau khi co ảnh ở bước trước.

Quá trình trích biên được thực hiện bằng công thức sau :

$$E = A - (A \ominus B)$$



Hình 6.6. Trích đường biên bằng hình thái học

6.7.2 Tô vùng – Region fill

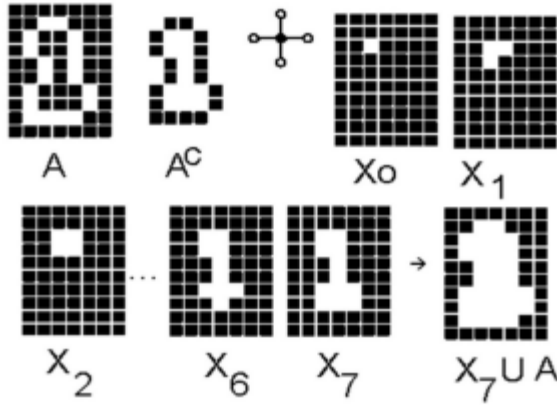
Sau khi phân ngưỡng, kết quả trả về là một ảnh nhị phân với kết quả phân vùng thô. Kết quả phân ngưỡng thường không ‘hoàn hảo’ do những yếu tố bên ngoài dẫn đến việc lấy mẫu, xử lý ảnh không kiểm soát được như cường độ sáng hay độ chói xuất hiện trong ảnh, nhiễu hoặc có thể đối tượng có kết cấu phức tạp, khi phân vùng không thể trích xuất được toàn bộ đối tượng mà có những ‘lỗ’ phía trong đối tượng hoặc khi sử dụng các kỹ thuật trích chọn cạnh, kết quả trả về là đường biên của đối tượng. Do vậy, để lấp đầy vùng rỗng trong đối tượng, các phép xử lý hình thái học có thể giúp lấp đầy các vùng rỗng một cách hiệu quả với độ phức tạp thấp.

Cho trước A là một vùng của đối tượng gồm các điểm ảnh là đường biên đối tượng có giá trị bằng 1 và các điểm ảnh không thuộc biên của đối tượng quan tâm được gán giá trị 0. Cho trước B là một phần tử cấu trúc và $X_0(x)=1$ (điểm xuất phát) là một điểm ảnh bất kỳ nằm phía trong được bao bởi biên của đối tượng. Quá trình tô đầy đối tượng được thực hiện bằng thuật toán lặp như sau:

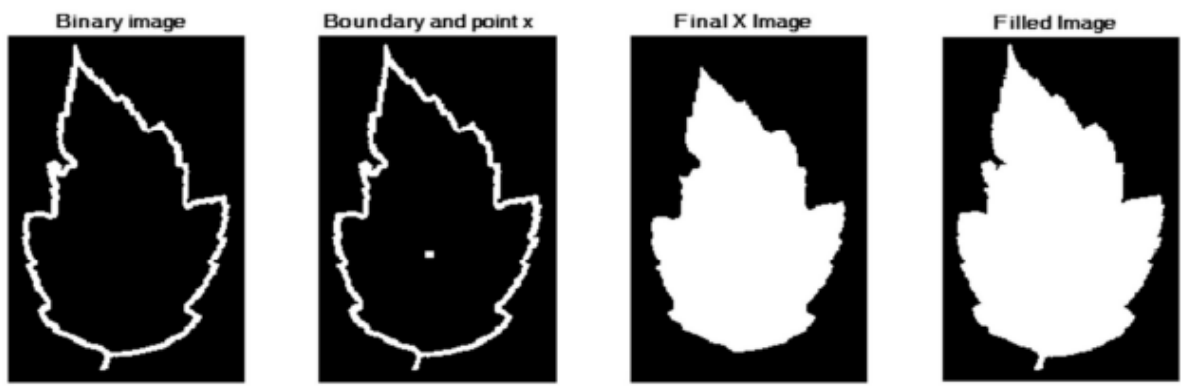
$$X_k = (X_{k-1} \oplus B) \cap A^c \quad \text{với } k=1,2,3,\dots \text{ và } A^c \text{ là phần bù của } A.$$

Quá trình kết thúc khi $X_k = X_{k-1}$

Kết quả cuối cùng thu được là $F = A \cup X_k$. Như vậy, kết quả vùng đối tượng được lấp đầy cuối cùng của A và X_k là đơn giản nhất, nó là một tập đã được làm đầy và đường biên.



Hình 6.7 Minh họa quá trình thực hiện tô vùng



Hình 6.8 Tô đầy vùng bằng kỹ thuật hình thái học

6.7.3 Gán nhãn cho các thành phần liên thông

Lọc thành phần liên thông là việc sử dụng các thuật toán xử lý ảnh hình thái học để phát hiện và kết nối các thành phần liên thông trong ảnh nhị phân. Thành phần liên thông có thể là các đốm (blob) và các vùng bất kỳ. Các thành phần riêng biệt là tập hợp giá trị các điểm ảnh được kết nối với nhau và phân biệt với các nhóm điểm ảnh khác thông qua việc gán nhãn khác nhau cho các nhóm liên thông. Nhãn của các nhóm này có thể là số nguyên, trong đó nên có giá trị bằng 0, nhóm điểm ảnh liên thông với nhau được gán nhãn từ 1 trở đi.

Việc trích các thành phần liên thông trong ảnh nhị phân là nhiệm vụ cơ bản trong xử lý ảnh tự động. Cho một tập A chứa một hoặc nhiều thành phần liên thông. X_0 là một tập chứa một điểm tương ứng cho mỗi thành phần liên thông trong A . Quá trình thực hiện tương tự như việc làm đầy vùng, nhưng sử dụng toán tử giao của A thay vì phần bù của nó. Mỗi thành phần liên thông thực hiện phép lặp sau để trích ra vùng liên thông X_k .

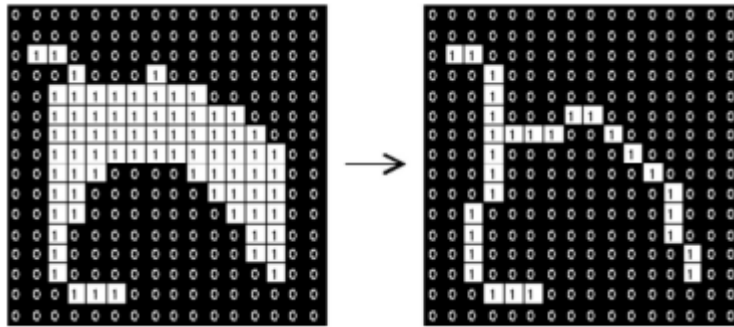
$$X_k = (X_{k-1} \oplus B) \cap A \text{ với } k=1, 2, 3$$

Kết thúc khi $X_k = X_{k-1}$, tập X_k chứa tất cả các thành phần liên thông của A .

Như vậy, để gán nhãn cho các đối tượng dựa vào hình dạng đường biên, trước tiên cần thực hiện lấp đầy phía trong của đối tượng, sau đó dùng phương pháp gán nhãn để gán nhãn cho đối tượng, mỗi đối tượng tách rời được gán một nhãn khác nhau.

