ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ

NGUYỄN VIỆT ANH

PHÁT HIỆN NGÃ SỬ DỤNG ĐẶC TRƯNG CHUYỂN ĐỘNG VÀ HÌNH DẠNG CƠ THỂ DỰA TRÊN CAMERA ĐƠN

LUẬN VĂN THẠC SĨ CÔNG NGHỆ THÔNG TIN

Hà Nội - 2016

ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ

NGUYỄN VIỆT ANH

PHÁT HIỆN NGÃ SỬ DỤNG ĐẶC TRƯNG CHUYỂN ĐỘNG VÀ HÌNH DẠNG CƠ THỂ DỰA TRÊN CAMERA ĐƠN

Ngành: Công nghệ thông tin

Chuyên ngành: Kỹ thuật phần mềm

Mã số: 60480103

LUẬN VĂN THẠC SĨ CÔNG NGHỆ THÔNG TIN

NGƯỜI HƯỚNG DẪN KHOA HỌC:

PGS. TS. Lê Thanh Hà

TS. Nguyễn Thị Thuỷ

Hà Nội - 2016

LỜI CAM ĐOAN

Tôi xin cam đoan các kết quả nghiên cứu, thực nghiệm được trình bày trong luận văn này do tôi thực hiện dưới sự hướng dẫn của Phó giáo sư, Tiến sĩ Lê Thanh Hà và Tiến sĩ Nguyễn Thị Thuỷ.

Tất cả những tham khảo từ các nghiên cứu liên quan đều được nêu nguồn gốc một cách rõ ràng từ danh mục tài liệu tham khảo của luận văn. Trong luận văn, không có việc sao chép tài liệu, công trình nghiên cứu của người khác mà không chỉ rõ về tài liệu tham khảo.

TÁC GIẢ LUẬN VĂN

Nguyễn Việt Anh

LỜI CẢM ƠN

Trước tiên, tôi xin gửi lời cảm ơn sâu sắc nhất đến thầy giáo, Phó giáo sư, Tiến sĩ Lê Thanh Hà và cô giáo, Tiến sĩ Nguyễn Thị Thuỷ, đã tận tình hướng dẫn tôi trong suốt quá trình thực hiện luận văn tốt nghiệp.

Cảm ơn thầy giáo - Tiến sĩ Trần Quốc Long, Tiến sĩ Nguyễn Đỗ Văn đã có những góp ý, nhận xét quý giá giúp cải thiện kết quả nghiên cứu của tôi trong luận văn này

Tôi xin bày tỏ lời cảm ơn chân thành tới trường Đại học Công Nghệ - ĐHQG Hà Nội và những thầy cô giáo tôi đã giảng dạy, truyền thụ kiến thức trong thời gian qua.

Cuối cùng, tôi xin cảm ơn tất cả gia đình, bạn bè đã luôn động viên giúp đỡ tôi trong thời gian nghiên cứu đề tài. Tuy đã có những cố gắng nhất định nhưng do thời gian và trình độ có hạn nên luận văn còn nhiều thiếu sót và hạn chế. Kính mong nhận được sự góp ý của thầy cô và các bạn.

TÁC GIẢ LUẬN VĂN

Nguyễn Việt Anh

MỤC LỤC

Danh mục các ký hiệu và chữ viết tắt

Viết tắt	Tiếng Anh	Tiếng Việt
GMM	Gaussian Mixture Model	Mô hình Gaussian hỗn hợp
MHI	Motion History Image	Ảnh lịch sử chuyển động
SVM	Support Vector Machine	Máy vector hỗ trợ
KDE	Kernel Density Estimation	
CB	Code book	Bảng mã

Danh mục hình vẽ

Danh mục bảng

MỞ ĐẦU

Theo dự báo về vấn đề dân số của Liên hợp quốc năm 2008, tỷ lệ người cao tuổi sẽ tăng từ 10% năm 2010 lên đến 23% vào năm 2050. Đó là hệ quả của tỷ suất sinh giảm, tỷ suất chết giảm và tuổi thọ trung bình tăng nhanh.

Không nằm ngoài kịch bản chung của biến đổi cơ cấu tuổi dân số thế giới, tình trạng già hóa dân số ở Việt Nam đang diễn ra nhanh chóng. Thậm chí theo thống kê, tốc độ già hóa dân số nước ta là nhanh chưa từng có trong lịch sử. Thời gian quá độ từ giai đoạn già hóa sang giai đoạn dân số già chỉ từ 18 đến 20 năm, trong khi Pháp mất 115 năm, Thụy Điển là 85 năm, Mỹ là 70 năm.

Theo số liệu của bộ Y tế [1], tỷ lệ người cao tuổi hiện chiếm 10,5%, dự đoán tăng lên 23% dân số cả nước năm 2040. Và tuy tuổi thọ trung bình tăng nhanh nhưng do chất lượng cuộc sống, chế độ dinh dưỡng và điều kiện chăm sóc y tế, số người cao tuổi có sức khỏe tốt chỉ chiếm khoảng 5% trong khi 95% còn lại không khỏe mạnh. Người cao tuổi thường mắc các chứng bệnh như tim mạch; phổi – phế quản; đái tháo đường; suy giảm trí tuệ… Đó là những chứng bệnh dễ dẫn đến đột quỵ.

Theo [2], đời sống gia đình của người cao tuổi đang thay đổi. Tỷ lệ người cao tuổi sống cùng con cái đang giảm nhanh, tỷ lệ hộ gia đình người cao tuổi sống cô đơn hoặc chỉ có vợ chồng người cao tuổi tăng lên đáng kể. Đó là hệ quả của việc di cư khi người trong độ tuổi lao động tập trung tại các thành phố lớn để tìm kiếm cơ hội việc làm hoặc thậm chí là di cư quốc tế.

Trong bối cảnh người cao tuổi sống cô đơn và không khỏe mạnh, người già gặp rất nhiều nguy hiểm khi đột quỵ hay ngã mà không được phát hiện, cấp cứu kịp thời. Luận văn này nghiên cứu về các phương thức phát hiện ngã tự động nhằm góp phần tìm ra giải pháp gia tăng sự an toàn cho người cao tuổi sống một mình.

Một thực trạng về điều kiện y tế khác là sự quá tải của bệnh viện khi thường xuyên xảy ra việc nhiều người bệnh nằm chung một giường. Phòng bệnh vốn chật chội lại càng chật chội bởi cứ mỗi một người ốm cần ít nhất một người nhà chăm sóc. Điều này gây ra mệt mỏi cho cả người bệnh và người chăm sóc, làm lãng phí sức lao động của xã hội khi người khoẻ mạnh phải nghỉ làm, cũng như gây cản chở các y bác sĩ trong khi thăm khám. Nếu có một hệ thống giám sát bệnh nhân tự động sẽ giúp giảm bớt số người chăm sóc, dẫn đến giảm tải cho bệnh viện. Một phương thức hiệu quả giúp tự động giám sát, phát hiện ngã cũng sẽ góp phần giải quyết bài toán trên.

Mục đích nghiên cứu

Mục tiêu nghiên cứu của luận văn là tìm hiểu, quan sát để tìm ra các đặc điểm của việc ngã, định nghĩa được sự kiện ngã. Từ đó đề xuất một phương thức phát hiện ngã dựa trên các quan sát quá trình ngã.

Đối tượng và phạm vi nghiên cứu

Do đặc điểm là một quốc gia đang phát triển với mức thu nhập bình quân thấp, các phương thức phát hiện ngã tự động phải là các giải pháp chi phí thấp, dựa trên các tài nguyên phổ biến, luận văn này tập trung vào các phương thức phát hiện ngã dựa trên phân tích dữ liệu video thu được từ camera giám sát. Đối tượng nghiên cứu bao gồm lý thuyết về xử lý ảnh số, xử lý video số, thị giác máy tính, các đặc điểm của hành động ngã và cách thức phát hiện việc ngã.

Phương pháp nghiên cứu

Phương pháp nghiên cứu khi thực hiện luận văn là tìm hiểu từ cơ sở lý thuyết chung về xử lý ảnh số, video số, thị giác máy tính, sau đó tìm hiểu về bài toán phát hiện ngã tự động từ các nghiên cứu đã được công bố và các kết quả đã đạt được. Từ đó cải tiến, đề xuất các kỹ thuật nhằm nâng cao hiệu quả phát hiện ngã.

Đóng góp mới của luận văn

Luận văn này đã cải tiến một số kỹ thuật và đề xuất một phương thức phát hiện ngã tự động dựa trên phân tích dữ liệu video; cài đặt thành công thuật toán phát hiện ngã với kết quả rất khả quan với tốc độ tính toán đảm bảo hoạt động thời gian thực; công bố kết quả nghiên cứu với tiêu đề "Single camera based Fall detection using Motion and Human shape Features" tại hội thảo quốc tế The Seventh International Symposium on Information and Communication Technology – SoICT 2016 (Đã được chấp nhận đăng trong kỉ yếu và trình bày tại hội thảo). Chi tiết kỹ thuật sẽ được trình bày ở các mục tiếp theo.

Kết cấu luận văn

Ngoài phần mở đầu và phần tham khảo, luận văn này được tổ chức thành 5 chương với các nội dung chính như sau:

- **Chương 1: Tổng quan bài toán phát hiện ngã tự động**
    - Giới thiệu chung về bài toán
    - Các nghiên cứu đã công bố liên quan đến bài toán
- **Chương 2: Cơ sở lý thuyết**
    - Tổng quan về xử lý ảnh số
    - Tổng quan về thị giác máy tính
    - Tổng quan về phân tích video
    - Tổng quan bài toán phát hiện hành động trong dữ liệu video
- **Chương 3: Phương thức đề xuất**
    - Tổng quan về phương thức đề xuất

- Trình bày phương thức tách vùng chuyển động trong video
- Trình bày về trích rút đặc trưng chuyển động
- Trình bày về trích rút đặc trưng hình dạng cơ thể
- Trình bày về quan sát các đặc trưng, đưa ra kết luận về việc ngã

- **Chương 4: Thí nghiệm và đánh giá**
  - Mô tả tập dữ liệu dùng để thí nghiệm
  - Trình bày phương pháp đánh giá độ hiệu quả của phương thức
  - Trình bày về cài đặt cấu hình thí nghiệm
  - Trình bày về kết quả thí nghiệm, giải thích về kết quả thí nghiệm
- **Chương 5: Kết luận và hướng phát triển**

CHƯƠNG 1.
TỔNG QUAN BÀI TOÁN PHÁT HIỆN NGÃ TỰ ĐỘNG

Theo tổ chức y tế thế giới [53], xấp xỉ 28 – 35% người có độ tuổi trên 65 bị ngã hằng năm. Tỷ lệ này tăng nhanh đến 32 – 42% đối với nhóm người già trên 70 tuổi. Tần suất ngã tăng theo tuổi và mức bệnh yếu. Thực tế, việc ngã tăng theo hàm mũ với thay đổi về mặt sinh học liên quan đến độ tuổi, dẫn đến một tỷ lệ cao các ca chấn thương liên quan đến ngã ở người già. Số ca chấn thương và tử vong do ngã chiếm đến khoảng 40% đối với người già. Trong bối cảnh đó, các phương thức giúp giảm bớt hậu quả của vấn đề sức khỏe này là rất cần thiết cho xã hội. Trong nhiều năm gần đây, các phương thức, thiết bị giúp phát hiện ngã đang được nghiên cứu tích cực.

Việc ngã có thể được xác định bởi các đặc điểm như sau:

- Xuất hiện chuyển động nhanh bất thường: Việc xuất hiện chuyển động nhanh rất có thể báo hiệu việc ngã, nhất là đối với người già. Và việc ngã gần như chắc chắn xuất hiện chuyển động nhanh tại một thời điểm nào đó
- Chuyển động theo chiều dọc: Khi ngã, cơ thể chuyển động theo chiều dọc, hoặc thành phần chuyển động theo chiều dọc chiếm ưu thế do tác dụng của trọng lực. Tuy nhiên, hành động ngồi, nằm nhanh cũng có đặc điểm này
- Thay đổi hình dạng, tư thế cơ thể: Với các hoạt động thông thường, hình dáng cơ thể thay đổi chậm. Trong một khoảng thời gian ngắn có thể xem như không thay đổi. Nhưng với việc ngã, hình dạng cơ thể có thể thay đổi rất nhanh, hoặc ngay lập tức
- Không xuất hiện chuyển động sau khi ngã: Sau khi ngã, thông thường người ngã sẽ không có chuyển động cơ thể. Hoặc cũng có thể xuất hiện chuyển động rất nhanh như lăn qua lăn lại do bị đau. Nhưng với người già, có thể xem như không sảy ra kịch bản này

Một hệ thống phát hiện ngã tự động có thể được định nghĩa như một hệ thống trợ giúp với nhiệm vụ chính là báo động khi có sự kiện ngã xảy ra. Hệ thống này phải đảm bảo hoạt động thời gian thực để giảm thiểu thời gian người ngã nằm trên sàn từ sau thời điểm ngã đến khi được người chăm sóc phát hiện. Khoảng thời gian này là yếu tố chủ chốt quyết định mức độ nghiêm trọng sau ngã. Rất nhiều người già không thể tự di chuyển hoặc gọi trợ giúp sau khi ngã và đối mặt với các mối nguy hiểm cho sức khỏe. Trong các nghiên cứu được công bố gần đây, có thể phân loại các hướng nghiên cứu về bài toán phát hiện ngã thành các nhóm chính: Phát hiện ngã dựa trên thiết bị cảm biến mang theo người; dựa trên cảm biến tích hợp trên điện thoại di động thông minh; dựa trên camera độ sâu (depth camera); và dựa trên camera thông thường. Phần tiếp theo của chương này sẽ tóm lược khái quát các hướng nghiên cứu chính kể trên.

1.1. Phát hiện ngã sử dụng thiết bị mang theo người

Thiết bị mang theo người có thể được định nghĩa là các thiết bị cảm biến điện tử nhỏ có thể cầm theo, hoặc đính trên quần áo. Phần lớn các thiết bị phát hiện ngã mang theo người sử dụng cảm biến đo gia tốc. Trong đó có thể kết hợp cảm biến khác như con quay hồi chuyển để thu thập thông tin về vị trí của người mang. Việc sử dụng các cảm biến kể trên có thể giúp đánh giá dáng đi, sự cân bằng, mức độ chuyển động và vị trí cơ thể của người mang, giúp dự đoán về việc ngã. Xu hướng sử dụng thiết bị đeo được tăng lên trong những năm gần đây do sự phổ biến của các cảm biến giá rẻ được tích hợp sẵn trong điện thoại thông minh.

Hình 1.1. Thiết bị có tích hợp cảm biến như điện thoại hay gậy thông minh

1.1.1. Gia tốc kế gắn trên cơ thể

Thông tin về sự gia tăng tốc độ chuyển động trong quá trình ngã được thu thập dựa trên sử dụng các gia tốc kế ba trục độc lập được gắn trên các vị trí khác nhau của cơ thể. Sau đó, các kỹ thuật thường được áp dụng để xác định ngã bao gồm: i) sử dụng ngưỡng, trong đó việc ngã được ghi nhận nếu độ gia tăng vận tốc đạt ngưỡng xác định trước; ii) sử dụng học máy (machine learning) để phân loại giữa ngã và không phải ngã.

Một số nghiên cứu áp dụng kỹ thuật phân ngưỡng như [3, 11, 21, 22, 29, 36, 37, 50]. Trong khi đó, hướng tiếp cận sử dụng học máy bắt đầu xuất hiện từ năm 2010 sử dụng SVM (Support Vector Machine) [10, 26, 40, 48, 49]; multi-layer perceptron, Naïve Bayes, decision tree [26, 30]. Mặc dù vậy cho đến nay không có một kỹ thuật nào được chấp nhận như là một kỹ thuật tiêu chuẩn từ cộng đồng các nhà khoa học.

1.1.2. Cảm biến tích hợp trên điện thoại thông minh

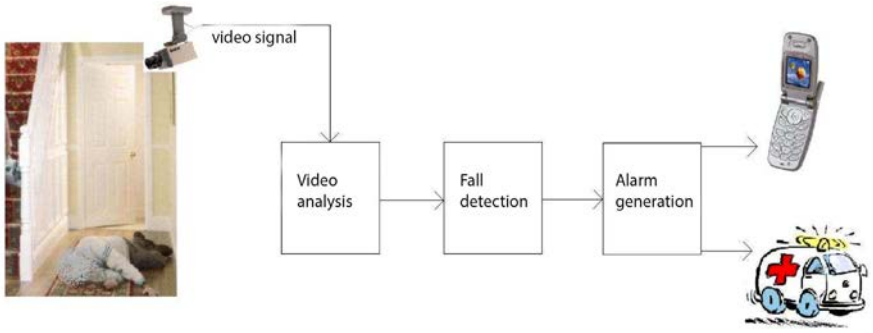
Ngày nay, điện thoại di động thông minh dần trở lên rất phổ biến và thường được tích hợp sẵn một cách phong phú các loại cảm biến như gia tốc kế, la bàn số, GPS, con quay hồi chuyển, micro và camera. Một số nghiên cứu đã khai thác ưu thế kể trên để đưa ra các phương thức phát hiện ngã dựa trên sử dụng điện thoại thông minh. Trong đó, một số thuật toán đơn giản sử dụng kỹ thuật phân ngưỡng như [23, 27, 31, 44, 46]. Một số nghiên cứu khác sử dụng các kỹ thuật học máy như [39, 45].

1.1.3. Xu hướng, ưu điểm và hạn chế

Nhìn chung, hướng tiếp cận sử dụng thiết bị mang theo người có xu hướng dịch chuyển sang sử dụng di động thông minh vì các ưu thế của nó, cộng với áp dụng các kỹ thuật học máy. Việc sử dụng thiết bị đeo được trong bài toán phát hiện ngã có ưu điểm là không bó hẹp ở phạm vi trong nhà mà có thể sử dụng cả ở môi trường bên ngoài. Tuy nhiên một nhược điểm lớn của cách tiếp cận này khi hướng đến đối tượng người sử dụng là người cao tuổi đó là người già thường xuyên bỏ quên thiết bị. Việc mang thiết bị theo người cũng gây những phiền phức nhất định. Một nhược điểm khác của việc sử dụng điện thoại thông minh là, chúng không được thiết kế cho mục đích chạy các ứng dụng đảm bảo an toàn mà ưu tiên được dành cho các chức năng nguyên thủy của điện thoại và thời gian sử dụng pin, dẫn đến các cảm biến không phải khi nào cũng hoạt động chính xác như mong muốn. Các nhà sản xuất điện thoại thông minh với các thiết kế kiến trúc khác nhau cho các cảm biến cũng dẫn đến tình trạng sự hoạt động của thuật toán có thể khác nhau trên các loại di động thông minh khác nhau.

1.2. Phát hiện ngã dựa trên phân tích dữ liệu video

Ngày nay, các thiết bị camera ngày càng trở lên phổ biến với giá thành thấp, vì vậy hệ thống phát hiện ngã dựa trên camera có chi phí không cao và dễ dàng triển khai. Hướng tiếp cận này dựa trên phân tích dữ liệu video thu được qua một hoặc nhiều camera giám sát. Các các camera này được lắp đặt xung quanh môi trường sinh hoạt thường ngày của người già. Có thể thấy, lợi thế lớn nhất là người sử dụng không cần phải mang theo thiết bị. Tuy vậy các phương pháp sử dụng camera giám sát bị giới hạn bởi môi trường trong nhà và không có cách nào hoạt động khi người già rời khỏi phòng, nơi có triển khai các camera. Các phương thức phát hiện ngã dựa trên camera có thể được coi là nhóm các phương thức sử dụng thị giác máy, phân biệt với các phương thức còn lại. Các phương thức sử dụng thị giác máy lại có thể chia thành ba nhóm nhỏ: nhóm sử dụng camera RGB đơn; nhóm dựa trên phân tích dữ liệu 3-D sử dụng hệ nhiều camera RGB; nhóm dựa trên phân tích dữ liệu 3-D sử dụng camera độ sâu (depth camera).



Hình 1.2. Minh hoạ hệ thống phát hiện ngã tự động dựa trên phân tích video

1.2.1. Phát hiện ngã sử dụng camera đơn

Phát hiện ngã sử dụng camera RGB đơn được nghiên cứu rộng rãi do việc cài đặt hệ thống rất dễ dàng với chi phí thấp. Các đặc trưng phổ biến được khai thác là đặc trưng hình dạng cơ thể, đặc trưng chuyển động, và việc thiếu vắng chuyển động sau ngã.

Đặc trưng hình dáng cơ thể được áp dụng rộng rãi cho việc phát hiện ngã như [5, 9, 14, 32, 35, 47, 52]. Các nghiên cứu [32, 47] sử dụng tỉ lệ giữa chiều cao và chiều rộng của cơ thể để xác định ngã. Mirmahboub và cộng sự [9] sử dụng một kỹ thuật trừ nền để tách vùng chuyển động trong chuỗi video, từ đó trích rút một số các đặc trưng hình dáng. Cuối cùng, một bộ phân lớp SVM được sử dụng để xác định việc ngã. Trong khi Rougier và các cộng sự của bà [14] sử dụng kỹ thuật so khớp hình dạng để theo vết vùng chuyển động tương ứng với cơ thể. Hình dáng cơ thể bị biến dạng trong khi ngã. Một số nghiên cứu đã sử dụng đặc điểm này bằng các kỹ thuật sử dụng bộ phân lớp dựa trên biến dạng hình dáng như [35], hoặc xây dựng một ellipse xấp xỉ vùng chuyển động thu được từ kỹ thuật trừ nền để mô hình hình dạng cơ thể.

Các đặc điểm chuyển động khi ngã thường rất khác biệt so với chuyển động trong các hoạt động thường nhật như đi lại, ngồi, nằm chủ động, làm việc nhà, etc. Vì thế có nhiều nghiên cứu dựa trên phân tích sự khác biệt này để phát hiện ngã, phân biệt ngã với các hoạt động thông thường khác, như [13, 25, 54, 56]. Liao và cộng sự [54] sử dụng kỹ thuật phân tích chuyển động cơ thể kết hợp đặc trưng hình dạng cơ thể để phân biệt giữa chủ động nằm với ngã. Trong khi Homa và cộng sự [25] áp dụng Integrated Time Motion Image (ITMI) cho phát hiện ngã. ITMI là một dạng dữ liệu không – thời gian bao gồm chuyển động và thông tin về thời gian của chuyển động. Cho trước một chuỗi video, ITMI sẽ tính toán và biểu diễn thông tin chuyển động xuất hiện trong video, sau đó áp dụng kỹ thuật phân tích thành phần chính (PCA) để giảm số chiều của thông tin đã biểu diễn được. Cuối cùng áp dụng mạng neural MLP để phân loại chuyển động và xác định ngã. Cũng có nghiên cứu sử dụng thông tin 3-D thu được từ camera đơn được hiệu chuẩn (calibrated) cho việc phát hiện ngã như [13]. Caroline và các cộng sự trích rút thông tin 3-D về quỹ đạo chuyển động của vùng đầu người, từ đó tính toán thông tin vận tốc chuyển động của đầu để phát hiện việc ngã.

Nhìn chung, vì những ưu điểm đã nêu, số lượng nghiên cứu phát hiện ngã dựa trên phân tích dữ liệu chuỗi video thu được từ một camera đơn là rất lớn, áp dụng nhiều kỹ thuật đa dạng. Các đặc trưng được sử dụng thường tập trung vào thông tin hình dạng cơ thể và thông tin chuyển động.

1.2.2. Phát hiện ngã sử dụng hệ multi camera

Một nhóm các phương pháp phát hiện ngã dựa trên thị giác là sử dụng thông tin 3-D thu được từ một hệ các camera được kết hợp cùng với nhau. Nhiều nghiên cứu thực hiện việc cân chỉnh các camera như [16-19] giúp việc tái tạo lại mô hình 3-D của đối tượng

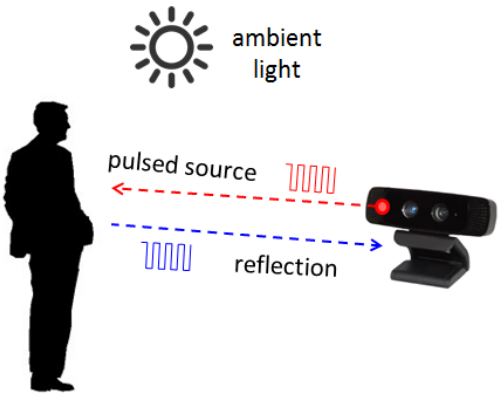
một cách chính xác. Tuy nhiên tiến trình cân chỉnh camera thường phức tạp và tiêu tốn thời gian. Auvinet và cộng sự [18,19] sử dụng một mạng các camera đã được cân chỉnh để tái tạo hình dạng ba chiều của cơ thể. Sau đó phân tích sự phân bố theo chiều dọc, nếu sự phân bố này là bất thường và gần mặt sàn, sẽ xác định là ngã. Còn Anderson và cộng sự [16, 17] lại áp dụng một nhánh của logic mờ cho phát hiện ngã.

Nhìn chung, việc sử dụng hệ multi camera mang đến lợi thế là cho phép dựng lại mô hình 3-D của đối tượng, trích rút được các thông tin 3 chiều, vốn là các thông tin hết sức phù hợp cho việc phát hiện ngã.

Bên cạnh việc tái tạo thông tin 3 chiều, hệ multi camera còn được sử dụng với mục đích như giám sát, phát hiện ngã ở nhiều phòng khác nhau [41]; phát hiện ngã từ các góc nhìn độc lập nhau [42]; và phát hiện ngã từ nhiều camera đơn độc lập rồi dung hợp kết quả với nhau [8].