6.7.4. Làm mỏng đối tượng

Làm mỏng đối tượng (thinning) là quá trình biến đổi ảnh trở nên đơn giản hơn bằng cách làm ‘mỏng’ đối tượng nhưng vẫn giữ được cấu trúc hình học của đối tượng.



Hình 6.9 Làm mỏng đối tượng bằng hình thái học

Để làm mỏng đối tượng trong ảnh nhị phân A với phần tử cấu trúc B được xác định theo công thức sau :

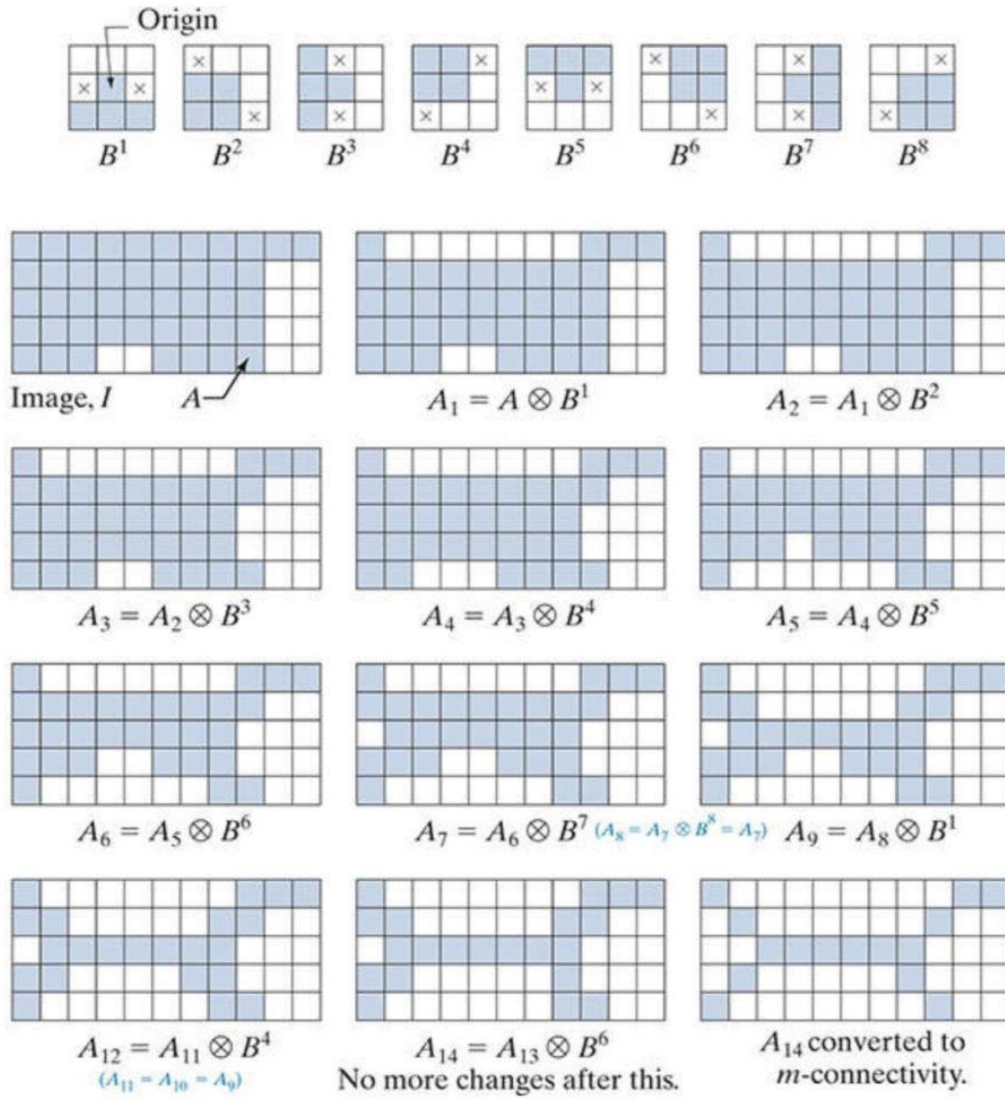
$$A \otimes B = A \setminus (A \odot B_i) = A \cap (A \odot B)^c$$

Các toán tử làm mỏng đối xứng sử dụng dãy các phần tử cấu trúc

$$\{B\} = \{B_1, B_2, \dots, B_n\}$$

Với B_i là kết quả của phép quay B_{i-1} .

Quá trình thực hiện theo các bước $A \odot B = (\dots((A \odot B_1) \odot B_2) \dots) \odot B_n$



Hình 6.8. Quá trình làm mỏng đối tượng với các phần tử cấu trúc



Hình 6.9. Kết quả làm mỏng đối tượng a) ảnh gốc b) ảnh làm mỏng

6.7.5 Làm dày đối tượng

Kỹ thuật làm dày đối tượng (thickening) cũng được thực hiện tương tự như phép giãn ảnh, tuy nhiên kỹ thuật này không gộp các đối tượng không liên thông nhau của ảnh đầu vào, nghĩa là các đối tượng tách rời nhau sẽ không bị gộp lại với nhau. Kỹ thuật này thường sử dụng để làm dày các đối tượng bị lõm. Làm dày đối tượng được thực hiện qua công thức sau :

$$A \odot B = A \cup (A \otimes B)$$

Các toán tử làm dày đối tượng sử dụng dãy các phần tử cấu trúc sau :

$\{B\} = \{B_1, B_2, \dots, B_n\}$ Với B_i là kết quả của phép quay B_{i-1} .

Quá trình thực hiện theo các bước

$$A \odot B = (\dots((A \odot B_1) \odot B_2) \dots) \odot B_n$$



Hình 6.10. Kết quả làm dày đối tượng a) ảnh gốc, b) ảnh được làm dày

Câu hỏi

1. Thực hiện phép đóng, phép mở ảnh I theo phần tử cấu trúc T đã cho (x ký hiệu vị trí gốc của T).

$$I = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix} \quad T = \begin{bmatrix} 1 & x \\ & 1 \end{bmatrix}$$

2. Thực hiện phép đóng, phép mở ảnh I theo phần tử cấu trúc T đã cho (x ký hiệu vị trí gốc của T).

$$\begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

$$\begin{array}{cccccc}
 I = & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\
 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\
 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\
 & 0 & 0 & 0 & 1 & 1 & 1 & 1
 \end{array}
 \quad
 \begin{array}{cc}
 T = & 1x & 0 \\
 & 0 & 1
 \end{array}$$

3. Thực hiện phép đóng, phép mở ảnh I theo phần tử cấu trúc T đã cho (x ký hiệu vị trí gốc của T).

$$\begin{array}{cccccc}
 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\
 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\
 I = & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\
 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\
 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\
 & 0 & 0 & 0 & 1 & 1 & 1 & 1
 \end{array}
 \quad
 \begin{array}{cc}
 T = & 1 & 0 \\
 & 1x & 1
 \end{array}$$

4. Thực hiện phép đóng, phép mở ảnh I theo phần tử cấu trúc T đã cho (x ký hiệu vị trí gốc của T).

$$\begin{array}{cccccc}
 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\
 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\
 I = & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\
 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\
 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\
 & 0 & 0 & 0 & 1 & 1 & 1 & 1
 \end{array}
 \quad
 \begin{array}{cc}
 T = & 1 & 1 \\
 & 1x & 1
 \end{array}$$

5. Trình bày phép biến đổi Hit or miss và cho ví dụ.

6. Trình bày ứng dụng trích chọn biên ảnh.

7. Trình bày ứng dụng tô vùng.

8. Trình bày ứng dụng gán nhãn cho các thành phần liên thông.

9. Trình bày ứng dụng làm mỏng đối tượng.

10. Trình bày ứng dụng làm dày đối tượng.

CHƯƠNG 7. KỸ THUẬT NHẬN DẠNG

Nội dung chính của chương sẽ giới thiệu về kỹ thuật nhận dạng và phương pháp véc tơ hỗ trợ.

Mục tiêu cần đạt được của chương

Sinh viên hiểu các khái niệm về kỹ thuật nhận dạng, phương pháp véc tơ hỗ trợ.

Bài 9: KỸ THUẬT NHẬN DẠNG (Số tiết: 5 tiết)

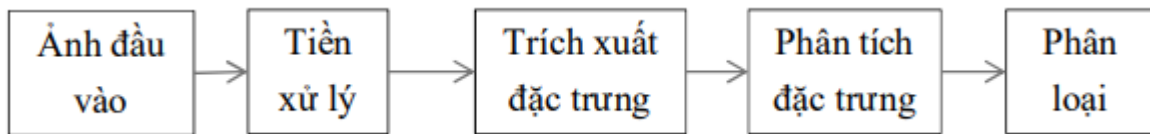
Một trong những vấn đề quan trọng của thị giác máy tính là khả năng kết hợp các phương pháp trí tuệ nhân tạo với xử lý ảnh đặc trưng ảnh vào việc phát hiện, nhận dạng và định danh đối tượng. Nhận dạng mẫu có nhiều ứng dụng trong các hệ thống tương tác người máy

7.1-Giới thiệu

Thuật ngữ phân loại trong lĩnh vực khai phá dữ liệu được hiểu là việc gán đối tượng vào một trong các lớp được xác định trước. Phân loại là dạng bài toán phổ biến và có nhiều ứng dụng khác nhau trong các hệ thống tự động. Trong lĩnh vực xử lý ảnh, bài toán phân loại mẫu và nhận dạng đối tượng được hiểu là việc phân loại, xác định các mẫu thuộc lớp nào đó một cách chính xác dựa vào việc đo lường về đối tượng dựa trên các đặc trưng của đối tượng. Nói cách khác, nhận dạng đối tượng nhằm mục đích phân loại các mẫu (dữ liệu) dựa trên kiến thức có trước về đối tượng hoặc dựa vào thông tin thống kê, đo lường được trích rút ra từ các mẫu trong tập dữ liệu huấn luyện. Hiện nay, có nhiều cách tiếp cận như nhận dạng người trong các hệ thống tương tác người máy; nhận dạng khuôn mặt và định danh cá thể, nhận dạng phương tiện giao thông trong các hệ thống giao thông thông minh, hỗ trợ lái xe; nhận dạng hành động trong các hệ thống giám sát thông minh, chăm sóc y tế; nhận dạng chữ viết,... Lĩnh vực nhận dạng đối tượng liên quan đến các phương pháp, kỹ thuật của xử lý ảnh và học máy. Để huấn luyện mô hình và nhận dạng đối tượng, các mẫu dữ liệu thường được số hóa và biểu diễn thành vector đặc trưng trong một không gian đa chiều tương ứng, được gọi là không gian đặc trưng. Nhờ vào vector đặc trưng này để có thể phân tích, đánh giá nhằm xây dựng được mô hình đối tượng phục vụ phân loại, nhận dạng chúng. Trong nhận dạng, các phương pháp học máy có giám sát và không giám sát đều được nghiên cứu và ứng dụng. Các kỹ thuật học máy có giám sát thường được sử dụng như cây quyết định, mạng nơron, SVM, Boosting, rừng ngẫu nhiên. Nhận dạng theo học máy có giám sát thì việc phân loại thường dựa vào tập dữ liệu mẫu đã được gán nhãn theo các lớp bởi các ‘chuyên gia’ để phân tích và xây dựng mô hình nhận dạng. Tập dữ liệu mẫu để học được gọi là tập dữ

liệu huấn luyện và quá trình phân tích, xây dựng mô hình đối tượng được gọi là quá trình huấn luyện máy nhận dạng hay huấn luyện mô hình. Ngược lại, các phương pháp học không có giám sát, tập dữ liệu phân loại không cần được gán nhãn trước mà bản thân thuật toán phải tự phân loại, xác định lớp của đối tượng dựa vào phân tích, thống kê từ các đặc trưng của tập mẫu dữ liệu đưa vào.

Quá trình nhận dạng đối tượng được mô tả bằng sơ đồ tổng quát như sau:



Hình 7.1. Sơ đồ tổng quát quá trình nhận dạng đối tượng

Phần dưới trình bày một số kỹ thuật học máy

Cây quyết định

Cây quyết định (decision tree) là một lĩnh vực nghiên cứu cụ thể của ngành học máy. Các kỹ thuật cây quyết định được sử dụng khá nhiều trong các lĩnh vực khai phá tri thức và nhận dạng mẫu. Cây quyết định là một kiểu mô hình dự báo (predictive model), được xây dựng trên cơ sở cấu trúc cây, dùng để phân lớp các mẫu dữ liệu dựa vào chuỗi các luật suy diễn (rules). Trong mô tả cấu trúc cây của cây quyết định, các nút lá đại diện cho các quyết định phân loại, các nhánh đại diện cho luật kết hợp các thuộc tính để dẫn tới phân loại nào đó. Một cây quyết định có thể được huấn luyện bằng cách chia tập dữ liệu huấn luyện thành các tập con để kiểm tra theo từng giá trị thuộc tính đơn hoặc một nhóm các thuộc tính. Việc phân loại có thể được mô tả như là các kết hợp phân loại đơn giản bằng cách sử dụng kỹ thuật suy diễn toán học. Quá trình huấn luyện mô hình phân loại là quá trình xây dựng cây quyết định. Hiện nay có nhiều phương pháp xây dựng cây quyết định cho bài toán phân loại đối tượng như ID3, C4.5,...