1.2.3. Phát hiện ngã sử dụng Camera độ sâu

Camera độ sâu là loại camera có khả năng ghi nhận thông tin khoảng cách từ đối tượng đến bề mặt cảm biến, tương tự như mắt người. Nguyên lý hoạt động của nó dựa trên vận tốc chuyển động của ánh sáng được mô tả như Hình 1, như sau: Nguồn phát phát đi một trùm tia hồng ngoại được cấu trúc thành lưới, trùm tia này phản xạ trên bề mặt vật thể. Bộ phận cảm biến độ sâu (depth sensor) đặt gần nguồn phát sẽ thu nhận trùm tia dội lại và suy ra khoảng cách đến bề mặt vật thể dựa trên thời gian từ khi tia ra khỏi nguồn phát đến khi depth sensor nhận được. Từ đó xây dựng được đám mây điểm (points cloud) mang thông tin 3-D của vật thể.



Hình 1.3. Hoạt động của camera độ sâu

Phương thức phát hiện ngã sử dụng camera độ sâu lần đầu tiên được đề suất bởi G. Diraco và cộng sự [24] vào năm 2010 khi mà giá thành loại thiết bị này là rất cao. Có rất ít nhà nghiên cứu sau đó quan tâm đến việc áp dụng loại camera này vào bài toán phát hiện ngã. Tuy nhiên tình thế đó đang thay đổi khi gần đây camera độ sâu dần trở lên phổ biến với mức giá ngày càng được cải thiện. Đặc biệt là sau khi Microsoft ra mắt thiết bị Kinect, đã có rất nhiều nghiên cứu sử dụng Kinect cho phát hiện ngã.

Với sự trợ giúp của camera độ sâu, việc tính toán khoảng cách từ đầu người tới mặt sàn là tương đối đơn giản. Một số nghiên cứu đã sử dụng khoảng cách này như là một đặc trưng để xác định ngã như [6, 12, 24, 38]. Diraco và cộng sự [24] sử dụng camera độ sâu treo trên tường để giám sát. Hệ thống xác định việc ngã xảy ra khi trọng tâm cơ thể ở gần sàn quá một ngưỡng cho trước, và sau đó người ngã không chuyển động trong một vài giây. Trong khi đó Leone và công sự [6] xác định ngã dựa trên hai tiêu chí: khoảng cách từ trọng tâm cơ thể đến mặt sàn giảm xuống dưới ngưỡng xác định trước quá 900ms; sau đó người ngã không chuyển động hoặc chuyển động không đáng kể trong khoảng thời gian 4s. Rougier và cộng sự [12] sử dụng Kinect để thu nhận chuỗi ảnh độ sâu. Sau đó sử dụng ngưỡng khoảng cách trọng tâm đến sàn và tốc độ chuyển động để xác định ngã. Còn Michal và các cộng sự của ông [38] lại sử dụng một camera độ sâu gắn trên trần, sử dụng một bộ phân lớp KNN để phân biệt tư thế nằm trên mặt sàn khi ngã với các hoạt động thường ngày. Đặc trưng được sử dụng là khoảng cách đầu tới sàn; chiều dài và chiều rộng của vùng diện tích cơ thể.

Các hệ thống phát hiện ngã dựa trên camera độ sâu có cùng lợi thế về khai thác thông tin ba chiều như khi sử dụng hệ multi camera, nhưng khác với hệ multi camera, sử dụng camera độ sâu không cần cấu hình phức tạp, không tốn chi phí tính toán cho tiến trình cân chỉnh. Với việc loại thiết bị này đang dần trở lên phổ biến, ngày càng nhiều các nghiên cứu đề xuất phương thức phát hiện ngã áp dụng camera độ sâu. Tuy nhiên ở Việt Nam hiện tại loại camera này ít được biết đến.

CHƯƠNG 2.
CƠ SỞ LÝ THUYẾT

Hướng tiếp cận sử dụng các phương pháp phân tích dữ liệu chỗi video thu được qua camera để phát hiện ngã tự động nằm trong lớp bài toán phát hiện hành động (action detection) của lĩnh vực thị giác máy (computer vision), thuộc ngành khoa học máy tính (computer science). Lĩnh vực thị giác máy cố gắng mô phỏng lại những gì bộ não con người làm được với dữ liệu hình ảnh gửi về từ võng mạc, nghĩa là hiểu được ngữ cảnh dựa trên dữ liệu hình ảnh. Nó chủ yếu liên quan đến việc phân đoạn (segmentation), nhận diện (recognition), tái xây dựng mô hình 3D của đối tượng (reconstruction) và việc kết hợp các công việc đó cho mục đích hiểu ngữ cảnh.

Thị giác máy ứng dụng các kỹ thuật của xử lý ảnh số (digital image processing) với các mô hình học máy (machine learning) cũng như một số phương thức toán học để thực hiện mục tiêu nói trên. Có thể nói, Thị giác máy cùng với xử lý ảnh và trí tuệ nhân tạo, mà cụ thể là học máy, có rất nhiều phần giao thoa với nhau. Ranh giới giữa các lĩnh vực này rất khó để phân định rõ ràng và còn gây nhiều tranh cãi. Tuy nhiên, xử lý ảnh, có thể được xem như lĩnh vực tập trung chủ yếu vào vấn đề xử lý dữ liệu ảnh thô mà không thu lại bất kỳ tri thức nào từ chúng. Ví dụ, trong bài toán phân đoạn ảnh dựa trên ngữ nghĩa, như xác định vị trí con mèo trong chuỗi video, một số bộ lọc cần được áp dụng trên ảnh trong quá trình xử lý. Đó là công việc của xử lý ảnh số. Còn việc nhận diện đối tượng (con mèo) trong khung cảnh của ảnh lại là nhiệm vụ của thị giác máy. Kết quả đầu ra của xử lý ảnh thường là một ảnh khác (gọi là ảnh đã được xử lý), còn thị giác máy nhận dữ liệu đầu vào là ảnh (kết quả của quá trình xử lý ảnh) và đầu ra là sự phân lớp (classifying), là tri thức về ngữ cảnh trong ảnh, là thông tin ngữ nghĩa. Phần cơ sở lý thuyết sẽ trình bày một cách khái quát về xử lý ảnh số và thị giác máy, đồng thời giới thiệu một số kỹ thuật, giải thuật cơ bản của các lĩnh vực này mà có liên quan trực tiếp hoặc gián tiếp đến bài toán của luận văn này.

2.1. Tổng quan về xử lý ảnh số

Ngày nay, các lĩnh vực như y tế, thiên văn học, vật lý, hóa học, viễn thám, chế tạo, v.v.. và rất nhiều lĩnh vực khác nữa ngày càng lưu trữ, hiển thị, cung cấp ảnh số với số lượng vô cùng lớn. Thách thức đặt ra cho giới khoa học là làm sao trích rút ra được các thông tin có giá trị từ ảnh số nguyên gốc một cách nhanh chóng. Đó là mục đích chính của lĩnh vực xử lý ảnh số: chuyển đổi ảnh số thành thông tin.

2.1.1. Ảnh kỹ thuật số

Ảnh kỹ thuật số là dữ liệu được các thiết bị ghi hình kỹ thuật số như máy ảnh số, camera số ghi lại từ phép chiếu hình ảnh ba chiều của vật thể từ thế giới thực lên mặt phẳng hai chiều. Ánh sáng từ nguồn sáng phản xạ trên bề mặt vật thể, đi qua thấu kính đến bề mặt

cảm biến điện tử. Cảm biến này tiếp nhận ánh sáng và chuyển đổi thành tín hiệu điện tử dạng tương tự. Sau đó bộ phận chuyển đổi tương tự - kỹ thuật số thực hiện việc lấy mẫu (sampling) để chuyển tín hiệu tương tự sang tín hiệu số và lưu xuống thiết bị lưu trữ. Ảnh số bao gồm một lưới các điểm ảnh (pixel), được lưu trữ dưới dạng mảng hai chiều. Trong đó, mỗi điểm ảnh là một thành phần ảnh nhỏ nhất biểu diễn giá trị cường độ sáng tại vị trí của nó. Giá trị của mỗi điểm ảnh là rời rạc. Mảng hai chiều lưu trữ dữ liệu ảnh gồm một số lượng hữu hạn số hàng và số cột.

Ảnh nhị phân

Mỗi điểm ảnh chỉ là màu đen hoặc trắng, được biểu diễn bằng 0 và 1. Vì chỉ có hai giá trị có thể cho mỗi điểm ảnh, chúng ta chỉ cần một bit cho mỗi điểm ảnh. Như vậy, việc lưu trữ khá hiệu quả. Ảnh nhị phân có thể phù hợp với văn bản (in hoặc viết tay), dấu vân tay, thiết kế kiến trúc. (Phân tích và xử lý ảnh – TS. Đào Nam Anh, Nhà xuất bản Bách Khoa Hà Nội).

Ảnh đa mức xám (Grayscale)

Giá trị cường độ điểm ảnh được mã hóa trong L mức. Trong đó mức độ đen hay trắng được chia thành L khoảng đều nhau. Giá trị mỗi điểm ảnh nằm trong L khoảng này, là giá trị rời rạc biểu diễn mức cường độ sáng tại vị trí điểm ảnh. Giá trị điểm ảnh càng cao, cường độ sáng càng lớn và ngược lại. Ngày nay các thiết bị thường sử dụng mỗi 8bit để mã hóa giá trị một điểm ảnh, nghĩa là L = 256 khoảng. Giá trị điểm ảnh nằm giữa 0 và 255. Trong trường hợp L = 2, một điểm ảnh chỉ có 2 mức giá trị 0 và 1, ta được ảnh nhị phân với mức 0 biểu diễn màu đen tuyệt đối và mức 1 biểu diễn màu trắng tuyệt đối. Nếu L > 2 ta được ảnh đa mức xám. Ảnh nhị phân có thể thu được qua phép tách ngưỡng ảnh đa mức xám: Giá trị điểm ảnh lớn hơn ngưỡng cho trước tương ứng với giá trị 1, nhỏ hơn ngưỡng tương ứng với giá trị 0 trên ảnh kết quả.

Ảnh đa mức xám được lưu trữ trên một mảng hai chiều duy nhất. Rất nhiều kỹ thuật trong xử lý ảnh số được thực hiện trên ảnh đa mức xám khi không cần thiết phải quan tâm đến thông tin màu sắc của ảnh, giúp giảm độ phức tạp tính toán.

Ảnh màu

Qua nghiên cứu thị lực của người với màu sắc, James Clerk Maxwell đã phát hiện ra rằng các tế bào hình nón chia thành 3 loại: một loại nhạy cảm với ánh sáng đỏ, một loại nhạy cảm với xanh lá, loại còn lại nhạy cảm với xanh dương, ông phán đoán rằng mắt người có thể tổng hợp một màu sắc bất kỳ dựa trên ba màu cơ bản trên. Các cuộc thử nghiệm thành công sau đó đã mở ra một kỉ nguyên mới về nhiếp ảnh màu.

Máy ảnh màu kỹ thuật số có bộ phận phân tách ánh sáng thành ba phổ màu cơ bản riêng biệt: đỏ (R); xanh lá (G); xanh dương (B). Mỗi phổ màu này được biến đổi thành tín hiệu số và lưu trữ riêng biệt. Mỗi kênh màu được lưu trữ tương tự như ảnh đa mức xám. Nếu dùng 8bit để mã hóa giá trị một kênh màu của điểm ảnh, với 3 kênh màu, để biểu

diễn một điểm ảnh cần 24bit. Nghĩa là cần gấp 3 lần không gian lưu trữ cho ảnh màu so với ảnh đa mức xám. Các kỹ thuật phân đoạn ảnh dựa trên màu sắc được thực hiện trên ảnh màu.

Ảnh đa phổ

Với thông tin về 3 màu cơ bản, ta có thể tổng hợp lên bất kì màu sắc nào trong dải nhìn thấy của mắt người. Tuy nhiên trong các lĩnh vực như viễn thám, y học, người ta còn quan tâm đến thông tin của các dải ánh sáng không nhìn thấy được. Vì vậy người ta cần lưu trữ các phổ khác, ngoài ba phổ màu cơ bản trên. Ảnh trong trường hợp này gọi là ảnh đa phổ. Trong khuôn khổ bài toán phát hiện ngã, luận văn này không đề cập đến loại ảnh này.

2.1.2. Xử lý ảnh số

Xử lý ảnh số là quá trình áp dụng các phương thức, thuật toán để tác động vào và biến đổi ảnh ban đầu thành ảnh mới có chất lượng tốt hơn theo một tiêu chí xác định trước, hoặc trích rút các thông tin có ích từ dữ liệu ảnh.

Về mặt toán học, ảnh số có thể được coi là một hàm rời rạc hai biến $f(x,y)$ với x, y là tọa độ của điểm ảnh. Giá trị hàm số $f(x,y)$ chính là giá trị cường độ điểm ảnh tại vị trí x,y. Miền giá trị của f là: $0 \le f \le f_{max.}$ Với f_{max} là giá trị lớn nhất của điểm ảnh. Với mã hóa 8bit, $f_{max} = 255$. Quá trình xử lý ảnh là quá trình thực hiện các phép biến đổi trên $f(x,y)$. Vì vậy có thể nói xử lý ảnh số là một dạng của xử lý tín hiệu số.

Xử lý ảnh thông thường bao gồm các bước sau đây:

- Quá trình thu nhận ảnh
- Phân tích và biến đổi ảnh gồm tiền xử lý, phân đoạn và trích rút đặc trưng ảnh.
- Biểu diễn kết quả như là ảnh kết quả, hoặc các báo cáo thu được từ việc phân tích ảnh

Một hệ thống xử lý ảnh gồm 5 thành phần: Thu nhận ảnh; tiền xử lý; phân đoạn ảnh; trích rút đặc trưng (mức thấp) ảnh; và mô tả, phân loại ảnh. Sơ đồ hệ thống được minh họa như Hình 2.1 dưới đây.

Hình 2.1. Hệ thống phân tích ảnh số

Trong đó, thu nhận ảnh (Acquisition) là ghi nhận hình ảnh và lưu dưới dạng thức phù hợp cho mục đích phân tích, xử lý. Còn tiền xử lý (Preprocessing) là quá trình nâng cao chất lượng ảnh và khử nhiễu. Phân đoạn ảnh (Segmentation) thực hiện việc gom nhóm các điểm ảnh thành các vùng, từ đó định ra các đường bao quanh khu vực ảnh chứa

thông tin cần quan tâm. Còn trích rút đặc trưng (Feature Extraction) là công việc làm nổi bật một đặc trưng cần quan tâm của ảnh. Tiếp theo là việc biểu diễn kết quả (Presentation) dưới một dạng thức thích hợp cho các tiến trình xử lý tiếp theo của máy tính.

1) Thu nhận ảnh

Phần lớn ảnh kỹ thuật số được thu bằng nguồn ánh sáng trong vùng nhìn thấy bởi ưu điểm là an toàn, giá thành thấp và có thể được xử lý bởi các phần cứng thích hợp. Có hai phương thức phổ biến để tạo ra ảnh số là sử dụng camera kỹ thuật số và máy quét ảnh (scanner). Nói chung, giai đoạn thu nhận hình ảnh có liên quan phần nào tới giai đoạn tiền xử lý, chẳng hạn như việc thu phóng kích thước ảnh (scaling) có thể được thực hiện ở bước này.

Quá trình thu nhận ảnh là quá trình biến đổi tín hiệu liên tục trong thế giới thực thành tín hiệu số rời rạc, gọi là số hóa, gồm hai bước là lấy mẫu (sampling) và lượng tử hóa (quantization).

a) Lấy mẫu

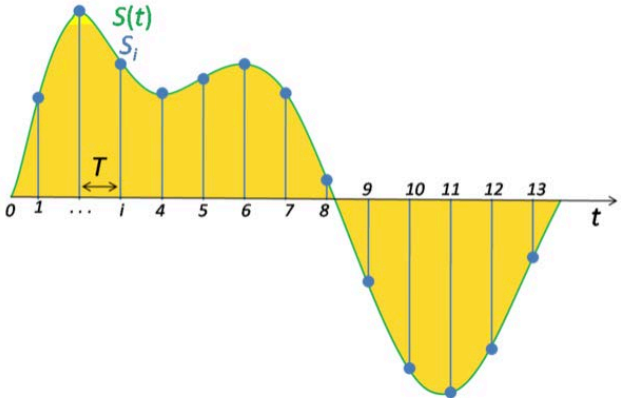
Giá trị cường độ sáng của hình ảnh trong thế giới thực có tính liên tục. Tuy nhiên vì nguyên lý mã hóa dữ liệu bởi các bit 0 và 1, máy tính chỉ có thể lưu trữ và xử lý các dữ liệu rời rạc. Việc lấy mẫu là quá trình chuyển đổi một tín hiệu liên tục thành một chuỗi số (một tín hiệu rời rạc). Yêu cầu đặt ra là tín hiệu ban đầu phải được khôi phục chính xác từ chuỗi số thu được qua lấy mẫu. Định lý lấy mẫu Nyquist – Shannon cung cấp điều kiện đủ để đạt được yêu cầu đó. Định lý lấy mẫu đảm bảo rằng một tín hiệu có thể được tái tạo hoàn toàn từ tín hiệu sau lấy mẫu nếu tần số lấy mẫu lớn hơn hai lần thành phần tần số lớn nhất của tín hiệu ban đầu:

$$f_s \geq 2f_{max}$$

Hay chu kỳ lấy mẫu phải đảm bảo:

$$T \geq \frac{1}{2f_{max}}$$

Chu kỳ lấy mẫu *T* là khoảng thời gian giữa hai thời điểm lấy mẫu liên tiếp. Giá trị biên độ của tín hiệu tại mỗi thời điểm lấy mẫu được lưu lại, tạo thành chuỗi số rời rạc. Chuỗi số này chính là kết quả của việc lấy mẫu. Hình 2.2 là một ví dụ minh họa cho chu kỳ lấy mẫu và giá trị biên độ thu được.



Hình 2.2. Minh họa chu kỳ lấy mẫu tín hiệu

b) Lượng tử hóa

Giá trị các điểm ảnh thu được từ bước lấy mẫu là rời rạc. Tuy nhiên miền giá trị của nó rất rộng. Như đã trình bày, cường độ sáng của điểm ảnh thường được chia thành 256 khoảng, nhận giá trị từ 0 đến 255. Vì thế cần phải xấp xỉ giá trị lấy mẫu bằng một đại lượng thuộc 256 khoảng đó.

c) Nén ảnh

Dữ liệu ảnh thu được qua lấy mẫu và lượng tử hóa vẫn có kích thước khá lớn. Vì thế để giảm chi phí lưu trữ và truyền tải dữ liệu, cần phải có các kỹ thuật làm giảm kích thước ảnh, gọi là nén ảnh. Nén ảnh là quá trình loại bỏ các thông tin dư thừa, sử dụng các dạng thức biểu diễn dữ liệu phù hợp làm giảm kích thước ảnh.

Các kỹ thuật nén ảnh có thể phân chia vào hai nhóm: nén bảo toàn thông tin và không bảo toàn thông tin. Trong đó nén bảo toàn thông tin giúp khôi phục hoàn toàn dữ liệu qua giải nén nhưng hiệu quả nén không cao, còn nén không bảo toàn thông tin cho hiệu quả nén cao nhưng lại gây mất mát dữ liệu. Các hướng tiếp cận chính của nén ảnh gồm có: dựa trên thống kê tần xuất xuất hiện của giá trị điểm ảnh; dựa vào vị trí không gian của điểm ảnh, khai thác sự giống nhau của các điểm ảnh gần nhau; thực hiện các phép biến đổi ảnh; và khai thác sự lặp lại của các chi tiết ảnh.

Phụ thuộc vào kỹ thuật nén được xử dụng, ảnh số có các định dạng khác nhau như BMP, GIF, JPEG, PNG, v.v..

Cấu trúc trung của các định dạng biểu diễn ảnh gồm 3 phần:

- Phần header: Chứa các thông tin về phương thức mã hóa; số bit dùng để mã hóa một điểm ảnh; kích thước và độ phân giải ảnh; v.v..
- Dữ liệu nén của ảnh: Dữ liệu hình ảnh đã được mã hóa theo phương thức đã chỉ ra ở header.
- Bảng màu: Cho biết thông tin về bảng màu mà ảnh sử dụng để hiển thị.

2) Tiền xử lý

Mục đích của bước tiền xử lý ảnh là loại bỏ các thông tin dư thừa, không có giá trị cho tiến trình phân tích, loại bỏ nhiễu, nâng cao độ tương phản,v.v.. Tiền xử lý bao gồm các bước cơ bản như thay đổi kích thước của ảnh và lọc ảnh.

a) Thay đổi kích thước ảnh (Re-sizing)

Thay đổi kích thước ảnh khi cần tăng hoặc giảm số lượng điểm ảnh của ảnh gốc, được thực hiện dựa trên nội suy ảnh. Đó là một tiến trình tái lấy mẫu để xác định giá trị nằm giữa hai điểm ảnh cho trước. Ảnh kết quả thu được có thể có số lượng điểm ảnh nhiều hơn hoặc ít hơn so với ảnh gốc. Giá trị cường độ các điểm ảnh bổ xung vào ảnh gốc thu được qua phép nội suy nếu độ phân giải không gian của ảnh tăng thêm.

b) Lọc ảnh (Filtering)

Ảnh thu được từ các thiết bị thu nhận thường bao gồm nhiều loại nhiễu. Hoặc vì lý do camera bị rung khi ghi hình, ảnh thu được bị nhòe, dẫn đến mất mát thông tin, suy giảm độ chi tiết ảnh và vùng biên ảnh. Lọc ảnh giúp giảm bớt nhiếu và làm tăng cường hoặc suy giảm mức độ chi tiết của ảnh. Kỹ thuật lọc ảnh có thể chia làm hai loại: Lọc trên miền không gian, dựa trên mối tương quan giá trị điểm ảnh với các điểm ảnh trong vùng lân cận của nó. Ví dụ như lọc trung vị, lọc trung bình, etc; Và lọc trên miền tần số, dựa trên việc thực hiện phép biến đổi Fourier để biểu diễn hàm ảnh $I = f(x,y)$ trên miền tần số. Sau đó loại bỏ tần sồ thấp (High Pass Filter) trong trường hợp cần tăng cường chi tiết ảnh hoặc loại bỏ thành phần tần số cao (Low Pass Filter) trong trường hợp cần làm trơn ảnh (smooth).

3) Phân đoạn ảnh

Phân đoạn ảnh là thao tác chia nhỏ ảnh thành các vùng đồng tính (cùng tính chất về màu sắc, kết cấu), hay nói cách khác là xác định đường biên giữa các vùng ảnh. Các vùng này tương ứng với toàn bộ hoặc một phần của đối tượng trong ảnh. Quá trình này giúp phân chia ảnh thành các vùng mang nhiều ý nghĩa. Việc phân đoạn ảnh dựa trên các đặc tính của ảnh như mức cường độ sáng, cạnh, màu sắc hay kết cấu ảnh. Mức độ chi tiết của việc phân chia các vùng phụ thuộc vào việc bài toán đã được giải quyết hay chưa. Việc phân đoạn kết thúc khi đối tượng hoặc vùng cần quan tâm trong bài toán đó được phát hiện. Ví dụ như trong bài toán tự động kiểm tra dây truyền lắp ráp thiết bị điện tử, cần phân tích ảnh chụp sản phẩm để xác định xem các bất thường có xuất hiện hay không như thiếu mất một thành phần nào đó, hoặc các mạch hàn bị gián đoạn. Không có cách nào định trước mức độ chi tiết của tiến trình phân đoạn mà chỉ kết thúc khi đối tượng cần quan tâm được phát hiện.

Phân đoạn ảnh là một trong những công việc khó khăn nhất trong xử lý ảnh. Mức độ chính xác của việc phân đoạn có vai trò quyết định đến sự thành công hoặc thất bại của toàn bộ quá trình phân tích ảnh.

Đa phần các phương pháp phân đoạn ảnh là dựa trên thuộc tính của giá trị cường độ sáng, chia thành hai nhóm: sự gián đoạn và tính tương tự. Các phương pháp thuộc nhóm thứ nhất dựa trên sự thay đổi đột ngột cường độ sáng, ví dụ như tại cạnh của đối tượng (edge). Còn hướng tiếp cận chính của nhóm thứ hai là phân chia ảnh thành các vùng có sự tương đồng theo một tập các điều kiện xác định trước. Phân ngưỡng, mở rộng vùng, tách và trộn vùng là các ví dụ về phương thức thuộc nhóm này.

4) Trích rút đặc trưng ảnh

Trong xử lý ảnh, trích rút đặc trưng là quá trình xử lý ảnh mức thấp (low level), khi kết quả thu được chỉ sau một bước thực hiện trên ảnh. Một đặc trưng có thể được định nghĩa là một phần thông tin cần quan tâm trong ảnh. Kết quả mong muốn của tiến trình phát hiện đặc trưng có thể lặp đi lặp lại. Ví dụ, trong cùng một ngữ cảnh, có hay không có các đặc trưng giống nhau phát hiện được trên các ảnh khác nhau. Các cạnh, đường, góc, điểm giao nhau thường chứa các thông tin giá trị trong ảnh, vì vậy cần phải có các kỹ thuật đáng tin cậy để phát hiện các đặc trưng liên quan đến chúng. Đặc trưng thu được trong phạm vi của xử lý ảnh số là đặc trưng mức thấp, nghĩa là không mang thông tin về ngữ nghĩa.