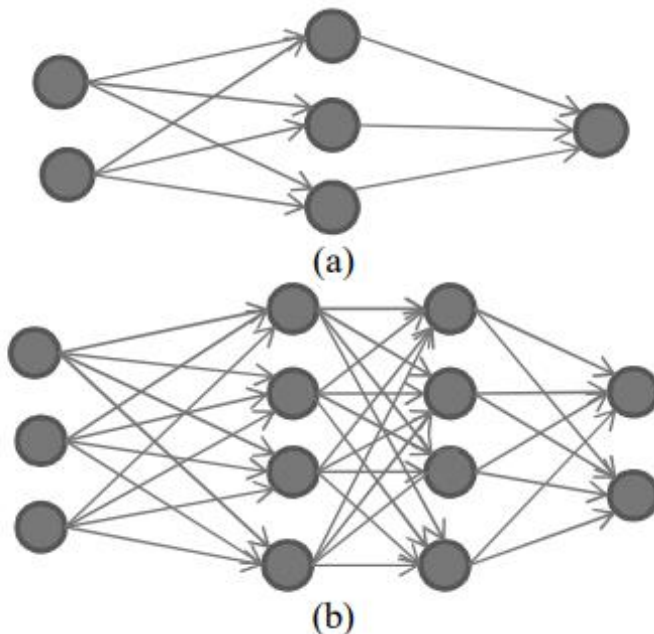
Kỹ thuật Boosting

Kỹ thuật Boosting là một dạng thuật toán học máy quần thể (machine learning ensemble algorithm) bằng cách xây dựng nhiều bộ phân loại cùng lúc (tương tự như rừng ngẫu nhiên) và sau đó kết hợp chúng lại theo trọng số xác định của từng phân loại thành phần. Mỗi bộ phân loại thành phần được gọi là phân loại yếu (weak classifier), các bộ phân loại yếu được hợp lại với nhau tạo thành một phân loại mạnh (strong classifier). Một trong những thuật toán phổ biến của boosting là AdaBoost (Adaptive boosting), được đề xuất bởi Freund và Schapire năm 1999. AdaBoost là bộ phân loại mạnh phi tuyến, hoạt động trên nguyên tắc kết hợp các bộ phân loại yếu theo trọng số để tạo ra một bộ phân loại mạnh hơn theo kiểu thích ứng (với mẫu dữ liệu). Theo đó, AdaBoost sử dụng các trọng số để đánh dấu các mẫu khó phân loại. Càng vào những mức sâu của phân loại yếu, bộ phân loại càng tập trung vào những mẫu khó phân loại, trong khi những mẫu phân loại dễ sẽ có giá trị ảnh hưởng nhỏ hơn. Nghĩa là trong quá trình huấn luyện, mỗi

bộ phân loại yếu tiến hành cập nhật lại trọng số theo hướng giảm dần trọng số các mẫu phân loại đúng (các mẫu dễ) và tăng trọng số của những mẫu phân loại sai (các mẫu khó) để áp dụng cho việc xây dựng các bộ phân loại yếu sau đó. Dựa trên ý tưởng này, các bộ phân loại yếu có thể tập trung chủ yếu vào các mẫu khó mà bộ phân loại trước đó phân loại không đúng. Cuối cùng, các bộ phân loại yếu được kết hợp lại theo trọng số tùy thuộc độ chính xác phân loại của chúng để tạo ra bộ phân loại mạnh.

Mạng neural nhân tạo

Mạng neural nhân tạo (ANN- artificial neural network) thường được gọi tắt là mạng neural (NN- neural network). Mạng neural là mô hình tính toán được xây dựng dựa trên nguyên tắc hoạt động và phát triển của mạng neural sinh học. Kiến trúc mạng neural gồm tập các nút gọi là các neural và tập các cung (hay gọi là cạnh) để kết nối các neural với nhau. Tập các nút được tổ chức thành các lớp, gọi là các layer. Trong đó có một lớp đầu vào (input layer) và một lớp đầu ra (output layer), các lớp giữa gọi là các lớp ẩn (hidden layer). Mỗi cung kết nối 2 cặp neural với nhau, trong đó có một nút vào và một nút ra, nhằm mục đích truyền thông tin và xử lý tính giá trị mới cho các nút ra. Mỗi liên hệ giữa các nút được thể hiện thông qua hàm biến đổi (gọi là hàm truyền) cùng với bộ trọng số tương ứng cho hàm truyền. Thông thường, kiến trúc của một mạng neural được xây dựng trước và các trọng số được xác định trong quá trình huấn luyện, tuy nhiên một số kiểu mạng có khả năng thay đổi để thích ứng (adaptive) với dữ liệu thực tế và có thể tự thay đổi cấu trúc của mạng dựa trên các thông tin trong quá trình học, một số kiểu như mạng neural đa lớp (MLNN- Multi layer neural network) và mạng Neural tự tổ chức (SOM- Self organizing maps)



Hình 7.1 Minh họa kiến trúc mạng neural: (a) mạng có một lớp ẩn, (b) mạng có hai lớp ẩn

Một trong những phần tử quan trọng của NN là khả năng học của nó. Một NN không chỉ là hệ thống phức tạp mà còn là một hệ thống thích nghi phức tạp, có nghĩa là có thể thay đổi cấu trúc bên trong dựa vào luồng thông tin truyền qua nó. Đặc biệt, những thay đổi này thu được thông qua việc điều chỉnh bộ trọng số. Trong hình minh họa trên, mỗi cung thể hiện một kết nối giữa hai neural và chỉ hướng của luồng dữ liệu đi qua trong mạng. Mỗi kết nối có một trọng số nhất định, là giá trị có chức năng điều khiển tín hiệu nữa hai neural với nhau. Nếu mạng cho kết quả phân lớp đủ tốt thì không cần thiết phải điều chỉnh lại trọng số nữa. Tuy nhiên nếu mạng cho kết quả chưa đủ tốt (với sai số trên ngưỡng cho phép) thì hệ thống phải thích nghi bằng cách điều chỉnh trọng số để thay đổi kết quả đầu ra theo hướng tốt hơn.

Một trong những lợi thế của NN là khả năng được sử dụng như một xấp xỉ hàm tùy ý học từ các mẫu dữ liệu quan sát được. Một số yêu cầu chính trong xây dựng mô hình mạng neural là vấn đề chọn mô hình phù hợp bài toán ứng dụng phụ thuộc vào cách mô tả mẫu dữ liệu. Nghĩa là một mô hình quá phức tạp sẽ dẫn đến khó khăn trong quá trình huấn luyện mô hình, trong khi đó những mô hình đơn giản lại không giải quyết được bài toán phức tạp. Vấn đề lựa chọn thuật toán học cũng có ý nghĩa hết sức quan trọng. Thuật toán học chủ yếu liên quan đến việc xây dựng bộ tham số (hyperparameter) và làm sao ước lượng các tham số tối ưu và hội tụ nhanh nhất có thể để phân lớp tốt nhất. Trong thực tế, lựa chọn và điều chỉnh một thuật toán để huấn luyện trên dữ liệu quan sát yêu cầu một số lượng lớn đáng kể trong thực nghiệm.