5) Biểu diễn ảnh, phân lớp ảnh

Biểu diễn ảnh sử dụng kết quả đầu ra của bước trích rút đặc trưng ảnh như dữ liệu điểm ảnh, đường bao phân chia các vùng, hoặc toàn bộ điểm ảnh nằm bên trong một vùng. Trong một số trường hợp, cần chuyển đổi các dữ liệu đầu ra trên thành dạng thức phù hợp cho việc tính toán của máy tính. Cách thức biểu diễn kết quả được quyết định dựa trên mục đích khác nhau: Nếu quan tâm đến thông tin về hình dạng bên ngoài của đối tượng, như góc cạnh hay độ cong, cần biểu diễn các đường bao; Còn nếu tập trung vào các thuộc tính bên trong như kết cấu, hoặc khung xương hình dạng, cần biểu diễn ảnh theo vùng. Việc lựa chọn hình thức biểu diễn chỉ là một phần của quá trình chuyển đổi dữ liệu thô của ảnh thành dạng thức phù hợp cho tiến trình xử lý tiếp theo. Cần thiết phải có một phương thức mô tả dữ liệu sao cho các đặc trưng cần quan tâm được làm nổi bật lên.

Phân lớp là quá trình sử dụng các đặc tính thu được từ việc định lượng các thông tin cần quan tâm từ ảnh để phân biệt một lớp các đối tượng từ các đối tượng khác.

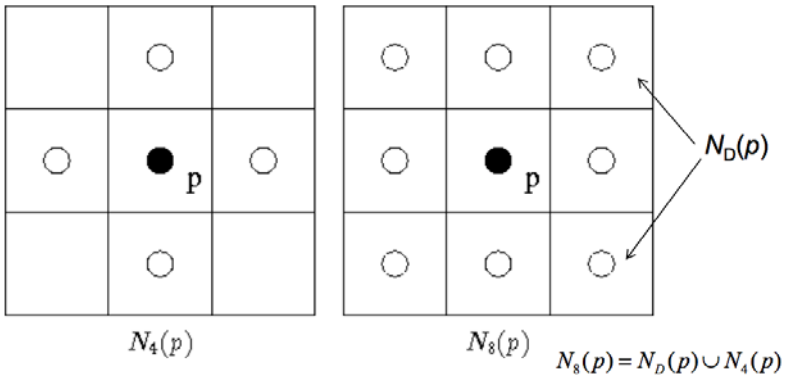
2.1.3. Các phép toán chính trong xử lý ảnh

Đa phần các kỹ thuật xử lý ảnh là các quá trình tác động lên ảnh đầu vào để thu được kết quả dựa trên phân tích mối tương quan về giá trị một điểm ảnh so với các điểm ảnh lân cận.

1) Điểm ảnh và quan hệ giữa các điểm ảnh

Quan hệ láng riềng

Ảnh được biểu diễn bằng một ma trận hai chiều như đã giới thiệu trong các phần trên. Trong đó, mỗi phần tử thuộc ma trận có giá trị là mức độ xám của điểm ảnh. Chỉ số của phần tử đó cũng là tọa độ của điểm ảnh tương ứng. Nếu biểu diễn các điểm ảnh bằng những ô vuông kích thước bằng nhau thì mỗi điểm ảnh có 4 láng riềng theo phương đứng và ngang (có chung cạnh), 4 láng riềng khác theo đường chéo (có chung góc). Một số điểm ảnh bị khuyết điểm láng riềng trong trường hợp nó nằm gần đường biên ảnh.



Hình 2.3. Các láng riềng của một điểm ảnh: 4 láng riềng theo phương đứng và phương ngang; 8 láng riềng bao gồm cả theo đường chéo

Ký hiệu $N_4(p)$ là tập các láng riềng theo chiều ngang và chiều đứng; còn $N_D(p)$ là tập các láng riềng theo đường chéo; $N_8(p)$ là tập các láng riềng của điểm ảnh p.

Quan hệ phụ cận (Adjacency)

Hai điểm ảnh được liên kết với nhau, hay có quan hệ phụ cận nếu chúng là láng riềng và giá trị cường độ sáng của chúng thỏa mãn điều kiện cho trước hoặc là bằng nhau. Ví dụ, với ảnh nhị phân, hai điểm ảnh được kết nối với nhau nếu chúng là láng riềng và có cùng giá trị (0/1).

Gọi V là tập các giá trị cường độ dùng để xác định quan hệ phụ cận và liên kết. Đối với ảnh nhị phân, V = {1} nếu xét quan hệ phụ cận của một điểm ảnh có giá trị là 1. Trong ảnh đa mức xám, V có nhiều phần tử hơn, ví dụ V = {180, 181, .., 200}. Nếu giá trị cường độ nằm trong khoảng từ 0 đến 255, V là tập con của tập gồm 256 giá trị.

Quan hệ phụ cận được chia thành 3 dạng:

- Phụ cận - 4: 2 điểm ảnh p và q với giá trị thuộc tập V là phụ cận - 4 nếu q thuộc tập $N_4(p)$.
- Phụ cận - 8: Hai điểm ảnh p và q với giá trị thuộc tập V là phụ cận - 8 nếu q thuộc tập $N_8(p)$.
- Phụ cận - m: Hai điểm ảnh p và q với giá trị thuộc tập V là phụ cận – m nếu q thuộc $N_4(p)$ hoặc thuộc $N_D(p)$ và tập $N_4(p) \cap N_4(q)$ không có điểm ảnh nào có giá trị nằm trong tập V. Ở đây "*m*" là viết tắt của *mixed.*

Đường đi (path)

Đường đi từ điểm ảnh p tọa độ (x,y) tới điểm ảnh q tọa độ (s,t) là một chuỗi các điểm ảnh không lặp lại có tọa độ (x_0, y_0); (x_1, y_1);…; (x_n, y_n), với $(x_0, y_0) = (x, y)$ và $(x_n, y_n) = (s, t)$. Trong đó: (xi, yi) là phụ cận của (xi-1, yi-1) $\forall$ $1 \leq i \leq n$, n là độ dài đường đi.

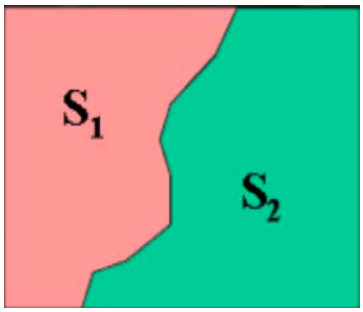
Trong trường hợp $(x_0, y_0) = (x_n, y_n)$, đường đi được gọi là đường khép kín.

Đường đi được phân loại thành đường đi – 4; đường đi – 8; và đường đi – m tương ứng với kiểu phụ cận giữa hai điểm ảnh.

Kết nối (Connectivity)

Gọi S là tập con của tập các điểm ảnh thuộc một ảnh. Hai điểm ảnh p và q gọi là được kết nối trong S nếu tồn tại một đường đi giữa chúng.

Hai tập S_1 và S_2 gọi là phụ cận nếu một vài điểm ảnh thuộc S_1 phụ cận với một vài điểm ảnh thuộc S_2. Ví dụ minh họa như Hình 4.



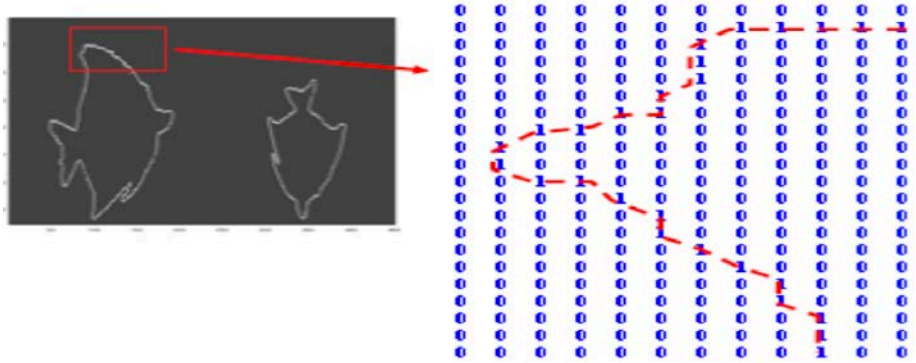
Hình 2.4. Hai tập điểm ảnh phụ cận với nhau

Vùng (Region)

Gọi R là tập con thuộc tập các điểm ảnh của ảnh. R được gọi là một vùng ảnh nếu R là một tập kết nối. Vùng ảnh không có phụ cận gọi là vùng cô lập. Không tồn tại kiểu đường đi – 4 giữa hai vùng khác nhau.

Nếu ký hiệu R_u là hợp nhất của tất cả các vùng, còn $(R_u)^c$ là phần còn lại của ảnh, thì R_u được gọi là vùng nổi lên (foreground), còn $(R_u)^c$ là nền (background) của ảnh.

Đường bao (contour) của một vùng R là tập các điểm thuộc R và là phụ cận của một số điểm thuộc phần bù của R, hay thuộc $(R_u)^c$, được minh họa như Hình 2.5 dưới đây.



Hình 2.5. Minh họa đường bao của vùng ảnh

2) Một số phép toán cơ bản trong xử lý ảnh số

Các phép toán với toán tử điểm

Các biến đổi đơn giản nhất trong xử lý ảnh số là các phép toán với toán tử điểm, trong đó giá trị điểm ảnh đầu ra chỉ phụ thuộc vào giá trị của điểm ảnh đầu vào tương ứng. Ví dụ như các phép biến đổi về độ sáng, hay điều chỉnh độ tương phản v.v..

a) Phép toán điều chỉnh độ sáng

Là phép toán làm thay đổi tổng thể độ sáng của ảnh. Phép toán này chỉ đơn giản là cộng thêm hoặc trừ bớt một số nguyên c vào giá trị mọi điểm ảnh thuộc ảnh:

$$I_{(x,y)} = I_{(x,y)} \pm c$$

Nếu cộng thêm c, độ sáng tổng thể của ảnh tăng và ngược lại, nếu trừ đi c, độ sáng tổng thể sẽ giảm.

b) Tách ngưỡng

Tách ngưỡng là phép toán biến đổi giá trị mức xám để phân hoạch các điểm ảnh thành các nhóm riêng biệt. Kỹ thuật này sử dụng một giá trị làm ngưỡng là θ và các giá trị V_1, V_2, $V_2 > V_1$. Xét mọi điểm ảnh p, nếu giá trị của p ≥ θ, thì gán giá trị mới cho p là V_2, nếu không giá trị mới của p là V_1:

$$p(x,y) = \begin{cases} V_1, & x < \theta \\ V_2, & x \geq \theta \end{cases}$$

Khi $V_1 = 0$ và $V_2 = 1$, ta thu được ảnh nhị phân sau phép tách ngưỡng.

c) Phép toán điều chỉnh độ tương phản (Contrast)

Điều chỉnh độ tương phản là kỹ thuật cải thiện chất lượng ảnh bằng cách làm biến đổi khoảng giá trị điểm ảnh sang khoảng giá trị mong muốn, ví dụ như khoảng đầy đủ các giá trị có thể của điểm ảnh mà kiểu ảnh đó cho phép. Ví dụ, với ảnh xám mã hóa bởi 8 bit, khoảng giá trị này từ 0 đến 255.

Độ tương phản là mức độ chênh lệch giữa giá trị lớn nhất và nhỏ nhất của các điểm ảnh trong cùng một ảnh. Độ tương phản càng cao, chi tiết ảnh càng nổi bật, ảnh càng sắc nét. Để làm thay đổi độ tương phản, tăng hoặc giảm tùy theo các mục đích khác nhau của bước tiếp theo trong tiến trình xử lý ảnh, các kỹ thuật được áp dụng là hiệu chỉnh min – max, hiệu chỉnh histogram, hiệu chỉnh gamma, v.v. Các kỹ thuật trên làm thay đổi khoảng giá trị điểm ảnh trong ảnh, điều chỉnh lại các giải giá trị để đạt được mục đích mong muốn. Ví dụ, với hiệu chỉnh min – max, công thức hiệu chỉnh như sau:

$$I_{new} = \frac{I_{old} - I_{min}}{I_{max} - I_{min}} \times 256$$

Trong đó, I_{max} và I_{min} là giá trị lớn nhất và nhỏ nhất của điểm ảnh. I_{old} và I_{new} là giá trị cường độ sáng của điểm ảnh trước và sau khi điều chỉnh.

Hình 2.6. Ví dụ minh họa điều chỉnh độ tương phản: Ảnh bên trái là ảnh ban đầu; ảnh bên phải là ảnh kết quả thu được sau điều chỉnh tương phản

d) Cân bằng biểu đồ mức xám (Histogram Equalization)

Thực chất cân bằng histogram cũng là một kỹ thuật điều chỉnh độ tương phản toàn cục. Trong đó, histogram là biểu đồ thể hiện tần suất suất hiện của mức xám *g* trong ảnh, hay là tổng số điểm ảnh có mức xám bằng *g*. Ký hiệu là *h(g)*. Một biểu đồ mức xám lý tưởng là biểu đồ phẳng, nghĩa là giá trị điểm ảnh phân bố đồng đều trên mọi khoảng mức xám.

Giả sử ảnh đa mức xám *I* có giải giá trị không phân bố đều mà co cụm trong một khoảng ngắn nào đó, thì ảnh *I* có khả năng có độ tương phản thấp. Mục tiêu của việc cân bằng biểu đồ mức xám là trải đều giá trị điểm ảnh trên một vùng giá trị rộng hơn, qua đó ảnh sẽ có độ tương phản cao hơn. Để thực hiện điều đó, cần phải xác định một hàm tham chiếu mức cường độ sáng *f(I)* để biểu đồ kết quả có tính bằng phẳng. Hàm tham chiếu này được xác định như sau:

- Xác xuất để một điểm ảnh thuộc ảnh đa mức xám *x* nhận giá trị mức xám *i* là:

$$p_x(i) = p_x(x = i) = \frac{n_i}{n}, 0 \leq i < L$$

  trong đó, *L* là tổng số mức xám của ảnh, thông thường bằng 256; *n* là tổng số điểm ảnh thuộc ảnh; n_i là tổng số điểm ảnh có mức xám = i. Giá trị của $p_x(i)$ nằm trong khoảng [0,1].

- Hàm phân bố tích lũy tương ứng với p_x được xác định:

$$cdf_x(i) = \sum_{j=0}^{i} p_x(j)$$

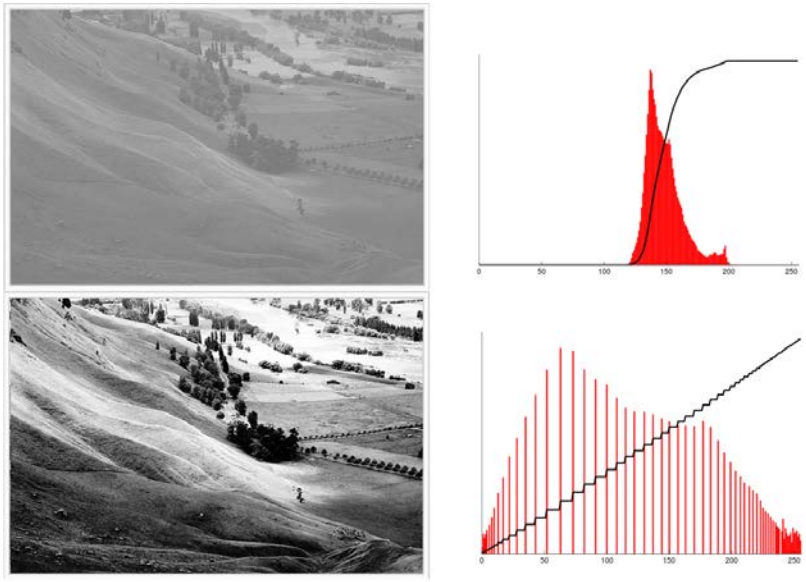
- Cần xác định một phép biến đổi $y = T(x)$ trên ảnh *x* để tạo ra ảnh *y* có biểu đồ mức xám bằng phẳng, nghĩa là hàm phân bố tích lũy *CDF* trên *y* có tính tuyến tính trong khoảng giá trị mức xám của *y*: $cdf_y(i) = iK$, với *K* là hằng số. Mặt khác, vì tính chất của hàm phân bố tích lũy:

$$cdf_y(i') = cdf_y\big(T(i)\big) = cdf_x(i), \qquad i \in [0, L]$$

Vậy, số điểm ảnh có giá trị $\leq i'$ thuộc ảnh *y* thu được qua phép biến đổi *T(i)* bằng với số điểm ảnh có giá trị $\leq i$ trong ảnh ban đầu *x*, và hàm phân bố tích lũy của *y* có tính chất tuyến tính. Suy ra:

$$i' = cdf_x(i) * (L - 1)$$

Hình 2.7 minh họa kết quả của kỹ thuật cân bằng biểu đồ mức xám: Ảnh trước khi cân bằng tương phản thấp, biểu đồ mức xám co cụm, hàm phân bố tích lũy không tuyến tính; Ảnh sau phép cân bằng sắc nét, độ tương phản cao, biểu đồ mức xám trải đều trên toàn giải giá trị và hàm phân bố tích lũy có dạng thức tuyến tính.



Hình 2.7. Minh họa cân bằng biểu đồ mức xám

Các phép toán với toán tử không gian

Là các phép toán mà giá trị điểm ảnh kết quả không chỉ phụ thuộc vào giá trị điểm ảnh đầu vào mà còn phụ thuộc vào các điểm ảnh lân cận, hay nói cách khác là phụ thuộc vào vị trí của điểm ảnh đầu vào.

a) Phép toán cửa sổ trượt

Một cửa sổ hay còn được gọi là nhân (kernel) hay mặt nạ (mark), là một ma trận kích thước nhỏ, ví dụ ma trận 3x3, 5x5, 7x7, v.v. Khi thực hiện các phép biến đổi trên từng điểm ảnh của ảnh cần xử lý, đặt điểm neo (anchor) của toán tử cửa sổ (thường được chọn là phần tử chính giữa của ma trận) vào vị trí của điểm ảnh đó và thực hiện các phép tính toán đối với điểm ảnh thuộc tâm và các điểm ảnh khác nằm trong phạm vi của toán tử cửa sổ để làm thay đổi giá trị của điểm ảnh đang xử lý. Tùy vào mục đích, kích thước cửa sổ và giá trị các phần tử thuộc cửa sổ cũng như các phép tính toán là khác nhau.

b) Phép toán nhân chập (Convolution)

Nhân chập là phép toán quan trọng trong xử lý ảnh, được xử dụng trong kỹ thuật tạo ảnh tích phân, làm trơn ảnh (smooth), tách cạnh, lọc tuyến tính, v.v..

Nhân chập là phép cửa sổ trượt, được định nghĩa bởi ma trận hạt nhân (kernel) K kích thước m x n. Ảnh I nhân chập với nhân K bởi công thức:

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

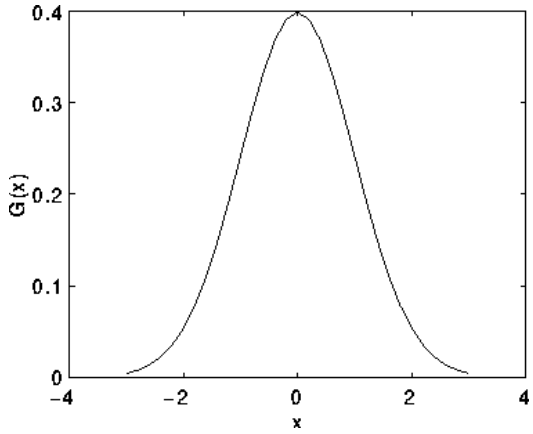
c) Lọc Gaussian

Lọc Gaussian có tác dụng làm trơn ảnh dựa trên ý tưởng rằng giá trị điểm ảnh (giá trị hàm hai chiều *I(x,y)*) là một biến xác xuất hai chiều tuân theo phân bố Gaussian. Điều này đồng nghĩa với việc trong một lân cận nào đó, giá trị điểm ảnh không có sự chênh lệch bất thường.

Phân bố Gaussian hay còn được gọi là phân bố chuẩn, có hàm phân bố xác xuất đối với biến ngẫu nhiên một chiều sau:

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

Trong đó, σ là độ lệch chuẩn của phân bố. Đồ thị của $G(x)$ có hình quả chuông như minh họa ở Hình 2.8.

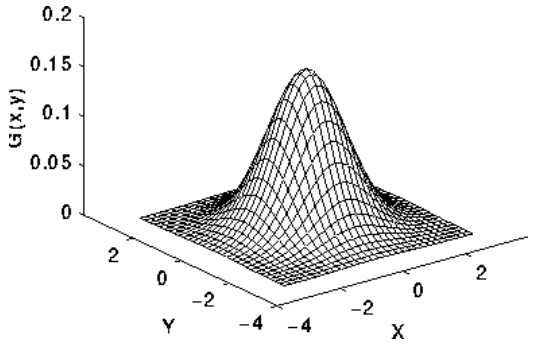


Hình 2.8. Minh họa phân bố Gaussian hàm một chiều

Đối với biến ngẫu nhiên hai chiều, hàm phân bố xác xuất trở thành:

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

Biểu đồ của $G(x,y)$ được minh họa trong Hình 2.9



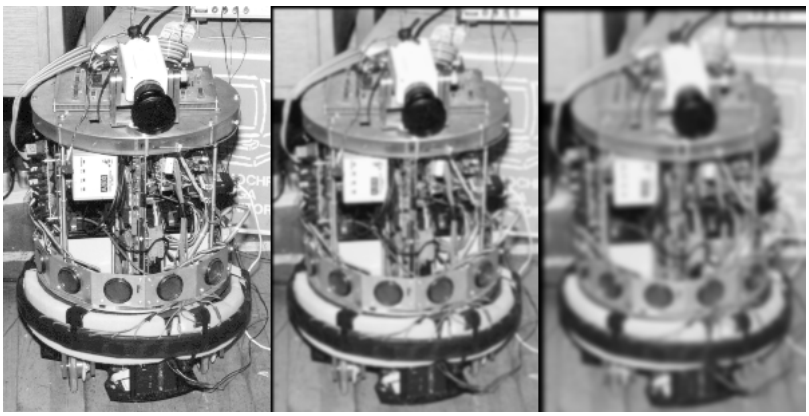
Hình 2.9. Minh họa phân bố Gaussian hai chiều

Đối với ảnh số, vốn là tập hợp của các điểm ảnh có giá trị rời rạc, cần thiết phải xấp xỉ hàm Gaussian bởi một ma trận rời rạc để có thể thực hiện phép nhân chập. Theo lý thuyết, phân bố Gaussian là khác 0 tại mọi điểm, dẫn đến để xấp xỉ phân bố này cần một ma trận với kích thước không giới hạn. Nhưng trong thực tế có thể coi như phân bố là bằng 0 với mọi giá trị sai khác hơn 3 lần độ lệch chuẩn, nên chỉ cần một ma trận có kích thước xác định làm nhân (kernel) của phép nhân chập. Ví dụ Hình 2.10 sau minh họa một xấp xỉ của hàm Gaussian với độ lệch chuẩn $\sigma = 1$.

$$\frac{1}{273}\begin{bmatrix} 1 & 4 & 7 & 4 & 1 \\ 4 & 16 & 26 & 16 & 4 \\ 7 & 26 & 41 & 26 & 7 \\ 4 & 16 & 26 & 16 & 4 \\ 1 & 4 & 7 & 4 & 1 \end{bmatrix}$$

Hình 2.10. Xấp xỉ rời rạc cho hàm Gaussian với $\sigma = 1$

Sau khi xác định được ma trận xấp xỉ trên, phép lọc Gaussian có thể thực hiện bằng cách tiến hành nhân chập ma trận xấp xỉ với ảnh gốc. Kết quả là ảnh mới có tính trơn hơn ảnh ban đầu, như minh họa ở Hình 2.11.



Hình 2.11. Minh họa lọc Gaussian: Ảnh đầu tiên bên trái là ảnh gốc; ảnh thứ hai là kết quả lọc Gaussian với độ lệch chuẩn = 1; ảnh thứ 3 là kết quả lọc Gaussian độ lệch chuẩn = 2

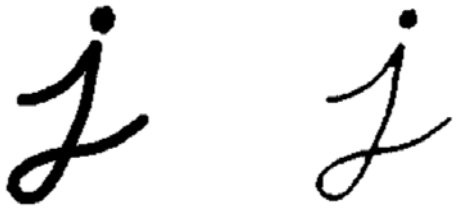
Các phép toán hình thái học

Phép toán hình thái học là tập hợp các phép toán xử lý ảnh dựa trên hình dạng. Phép toán hình thái áp dụng một yếu tố cấu trúc (structuring element) lên ảnh đầu vào để tạo ra ảnh kết quả. Các phép toán hình thái cơ bản nhất gồm có khép giãn nở và phép xói mòn. Chúng được ứng dụng rất rộng trong các vấn đề như:

- Khử nhiễu.
- Cô lập các thành phần riêng biệt hoặc kết nối các thành phần khác nhau lại trong ảnh.
- Tìm kiếm các điểm nhô lồi, lõm (về cường độ sáng) trên ảnh.

a) Phép giãn nở (Dilation)

Phép toán này thực hiện việc nhân chập ảnh *A* với nhân (kernel) *B,* là ma trận có hình dạng và kích thước nào đó, thường là hình vuông hoặc tròn. Ma trận *B* được xác định một điểm neo, thường là tâm của nhân. Tiến hành trượt *B* trên toàn bộ ảnh, tìm điểm ảnh có giá trị lớn nhất của ảnh *A* trong vùng cửa sổ *B* và thay thế giá trị điểm ảnh ở vị trí điểm neo bằng giá trị lớn nhất đó. Có thể thấy rằng, phép trượt này khiến cho vùng sáng thuộc ảnh được mở rộng. Ví dụ được minh họa ở Hình 2.12 dưới đây:

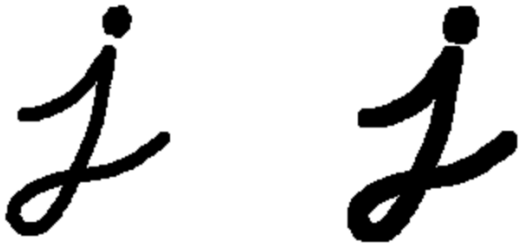


Hình 2.12. Phép giãn nở: Bên trái là ảnh ban đầu; bên phải là ảnh kết quả sau giãn nở

Phần nền (màu trắng) được mở rộng xung quanh ký hiệu.

b) Phép xói mòn (Erosion)

Phép xói mòn là phép toán ngược lại với phép giãn nở. Cách thức thực hiện hoàn toàn giống với phép giãn nở, nhưng thay vì thay thế giá trị điểm ảnh lớn nhất trong vùng cửa sổ cho điểm ảnh tại điểm neo, phép xói mòn sử dụng giá trị điểm ảnh nhỏ nhất. Ví dụ trong Hình 2.13 thể hiện kết quả của phép xói mòn.



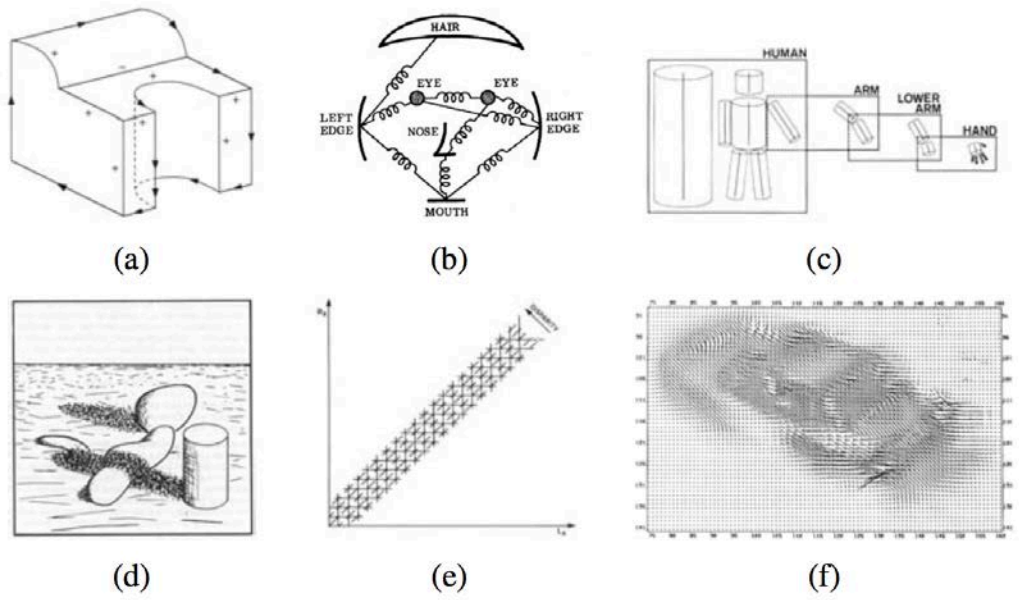
Hình 2.13. Phép xói mòn: Bên phải là ảnh ban đầu; bên phải là ảnh kết quả sau xói mòn

Vùng ảnh màu trắng trở lên nhỏ hơn, trong khi vùng màu tối lớn hơn sau phép xói mòn.

2.2. Tổng quan về thị giác máy tính

Bộ não con người dễ dàng nhận thức về các đối tượng tồn tại xung quanh trong không gian ba chiều nhờ vào tín hiệu thu được thông qua hệ thống thị giác. Con người có thể phân biệt, nhận thức về khoảng cách, hình dáng, chất liệu, số lượng và nhiều thuộc tính khác của vật thể, cũng như nhận diện một đối tượng từ các đối tượng khác cùng loại. Điều này đạt được nhờ vào quá trình tái tạo hình ảnh ba chiều và xác định các đặc trưng của đối tượng trong bộ não thông qua tín hiệu thu được từ mắt người. Các nhà tâm lý học nhận thức đã dành nhiều thập kỷ để tìm hiểu cách thức hoạt động của hệ thống thị giác và não người.

Từ những năm 70, những nghiên cứu đầu tiên về hệ thống thị giác máy đã xuất hiện. Thời điểm đó, giới khoa học coi thị giác máy là một thành phần nhận thức trực quan của kế hoạch đầy tham vọng nhằm cho máy móc bắt trước con người, thực hiện những hành vi thông minh, tự động đưa ra các quyết định chính xác và hợp lý. Bài toán được đặt ra là, khi kết nối camera với máy tính, làm thế nào để máy tính có thể mô tả những gì nó nhìn thấy? Ngày nay, chúng ta biết rằng vấn đề đó là không hề đơn giản.



Hình 2.14. Một số ví dụ về các thuật toán thị giác máy xuất hiện sớm nhất: a) gán nhãn cho đường trong hình vẽ kỹ thuật; b) kỹ thuật cấu trúc ảnh, hay mô hình biến dạng cho nhận diện đối tượng; c) mô hình khớp nối cho bài toán ước lượng tư thế người; d) ảnh nội tại; e) stereo correspondence – xác định các điểm tương ứng trong hai bức ảnh khác nhau; f) kỹ thuật optical flow, biểu diễn chuyển động tương đối của điểm trên bề mặt vật thể

Thị giác máy có liên quan chặt chẽ và có nhiều phần giao thoa với các lĩnh vực như xử lý ảnh số, trí tuệ nhân tạo và toán học. Vậy thị giác máy là lĩnh vực liên ngành nhằm giúp cho máy tính có thể hiểu các thông tin mức cao (thông tin ngữ nghĩa) trong ảnh và video số. Từ góc độ khoa học ứng dụng, nó được coi là nghiên cứu để xây dựng lên các tác vụ tự động thực hiện các công việc như hệ thống thị giác con người có thể làm.

Hiểu ảnh trong ngữ cảnh thị giác máy là việc biến đổi ảnh thành các mô tả của thế giới thực. Các mô tả này phải phù hợp để làm đầu vào cho các tiến trình tiếp theo, hoặc giúp gợi ý các hành động thích hợp. Hiểu ảnh được xem như là sự giải thích cho các thông tin mang tính biểu tượng từ dữ liệu hình ảnh, được xây dựng với sự trợ giúp của hình học, vật lý, xác suất thống kê và học máy.

Nếu việc chụp ảnh là tiến trình biến đổi khung cảnh 3 chiều của thế giới thực thành hình ảnh hai chiều thì các kỹ thuật thuộc thị giác máy lại cố gắng đảo ngược tiến trình ấy, tái tạo các thuộc tính như hình dạng, sự chiếu sáng, phân bố màu sắc của vật thể trong khung cảnh 3 chiều từ hình ảnh 2 chiều trên bề mặt ảnh.

Thị giác máy hiện nay bao gồm nhiều nhánh như: tái tạo khung cảnh; phát hiện sự kiện; theo vết; nhận diện đối tượng; ước lượng chuyển động; khôi phục hình ảnh, v.v.. Đối tượng nghiên cứu của thị giác máy rất rộng. Một đặc điểm chung của các bài toán thuộc lĩnh vực này là có rất nhiều cách giải quyết theo nhiều cách tiếp cận khác nhau. Mỗi giải pháp đạt được một kết quả nhất định cho một trường hợp cụ thể. Rất khó để so sánh giải pháp nào là tốt hơn.

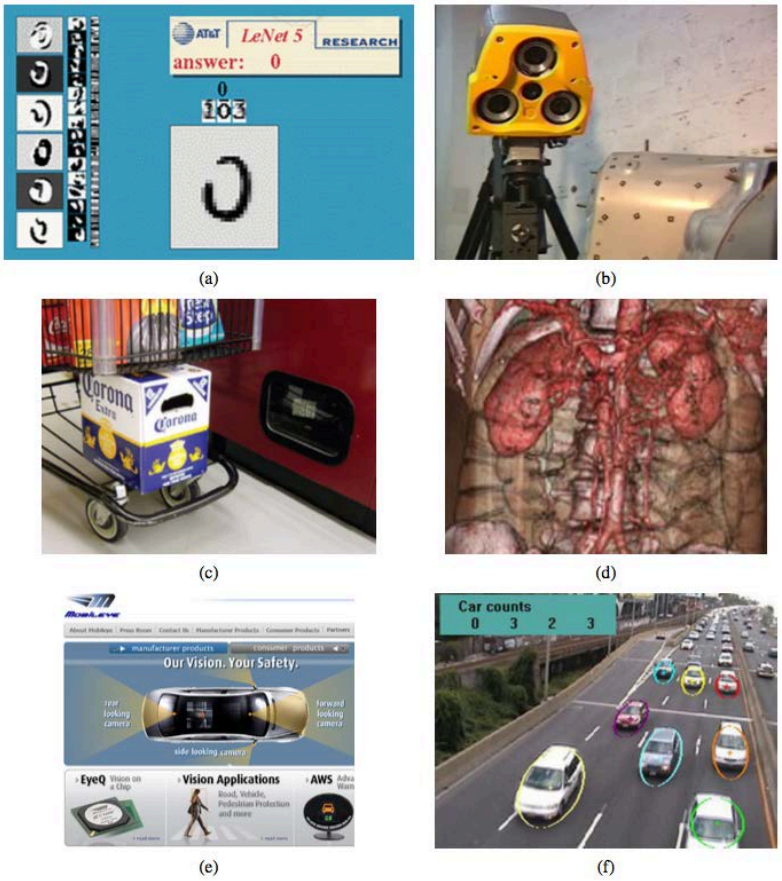
Gần đây, với việc năng lực tính toán ngày càng tăng cao của máy tính và sự phát triển mạnh mẽ của lĩnh vực trí tuệ nhân tạo mà cụ thể là học máy, ngành thị giác máy đang phát triển nhanh chóng chưa từng có. Đến giờ, chúng ta đã có những kỹ thuật đáng tin cậy cho việc tính toán mô hình không gian 3 chiều của một môi trường từ hàng trăm bức ảnh trùng nhau một phần; Với một tập đủ lớn các góc nhìn của vật thể, có thể tạo ra bề mặt 3D chính xác của đối tượng dựa trên việc so khớp (matching); Chúng ta có thể theo viết chuyển động của một người trong bối cảnh phức tạp; Thậm chí, có thể cố gắng xác định tên của tất cả mọi người xuất hiện trong một bức ảnh với độ chính xác vừa phải, sử dụng kết hợp các kỹ thuật phát hiện và nhận diện khuôn mặt, quần áo, tóc, v.v..

Ngày càng có nhiều ứng dụng thực tế áp dụng các kỹ thuật của thị giác máy, bao gồm:

- **Nhận dạng ký tự (ORC)**: Ví dụ nhận diện ký tự viết tay như mã bưu chính trên phong bì thư, hay tự động nhận diện biển số xe.
- **Kiểm tra tự động bằng máy (Machine inspection)**: Ví dụ tự động ước lượng dung sai kích thước cánh máy bay hay xác định các khiếm khuyết của vật liệu trong quản lý chất lượng.
- **Xây dựng mô hình 3D**: Tự động tạo mô hình 3D từ nhiều ảnh chụp từ trên không (viễn thám) như Bing Maps đã làm.
- **Áp dụng trong y tế**: Ví dụ tìm hiểu quá trình thay đổi hình thái bộ não người theo tuổi.
- **Giám sát:** Phát hiện xâm nhập trái phép; phân tích luồng giao thông; quan sát hồ nước để phát hiện nạn nhân rơi xuống hồ.

- **Nhận dạng vân tay và sinh trắc học:** Cho bài toán tự động xác thực quyền truy cập nội dung cũng như các ứng dụng trợ giúp cho tòa án trong xác định danh tính tội phạm.
- **Phát hiện khuôn mặt:** Cho việc cải thiện khả năng tự động lấy nét của camera, cũng như cho bài toán tìm kiếm ảnh.

Và còn rất nhiều ứng dụng khác nữa.



Hình 2.15. Một số ứng dụng trong công nghiệp của thị giác máy: a) Nhận dạng chữ viết tay; b) Kiểm tra cơ khí; c) Ứng dụng trong hệ thống bán lẻ hàng hoá; d) Ảnh y tế; e) Thị giác máy trong hệ thống lái xe tự động; f) Hệ thống camera giám giao thông.

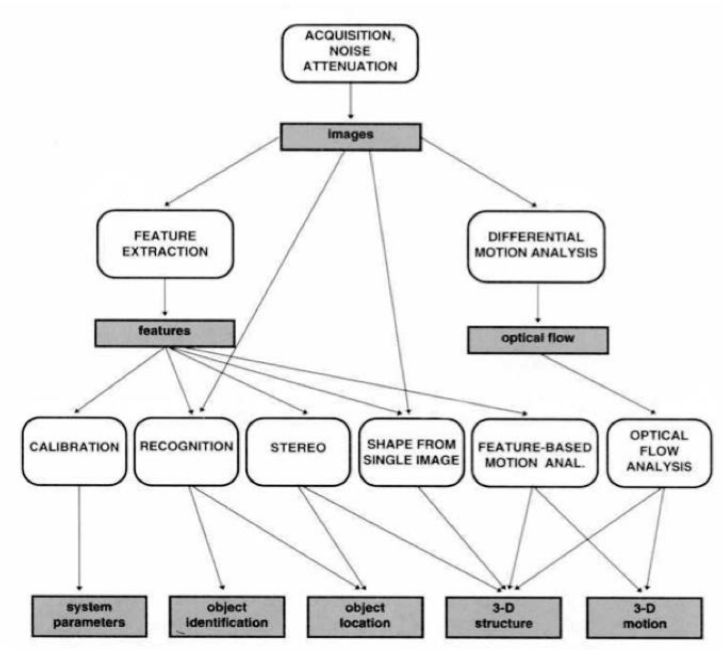
2.2.1. Hệ thống các kỹ thuật thị giác máy

Hình ảnh thu được qua bước thu nhận, khử nhiễu nhờ các kỹ thuật thuộc xử lý ảnh sẽ là đầu vào cho các bước tiếp theo của hệ thống thị giác máy. Các kỹ thuật thị giác máy có thể chia thành các mức sau:

- Thị giác mức thấp (low-level): Là các kỹ thuật trích rút đặc trưng mức thấp của ảnh như đặc trưng cạnh (edge), góc (corner), hay luồng quang học (optical flow).
- Thị giác mức trung (middle-level): Các kỹ thuật như nhận diện đối tượng (object recognition), phân tích chuyển động hay tái tạo mô hình 3D (3D reconstruction), v.v.. sử dụng các đặc trưng có được từ các kỹ thuật mức thấp.

- Thị giác mức cao (high-level): Giải thích các thông tin thu được qua các kỹ thuật thị giác mức trung và mức thấp, bao gồm mô tả khái nệm của khung cảnh như hoạt động, ý định, hành vi, v.v..

Các kỹ thuật cùng với kết quả đầu ra được khái quát lại bởi Hình 2.16 dưới đây.



Hình 2.16. Hệ thống các kỹ thuật thị giác máy

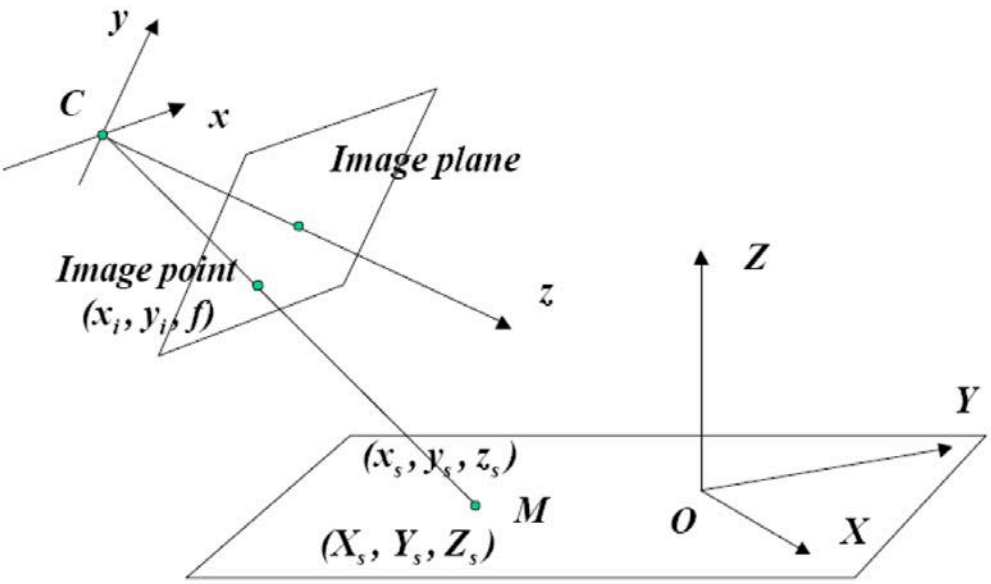
2.2.2. Các khái niệm quan trọng

1) Hình học (Geometry)

Hình học trong thị giác máy là công cụ giải quyết mối quan hệ hình học giữa hình ảnh 3D trong thế giới thực với hình chiếu của nó trên ảnh 2D. Các bài toán phổ biến đối với hình học trong thị giác máy bao gồm:

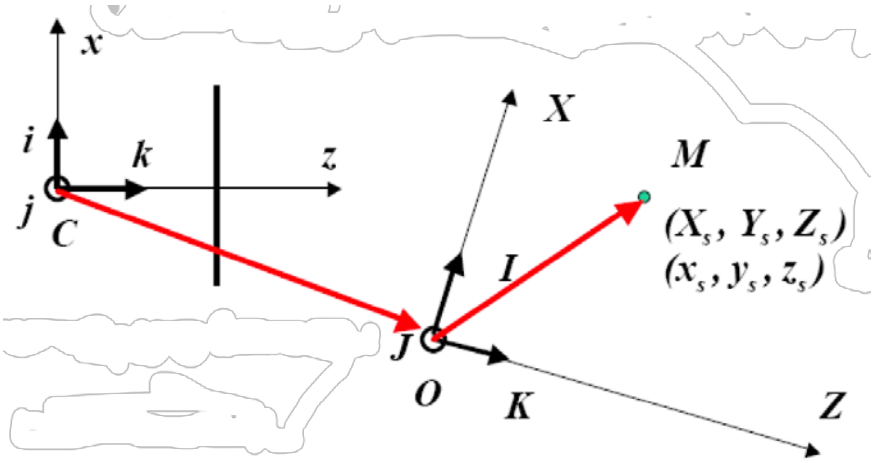
- Tái xây dựng cấu trúc hình học của đối tượng dựa trên việc đo lường, ước lượng hình chiếu của nó trên các ảnh hai chiều.
- Đối sánh các điểm giữa các ảnh chụp cùng một đối tượng ở nhiều góc nhìn khác nhau.
- Bài toán cân chỉnh camera tự động.

Để giải quyết các vấn đề nêu trên, cần thiết phải xác định được mối liên hệ giữa hệ toạ độ trong thế giới thực với hệ toạ độ trong ảnh, hay nói cách khác là việc tìm ra các tham số của hàm biến đổi toạ độ một điểm trong thế giới thực sang toạ độ của điểm đó trong hệ toạ độ của máy ảnh, như minh hoạ trong hình 2.17 dưới đây:



Hình 2.17. Hệ toạ độ trong thế giới thực và hệ toạ độ của camera

Trong đó, M là điểm đang xét, (X_s,Y_s,Z_s) là toạ độ của M trong hệ toạ độ OXYZ của thế giới thực, còn (x_s,y_s,z_s) là toạ độ của M đối với hệ toạ độ của camera Cxyz, C là tâm của camera. Ảnh của M trên mặt phẳng ảnh có toạ độ (x_s,y_s, f) với f là độ dài tiêu cự của ống kính. Việc chuyển hệ hoạ độ từ O về C được thực hiện thông qua phép dịch chuyển với vector dịch chuyển T = $\overrightarrow{CO}$; và phép xoay trục, sẽ được làm rõ sau đây.



Hình 2.18. Phép chuyển trục toạ độ

Theo mô tả ở hình 2.18 ta có: $\overrightarrow{CM} = \overrightarrow{CO} + \overrightarrow{OM}$.

Với *i, j, k* và *I, J, K* lần lượt là các vector đơn vị của hệ Cxyz và OXYZ, phương trình trên trở thành:

$$x_s i + y_s i + z_s i = T_s i + T_s i + T_s i + X_s i + Y_s i + Z_s i$$

Thực chất, x_s chính là hình chiếu của $\overrightarrow{OM}$ lên Cx sau phép dịch chuyển T. Vậy:

$$x_s = Tx + XsI.i + YsJ.i + ZsK.i$$

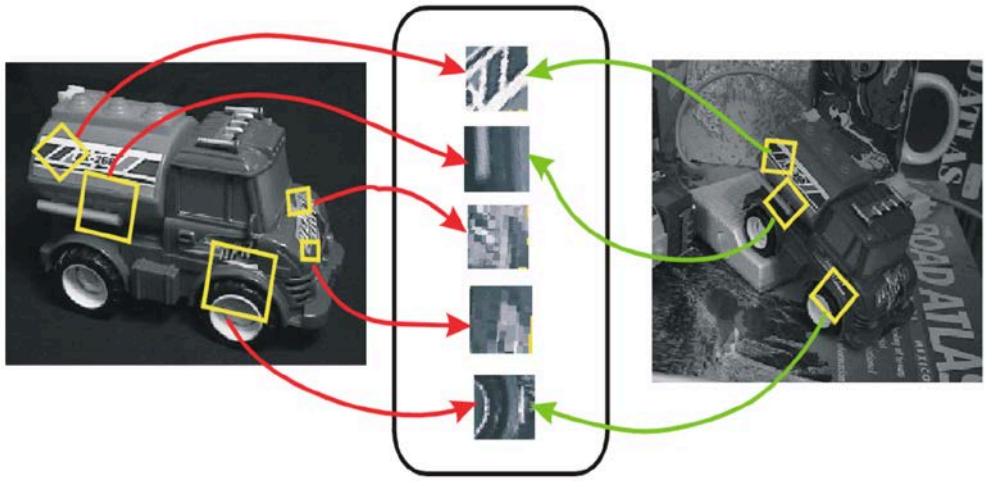
Tương tự đối với y_s và z_s ta có:

$$\begin{bmatrix} x_s \\ y_s \\ z_s \end{bmatrix} = \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix} + \begin{bmatrix} Ii & Ji & Ki \\ Ij & Jj & Kj \\ Ik & Jk & Kk \end{bmatrix} \begin{bmatrix} X_s \\ Y_s \\ Z_s \end{bmatrix}$$

Vậy, ma trận của phép xoay là: R = $\begin{bmatrix} Ii & Ji & Ki \\ Ij & Jj & Kj \\ Ik & Jk & Kk \end{bmatrix}$

2) Đối sánh (Matching)

Đối sánh là việc so khớp một vùng ảnh trong ảnh với vùng ảnh tương ứng trên một ảnh khác. Ví dụ như hình 2.19 sau:

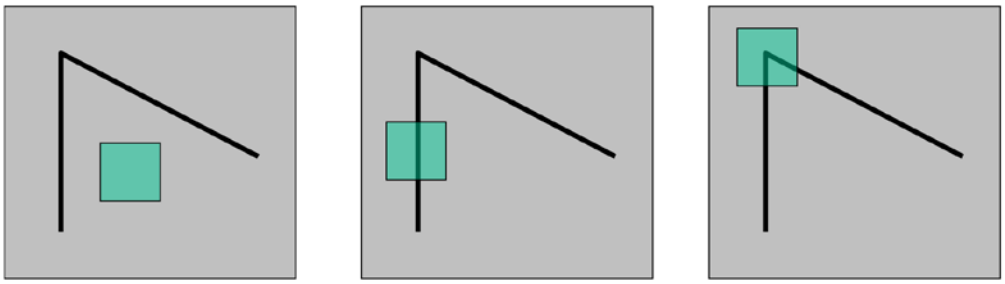


Hình 2.19. Đối sánh vùng ảnh giữa các ảnh

Việc đối sánh có thể là giữa các ảnh được chụp đồng thời của đối tượng dưới các khung nhìn khác nhau đối với bài toán stereo matching; hoặc trong hai khung hình liên tiếp trong chuỗi video trong các bài toán theo vết đối tượng; hoặc trong các bức ảnh khác nhau của các đối tượng giống nhau trong bài toán nhận dạng đối tượng.

Có thể nói, đối sánh là một trong những khái niệm quan trọng bậc nhất của thị giác máy với rất nhiều ứng dụng trong các bài toán khác nhau. Những vấn đề mà một kỹ thuật đối sánh gặp phải bao gồm việc thay đổi tỉ lệ, thay đổi góc nhìn (phép xoay), thay đổi điều kiện chiếu sáng, hay biến dạng gây ra bởi camera. Để giải quyết các vấn đề đó, cần phải xây dựng được mô hình biểu diễn một vùng ảnh (representation) bất biến với các phép dịch chuyển, phép xoay ảnh, phép thay đổi tỉ lệ, thay đổi cường độ sáng hay các phép biến đổi affine. Ý tưởng phổ biến là sử dụng các điểm hấp dẫn (interest points) trong ảnh và mối tương quan cục bộ giữa các điểm hấp dẫn đó trong vùng ảnh, từ đó xác định được các đặc trưng cục bộ của vùng ảnh mà bất biến với các phép biến đổi kể trên. Điển hình như các kỹ thuật SIFT, SUFT hay HOG. Trong đó, một cách trực quan, điểm hấp dẫn là những điểm trên ảnh mà người xem có thể dễ dàng ghi nhớ nó và xác định lại

được sau khi ảnh bị xoay đi, thay đổi tỉ lệ hay áp dụng các phép biến đổi làm biến dạng ảnh. Những điểm này là các cực trị trên ảnh.