Phân theo phương pháp huấn luyện, mạng neural có thể thuộc dạng học có giám sát hoặc học không có giám sát, học tăng cường. Với cách tiếp cận học có giám sát, quá trình huấn luyện được thực hiện trên tập dữ liệu đã có gán nhãn (bởi chuyên gia). Học không giám sát, mạng neural xử lý trên bộ mẫu dữ liệu không cần gán nhãn trước. Quá trình thực hiện được xem như quá trình tìm kiếm các đặc trưng tiềm năng (tiềm ẩn) trong tập dữ liệu. Ví dụ thuật toán phân cụm dữ liệu bằng cách chia tập dữ liệu ban đầu vào các nhóm theo một số tính chất nào đó chưa biết trước. Bên cạnh đó, học tăng cường là chuỗi các kỹ thuật liên quan đến xây dựng mô hình dựa vào quan sát. Ví dụ một con chuột chạy qua một mê cung. Nếu nó quay trái, nó sẽ gặp được một miếng phô mát, nếu nó quay phải nó sẽ bị một số rủi ro nào đó. Giả sử rằng con chuột sẽ học theo thời gian để quyết định qua trái. Mạng thần kinh của chuột đưa ra quyết định với một kết quả (qua trái hoặc qua phải) và quan sát môi trường quanh nó. Nếu như quan sát thấy vấn đề đó là tiêu cực, mạng thần kinh có thể điều chỉnh trọng số để tạo ra một quyết định khác cho những lần tiếp theo. Học tăng cường ngày nay được sử dụng khá phổ biến trong các hệ thống tự động hóa và hệ thống tương tác người máy. Ví dụ tại thời điểm t , một robot thực hiện một nhiệm vụ và quan sát kết quả. Nó có thể va chạm vào tường, bị rơi khỏi bàn, va chạm với các đối tượng khác hoặc nó không gặp vấn đề gì. Mạng neural xem xét việc học tăng cường trong ngữ cảnh của các mô hình hệ thống xe tự động. Những khả năng của mạng neural giúp học và tạo ra những thay đổi đối với cấu trúc theo thời gian và làm cho mạng neural trở nên thông minh hơn.

Mạng neural có nhiều ứng dụng và đạt được những kết quả nhất định trong lĩnh vực trí tuệ nhân tạo: Nhận dạng mẫu là một trong những ứng dụng phổ biến của mạng neural như nhận dạng ký tự quang học, nhận dạng mặt người, nhận dạng dáng người, định danh đối tượng và con người qua hình ảnh, phát hiện và nhận dạng tín hiệu biến báo giao thông, phát hiện các hành động bất thường trong các hệ thống giám sát thông minh. Dự đoán theo chuỗi thời gian, mạng neural có thể được sử dụng để tạo ra các dự đoán ví dụ như giá trị chứng khoán thay đổi như thế nào, dự báo thời tiết,... Xử lý tín hiệu: Như các hệ thống lọc âm thanh, khử nhiễu tín hiệu thu vào, khuếch đại các âm thanh quan trọng. Trong điều khiển như việc dự đoán chuyển động để điều khiển tay lái điều hướng.

7.2. Phương pháp máy véc tơ hỗ trợ

7.2.1 Giới thiệu

Phương pháp máy vector hỗ trợ (Support vector machine-SVM) là thuật toán phân loại, thuộc phương pháp học có giám sát, được đề xuất bởi Corinna và Vapnik vào năm 1995. Ban đầu SVM được thiết kế cho bài toán phân loại nhị phân, sau này được mở rộng để áp dụng cho phân loại đa lớp khác nhau. Với một mẫu dữ liệu huấn luyện thuộc hai loại xác định trước, thuật toán SVM tiến hành huấn luyện để xây dựng mô hình nhằm phân loại các mẫu dữ liệu theo các lớp tương ứng như tập dữ liệu huấn luyện. Một mô hình SVM là một cách biểu diễn các vector hỗ trợ phân loại trong không gian nhiều chiều và lựa chọn siêu phẳng (hyperplane) phân loại giữa hai lớp sao cho cực đại khoảng cách từ các mẫu dữ liệu huấn luyện (các điểm trong không gian n chiều) tới mặt phẳng phân loại. Các mẫu mới đưa vào để phân loại phải được biểu diễn trong cùng một không gian như tập huấn luyện và SVM phân loại vào một trong hai lớp tùy thuộc vào giá trị cụ thể của mẫu dữ liệu thuộc phía nào của mặt siêu phẳng phân loại.

Đến nay, SVM là một trong những phương pháp phân lớp được sử dụng rộng rãi nhất trong lĩnh vực khoa học máy tính và phân tích số liệu. SVM thực hiện hiệu quả trên tập dữ liệu lớn và xử lý hiệu quả trên không gian có số chiều lớn, đặc biệt áp dụng cho các bài toán phân loại dữ liệu hình ảnh, văn bản, và tiếng nói,... SVM hoạt động khá linh hoạt và có khả năng áp dụng nhiều hàm nhân khác nhau và có thể phân loại theo phương pháp tuyến tính hoặc phi tuyến. Trong thực tế, SVM đạt độ chính xác khá cao so với kỹ thuật học máy truyền thống khác. Phương pháp SVM có nhiều phiên bản được phát triển theo hướng tiếp cận phân loại tuyến tính, phân loại phi tuyến và đã được lập trình thành công cụ. Thư viện LibSVM của nhóm tác giả Chang và Lin là một trong những công cụ được sử dụng khá phổ biến hiện nay.

7.2.2 Phân loại tuyến tính

Nội dung cơ bản của kỹ thuật SVM tuyến tính được mô tả như sau : Cho trước một tập dữ liệu huấn luyện S gồm n mẫu $(x_1, y_1), \dots, (x_n, y_n)$ với x_i là mẫu dữ liệu được

biểu diễn dưới dạng các điểm trong không gian p chiều (hoặc xem nó là vector có p phần tử) và $y_i \in \{-1, 1\}$ là nhãn tương ứng để chỉ lớp của mẫu dữ liệu đó. Mục tiêu của SVM là tìm một siêu phẳng có khả năng phân tách tập mẫu thành 2 tập theo nhãn của nó với lề đạt khoảng cách cực đại, lề là khoảng cách từ siêu phẳng đến các điểm (trong p chiều) gần nhất.

Các siêu phẳng cần tìm ở đây chính là các đường thẳng trong trường hợp không gian hai chiều, mặt phẳng trong trường hợp không gian ba chiều và tổng quát hơn là các siêu phẳng không gian con R^{d-1} chiều trong không gian R^d .

Siêu phẳng trong không gian đa chiều có thể được viết dưới dạng

$$w \bullet x - b = 0$$

Trong đó, w là vector pháp tuyến của siêu phẳng, w không nhất thiết là vector đơn vị. Tham số $b/\|w\|$ được xác định là khoảng cách từ siêu phẳng đến gốc tọa độ theo vector pháp tuyến w .

Lề cực đại (maximal margin) là khoảng cách lớn nhất từ siêu phẳng đến các điểm gần nhất, nghĩa là khoảng cách xa nhất từ siêu phẳng đến đường thẳng song song với nó đi qua điểm gần nhất, mà không chứa điểm dữ liệu nào bên trong nó. Các vector hỗ trợ (support vectors) là các điểm dữ liệu gần nhất nằm trên lề cực đại.

Hình minh họa cho thấy có vô số đường thẳng có thể phân loại hai lớp dữ liệu thành hai miền khác nhau, các đường này được gọi là siêu phẳng. Mục tiêu là cần tìm ra siêu phẳng ‘tốt nhất’ để phân lớp dữ liệu. Siêu phẳng SVM tìm kiếm là siêu phẳng có giá trị lề lớn nhất, còn gọi là siêu phẳng với lề cực đại (maximal margin hyperplane). Để phân loại tốt nhất cần xây dựng siêu phẳng tối ưu. Lề càng lớn thì sai số tổng quát của thuật toán phân loại càng bé. Nghĩa là nó nằm càng xa các điểm dữ liệu của tất cả các lớp càng tốt. Muốn các điểm dữ liệu có thể được chia tách một cách tuyến tính, cần xác định hai siêu phẳng sao cho không có điểm nào ở giữa chúng và khoảng cách giữa chúng là lớn nhất.