Hình 2.20. Điểm hấp dẫn trong ảnh

Như hình 2.20 mô tả, dễ thấy rằng điểm quan tâm trong ảnh thứ 3 từ trái sang – nằm trên góc tạo bởi hai đường thẳng là ứng viên tiềm năng nhất để xem xét là điểm hấp dẫn, trong khi điểm nằm trên cạnh ở hình thứ hai hoặc trên vùng đồng nhất ở ảnh đầu tiên là không phù hợp.

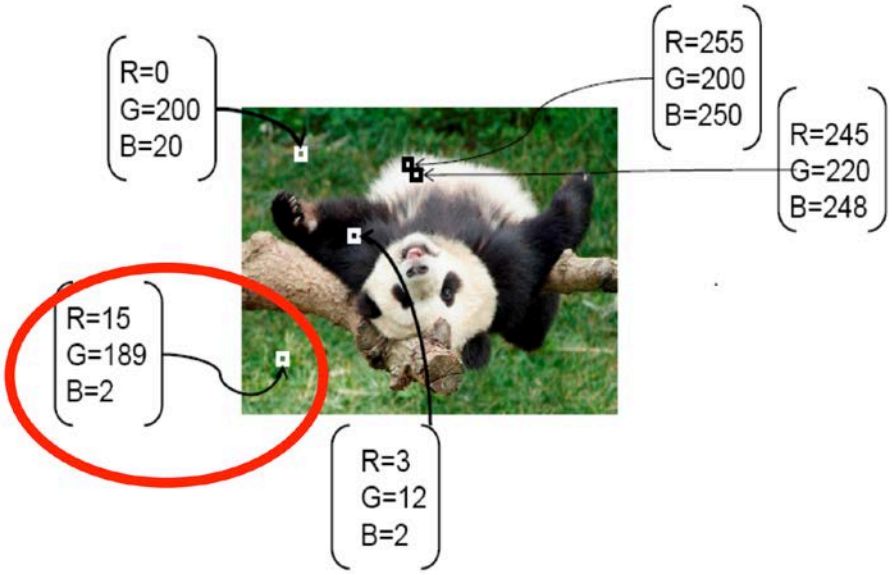
3) Điều chỉnh ảnh (Image alignment)

Điều chỉnh ảnh là việc thực hiện các phép biến đổi trên một tập các dữ liệu ảnh khác nhau để đưa chúng về một hệ toạ độ duy nhất. Tập dữ liệu ảnh này có thể là các ảnh được chụp bởi nhiều máy ảnh khác nhau, thời gian, góc nhìn, khoảng cách đến vật thể khác nhau.

4) Phân cụm (Clustering)

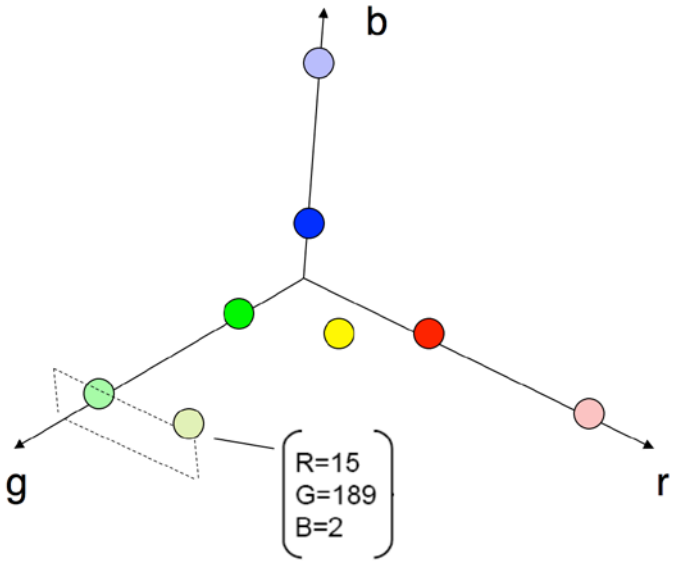
Phân cụm là việc nhóm các đối tượng vào những các nhóm có nhiều ý nghĩa, cung cấp một phương pháp biểu diễn N đối tượng vào k cụm khác nhau dựa trên một độ đo sự tương đồng giữa các đối tượng. Phân cụm rất quan trọng đối với thị giác máy trong các bài toán phân tách vùng ảnh thành các vùng (regions) có nhiều ý nghĩa, hoặc có sự đồng nhất về mặt cảm nhận thị giác; trong các bài toán mô hình kết cấu (texture) của ảnh; ngoài ra còn hữu ích trong khai phá dữ liệu, nén, và phân lớp. Với hơn 1500 bài báo liên quan được công bố, phân cụm thực sự là chủ đề quan trọng và rất được quan tâm ngày nay.

Để thực hiện phân cụm, ảnh được chuyển sang biểu diễn trong không gian đặc trưng, trong đó dấu hiệu của mỗi đối tượng quan tâm được xác định bởi một tập các đặc điểm phi trực quan. Ví dụ như vị trí, màu sắc, kết cấu, vector chuyển động, kích thước, hướng, v.v..



Hình 2.21. Ví dụ không gian đặc trưng của ảnh

Trong không gian đặc trưng, mỗi dấu hiệu được biểu diễn bằng một điểm như minh hoạ tại hình 2.22.

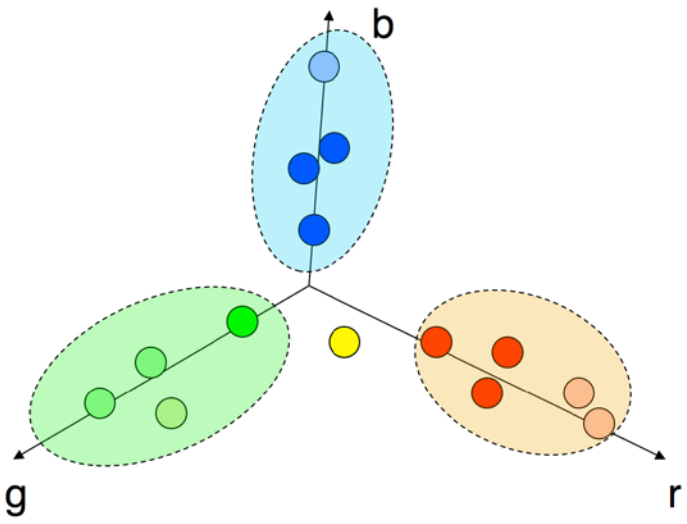


Hình 2.22. Biểu diễn dấu hiệu của đối tượng trong không gian đặc trưng

Sự tương đồng của các dấu hiệu có thể đo lường được bằng khoảng cách giữa các điểm (hay các vector đặc trưng) trong không gian đặc trưng:

$$D_{pq} = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2}$$

Trong đó, D_{pq} là khoảng cách giữa hai điểm p và q; n là số chiều của vector đặc trưng. Các điểm này được phân thành các cụm sao cho mỗi cụm có sự tương đồng cao.



Hình 2.23. Các điểm được phân cụm với sự tương đồng cao trong mỗi cụm

Các công cụ thường được sử dụng bao gồm kỹ thuật K-means, Mean-shift, và Spectral.

5) Phân loại/Nhận dạng (Categorization/Recognition)

Trong tất cả các nhiệm vụ mà một hệ thống thị giác máy cần thực hiện, nhận dạng là một nhiệm vụ còn tồn tại rất nhiều thách thức. Trong khi máy tính có thể thực hiện chính xác việc tái tạo khung cảnh 3D từ các ảnh thu được từ nhiều hướng nhìn khác nhau, nó vẫn không thể xác định được đối tượng, loài vật hiện diện trong ảnh dù chỉ là ở cấp độ của một đứa trẻ 2 tuổi.

Thế giới thực bao gồm hàng vạn đối tượng khác nhau, một số có các đặc điểm gần giống nhau, và mỗi đối tượng lại có nhiều dạng thù hình khác nhau. Trong cùng một lớp các đối tượng, lại có rất nhiều chủng loại. Ví dụ như có nhiều giống mèo, có các đặc điểm rất khác nhau nhưng chúng vẫn là mèo. Những điều đó khiến cho việc nhận dạng, phân loại đối tượng là rất khó khăn.

Bài toán nhận dạng có thể phân chia thành một số dạng. Ví dụ, nếu là tìm kiếm một đối tượng nhất định, thì bài toán gọi là phát hiện đối tượng (object detection), liên quan đến việc tìm kiếm trên toàn bộ ảnh để xác định vị trí của đối tượng trên ảnh nếu có. Nếu là một đối tượng cụ thể trong tập các đối tượng cùng loại, bài toán có thể trở thành việc nhận dạng, gọi tên đối tượng đó bằng cách tìm kiếm các điểm đặc trưng riêng biệt.

Thách thức lớn nhất đối với nhận dạng là bài toán phân loại tổng quát, hay là phân lớp đối tượng (classifying), liên quan đến việc phân một đối tượng vào một trong số hàng vạn lớp khác nhau trong thế giới thực.

2.2.3. Phân tích nội dung video (video content analysis)

Phân tích video được xem như là một nhánh của thị giác máy. Phân tích video là việc phân tích nội dung video để hiểu được ngữ cảnh mà nó mô tả. Đây là một thành phần đặc biệt của một số lĩnh vực công nghệ như camera giám sát (surveillance), robot

(robotics), và đa phương tiện (multimedia). Các phương thức trong phân tích video được thúc đẩy bởi các nhu cầu phát triển các thuật toán máy tính có thể bắt trước được khả năng của hệ thống thị giác con người. Người nghiên cứu trong lĩnh vực này cần phải có rất nhiều kiến thức nền tảng khác nhau như xử lý ảnh/tín hiệu số; khoa học máy tính; lý thuyết hệ thống; xác xuất thống kê; và toán ứng dụng.

1) Các tác vụ cơ bản trong phân tích video

Các tác vụ trong phân tích video có thể phân loại theo: mức thấp; mức trung; mức cao, mặc dù ranh giới giữa các nhóm là tương đối mơ hồ. Ví dụ, một số nhà nghiên cứu phân chia tách cạnh, phân đoạn vào nhóm mức thấp; tái tạo 3D là mức trung; trong khi tác vụ nhận diện (recognition) được xếp vào nhóm mức cao.

Ở mức xử lý một ảnh đơn trong chuỗi video, có 2 phép toán cơ bản áp dụng cho phân tích video. Bao gồm *gradients* – làm nổi bật các phần có sự thay đổi nhanh về giá trị cường độ sáng của bức ảnh, và phân tích giá trị cường độ sáng của ảnh, có thể là giá trị mức xám hoặc giá trị màu. Tính toán gradients dẫn đến nhận diện được cạnh trong ảnh (edges), sau đó kết hợp chúng lại để nhận diện các đặc trưng cấu trúc như đường và góc. Các đặc trưng cấu trúc này có thể được sử dụng để xác định các đặc trưng mức cao, chẳng hạn như hình dáng đối tượng. Ước lượng gradient cũng là nền tảng cho một số kỹ thuật phân đoạn ảnh và phát hiện đối tượng.

Phát hiện đối tượng (object detection) cũng là tác vụ mức thấp cơ bản của phân tích ảnh, giúp xác định đối tượng cần quan tâm trong khung cảnh, ví dụ con người trong ảnh, giúp hiểu ngữ cảnh hoặc hành động của đối tượng. Có rất nhiều kỹ thuật phát hiện khác nhau cho các kiểu đối tượng khác nhau, trong đó phổ biến nhất là phát hiện người, phát hiện phương tiện. Các kỹ thuật này phân tích cường độ sáng và sự thay đổi trong ảnh, xây dựng mô hình thống kê để xác định dấu hiệu nhận biết đối tượng.

Video bao gồm một chuỗi các ảnh đơn có sự tương quan cao. Một nhiệm vụ khác của phân tích video là tính toán chuyển động của đối tượng trong video. Rất nhiều kỹ thuật được đề xuất cho mục đích này. Luồng quang học (Optical flow) là kỹ thuật ước lượng sự chuyển động ở mỗi điểm ảnh độc lập, kết hợp với phân đoạn (segmentation) có thể cho biết thông tin từng phần khung cảnh thay đổi ra sao theo thời gian. Theo vết (tracking) là việc tính toán vị trí của đối tượng qua mỗi khung hình.

Một trường hợp đặc biệt cần được quan tâm là môi trường multi-camera. Miền ứng dụng của môi trường này đặt ra những thách thức riêng. Hình ảnh của cùng một đối tượng thu được từ các camera khác nhau cần phải được ghi nhận sao cho các đặc trưng giống nhau được kết hợp lại để dùng cho việc phân tích tiếp theo. Tái xác định đối tượng mục tiêu qua một mạng các camera có góc nhìn không chồng lấp sau khi nó không xuất hiện ở bất kì một camera nào trong một khoảng thời gian là một thách thức quan trọng khác. Hơn nữa, môi trường multi-camera còn dẫn đến một vấn đề cần nghiên cứu khác là xử lý ảnh phân tán, trong đó việc hiểu ngữ cảnh cần đạt được bởi sự hoạt động của mỗi

camera như là một tác tử độc lập với các camera xung quanh, sau đó gửi dữ liệu về bộ xử lý trung tâm.

Dựa trên các tác vụ nêu, ứng dụng phân tích video có thể hiểu được ngữ cảnh ở mức cao (ngữ nghĩa). Khi một bức ảnh 2D biểu diễn hình ảnh 3D của thế giới thực, câu hỏi đặt ra là, có thể hay không việc tái tạo khung cảnh 3D từ những bức ảnh cho trước. Đó là bài toán ngược. Quá trình chụp ảnh liên quan đến việc phát triển một mô hình toán học giúp camera tham chiếu hình ảnh 3D vào ảnh 2D. Cho trước một mô hình và một tập các điểm tương ứng trên ảnh và thế giới thực, hoàn toàn có thể thực hiện việc hiệu chỉnh camera (camera calibration), qua đó làm cơ sở để tái tạo hình ảnh 3D. Có nhiều phương pháp tái tạo 3D từ một tập các ảnh, như sử dụng 2 camera ghi hình cùng một khung cảnh. Nếu hình ảnh thu được từ một camera đơn, sự chuyển động giữa các khung hình có thể giúp ước lượng chiều sâu 3D của khung cảnh (structure from motion).

Mục tiêu cuối cùng của phân tích video là hiểu được ngữ cảnh. Các tác vụ đã mô tả ở trên cung cấp các công cụ để thực hiện mục tiêu này. Hiểu ngữ cảnh đòi hỏi phải nhận diện được đối tượng và các sự kiện. Nhận diện đối tượng có thể đạt được ở mức phân tích ảnh đơn, trong khi nhận diện sự kiện và hành động thường yêu cầu xử lý trên một chuỗi ảnh.

2) Ứng dụng của phân tích video

Các ứng dụng phổ biến của phân tích video bao gồm:

Camera giám sát (surveillance): Sử dụng camera để đảm bảo an ninh cho một khu vực hoặc giám sát môi trường xung quanh. Một ứng dụng khác nữa là nhận diện người dựa trên trích rút thông tin sinh trắc học (vân tay, mống mắt).

Phương tiện truyền thông và Internet: Lập chỉ mục và tìm kiếm kho dữ liệu video.

Thông tin di động: Ứng dụng phân tích video có thể giúp cải thiện yêu cầu 41ang thông và điện năng tiêu thụ trên thiết bị di động trong quá trình chia sẻ dữ liệu video giữa thiết bị và máy chủ lưu trữ.

Thực tế ảo: Môi trường ảo có thể trở lên giống thực tế hơn nếu sử dụng thông tin thu thập được từ các video cảnh quan thiên nhiên để đưa vào quá trình dựng hình. Việc này đòi hỏi sự phân tích tự động nội dung video.

Robot hoạt động dựa trên thị giác: Robot được tích hợp camera, có thể làm việc độc lập hoặc bên cạnh con người, giúp dẫn đường qua các môi trường phức tạp, ví dụ như vùng xảy ra thảm hoạ. Ứng dụng dạng này đòi hỏi ứng dụng nhiều khía cạnh khác nhau của phân tích video như theo vết, nhận diện, xử lý ảnh phân tán.

Ứng dụn trong sinh học: Hỗ trợ chuẩn đoán trong y học hay tự động phân tích dữ liệu lớn trong các nghiên cứu sinh học.

2.2.4. Bài toán phát hiện hành động (action detection)

Phát hiện hành động trong chuỗi video là công việc không chỉ nhận diện loại hành động mà còn phải xác định nó xảy ra ở đâu (vị trí không gian trên ảnh) và khi nào (vị trí thời gian trên chuỗi video).

Bài toán phát hiện hành động đặt ra rất nhiều thách thức. Thứ nhất, các hành động của con người hết sức đa dạng như đi lại thông thường, chạy, nhẩy, làm việc nhà, múa, v.v.. mà trong số đó có rất nhiều hành động khó mà phân định rõ ràng với nhau. Các hành động, không những là tự thân mà còn có các hành động mang tính tương tác như ôm, bắt tay, hay các hành vi bạo lực tấn công người khác. Ngoài việc tương tác giữa người và người, còn có các hành động có sự tương tác giữa người với đối tượng bất kỳ, như lau nhà, bê đồ, v.v.. Thứ hai, trong chuỗi video rất dễ sảy ra việc che khuất một phần cơ thể, gây khó khăn trong việc xác định tư thế, hình dạng. Mặt khác, trang phục cũng có sự khác biệt, đa dạng. Với trang phục là quần âu, hình dáng và sự chuyển động sẽ rất khác so với mặc váy rộng. Thứ ba, rất khó để phát triển một giải thuật phù hợp cho đa số các trường hợp phạm vi bài toán phát hiện hành động do sự khác nhau về môi trường như điều kiện chiếu sáng, sự đổ bóng, sự phản xạ ánh sáng và kết cấu của hậu cảnh. Thứ tư, ngoài những thách thức thuộc vấn đề thị giác máy, còn có các thách thức mang tính ngữ cảnh và tính triết học khi con người có thể thực hiện một hành động bằng cách không hành động gì cả.

Phát hiện hành động từ dữ liệu video – mà trọng tâm là hành động của con người, là công việc quan trọng trong rất nhiều các ứng dụng như video giám sát, phân tích nội dung video thông minh. Nhằm làm giảm bớt khó khăn của bài toán này, các hướng tiếp cận đa phần dựa trên giả định rằng vị trí không gian sảy ra hành động là biết trước, không có hoặc có rất ít sự thay đổi về tỉ lệ và góc nhìn cùng với hậu cảnh tĩnh và rõ ràng. Do đó, thông tin hình dạng cơ thể người có thể được trích rút một cách đáng tin cậy.

Chìa khoá cho bài toán phát hiện hành động là việc biểu diễn một hành động theo cách nào. Các đề xuất sử dụng đặc điểm tư thế và sự chuyển động tại các khớp của cơ thể cho việc biểu diễn hành động là rất hữu dụng, tuy nhiên, có rất nhiều khó khăn khi thực hiện.

Có hai dạng bài toán phát hiện hành động, bao gồm ngoại tuyến (offline) và trực tuyến (online), cho các miền ứng dụng khác nhau. Hai dạng bài toán này đặt ra các yêu cầu và các hướng tiếp cận khác nhau.

1) Phát hiện hành động ngoại tuyến

Trong phát hiện ngoại tuyến, mục tiêu là tìm ra khung hình đầu tiên và cuối cùng mà trong đó hành động xảy ra trong chuỗi video. Mọi thông tin của chuỗi video là có sẵn ngay từ lúc đầu (video đã được ghi lại đầy đủ, không phải ghi hình trực tiếp thời gian thực). Thời gian tính toán hay thậm chí chi phí tính toán ở một mức độ nào đó không là vấn đề cần quan tâm. Miền ứng dụng của phát hiện ngoại tuyến chủ yếu là trong phân

tích nội dung video, gán nhãn, tạo chú thích, xây dựng cơ sở dữ liệu cho truy vấn video dựa trên nội dung, hay tóm tắt nội dung video (video summarization).

2) Phát hiện hành động trực tuyến

Khác với ngoại tuyến, chuỗi video đầu vào thường đang được ghi hình trực tiếp từ môi trường và theo thời gian thực. Mục tiêu là phát hiện, dự đoán một hành động ngay khi nó đang xảy ra, trước khi hành động hoàn toàn kết thúc. Nó được ứng dụng một cách hữu ích trong các ứng dụng thực tế như robot tự hành, camera giám sát có báo động, camera thông minh tự động phóng to và tập trung vào một số sự kiện chọn trước khi nó xảy ra, hay ứng dụng cho ôtô tự hành.

Với một luồng video (stream) đầu vào, hệ thống cần phải xuất đầu ra nhanh chóng, hay lý tưởng là trong thời gian thực tại thời điểm hành động diễn ra. Hệ thống không chỉ thực hiện việc phát hiện hành động mà còn phải đảm bảo yêu cầu phân biệt được hành động cần quan tâm từ rất nhiều dữ liệu hỗn hợp khác như hậu cảnh, các hoạt động không mang ý nghĩa, và các hành động không có liên quan.

Nói chung, khó khăn của hệ thống phát hiện hành động trực tuyến thời gian thực bao gồm, thứ nhất, hành động cần phải được phát hiện sớm nhất có thể, trong trường hợp lý tưởng là ngay khi một phần của hành động quan sát được. Thứ hai, hành động cần được phát hiện từ rất nhiều các hành động không liên quan khác. Thứ ba, công việc này làm việc với dữ liệu thực, không phải dữ liệu nhân tạo. Có rất nhiều các tình huống xảy ra nằm ngoài dự liệu. Cuối cùng, yêu cầu tính toán thời gian thực đặt ra nhiều thách thức đối với hiệu năng các kỹ thuật được áp dụng.

CHƯƠNG 3.
PHƯƠNG THỨC ĐỀ XUẤT

3.1. Tổng quan

Chương đầu tiên, tổng quan về bài toán phát hiện ngã đã tóm lược lại các hướng tiếp cận, phương pháp, kỹ thuật giải quyết bài toán tự động phát hiện ngã với các ưu điểm, nhược điểm và xu hướng nghiên cứu của từng hướng. Đối với sử dụng camera đơn, mặc dù có hạn chế là làm mất mát thông tin về không gian nhưng bù lại, việc cài đặt là đơn giản và chi phí thấp. Luận văn này tập trung vào bài toán phát hiện ngã sử dụng một camera giám sát.

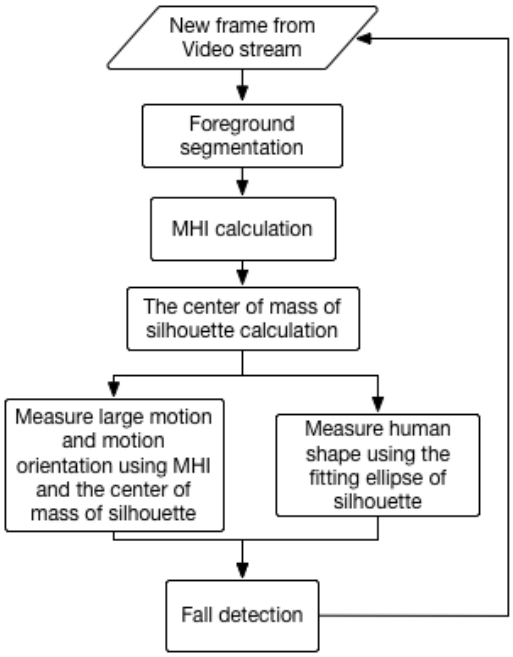
Phương thức phát hiện ngã đề xuất trong luận văn này dựa trên quá trình quan sát kỹ lưỡng các đặc điểm, các tình huống khác nhau của hành động ngã. Một hành động ngã, theo quan sát, thường có những đặc điểm sau:

- Xuất hiện chuyển động nhanh bất thường: Đây là một trong những dấu hiệu quan trọng để nhận biết việc ngã. Khi ngã, luôn luôn xuất hiện chuyển động nhanh bất thường ở một thời điểm nào đó của quá trình ngã. Qua quan sát, chuyển động trong giai đoạn đầu tiên của ngã có thể theo phương ngang hoặc hướng xuống, nhưng chắc chắn không có chuyển động hướng lên. Trong các giai đoạn tiếp theo, chuyển động có thể chậm lại do sự kháng cự của người ngã.
- Chuyển động hướng xuống theo phương dọc: Chuyển động của cơ thể khi ngã trong thời điểm cuối là hướng xuống do tác dụng của trọng lực. Tuy nhiên một số hành động thông thường khác cũng có đặc điểm này như chủ động ngồi và nằm.
- Thay đổi tư thế, hình dạng cơ thể: Với các hoạt động thông thường, hình dạng người thay đổi rất chậm, trong một khoảng thời gian ngắn có thể xem như không đổi. Nhưng khi ngã, hình dạng người thay đổi rất nhanh chóng, thậm chí là ngay lập tức.
- Thiếu vắng chuyển động sau khi ngã: Thông thường, sau khi ngã người ngã nằm bất động trên sàn. Cũng có khi cơ thể chuyển động nhanh, lăn đi lăn lại vì đau. Nhưng tình huống này có thể coi như không xuất hiện đối với người già. Mặt khác, có thể tự di chuyển được sau khi ngã nghĩa là người ngã không rơi vào tình huống nguy hiểm.