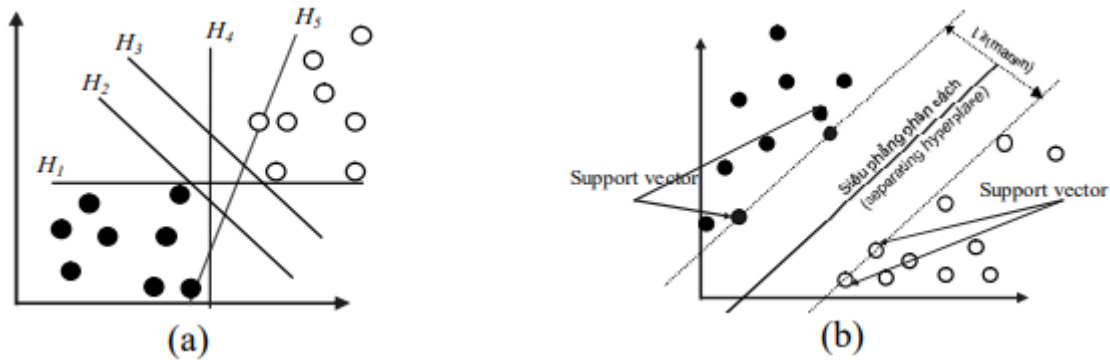
Đối với dữ liệu có khả năng phân loại tuyến tính, các vector hỗ trợ tìm các siêu phẳng phân tách có giá trị lề càng lớn càng tốt. Giả sử tất cả dữ liệu huấn luyện thỏa mãn điều kiện :

$$w \cdot x_i + b \geq +1 \text{ khi } y_i = +1$$

$$w \cdot x_i + b \leq -1 \text{ khi } y_i = -1$$

Kết hợp hai công thức lại ta có :

$$y_i(w \cdot x_i + b) \geq 1 \text{ với } i = 1, \dots, n$$



Hình 7.2. Siêu phẳng phân tách mẫu dữ liệu thành 2 lớp: a) các siêu phẳng ứng cử viên; b) lề phân loại và các vector hỗ trợ tìm được

Đối với vector hỗ trợ x_i là các điểm nằm trên đường biên giới hạn của siêu phẳng H_+ : $w \cdot x_i + b = +1$ với khoảng cách đến gốc tọa độ là $|1-b|/\|w\|$ và các điểm nằm trên đường biên giới hạn của siêu phẳng H_- : $w \cdot x_i + b = -1$ với khoảng cách đến gốc tọa độ là $|-1-b|/\|w\|$. Bài toán tối ưu tương đối khó giải vì hàm mục tiêu phụ thuộc vào $\|w\|$, là một hàm có khai căn. Tuy nhiên có thể thay $\|w\|$ bằng hàm mục tiêu $\frac{1}{2} \|w\|^2$ mà không làm thay đổi lời giải đối với bài toán. Như vậy, vấn đề cần thiết là cực tiểu hóa mục tiêu $\frac{1}{2} \|w\|^2$ theo w và b với điều kiện $y_i(w \cdot x_i + b) \geq 1$.

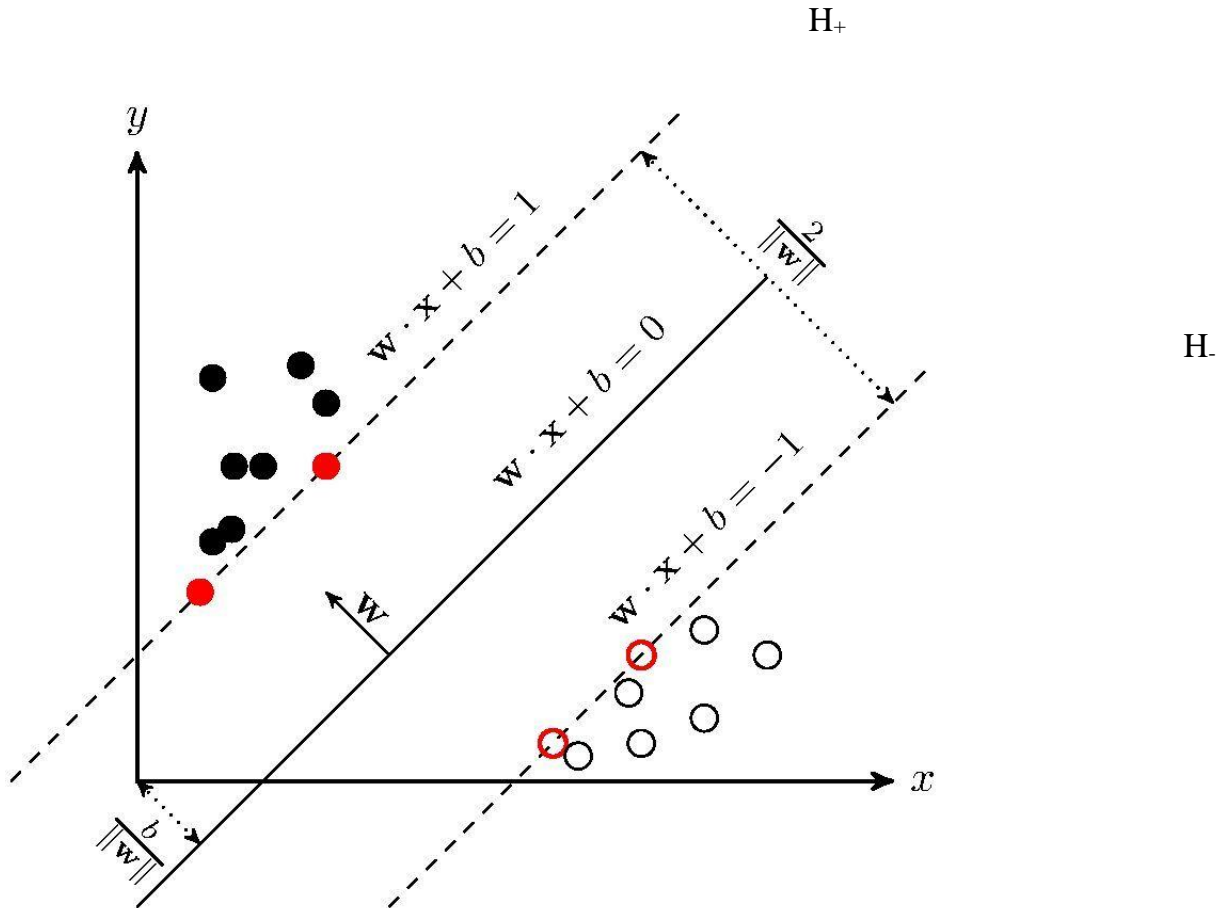
Bằng cách thêm các nhân tử Lagrange α , bài toán trở thành dạng sau:

$$\min_{w,b} \max_{\alpha \geq 0} \left\{ \frac{1}{2} \|w\|^2 - \sum_{i=1}^n \alpha_i [y_i(w \cdot x_i - b) - 1] \right\}$$

Nghĩa là cần tìm điểm cực tiểu cục bộ. Mỗi mẫu huấn luyện x_i tương ứng với một hệ số Lagrange α_i , khi đó tất cả các điểm không nằm trên lề $y_i(w \cdot x_i + b) > 1$ đều không ảnh hưởng đến giá trị hàm mục tiêu vì lúc đó hệ số Lagrange α_i có thể thiết lập bằng không ($\alpha_i = 0$). Nghĩa là những điểm không nằm trên lề thì có hệ số Lagrange bằng không và không tham gia vào việc xác định hàm mục tiêu.