Qua những đặc điểm của hành động ngã đã kết luận được từ thực nghiệm, luận văn đề xuất một phương thức phát hiện ngã dựa vào việc phân tích thông tin về chuyển động và sự thay đổi hình dáng cơ thể như sau: Áp dụng kỹ thuật Motion History Image (MHI) cho trích rút các đặc trưng chuyển động; Và phân tích hình dáng cơ thể bằng kỹ thuật Ellipse Fitting, đo lường tỉ lệ giữa chiều cao và bề ngang, góc nghiêng so với mặt phẳng nằm ngang của cơ thể. Luận văn cũng đề xuất sử dụng kỹ thuật Moment ảnh để xác định trọng tâm người, sử dụng tốc độ di chuyển của trọng tâm hỗ trợ cho việc trích rút đặc

trưng chuyển động. Mô tả chi tiết các kỹ thuật sẽ được trình bày trong các phần tiếp theo.

Luồng hoạt động của phương thức đề xuất được mô tả như Hình 3.1 dưới đây:



Hình 3.1. Luồng hoạt động của hệ thống phát hiện ngã được đề xuất

Trong đó, khối Foreground Segmentation là khối phân đoạn vùng chuyển động sử dụng kỹ thuật trừ nền, sẽ được trình bày ở mục 3.2; Khối MHI calculation là khối xây dựng ảnh lịch sử chuyển động dựa trên kỹ thuật MHI, thuộc nội dung mục 3.3; Khối xác định tâm khối lượng cơ thể sử dụng kỹ thuật Image Moments, khối trích rút đặc trưng chuyển động cũng sẽ được giới thiệu chi tiết tại mục 3.3; Đo lường các đặc trưng hình dáng sử dụng kỹ thuật Ellipse fitting là nội dung mục 3.4; Cuối cùng, khối phát hiện ngã (fall detection) sử dụng các đặc trưng trích rút được để xác định việc ngã có xảy ra hay không, sẽ được trình bày tại mục 3.5 của chương này.

3.2. Phân tách vùng chuyển động

Để thu được các thông tin về hình dạng và sự chuyển động của cơ thể phục vụ cho việc phát hiện ngã, trước tiên cần phải phân tách được vùng ảnh tương ứng với cơ thể ra khỏi nền. Vùng ảnh này gọi là tiền cảnh (foreground). Từ hình thù của foreground và sự thay đổi vị trí của foreground trên khung hình có thể đo lường được mức độ chuyển động và thay đổi hình dáng của người cần giám sát.

Camera giám sát là loại camera được đặt cố định và liên tục ghi hình vùng môi trường trong phạm vi góc nhìn của nó. Điều này dẫn đến nếu không xuất hiện chuyển động trong vùng nhìn, các khung hình trong chuỗi video thu được là không đổi. Khi có chuyển động chỉ những điểm ảnh mà có sự chuyển động tại vị trí của nó thay đổi giá trị, các điểm ảnh còn lại không thay đổi giữa các khung hình. Tập các điểm ảnh bị thay đổi giá

trị do xuất hiện chuyển động được gọi là tiền cảnh (foreground), phần còn lại không đổi giá trị được gọi là hậu cảnh (background). Tách vùng chuyển động là công việc xác định tiền cảnh, đạt được thông qua các kỹ thuật trừ nền (background subtraction).

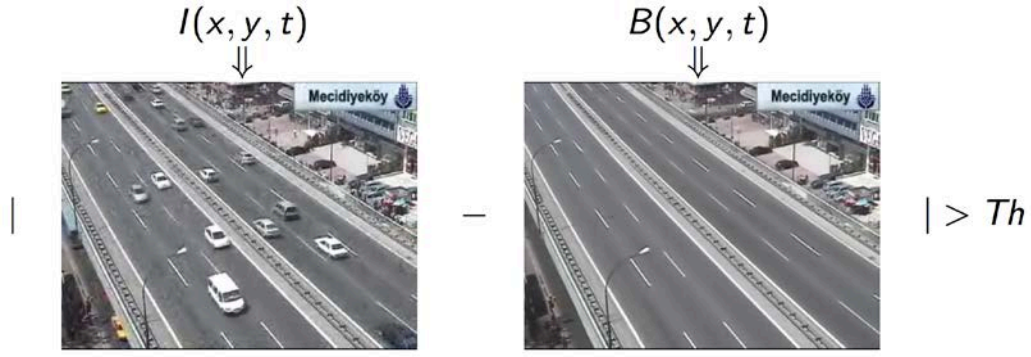
Trừ nền là phương thức so sánh khung hình hiện tại với mô hình nền, qua đó xác định được đối tượng chuyển động. Kỹ thuật này ngày càng được áp dụng rộng rãi cho việc phát hiện đối tượng chuyển động, nhất là trong ứng dụng giám sát bằng camera. Lý do chính để sử dụng kỹ thuật này là vì nó đơn giản, chính xác và chi phí tính toán thấp. Tuy nhiên các phương thức trừ nền đối mặt với một số thách thức trong trường hợp như môi trường thay đổi (độ chiếu sáng, nguồn sáng) hay sự di chuyển của bóng của vật thể, sự đồng nhất về màu sắc của đối tượng chuyển động.

3.2.1. Một số thuật toán trừ nền

Đa phần các kỹ thuật trừ nền có trung một ý tưởng sau: Giả sử rằng chuỗi video được quan sát *I* bao gồm một nền tĩnh *B*, và đối tượng di chuyển bên trên nền tĩnh này. Và giả thiết là mọi đối tượng di chuyển có màu sắc (hoặc phân bố màu sắc) khác biệt với nền *B*. Các phương thức trừ nền có thể tóm lược lại bởi công thức:

$$X_t(S) = \begin{cases} 1 & nếu\ d(I_{s,t}, B_s) > \tau \\ 0 & trong\ trường\ hợp\ khác \end{cases}$$

Trong đó, τ là ngưỡng, X_t là thể hiện (nhãn) của vùng chuyển động (motion mask), *d* là khoảng cách giữa $I_{s,t}$ – màu của điểm ảnh s tại thời điểm *t*, với B_s – mô hình nền tại vị trí điểm ảnh s. Sự khác biệt giữa các phương thức trừ nền nằm ở cách thức xây dựng mô hình nền *B* và khoảng cách *d* được sử dụng. Một số kỹ thuật mô hình nền được trình bày trong phần tiếp theo.



Hình 3.2. Minh họa trừ nền: Ảnh bên trái là khung hình hiện tại; ảnh bên phải là mô hình nền; Th là ngưỡng

1) Mô hình nền sử dụng ảnh đơn

Cách thức mô hình nền đơn giản nhất là sử dụng một ảnh (xám hoặc màu) không bao gồm đối tượng chuyển động. Ảnh đơn này có thể được ghi lại tại thời điểm môi trường

không xuất hiện chuyển động, hoặc sử dụng bộ lọc trung vị. Để giải quyết vấn đề thay đổi độ sáng và nền, mô hình được cập nhật bởi công thức:

$$B_{s,t+1} = (1-\alpha)B_{s,t} + \alpha * I_{s,t}$$

Trong đó α được chọn trước, giá trị nằm trong khoảng (0,1).

Với mô hình nền đơn giản này, điểm ảnh thuộc đối tượng chuyển động được xác định thông qua phân ngưỡng bởi một trong các hàm khoảng cách sau:

$$d_0 = \left|I_{s,t} - B_{s,t}\right|;$$
$$d_1 = \left|I^R_{s,t} - B^R_{s,t}\right| + \left|I^G_{s,t} - B^G_{s,t}\right| + \left|I^B_{s,t} - B^B_{s,t}\right|;$$
$$d_2 = \left(I^R_{s,t} - B^R_{s,t}\right)^2 + \left(I^G_{s,t} - B^G_{s,t}\right)^2 + \left(I^B_{s,t} - B^B_{s,t}\right)^2;$$
$$d_3 = max\left\{\left|I^R_{s,t} - B^R_{s,t}\right|, \left|I^G_{s,t} - B^G_{s,t}\right|, \left|I^B_{s,t} - B^B_{s,t}\right|\right\}$$

Với R, G, B là giá trị điểm ảnh lần lượt của các kênh màu đỏ, xanh lá và xanh dương.

2) Mô hình MinMax

Trong mô hình này, mỗi điểm ảnh s thuộc nền có giá trị nhỏ nhất m_s, giá trị lớn nhất M_s và giá trị lớn nhất của sự sai khác của giá trị điểm ảnh D_s trong các khung hình liên tiếp của chuỗi huấn luyện. Phương pháp MinMax gán nhãn tất cả các điểm ảnh thoả mãn điều kiện sau đây là nền:

$$|M_s - I_{s,t}| < \tau d_\mu \quad hoặc \quad |m_s - I_{s,t}| < \tau d_\mu$$

Trong đó, τ là ngưỡng đặt trước còn d_μ là giá trị trung bình của giá trị sai khác tuyệt đối giữa các khung hình của điểm ảnh trên toàn bộ ảnh.

3) Mô hình nền Gaussian Mixture Model (GMM)

Trong môi trường có xuất hiện các chuyển động nhỏ không có ý nghĩa cho bài toán phát hiện chuyển động như gợn sóng trên mặt hồ, lá cây đung đưa, hoặc sự thay đổi độ chiếu sáng khi giám sát trong thời gian dài, kỹ thuật mô hình nền phải loại bỏ được các chuyển động vô nghĩa đó và tự động thích nghi với các thay đổi của môi trường. GMM là một trong số các kỹ thuật mô hình nền nhằm mục đích ấy.

Stauffer và Grimson [15] mô hình mọi điểm ảnh riêng biệt bởi sự hòa trộn của K phân bố Gaussian (phân bố chuẩn). Phân bố Gaussian có hàm mật độ xác suất như sau:

- Một chiều: $\mathcal{N}(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- Nhiều chiều: $\mathcal{N}(X|\mu,\sum) = \frac{1}{(2\pi)^{N/2}} \frac{1}{|\sum|^{1/2}} e^{-\frac{1}{2}(X-\mu)^T\sum^{-1}(X-\mu)}$

Trong đó, μ là kỳ vọng (hay trung vị), σ là độ lệch chuẩn và σ^2 là phương sai của phân bố chuẩn. Đối với phân bố nhiều chiều, $X = [X_1, X_2,.., X_N]^T$ là vector ngẫu nhiên, $\mu = [\mu_1,$

$\mu_2,.., \mu_N]^T$ *và* $\sum$ là ma trận hiệp phương sai (xác định dương, kích thước *N* x *N*). $|\sum|$ là định thức của $\sum$.

Tại thời điểm *t* bất kỳ, vector ngẫu nhiên $X = [X_1, X_2,.., X_N]^T$ là giá trị của một điểm ảnh riêng biệt trong *t* khung hình, với $\{X_1, X_2,.., X_t\} = \{I_{s,i} : 1 \leq i \leq t\}$.

Chuỗi giá trị liên tiếp nhau của điểm ảnh (vector ngẫu nhiên *X*) được mô hình bởi sự hòa trộn của K phân bố Gaussian:

$$P(X_t) = \sum_{i=1}^{k} \omega_{i,t} \mathcal{N}(\mu_{i,t}, \sum_{i,t})$$

Trong đó, $N(\mu_{i,t}, \sum_{i,t})$ là phân bố Gaussian thứ *i* và $\omega_{i,t}$ là trọng số.

Vì các mục đích tính toán, ma trận hiệp phương sai $\sum_{i,t}$ có thể giả thiết là ma trận đường chéo theo gợi ý của Stauffer và Grimson.

Nếu giá trị mới của điểm ảnh X_{t+1} có thể so khớp (hay thuộc vào) một trong các phân bố Gaussian (sai khác không quá 2,5 lần độ lệch chuẩn σ) thì $\mu_{i,t+1}$ và $\sigma^2_{i,t+1}$ của phân bố đó được cập nhật bởi công thức:

$$\mu_{i,t+1} = (1-\rho)\mu_{i,t} + \rho X_{t+1}, và$$
$$\sigma^2_{i,t+1} = (1-\rho)\sigma^2_{i,t} + \rho(X_{t+1} - \mu_{i,t+1})^2$$

Trong đó, $\rho = \alpha \mathcal{N}(X_{t+1}|\mu_{i,t}, \sigma^2_{i,t})$, với α là hệ số học.

Trọng số của tất cả các phân bố được điều chỉnh bởi:

$$\omega_{i,t+1} = (1-\alpha)\omega_{i,t} + \alpha(M_{i,t+1})$$

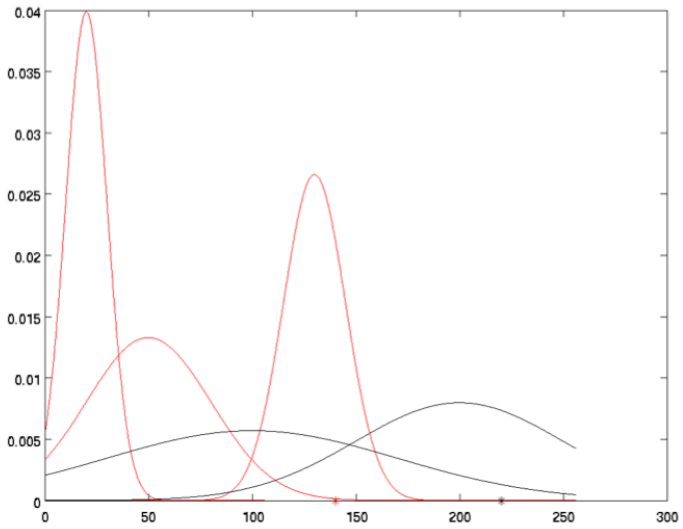
Trong đó $M_{i,t+1} = 1$ với phân bố đã được so khớp, bằng 0 đối với các phân bố còn lại.

Nếu X_{t+1} không thuộc vào bất kỳ một phân bố đã có nào, phân bố có xác suất ít xảy ra nhất được thay thế bởi một phân bố mới với $\mu_{t+1} = X_{t+1}$. Trong đó các phân bố được sắp xếp theo thứ tự về xác suất xảy ra dựa trên tỉ số ω/σ (tỉ lệ thuận với trọng số, tỉ lệ nghịch với độ lệch chuẩn).

Chọn *b* phân bố, Mô hình nền *B* là bài toán tìm cực tiểu:

$$B = argmin_b(\sum_{i=1}^{b} \omega_i > T)$$

Với *T* là ngưỡng tối thiểu. Hình 3.3 dưới đây minh họa cho các phân bố với giả thiết ảnh là ảnh đa mức xám (phân bố một chiều), với K = 5.



Hình 3.3. Minh họa mô hình nền: Sau khi ước lượng mô hình nền, phân bố màu đỏ trở thành mô hình nền, phân bố màu đen được xác định là foreground.

4) Mô hình KDE (Kernel Density Estimation)

Xác suất giá trị của một điểm ảnh thuộc nền được ước lượng như sau:

$$P(I_{s,t}) = \frac{1}{N}\sum_{i=t-N}^{t-1} K(I_{s,t} - I_{s,i})$$

Trong đó, K là nhân (kernel – thường là hàm Gaussian) và N là số khung hình liên tiếp dùng để ước lượng P(.). Đối với video màu:

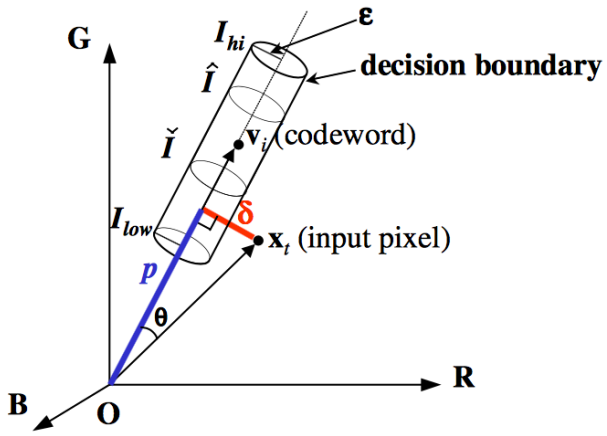
$$P(I_{s,t}) = \frac{1}{N}\sum_{i=t-N}^{t-1} \prod_{j=\{R,G,B\}} K(\frac{I_{s,t}^{j} - I_{s,i}^{j}}{\sigma_j})$$

Một điểm ảnh được xác định là tiền cảnh nếu giá trị của nó không thuộc phân bố trên, ví dụ khi $P(I_{s,t})$ nhỏ hơn một ngưỡng chọn trước. Giá trị σ_j là cố định hoặc được ước lượng tuỳ theo các kỹ thuật đã đề xuất.

5) Sử dụng bảng mã (Codebook)

Một tiếp cận khác nữa để đối phó với những hậu cảnh phức tạp, đề xuất bởi Kim và các cộng sự [34], được gọi là bảng mã (codebook). Hậu cảnh thực tế bao gồm tập các điểm ảnh có giá trị cố định và tập các điểm ảnh có sự chuyển động nhỏ như lá cây đung đưa, gợn sóng lăn tăn, v.v.. Các chuyển động thuộc nền thường là bán chu kỳ (giá trị màu của điểm ảnh có lặp lại trong một khoảng thời gian). Dựa trên chuỗi huấn luyện, mỗi điểm ảnh thuộc nền được gán với một chuỗi các giá trị màu (key values), được gọi là mã từ (code-words), mã từ được lưu trữ trong bảng mã. Các mã từ sẽ nhận giá trị màu cụ thể trong khoảng thời gian nhất định. Trong trường hợp một điểm ảnh thuộc một khu vực ổn định về giá trị màu, giá trị màu của điểm ảnh đó được tóm lược lại bởi duy nhất một mã từ. Còn nếu điểm ảnh thuộc vùng chịu ảnh hưởng bởi những biến động liên tục,

ví dụ như khu vực mà lá cây đung đưa vì gió, điểm ảnh đó được tóm lược lại bởi 3 giá trị màu: màu xanh của lá cây, màu xanh của da trời, và màu nâu của vỏ cây. Với giả định rằng bóng đổ tương ứng với sự thay đổi độ sáng và chuyển động thực của đối tượng gây ra sự thay đổi đó, phiên bản đầu tiên của phương thức này được thiết kế để loại bỏ nhận diện sai gây ra bởi sự thay đổi của độ sáng. Điều này đạt được nhờ vào việc xây dựng một mô hình đánh giá sự biến đổi giá trị màu theo giá trị thành phần cường độ sáng, được mô tả bởi Hình 3.4 dưới đây:



Hình 3.4. Đánh giá biến đổi màu sắc theo cường độ sáng

Trong đó, vùng bao quyết định (decision boundary) được xây dựng quanh giá trị màu của từ mã v_i theo dải cường độ sáng từ I_{low} đến I_{high}. Để xác định điểm ảnh đầu vào x_t có thuộc mô hình nền hay không, chỉ cần xác định xem x_t nằm trong hày ngoài vùng bao quyết định.

6) Mô hình Eigen Backgrounds

Trong khi các phương pháp đã trình bày sử dụng phương pháp thống kê ở mức điểm ảnh (pixel-level) thì Eigen là phương pháp sử dụng không gian riêng (eigenspace) để mô hình nền.

Giả sử $\{I_i\}_{i=1:N}$ là vector biểu diễn giá trị của một điểm ảnh trong N khung hình liên tiếp. Giá trị trung bình được tính bởi:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} I_i$$

Vector $\{X_i\}_{i=1:N}$ được xây dựng bởi: $X_i = I_i - \mu$. Sau đó, ma trận hiệp phương sai Σ được tạo bởi: $\Sigma = E[XX^T]$ với $X = [X_1, .., X_n]$. Ta có thể tìm vector riêng ϕ làm chéo hoá ma trận hiệp phương sai Σ:

$$D = \phi\Sigma\phi^T$$

Với D là ma trận đường chéo tương ứng.

Ma trận eigenbackground ϕ_M được tạo từ M vector riêng tương ứng với M giá trị riêng lớn nhất. Sau khi tính được ϕ_M và μ, mỗi giá trị ảnh đầu vào I_t được chiếu lên không gian con M chiều:

$$B_t = \phi_M(I_t - \mu)$$

Sau đó được tái tạo lại bởi:

$$I_t' = \phi_M^T B_t + \mu$$

Cuối cùng, một điểm ảnh thuộc vào tiền cảnh được phát hiện dựa trên việc tính toán khoảng cách giữa I_t và I_t'.

$$\chi_t(s) = \begin{cases} 1 & nếu\ d(I_t, I_t') > \tau \\ 0 & trong\ các\ trường\ hợp\ khác \end{cases}$$

Với τ là ngưỡng, còn d là khoảng cách Euclidian.

3.2.2. Áp dụng kỹ thuật trừ nền, phân tách vùng chuyển động

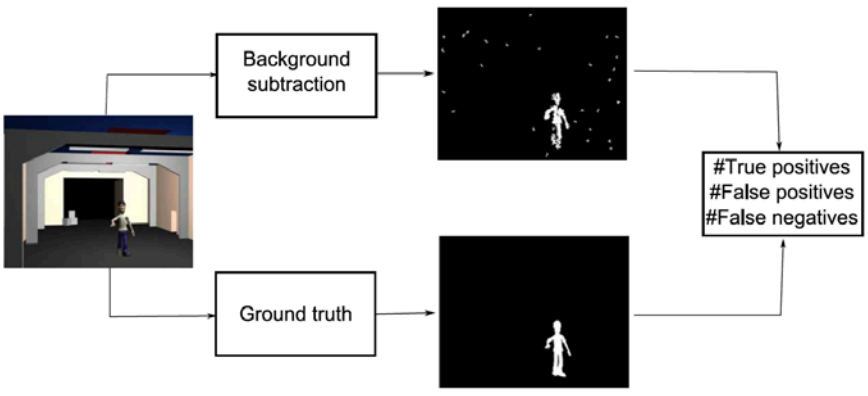
Để lựa chọn kỹ thuật trừ nền phù hợp cho bài toán của luận văn, tác giả đã tìm hiểu các thực nghiệm so sánh, đánh giá hiệu quả, tốc độ tính toán và mức sử dụng tài nguyên của các thuật toán trừ nền đã trình bày. Cụ thể theo [55] như sau:

Phương pháp đánh giá hiệu quả kỹ thuật trừ nền

Để đánh giá hiệu quả các phương pháp trừ nền, [55] sử dụng độ khôi phục (Recall) và độ chính xác (Precision) định nghĩa bởi:

$$Precision = \frac{TP}{TP + FP} \ và\ Recall = \frac{TP}{TP + FN}$$

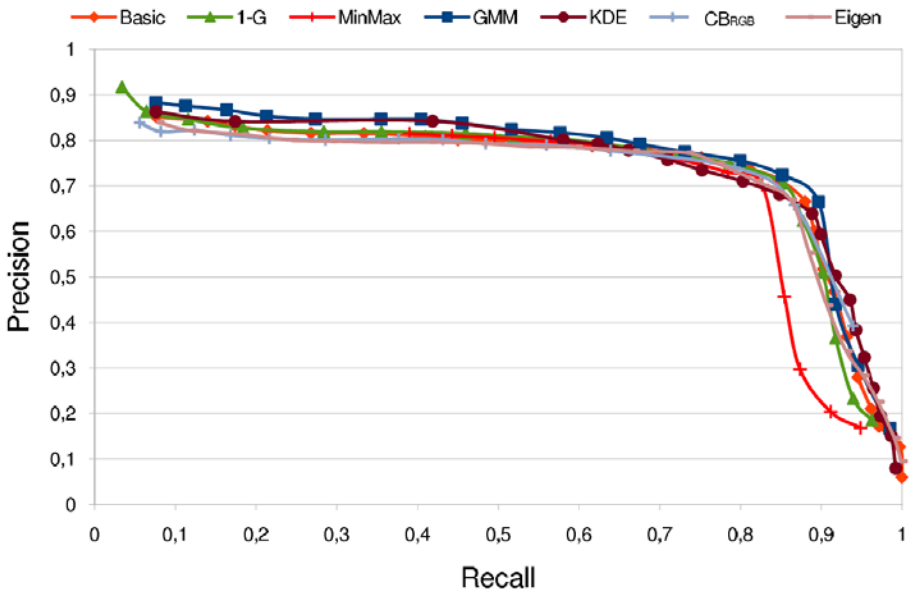
Trong đó, TP là số lượng điểm ảnh thuộc tiền cảnh phát hiện đúng, FP là số lượng điểm ảnh thuộc nền bị nhận diện nhầm thành tiền cảnh, FN là số lượng điểm ảnh thuộc tiền cảnh không được phát hiện, được minh hoạ trong hình 3.5.



Hình 3.5. Minh hoạ phương pháp đánh giá hiệu quả kỹ thuật trừ nền

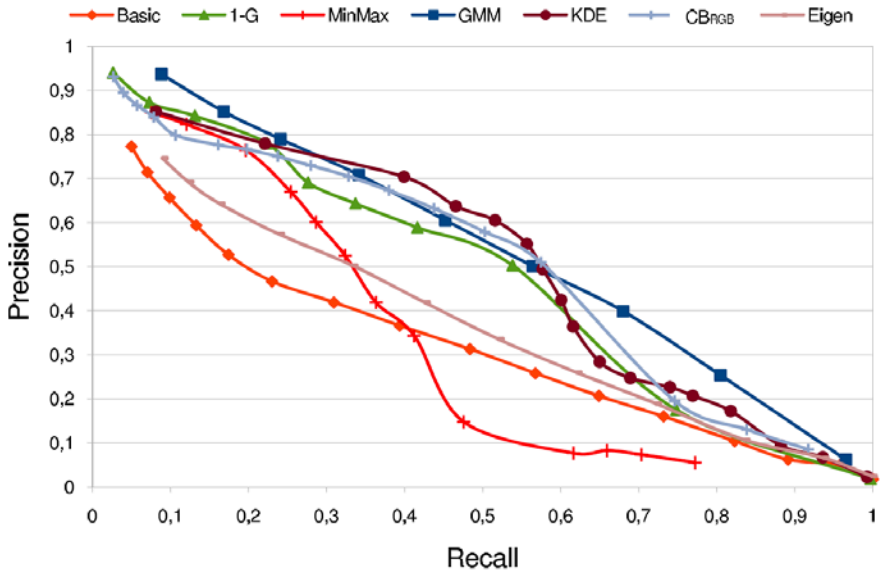
So sánh các kỹ thuật trừ nền

Để so sánh mức độ hiệu quả các kỹ thuật trừ nền trong các điều kiện khác nhau của từng tập dữ liệu thực nghiệm bao gồm: nền đơn; nền phức tạp; hình ảnh video bao gồm nhiều nhiễu, cần xem xét các đường cong Precision-Recall dưới đây:



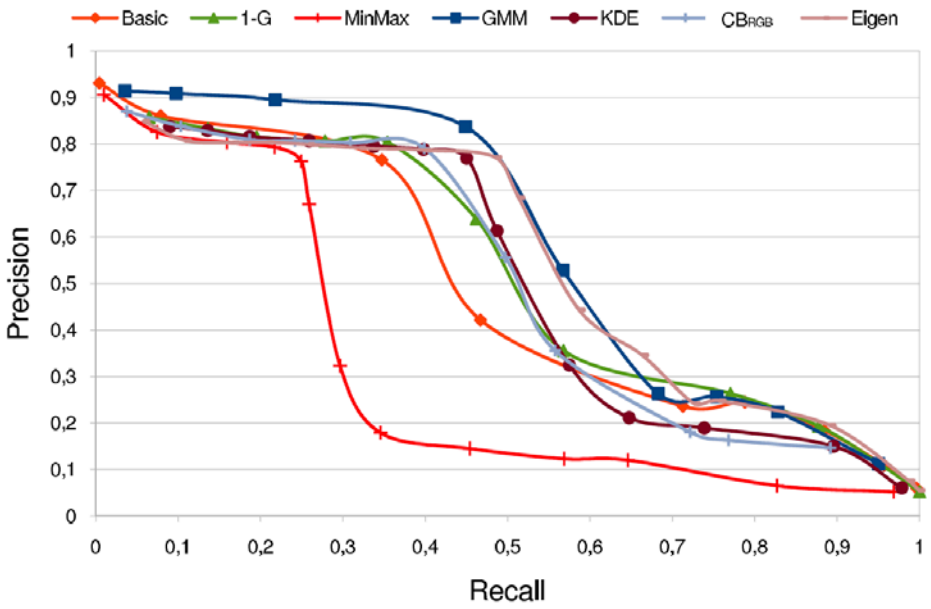
Hình 3.6. Đường cong Precision-Recall các kỹ thuật trừ nền khi thử nghiệm trên tập dữ liệu có nền tĩnh, không nhiễu

Hình 3.6 thể hiện hiệu quả phát hiện tiền cảnh của các kỹ thuật khác nhau. Trong đó, Basic là kỹ thuật sử dụng ảnh đơn để mô hình nền; 1-G là sử dụng duy nhất một phân bố Gaussian; MinMax là kỹ thuật mô hình nền MinMax; GMM là kỹ thuật sử dụng nhiều phân bố Gaussian để mô hình nền; KDE là kỹ thuật Kernel Density Estimation; CB_{RGB} là kỹ thuật sử dụng bảng mã với ảnh màu RGB; Eigen là kỹ thuật mô hình nền bằng Eigen Backgrounds. Có thể thấy đường cong của GMM nằm phía trên cùng, hay nói cách khác GMM hiệu quả nhất trong trường hợp nền đơn.



Hình 3.7. Đường cong Precision-Recall các kỹ thuật trừ nền khi thử nghiệm trên tập dữ liệu có nền phức tạp

Hình 3.7 thể hiện đường cong Precision-Recall của các kỹ thuật khi thử nghiệm trên tập dữ liệu có nền phức tạp. Trong trường hợp này, KDE, GMM và CB_{RGB} cho kết quả chính xác nhất.



Hình 3.8. Đường cong Precision-Recall các kỹ thuật trừ nền khi thử nghiệm trên tập dữ liệu rất nhiễu

Với tập dữ liệu video rất nhiễu, các phương pháp dựa trên thống kê cho kết quả tốt hơn, đặc biệt GMM chiếm ưu thế về độ chính xác.

Với yêu cầu hệ thống thời gian thực, việc lựa chọn kỹ thuật trừ nền cần cân nhắc nhiều tới thời gian tính toán và yêu cầu tài nguyên hệ thống. Bảng 3.1 dưới đây thể hiện tốc độ của các thuật toán. Cần lưu ý CB_{RGB} là kỹ thuật có thời gian tính toán phụ thuộc rất nhiều vào mức độ phức tạp của khung cảnh trong video. Vì vậy việc so sánh về mặt tốc độ đối với kỹ thuật này là không phù hợp. Vì thế bảng 3.1 không cung cấp số liệu cho thuật toán CB_{RGB}.

Bảng 3.1. Thời gian xử lý trung bình của các kỹ thuật trừ nền

Phương pháp	**Thời gian xử lý trung bình**
Mô hình nền bằng ảnh đơn	1
Mô hình nền bằng 1-Gaussian	1.32
MinMax	1.47
GMM	4.91
KDE	13.80
Eigen	11.98

Trong khi đó, mức độ chiếm dụng bộ nhớ của các thuật toán được trình bày trong bảng 3.2 dưới đây. Trong đó, với GMM, K là số phân bố Gaussian được dùng; với KDE thì N là số khung hình liên tiếp dùng để ước lượng P(.) – thường từ 100 đến 200; với CB_{RGB}, L là số lượng từ mã; còn M là số lượng vector riêng được giữ lại để tạo không gian con M chiều.

Bảng 3.2. Số phép tính dấu phẩy động của các kỹ thuật trừ nền

Phương pháp	Số phép tính
Mô hình nền bằng ảnh đơn	3
Mô hình nền bằng 1-Gaussian	6
MinMax	3
GMM	K×5
KDE	N×3+3
CB_{RGB}	L×6
Eigen	M×3+3

Cuối cùng, các thuật toán trừ nền được so sánh một cách tổng quát tại bảng 3.3, trong đó số lượng dấu * thể hiện mức độ hiệu quả của thuật toán.

Bảng 3.3. Bảng so sánh chung mức độ hiệu quả các kỹ thuật trừ nền

	Basic	1-G	MinMax	GMM	KDE	CB_{RGB}	Eigen
Nền tĩnh	***	***	**	***	***	***	***
Nền phức tạp	*	**	*	***	***	***	*
Nền rất nhiễu	*	***	*	***	***	***	***
Thời gian tính toán	***	***	***	**	*	-	*
Yêu cầu bộ nhớ	***	***	***	**	*	**	*

Nhìn chung, những kỹ thuật cho kết quả tốt thì thường yêu cầu tài nguyên lớn và tốc độ tính toán chậm và ngược lại, những kỹ thuật tính toán rất nhanh và yêu cầu ít tài nguyên thì cho kết quả kém hơn trong đa số các trường hợp. Có thể nhận thấy, GMM là kỹ thuật cho kết quả rất tốt và thời gian tính toán cũng như yêu cầu bộ nhớ ở mức chung bình. Vậy GMM là giải thuật phù hợp hơn cả đối với bài toán phát hiện ngã, đảm bảo được tốc độ xử lý thời gian thực. Vì vậy luận văn này lựa chọn áp dụng giải thuật GMM.

Chuỗi video thu được từ camera thường bao gồm nhiều nhiễu. Vì vậy qua giải thuật trừ nền thì phần tiền cảnh (foreground) bao gồm rất nhiều đốm nhỏ. Vùng đánh dấu chuyển

động tương ứng với cơ thể người bị phân chia thành nhiều đốm không kết nối với nhau. Vì thế tác giả đề xuất sử dụng một số kỹ thuật để cải thiện kết quả ở bước này như sau:

- Sử dụng bộ lọc Gaussian làm trơn ảnh đầu vào, giúp giảm bớt ảnh hưởng của nhiễu ngẫu nhiên.
- Thực hiện trừ nền, phân tách vùng chuyển động.
- Sử dụng các phép biến đổi hình thái học (morphological) để loại bỏ đốm nhỏ, kết nối các đốm lớn thành khối. Cụ thể, dùng phép co (erosion) để loại bỏ đốm nhỏ, sau đó áp dụng phép giãn nở (dilation) để mở rộng các đốm lớn, khiến các đốm này có phần giao nhau.

Hình 3.9. Một ví dụ phân tách vùng chuyển động. Từ trái qua phải, ảnh đầu tiên là ảnh đầu vào từ chuỗi video; ảnh thứ 2 là kết quả của giải thuật trừ nền GMM; ảnh thứ 3 thể hiện kết quả loại bỏ các đốm nhỏ bằng thực hiện phép co; ảnh cuối cùng sau khi thực hiện phép giãn nở, các đốm lớn đã được kết nối và thể hiện tốt hình dạng cơ thể

3.3. Trích rút đặc trưng chuyển động

Chuyển động là thông tin quan trọng bậc nhất về việc ngã, bởi vì không có ca ngã nguy hiểm nào mà không xuất hiện chuyển động lớn. Vì vậy cần thiết phải trích rút thông tin chuyển động trong chuỗi video. Các kỹ thuật trích rút thông tin chuyển động phổ biến gồm có kỹ thuật Optical flow và Motion History Image.

3.3.1. Optical flow

Trong chuỗi video, bài toán đặt ra là, chọn một tập điểm ảnh thuộc khung hình thứ *t*, làm sao để xác định vị trí của tập điểm đó trong các khung hình tiếp theo. Hay nói cách khác là xác định sự chuyển động tương đối của các điểm ảnh trên bề mặt một đối tượng qua các khung hình. Optical flow là khái niệm chỉ sự chuyển động tương đối đó. Các phương pháp ước lượng optical flow có mục đích là xấp xỉ các chuyển động của điểm ảnh giữa các khung hình.

Có rất nhiều hướng tiếp cận đối với ước lượng optical flow nhưng vì giới hạn thời gian, luận văn này chỉ tập trung vào phương pháp ước lượng dựa trên gradient.

Optical flow dựa trên gradient sử dụng giả thiết quan trọng là, bề mặt đối tượng không có nhiều sự thay đổi về giá trị cường độ sáng trong hai khung hình liên tiếp. Nghĩa là:

$$I_2(x+u, y+v) = I_1(x, y)$$

Trong đó, I_1, I_2 là giá trị của điểm ảnh đang xét tại khung hình thứ *t* và *t+1*; (x,y) là vị trí của điểm ảnh ở thời điểm *t*; (u,v) là vector chuyển động của điểm ảnh.

Hay là:

$$I(x+u, y+v, t+1) = I(x, y, t)$$

Sử dụng khai triển Taylor:

$$\frac{\partial_I}{\partial_x}u + \frac{\partial_I}{\partial_y}v + \frac{\partial_I}{\partial_t} + I(x, y, t) = I(x, y, t)$$

$$\Rightarrow \frac{\partial_I}{\partial_x}u + \frac{\partial_I}{\partial_y}v = -\frac{\partial_I}{\partial_t}$$

Vậy ta có phương trình cơ bản của optical flow:

$$\nabla I. \begin{bmatrix} u \\ v \end{bmatrix} = -\frac{\partial_I}{\partial_t}$$

Phương trình trên có 2 ẩn là u và v. Để giải được cần phải bổ xung thêm một phương trình nữa. Các phương pháp ước lượng đều cố gắng tìm ra phương trình thứ 2 này (ràng buộc thứ 2). Trong phạm vi luận văn chỉ trình bày phương pháp ước lượng bình phương nhỏ nhất (Least-squares), như sau:

Xét cửa sổ 3×3 các lân cận của điểm ảnh đang xét, đặt $f_x = \frac{\partial_I}{\partial_x}$; $f_y = \frac{\partial_I}{\partial_y}$ và $f_t = \frac{\partial_I}{\partial_t}$ ta có:

$$\begin{cases} f_{x1}u + f_{y1}v = -f_{t1} \\ \vdots \\ f_{x9}u + f_{y9}v = -f_{t9} \end{cases}$$

Hay:

$$\begin{bmatrix} f_{x1} & f_{y1} \\ \vdots & \vdots \\ f_{x9} & f_{y9} \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} -f_{t1} \\ \vdots \\ -f_{t9} \end{bmatrix}$$

Để bổ xung một ràng buộc nữa, cần sử dụng một giả thiết thứ 2 là các điểm ảnh lân cận điểm đang xét cũng chuyển động với cùng vận tốc. Nhưng thực tế là không phải vậy, vẫn có sự khác biệt trong chuyển động của các điểm ảnh lân cận. Vì vậy cần tìm vector vận tốc làm cực tiểu bình phương lỗi:

$$E(u, v) = \sum_i (f_{xi}u + f_{yi}v + f_{ti})^2$$

Vận tốc cần tìm là (u,v) làm cực tiểu E(u,v). Người ta chứng minh được rằng E(u,v) đạt cực tiểu tại điểm đạo hàm riêng của nó bằng không, hay là:

$$\begin{cases} \sum (f_{xi}u + f_{yi}v + f_{ti})f_{xi} = 0 \\ \sum (f_{xi}u + f_{yi}v + f_{ti})f_{yi} = 0 \end{cases}$$

Thực hiện phép nhân và chuyển vế ta được:

$$\sum f_{xi}^2 u + \sum f_{xi}f_{yi}v = -\sum f_{xi}f_{ti}$$

$$\sum f_{yi}^2 v + \sum f_{xi}f_{yi}u = -\sum f_{yi}f_{ti}$$

Viết lại dưới dạng ma trận:

$$\begin{bmatrix} \sum f_{xi}^2 & \sum f_{xi}f_{yi} \\ \sum f_{xi}f_{yi} & \sum f_{yi}^2 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} -\sum f_{xi}f_{ti} \\ -\sum f_{yi}f_{ti} \end{bmatrix}$$

Vậy vector vận tốc (u,v) được tính bằng:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \sum f_{xi}^2 & \sum f_{xi}f_{yi} \\ \sum f_{xi}f_{yi} & \sum f_{yi}^2 \end{bmatrix}^{-1} \begin{bmatrix} -\sum f_{xi}f_{ti} \\ -\sum f_{yi}f_{ti} \end{bmatrix}$$

Có thể thấy rằng, việc ước lượng optical flow như đã mô tả dựa trên hai giả thiết là cường độ sáng không đổi trong hai khung hình liên tiếp; và các điểm lân cận thì chuyển động giống nhau. Vì vậy kỹ thuật này chỉ phù hợp với các chuyển động không quá lớn và mượt mà. Khi chuyển động nhanh đột ngột, các giả thiết trên không còn phù hợp nữa.

3.3.2. Motion History Image (MHI)

MHI được giới thiệu lần đầu tiên tại [4], là một kỹ thuật biểu diễn chuyển động trong chuỗi ảnh đơn giản nhưng hết sức mạnh mẽ. MHI biểu diễn vị trí và quỹ đạo của chuyển động trong chuỗi video. Giá trị của mỗi điểm ảnh trong MHI là một hàm số của lịch sử chuyển động tại vị trí điểm ảnh đó. Giá trị điểm ảnh càng lớn, chuyển động diễn ra càng gần thời điểm hiện tại.

Cường độ của mỗi điểm ảnh trong MHI đặc chưng cho năng lượng chuyển động tại vị trí của điểm ảnh đó. Giả sử một điểm ảnh nhận mức năng lượng khi xuất hiện chuyển động tại vị trí của nó, điểm ảnh có giá trị cực đại. Năng lượng đó mất dần đi theo thời gian nếu không có chuyển động nào khác diễn ra. Giá trị điểm ảnh giảm dần. Vì vậy, MHI giữ được thông tin về lịch sử chuyển động.

1) Tạo ảnh MHI

Thông thường, MHI được xây dựng dựa trên ảnh nhị phân tạo bởi phép trừ 2 khung hình liên tiếp. Sau đó áp dụng ngưỡng:

$$\psi(x,y,t) = \begin{cases} 1 & nếu\ D(x,y,t) \geq \partial \\ 0 & trong\ trường\ hợp\ khác \end{cases}$$

Trong đó:

$$D(x,y,t) = |I(x,y,t) - I(x,y,t \pm \Delta)|$$

với $I(x,y,t)$ là giá trị của điểm ảnh *(x,y)* tại khung hình thứ *t*.

MHI H_τ được tính như sau:

$$H_\tau(x,y,t) = \begin{cases} \tau & nếu\ \psi = 1 \\ \max(0, H_\tau(x,y,t-1) - \partial) & nếu\ \psi \neq 1 \end{cases}$$

Với τ là khoảng thời gian tối đa mà lịch sử chuyển động được lưu lại. Còn ∂ là tham số độ phân rã. ∂ càng nhỏ thì MHI càng mịn.

2) Tính hướng của chuyển động

Từ MHI, phương pháp thông dụng để xác định hướng chuyển động như sau theo [33]:

- Với mỗi vùng lân cận kích thước 3x3 của một điểm ảnh, tìm giá trị lớn nhất *M(x,y)* và nhỏ nhất *m(x,y)*. Nếu thỏa mãn:

  $$\partial_1 \leq M(x,y) - m(x,y) \leq \partial_2$$

  với ∂_1, ∂_2 là các giới hạn cho phép của sự chênh lệch giá trị của các điểm ảnh trong vùng lân cận, thì hướng chuyển động của mỗi điểm ảnh tính bằng:

  $$orientation(x,y) = arctan\frac{d_{mhi}/d_y}{d_{mhi}/d_x}$$

- Hướng tổng thể của chuyển động được tính bằng giá trị trung bình có trọng số của hướng của tất cả các điểm ảnh. Trong đó, chuyển động càng gần thời điểm hiện tại thì trọng số càng lớn và ngược lại.

Hình 3.10. Ví dụ minh họa ảnh MHI: Ảnh bên trái là khung hình trong chuỗi video; ảnh bên phải là ảnh kết quả thu được từ giải thuật MHI

3.3.3. Image Moments

Moment của hàm liên tục *f(x,y)* được định nghĩa:

$$M_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x,y) dx dy$$

Đối với ảnh xám *I(x,y)*, công thức trên trở thành:

$$M_{pq} = \sum_x \sum_y I(x,y)$$

Áp dụng vào ảnh nhị phân:

$$M_{00} = \sum_x \sum_y I(x,y)$$
$$M_{10} = \sum_x \sum_y xI(x,y) = x_0 \sum_x \sum_y I(x,y)$$
$$M_{01} = \sum_x \sum_y yI(x,y) = y_0 \sum_x \sum_y I(x,y)$$

Với x_0, y_0 là trọng tâm của ảnh nhị phân. Vậy:

$$(x_0, y_0) = \left(\frac{M_{10}}{M_{00}}, \frac{M_{01}}{M_{00}}\right)$$

Như vậy trọng tậm của ảnh nhị phân có thể được tính thông qua các moment của ảnh như đã mô tả ở trên.

3.3.4. Áp dụng MHI, Image Moments trích rút đặc trưng chuyển động

Kỹ thuật Optical flow thường được sử dụng để phát hiện chuyển động trong chuỗi video. Tuy nhiên kỹ thuật này không mấy phù hợp với ứng dụng thời gian thực do tốc độ tính toán. Hơn nữa như đã trình bày, kỹ thuật này có thể gây lỗi trong trường hợp chuyển động nhanh, trong khi chuyển động nhanh hầu như sẽ sảy ra trong khi ngã. Vì vậy luận văn này lựa chọn MHI, một kỹ thuật đơn giản với tốc độ cao và rất hiệu quả trong việc trích rút thông tin chuyển động.

Luận văn sử dụng các đặc trưng chuyển động bao gồm độ lớn và hướng của chuyển động cho phát hiện ngã. Các đặc trưng được trích rút dựa trên kỹ thuật Motion History Image, kết hợp với chuyển động của trọng tâm cơ thể, xác định qua kỹ thuật Image Moments. Chi tiết về các kỹ thuật được áp dụng như sau:

1) Hướng của chuyển động

Với phương thức đo lường hướng tổng thể của chuyển động đã trình bày trong phần 2), mục 3.3.1 áp dụng cho tình huống ngã do vấp với đặc điểm chung là giai đoạn đầu chuyển động rất nhanh theo phương ngang, đến cuối hành trình mới chuyển động theo phương dọc thì hướng tổng thể vẫn theo phương ngang. Vì vậy luận văn này đề xuất

một phương thức tính hướng chuyển động nhạy cảm hơn với trường hợp kể trên như sau:

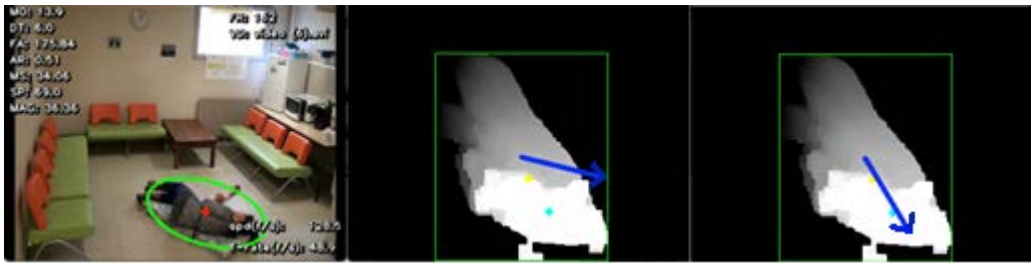
- Xác định trọng tâm của foreground ở khung hình hiện tại M_{recent}
- Xác định trọng tâm của MHI M_{MHI}
- Hướng chuyển động được xác định như sau:

$$dx = |M_{recent}[0] - M_{MHI}[0]|;$$
$$dx = |M_{recent}[1] - M_{MHI}[1]|;$$
$$angle = arctan\frac{dx}{dy}$$

Với độ lớn: $magn = \sqrt{dx^2 + dy^2}$

Trong đó, trọng tâm được xác định dựa trên Image Moments đã trình bày trong mục 3.3.2.

Hình 3.11 mô tả thực nghiệm để khẳng định tính đúng đắn của quan sát đã được đề cập và cho thấy hiệu quả tốt hơn của phương thức đề xuất trong ngữ cảnh bài toán. Có thể thấy, hướng chuyển động (được thể hiện bằng đường thẳng màu xanh đậm trong hình) xác định bằng phương thức được đề xuất là phù hợp hơn với thực tế chuyển động của hành động ngã trong chuỗi video.



Hình 3.11. So sánh phương thức xác định hướng chuyển động. Từ trái qua phải: Ảnh 1 là khung hình từ chuỗi video; ảnh 2 thể hiện kết quả xác định hướng chuyển động bằng kỹ thuật truyền thống; ảnh 3 là kết quả của kỹ thuật được đề xuất

2) Độ lớn của chuyển động

Một cách trực quan như ví dụ ở Hình 3.10 đã minh họa, ảnh MHI thể hiện chuyển động ở thời điểm hiện tại bằng vùng điểm ảnh có giá trị cực đại (màu trắng trong ảnh), còn vùng chuyển động quá khứ là vùng điểm ảnh có giá trị nhỏ hơn giá trị cực đại và giảm dần về 0 (vùng màu xám trong ảnh). Vùng không xuất hiện chuyển động có giá trị điểm ảnh bằng 0. Chuyển động càng nhanh, vùng màu xám càng lớn. Vì vậy tác giả đề xuất phương thức ước lượng độ lớn của chuyển động từ MHI bởi:

$$M_{rate} = \frac{\sum NZP_{MHI} - NZP_{MotionMask}}{\sum NZP_{MHI}} \times 100\%$$

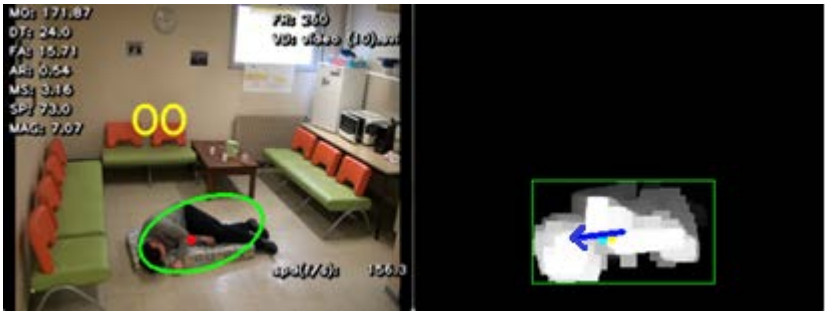
Trong đó, $\sum NZP_{MHI}$ là tổng số điểm ảnh có giá trị lớn hơn 0 của MHI; $NZP_{MotionMask}$ là tổng số điểm ảnh có giá trị lớn hơn 0 của foreground phân tách được ở khung hình hiện tại.

Giá trị của M_{rate} nằm trong khoảng từ 0% đến 100%, tương ứng với mức độ chuyển động từ nhỏ đến lớn.