Sau khi huấn luyện, các mẫu có $\alpha_i > 0$ được gọi là vector hỗ trợ và nằm trên một trong hai siêu phẳng H_+ hoặc H_- . Khi đã xác định được tập vector hỗ trợ phân loại nghĩa là đã thu được mô hình máy phân loại SVM, với mỗi mẫu thử (đánh giá, testing) có thể đơn giản được phân loại theo siêu phẳng H_+ và H_- tạo ra bằng cách dùng hàm dấu theo công thức sau:

$$\text{Sign}(w \cdot x + b)$$



Hình 7.3 Mô hình phân lớp SVM

Có thể giải bài toán tối ưu hóa trên bằng các kỹ thuật thông thường cho quy hoạch toàn phương. Theo điều kiện Karush-Kuhn-tucker (KKT), lời giải có thể được viết dưới dạng tổ hợp tuyến tính của các vector huấn luyện

$$w = \sum_{i=1}^n \alpha_i y_i x_i$$

Các điều kiện KKT đóng vai trò quan trọng trong giải quyết tối ưu các vấn đề lập trình phi tuyến. Điều kiện KKT được thể hiện như sau:

$$\alpha_i = 0 \leftrightarrow y_i(w \cdot x_i + b) \geq 1$$

$$\alpha_i = C \leftrightarrow y_i(w \cdot x_i + b) \leq -1$$

$$0 \leq \alpha_i \leq C \leftrightarrow y_i(w \cdot x_i + b) = -1$$

Với bài toán lồi cực đại cứng thì điều kiện $0 \leq \alpha_i$ được sử dụng để thay thế cho điều kiện $0 \leq \alpha_i \leq C$.

7.2.3. Phân loại tuyến tính lề mềm

Trong trường hợp không thể tìm được siêu phẳng phân tách lề mềm, thì kỹ thuật phân loại lề mềm (soft margin classification) được sử dụng để chọn một siêu phẳng phân tách tập mẫu dữ liệu huấn luyện tốt nhất có thể và cực đại hóa khoảng cách giữa siêu phẳng với các mẫu được phân loại đúng. Phương pháp lề mềm sử dụng thêm các biến bổ sung ξ_i (xem như giá trị bù cho phần tử sai) nhằm đo độ sai lệch của mẫu bị phân loại sai x_i qua đó giá trị này có tác dụng kéo phần tử x_i về đúng phía nhãn.

$$y_i(w \cdot x_i - b) \geq 1 - \xi_i \text{ với } i=1, \dots, n$$

Hay nói cách khác, hàm mục tiêu có bổ sung thêm hệ số mới để phạt (penalty) các điểm bị phân loại nhầm, giá trị ξ_i khác không và bài toán tối ưu hóa trở thành bài toán tối ưu giữa lề cực đại và mức phạt nhỏ nhất. Nếu hàm phạt là tuyến tính thì bài toán trở thành dạng sau:

$$\min_{w, \xi, b} \left\{ \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \right\} \text{ với } y_i(w \cdot x_i - b) \geq 1 - \xi_i, \quad \xi_i \geq 0$$

Có thể giải bài toán trên bằng nhân tử Lagrange tương tự trường hợp cơ bản phân loại lề cứng. Bài toán cần giải trở thành:

$$\min_{w, \xi, b} \max_{\alpha, \beta} \left\{ \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i [y_i(w \cdot x_i - b) - 1 + \xi_i] - \sum_{i=1}^n \beta_i \xi_i \right\}$$

Với $\alpha_i, \beta_i \geq 0$

7.2.4 Hàm nhân

Hàm nhân (hay còn gọi là hàm lõi, Kernel function) ký hiệu K là một hàm số trên không gian X , với mọi cặp vector $u, v \in X$, ta có:

$$K(u, v) = \Phi(u) \Phi(v)$$

Với Φ là ánh xạ từ không gian mẫu X tới một không gian thuộc tính F được xác định $\Phi: R^d \rightarrow F$

Khi đó, hàm phân loại có dạng:

$$f(x) = \sum_{e_i \neq 0} y_i \alpha_i K(x_i, x) + b$$

Trong thực tế, có hai loại hàm nhân được sử dụng phổ biến là hàm đa thức (polynomial function) và hàm cơ sở bán kính (RBF-radial basis function).

+ Hàm nhân đa thức trên không gian R^d có dạng:

$$K(u, v) = (u \cdot v)^p$$

Với $p=2$ và $d=2$, ta có thể xây dựng ánh xạ Φ từ không gian số thực hai chiều R^2 vào không gian số thực ba chiều R^3 như sau:

$$\Phi: R^2 \rightarrow R^3$$

$$(u_1, u_2) \rightarrow (u_1^2, \sqrt{2}u_1u_2, u_2^2)$$

Hàm nhân $K(u,v)$ là tích vô hướng của hai vector $\Phi(u)$ và $\Phi(v)$ trong không gian R^3 . Lưu ý rằng một số hàm nhân có thể tương ứng với nhiều ánh xạ khác nhau.

Với $d > 2$ ta có:

$$(u \cdot v)^2 = \left(\sum_{i=1}^d u_i v_i \right)^2 = \sum_{i=1}^d \sum_{j=1}^d u_i u_j v_i v_j = \sum_{(i,j)=(1,1)}^{(d,d)} (u_i u_j)(v_i v_j)$$

+ Hàm nhân RBF: Hàm cơ sở bán kính là một hàm giá trị thực mà giá trị của nó chỉ phụ thuộc vào khoảng cách từ gốc tọa độ, sao cho thỏa mãn điều kiện $\Phi(x) = \Phi(\|x\|)$.

Có một số dạng hàm RBF như hàm mũ $K(u, v) = e^{-\|u-v\|^2/2\sigma^2}$ với σ thường được gọi là độ rộng của hàm hay giá trị độ lệch chuẩn, số chiều trong không gian đặc trưng có thể là vô hạn hoặc hữu hạn với giá trị lớn. Do vậy, khả năng của hàm các hàm tuyến tính cũng trở nên không giới hạn.

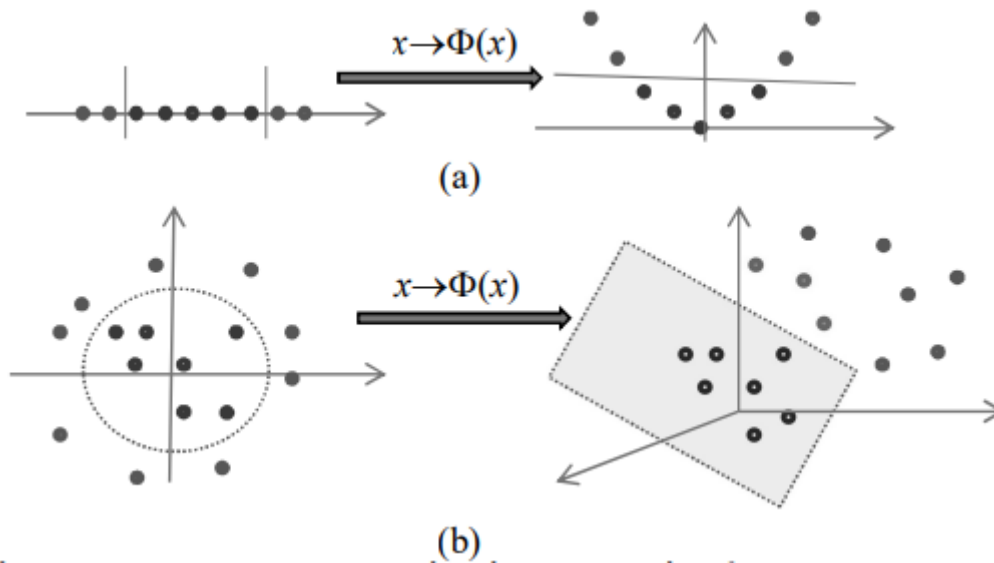
7.2.5 Tuyến tính hóa phân loại phi tuyến

Có một cách tiếp cận tương đối phổ biến là phân loại phi tuyến (nonlinear classification) có thể thực hiện ánh xạ vector đặc trưng phi tuyến vào không gian nhiều chiều hơn để có thể tuyến tính hóa rồi dùng phân loại tuyến tính trong không gian mới này. Đặc biệt, Φ là một hàm ánh xạ từ không gian mẫu d chiều ban đầu vào không gian đặc trưng F được định nghĩa bởi $\Phi: R^d \rightarrow F$.

Đối với một vector đặc trưng ban đầu $x \in R^d$ và vector đặc trưng đã biến đổi được xác định theo hàm biến đổi $\Phi(x)$ với nhân của nó vẫn giữ nguyên. Như vậy, mẫu huấn luyện ban đầu (x_i, y_i) được biến đổi thành $(\Phi(x_1), y_1), \dots, (\Phi(x_n), y_n)$.

Như vậy, phân loại phi tuyến thực hiện bằng cách biến đổi qua một không gian khác để có thể tuyến tính hóa phân loại. Nghĩa là trong trường hợp không thể phân chia các lớp dữ liệu một cách tuyến tính trong không gian ban đầu thì thực hiện biến đổi dữ liệu vào không gian mới nhiều chiều hơn, để việc phân loại trở nên dễ dàng hơn trong không gian mới theo cách tiếp cận tuyến tính đã có. Quá trình tuyến tính hóa được thể hiện như ví dụ minh họa trong hình dưới đây với trường hợp (a) là biến đổi trong không

gian 1D thành 2D có thể sử dụng hàm biến đổi $\Phi(x)=ax^2+b$ còn trường hợp (b) là biến đổi 2D thành 3D theo một hàm biến đổi cụ thể.



Hình 7.4. Tuyến tính hóa phân loại phi tuyến bằng cách biến đổi tập dữ liệu sang một không gian mới có khả năng phân loại tuyến tính

Thảo luận: Trình bày các bài toán thực tế sử dụng các phương pháp nhận dạng đã học

Câu hỏi

1. Học máy được chia thành những kỹ thuật học nào.
2. Trình bày kỹ thuật học có giám sát.
3. Trình bày kỹ thuật học không có giám sát.
4. Trình bày kỹ thuật cây quyết định, cho ví dụ.
5. Trình bày kỹ thuật Boosting.
6. Trình bày kỹ thuật mạng nơron.
7. Trình bày kỹ thuật phân loại SVM
8. Trình bày ứng dụng phân loại SVM
9. Lập trình kỹ thuật SVM

Tài liệu tham khảo

- [1] Hoàng Văn Dũng, “Giáo trình nhận dạng và xử lý ảnh”, Nhà xuất bản Khoa học và kỹ thuật, 2018
- [2] Lương Mạnh Bá, Nguyễn Thanh Thủy, “Nhập môn xử lý ảnh số”, Nhà xuất bản khoa học kỹ thuật, 2006
- [3] Đỗ Năng Toàn, “Xử lý ảnh”, Học viện bưu chính viễn thông, 2006.

Các câu hỏi thường gặp

1. Hình ảnh chiếm bao nhiêu phần trăm thông tin con người thu nhận hàng ngày?

Câu trả lời: Hình ảnh chiếm 80% phần trăm thông tin con người thu nhận hàng ngày.

2. Độ phân giải của ảnh là gì?

Câu trả lời: Độ phân giải của ảnh là số điểm ảnh trên một đơn vị đo. Độ phân giải càng cao thì ảnh càng rõ nét.

3. Độ phân giải của ảnh là gì?

Câu trả lời: Độ phân giải của ảnh là số điểm ảnh trên một đơn vị đo. Độ phân giải càng cao thì ảnh càng rõ nét.

4. Ảnh màu có mấy kênh màu?

Câu trả lời: Ảnh màu RGB có 3 kênh màu là màu đỏ, xanh lam và xanh lục.

5. Ảnh đen trắng có phải là ảnh nhị phân không?

Câu trả lời: Ảnh đen trắng đúng là ảnh nhị phân.

6. Hệ thống xử lý ảnh có các mức độ nào?

Câu trả lời: Hệ thống xử lý ảnh có 3 mức độ.

-Mức độ thấp: chỉ biết sử dụng các kỹ thuật xử lý ảnh đơn giản, thuần túy, không có tri thức như các thao tác tiền xử lý ảnh.

-Trung bình: có một chút về tri thức (trí tuệ nhân tạo) có nhiệm vụ tách và đặc trưng hóa các thành phần trong một ảnh nhận được từ quá trình xử lý mức độ thấp như các thao tác phân đoạn ảnh.

-Cao: nhận dạng, phân tích ảnh, ra quyết định. Cần kiến thức và sự hiểu biết.

7. Tế bào cảm nhận ánh sáng trong mắt là tế bào nào?

Câu trả lời: Tế bào cảm nhận ánh sáng trong mắt là tế bào hình nón và tế bào hình que.

8. Thế nào được gọi là 2 màu bù nhau?

Câu trả lời: Hai màu được gọi là màu bù nhau nếu hai màu kết hợp với nhau được màu trắng hay màu xám.

9. Màn hình máy tính sử dụng mô hình màu nào?

Câu trả lời: Màn hình máy tính sử dụng mô hình màu RGB.

10. Máy in sử dụng mô hình màu nào?

Câu trả lời: Máy in sử dụng mô hình màu CMY, CMYK.

11. Ảnh nhị phân thường dùng để làm gì?

Câu trả lời: Ảnh nhị phân thường được dùng để biểu diễn trạng thái đối tượng, phân biệt đối tượng trong ảnh với nền, hoặc để biểu diễn các đường biên đối tượng, vùng ảnh.