Tuy nhiên ở giai đoạn cuối của hành động ngã, phần lớn cơ thể đã dừng chuyển động, vùng điểm ảnh có giá trị cực đại thu hẹp lại trong khi vùng màu xám còn rất rộng. Điều đó dẫn đến giá trị M_{rate} là rất lớn, không mô tả đúng thực tế là tốc độ chuyển động nhỏ trong trường hợp này. Thực tế này được minh hoạ như ví dụ ở Hình 3.12. Vì vậy cần thiết phải sử dụng một ngưỡng cho độ dài vector chuyển động $magn$ tính được ở phần 1) mục này trong xác định độ lớn chuyển động. Cụ thể như sau:

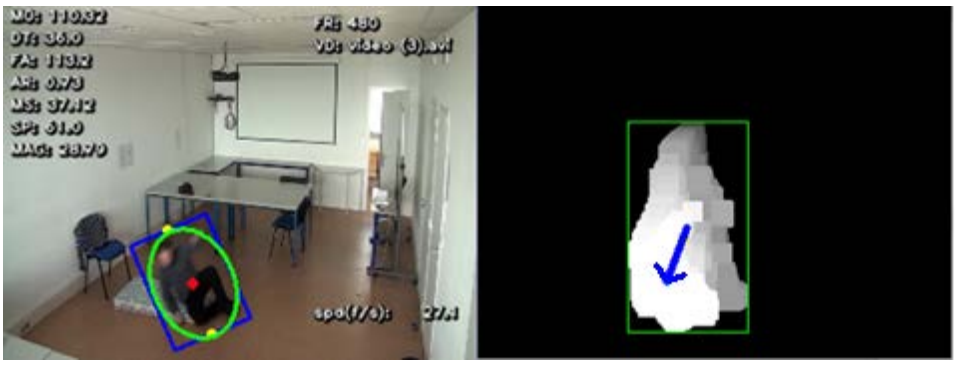
$$M_{rate} = \begin{cases} \dfrac{\sum NZP_{MHI} - NZP_{MotionMask}}{\sum NZP_{MHI}} \times 100\% & \text{nếu } magn > \tau \\ 0 & \text{nếu } magn \leq \tau \end{cases}$$

Trong đó, τ là ngưỡng cho trước, được xác định qua thực nghiệm để chọn giá trị phù hợp nhất cho từng tập dữ liệu khác nhau.



Hình 3.12. Minh hoạ xác định M_{rate} lỗi trong thời điểm gần kết thúc chuyển động: Khi người ngã đã nằm trên sàn và vừa kết thúc chuyển động, M_{rate} = 73% (rất lớn) nhưng magn chỉ là 7.07 (rất nhỏ)

Hạn chế lớn nhất của MHI khi áp dụng đo lường độ lớn chuyển động là, khi vùng chuyển động hiện tại chồng lấp phần chuyển động trong quá khứ, một phần thông tin về lịch sử chuyển động bị mất đi. Tương ứng với vùng màu xám bị nhỏ hơn thực tế phải có. Dẫn đến trong trường hợp xảy ra chồng lấp, độ lớn chuyển động đo được từ phương thức trên là nhỏ hơn thực tế. Để khắc phục vấn đề đó, tác giả đề xuất sử dụng thêm một đặc trưng về chuyển động nữa, đó là khoảng cách trọng tâm của cơ thể trong khung hình hiện tại so với vị trí trọng tâm ở trước đó 15 khung hình, gọi là mức độ chuyển động của trọng tâm. Nếu khoảng cách này lớn hơn ngưỡng chọn trước, rất có thể đã xuất hiện chuyển động nhanh bất thường. Trong trường hợp này, ngưỡng xác định chuyển động nhanh bất thường đối với M_{rate} và *magn* được điều chỉnh giảm xuống để không bỏ xót chuyển động nhanh trong trường hợp xảy ra chồng lấp.



Hình 3.13. Ví dụ cho ước lượng độ lớn chuyển động: Ảnh bên trái là khung hình trích ra từ chuỗi video; Ảnh bên phải là MHI. Khi ngã, $M_{rate} = 61\%$, magn=28, độ chuyển động của trọng tâm = 37

Bằng các quan sát trong quá trình thực nghiệm, tác giả nhận thấy rằng trong điều kiện chiếu sáng phức tạp, foreground thu được từ quá trình phân tách vùng chuyển động là rất nhiễu và không ổn định. Trong trường hợp camera được đặt quá thấp, ngang với chiều cao của người, khi đối tượng chuyển động đến gần sẽ khiến kích hoạt chức năng tự động đo sáng của camera, dẫn đến sự thay đổi đột ngột mức sáng trên toàn bộ khung hình. Điều đó dẫn đến sai lầm của giải thuật trừ nền. Để hạn chế bất lợi này, tác giả đề xuất sử dụng ngưỡng diện tích vùng chuyển động. Nếu diện tích vùng chuyển động quá lớn hoặc quá nhỏ, không phù hợp với vùng chuyển động tương ứng tạo ra bởi việc di chuyển của cơ thể người, chuyển động này sẽ được xem như là nhiễu và bị bỏ qua. Mặt khác, việc ứng dụng ngưỡng độ lớn diện tích trên còn giúp loại trừ các chuyển động nhỏ của một phần cơ thể trong các hoạt động thường nhật bình thường.

Cũng vì lý do nhiễu, vùng foreground có thể di chuyển đột ngột từ vị trí này tới vị trí khác. Trong trường hợp này, khoảng cách trọng tâm của foreground trong hai khung hình liên tiếp là lớn bất thường. Chuyển động này cũng được bỏ qua nếu khoảng cách kể trên lớn hơn một ngưỡng chọn trước.

Khi không xuất hiện chuyển động trong khung hình, phương thức được đề xuất yêu cầu phải xóa bỏ lịch sử chuyển động trước đó để không làm ảnh hưởng đến kết quả đo lường ở những khung hình tiếp theo.

3.4. Trích rút đặc trưng hình dạng cơ thể

Hình dạng cơ thể thay đổi rõ rệt trong quá trình ngã. Khi đứng thẳng hay đi lại, cơ thể người gần như vuông góc với phương nằm ngang, tỉ lệ chiều ngang so với chiều cao của người là nhỏ. Khi xảy ra ngã, người ngã nằm trên mặt sàn thường với tư thế co người lại và chân tay không ở sát người. Vì vậy tỉ lệ giữa chiều ngang và chiều cao của hình dạng tăng lên đáng kể. Các đặc trưng này có thể ước lượng bằng cách xác định các hình dạng hình học bao ngoài cơ thể như hình chữ nhật, hình ellipse.

3.4.1. Kỹ thuật fitting ellipse

Kỹ thuật này được Fitzgibbon giới thiệu tại [7]. Chi tiết như sau:

Tìm Fitting ellipse của tập điểm là bài toán xác định một ellipse sao cho tổng khoảng cách từ tập điểm đến ellipse là nhỏ nhất.

Hình ellipse được biểu diễn theo conic: Tập điểm *(x,y)* thỏa mãn

$$F(x,y) = ax^2 + bxy + cy^2 + dx + ey + f = 0$$

xác định một ellipse nếu:

$$b^2 - 4ac < 0$$

Với *a, b, c, d, e, f* là bộ hệ số của ellipse.

Viết dưới dạng vector, *F(x,y)* trở thành:

$$F(X) = X * \alpha = 0$$

Trong đó:

$$X = [x^2, xy, y^2, x, y, 1];$$
$$\alpha = [a, b, c, d, e, f]^T$$

Theo đề xuất ở [43], việc xác định fitting ellipse có thể đạt được bằng cách tìm ma trận hệ số α sao cho tổng bình phương khoảng cách đại số từ tập điểm đến ellipse là nhỏ nhất:

$$min \sum_{k=1}^{n} (F(X_k))^2 = min \sum_{k=1}^{n} (X_k * \alpha)^2$$

Với điều kiện:

$$b^2 - 4ac < 0$$

Trong đó, *n* là kích thước tập điểm.

Tuy nhiên, bởi vì nếu α là bộ hệ số biểu diễn một ellipse thì $k * \alpha, k \neq 0$ cũng biểu diễn chính ellipse đó. Hoàn toàn có thể chọn k sao cho:

$$b^2 - 4ac = 1$$

Vậy bài toán có thể biểu diễn bởi:

$$min \parallel D\alpha \parallel^2$$

với điều kiện:

$$\alpha^T C \alpha = 1$$

Trong đó,

$$D = \begin{pmatrix} x_1^2 & x_1 y_1 & y_1^2 & x_1 & y_1 & 1 \\ x_i^2 & x_i y_i & y_i^2 & x_i & y_i & 1 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_n^2 & x_n y_n & y_n^2 & y_n & y_n & 1 \end{pmatrix}$$

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

Đây là bài toán tìm cực trị có điều kiện. Bằng cách áp dụng phương pháp nhân tử Lagrange, ta có phương trình Lagrange:

$$L(\alpha, \lambda) = \| D\alpha \|^2 - \lambda(\alpha^T C\alpha - 1)$$

Trong đó λ là nhân tử Lagrange.

Điểm cực trị là nghiệm của hệ:

$$\begin{cases} \frac{\partial_L}{\partial_\alpha} = 0 \\ \alpha^T C\alpha = 1 \end{cases}$$

Hay:

$$\begin{cases} S\alpha = \lambda C\alpha \\ \alpha^T C\alpha = 1 \end{cases}$$

Với $S = D^T D$.

Sử dụng Generalized eigenvectors (Vector riêng tổng quát) để giải hệ trên có thể thu được 6 cặp nghiệm. Tuy nhiên, vì có:

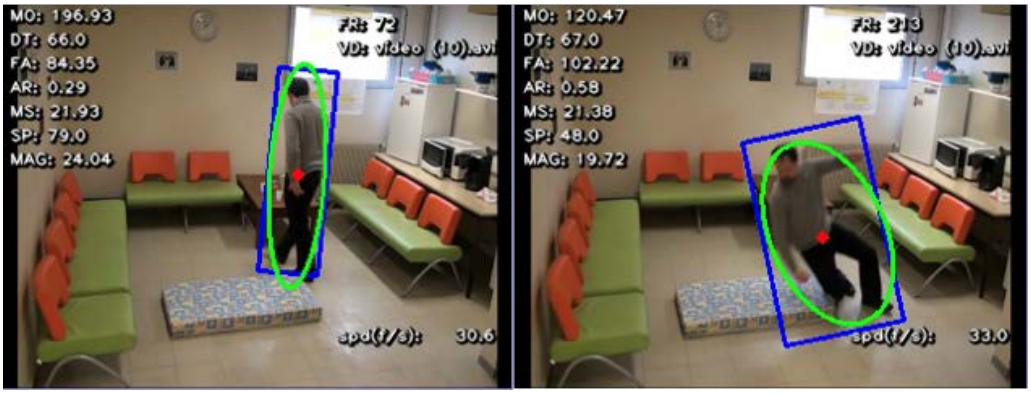
$$\| D\alpha \|^2 = \alpha^T D^T D\alpha = \alpha^T S\alpha = \lambda \alpha^T C\alpha = \lambda$$

Vậy ta chọn trị riêng và vector riêng sao cho trị riêng là nhỏ nhất và lớn hơn 0. Sau đó tìm hệ số để đảm bảo $\alpha^T C\alpha = 1$.

Với bài toán xác định fitting ellipse cho hình dáng người, hay fitting ellipse cho foreground, để giảm kích thước bài toán, hoàn toàn có thể quy về việc tìm fitting ellipse cho tập điểm thuộc bao lồi (convex hull) của foreground.

3.4.2. Áp dụng fitting ellipse đo lường đặc trưng hình dạng

Để đánh giá sự phù hợp của các kỹ thuật bounding box và fitting ellipse cho việc ước lượng đặc trưng hình dạng cơ thể, tác giả tiến hành các thực nghiệm cho cả hai kỹ thuật trên cùng tập dữ liệu. Thực tế trong hầu hết các trường hợp, bounding box cho kết quả kém hơn. Một ví dụ được minh hoạ trong Hình 3.14 dưới đây:



Hình 3.14. So sánh kỹ thuật bounding box với fitting ellipse: Ảnh bên trái, khi đi lại thì cả hai kỹ thuật cho kết quả tương đối giống nhau; ảnh bên phải, khi ngã, hai kỹ thuật cho kết quả rất khác biệt. Trong đó fitting ellipse mô tả sát hơn so với bounding box về hình dạng thực tế

Với cùng một tình huống ngã, có thể thấy hình chữ nhật bao quanh cơ thể không thể hiện chính xác góc nghiêng của người ngã, và vì vậy cả tỉ lệ chiều dài và chiều rộng của hình chữ nhật cũng không phù hợp với hình dạng cơ thể. Trong khi hình ellipse bao quanh tạo bởi kỹ thuật fitting ellipse mô tả rất sát các đặc trưng hình dáng của người ngã. Vì vậy luận văn này lựa chọn sử dụng kỹ thuật fitting ellipse cho việc trích rút đặc trưng hình dạng. Chi tiết như sau:

Từ hình ellipse xác định được qua kỹ thuật đã mô tả, chuyển đổi sang dạng biểu diễn parametric ta có các tham số biểu diễn một ellipse gồm: hai trục ellipse *a, b*; tâm ellipse *c(h,k)*; và góc quay *t*.

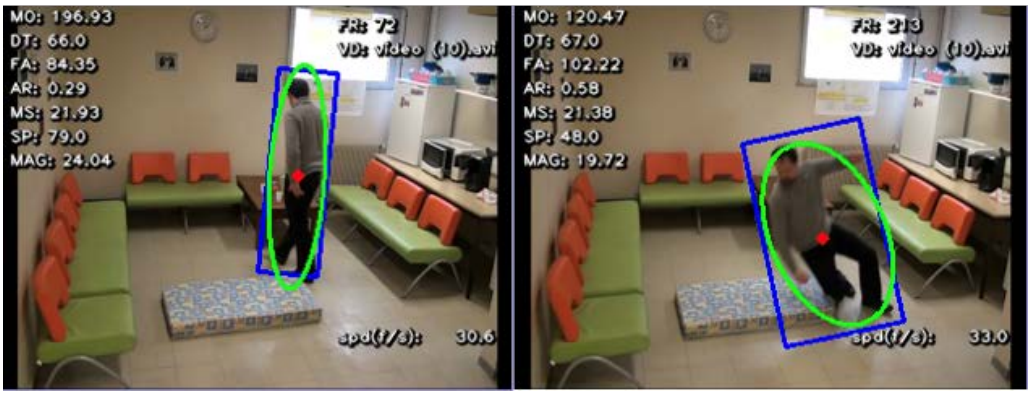
Tỉ lệ giữa bề ngang và chiều cao cơ thể AR tính bằng:

$$AR = \frac{b}{a}$$

Góc nghiêng cơ thể chính là góc nghiêng trục lớn của ellipse so với phương nằm ngang, hay chính là *t*.

Tuy nhiên, tác giả nhận thấy rằng vùng foreground thu được từ phân tách vùng chuyển động trong một số trường hợp bị ảnh hưởng rất nhiều bởi nhiễu. Một số trường hợp khác có thể không thu được foreground tại thời điểm người ngã vừa chấm dứt chuyển động, trong khi thông tin về hình dáng cơ thể tại ngay thời điểm này là rất quan trọng cho việc xác định ngã. Vì vậy luận văn này đề xuất sử dụng MHI cho xác định hình dáng cơ thể. Cụ thể như sau:

Từ MHI, sử dụng ngưỡng để tách lấy vùng điểm ảnh có giá trị lớn hơn ngưỡng. Ngưỡng được chọn qua thực nghiệm sao cho kết quả không quá sai khác so với hình dạng người chuyển động. Sau đó vùng điểm ảnh thu được sử dụng thay thế cho foreground trong quá trình xác định fitting ellipse. Kỹ thuật này giúp khai thác thông tin lịch sử chuyển động trong khoảng thời gian ngắn từ MHI, nhờ đó fitting ellipse thu được trở lên ổn định hơn rất nhiều.

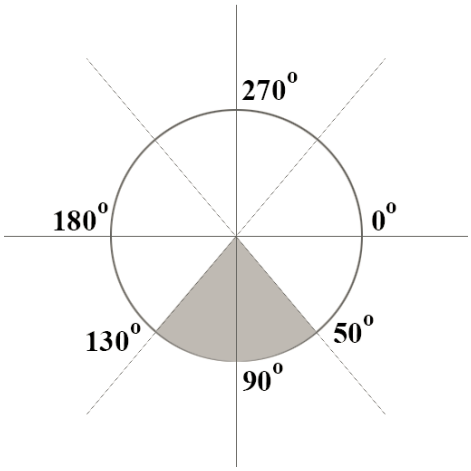


Hình 3.15. Minh họa sự thay đổi hình dạng cơ thể khi ngã: Ảnh bên trái, AR = 0.29 khi đi lại thông thường; ảnh bên phải, AR = 0.58 khi ngã

3.5. Phát hiện ngã

Theo quan sát của tác giả, thực tế rằng, quá trình ngã diễn ra trong khoảng thời gian nhỏ hơn 2s kể từ khi mất thăng bằng cho đến khi hoàn toàn nằm dưới sàn. Mặt khác, giai đoạn đầu của việc ngã cơ thể chuyển động nhanh. Nhưng nếu ngã về phía trước thì chuyển động có thể chậm lại sau đó do sự cố gắng kháng cự của người ngã. Ở giai đoạn đầu cơ thể cũng có thể chuyển động theo phương ngang (nhưng chắc chắn không có thành phần chuyển động hướng lên) trong trường hợp ngã do bị vấp, rồi sau đó mới chuyển động hướng xuống dưới. Vì vậy việc phát hiện ngã phải trải qua phân tích trong suốt hành trình ngã, kể từ khi xuất hiện chuyển động nhanh bất thường cho đến khi không xuất hiện chuyển động sau khi ngã. Trình tự như sau:

Khi phát hiện chuyển động nhanh đột ngột với hướng chuyển động < 180^0, hệ thống tiếp tục theo dõi các khung hình tiếp theo trong khoảng 50 khung hình (tương đương với 2s với tốc độ 25fps). Nếu có chuyển động hương xuống dưới sau đó (tương ứng với góc chuyển động nằm trong vùng màu xám minh hoạ ở Hình 3.16), tiếp theo là tỉ lệ AR vượt ngưỡng, góc nghiêng trục lớn vượt ngưỡng thì có thể xác định có thể là một tình huống ngã. Sau thời điểm đó, theo dõi tiếp trong 25 khung hình tiếp theo, nếu không phát hiện chuyển động hoặc chuyển động nhỏ thì kết luận xảy ra ngã. Nếu một trong các điều kiện kể trên không được thỏa mãn sau 50 khung hình thì không cảnh báo và tiếp tục theo dõi.



Hình 3.16. Quy ước về góc trong xác định hướng chuyển động và góc nghiêng cơ thể

Nếu sử dụng đặc trưng góc nghiêng trục lớn, thì một số trường hợp người ngã theo hướng song song với trục ống kính camera sẽ không được phát hiện. Nhưng đặc trưng này có thể loại bỏ được một số trường hợp nhận diện sai lầm khi có hành động chủ động ngồi nhanh.

CHƯƠNG 4.
THÍ NGHIỆM VÀ ĐÁNH GIÁ

4.1. Tập dữ liệu và phương pháp đánh giá hiệu quả thuật toán

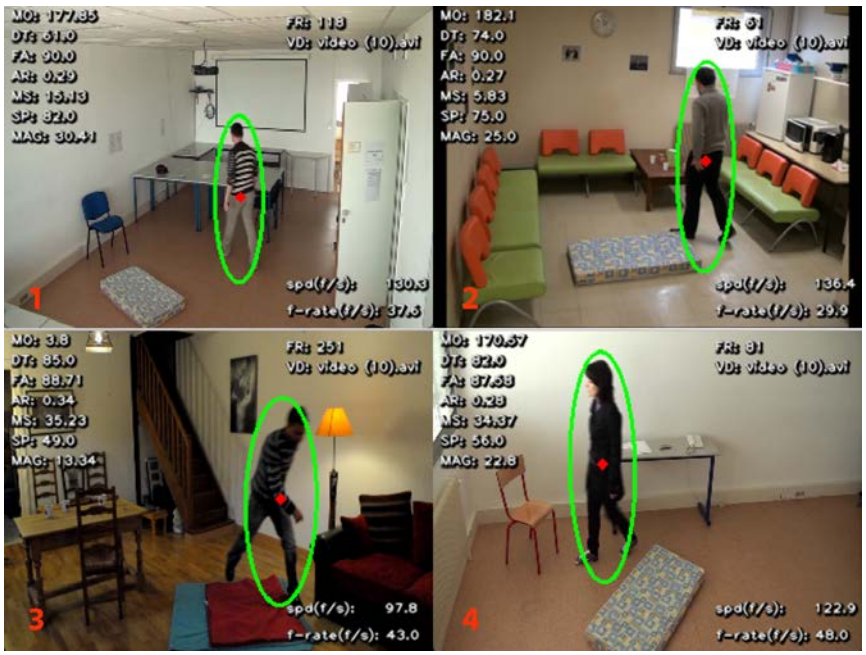
4.1.1. Tập dữ liệu thực nghiệm

Phương thức đề xuất được thiết kế để hoạt động ở môi trường trong nhà với một camera treo tường có góc nhìn cố định hướng xuống. Để đáp ứng hoạt động thời gian thực, hình ảnh thu được từ camera được điều chỉnh độ phân giải là 320x240 mà không làm ảnh hưởng đến kết quả tính toán. Tập dữ liệu thực nghiệm cho các nghiên cứu phát hiện ngã do phòng thí nghiệm Le2i cung cấp [28] bao gồm 221 videos, được quay ở môi trường trong nhà, kích thước 320x240 với tốc độ khung hình 25fps, rất phù hợp với môi trường mà phương thức này hướng đến. Mặt khác, tập dữ liệu bao gồm nhiều video đã mô phỏng được những khó khăn chính có thể gặp phải đối với môi trường thực tế trong nhà của người già. Các video có điều kiện chiếu sáng đa dạng, cũng như có nhiều vùng đổ bóng và vùng phản xạ ánh sáng, vốn là những yếu tố gây thách thức cho kỹ thuật phát hiện đội tượng chuyển động. Mặt khác, hậu cảnh trong các chuỗi video là rất lộn xộn và có kết cấu phức tạp. Những người thực hiện mô phỏng hành động ngã hay các hoạt động thường ngày trong bộ dữ liệu mặc quần áo đa dạng với nhiều màu sắc, hoạ tiết khác nhau, thực hiện rất nhiều các hoạt động thông thường như đi lại theo nhiều hướng, ngồi xuống, đứng lên, cúi người, lau nhà, v.v.. Các hành động ngã cũng phong phú như ngã do vấp, ngã do mất thăng bằng, ngã sấp, ngã ngửa, ngã trong trạng thái không kháng cự, ngã có kháng cự, ngã từ trên ghế, v.v.. Có thể nói tập dữ liệu thử nghiệm đã đáp ứng được việc cung cấp các thách thức đa dạng cho hệ thống phát hiện ngã, mô phỏng gần như mọi hoàn cảnh thực tế trong môi trường mà bài toán của luận văn này hướng đến. Vì những lý do đó, bộ dữ liệu này được sử dụng để tiến hành thực nghiệm cho giải thuật phát hiện ngã của luận văn.

Bộ dữ liệu thực nghiệm bao gồm 4 phần riêng biệt, tương ứng với 4 môi trường hoàn toàn khác nhau. Bảng 4.1 dưới sẽ đây mô tả về 4 tập đó.

Bảng 4.1. Bảng mô tả các tập dữ liệu thực nghiệm

Datasets	Videos	Falls
Lecture-room	27	14
Coffee-room	70	62
Home	60	33
Office	64	17



Hình 4.1. Một số hình ảnh của tập dữ liệu thực nghiệm. Ảnh 1: Lecture-room; ảnh 2: Coffee-room; ảnh 3: Home; ảnh 4: Office

4.1.2. Phương pháp đánh giá độ hiệu quả của giải thuật

Tính hiệu quả của giải thuật được đánh giá bởi chỉ số khôi phục (Recall), tỉ lệ chính xác (Precision), và hệ số điều hòa F-measure. Trong đó:

- $Recall = \frac{TP}{RE}$
- $Precision = \frac{TP}{TP+FN}$
- $F - measure = \frac{2*Precision*Recall}{Precision+Recall}$

Với:

- TP – True Positives: Tổng số hành động ngã được phát hiện chính xác
- FP – False Positives: Tổng số hành động thông thường bị nhận diện nhầm là ngã
- FN – False Negatives: Tổng số hành động ngã không được phát hiện
- RE – Relevant Elements: Tổng số hành động ngã thực tế

Một hệ thống hoạt động hoàn hảo khi nó phát hiện tất cả các trường hợp ngã trong khi không có trường hợp nhầm lẫn nào, tương ứng với chỉ số khôi phục bằng 100% và độ chính xác cũng đạt 100%. Tuy nhiên thực tế rất khó để đạt được kết quả đó. Trong trường hợp điều chỉnh ngưỡng theo chiều hướng ít khắt khe hơn, độ khôi phục sẽ được cải thiện nhưng đồng thời làm giảm độ chính xác. Ở chiều ngược lại, nếu tăng mức độ khắt khe thì có thể làm tăng độ chính xác nhưng sẽ dẫn đến giảm độ khôi phục. Việc sử dụng hệ số điều hoà được tính từ tỉ lệ khôi phục và tỉ lệ chính xác cho phép đánh giá hiệu quả của phương thức một cách hợp lý với sự hài hoà giữa hai tiêu chí trên.

4.2. Cài đặt thí nghiệm

Giải thuật của phương thức đề xuất trong luận văn này được cài đặt bằng ngôn ngữ Python với thư viện OpenCV trên môi trường MacOS. Hệ thống dùng để thử nghiệm là máy tính gồm CPU Core i5 2.9MHz, 8GB RAM, 512GB SSD.

Để xác thực hiệu quả của các kỹ thuật đề xuất, tác giả thực hiện 4 thí nghiệm riêng biệt, trong đó 3 thí nghiệm đầu tiên lần lượt không áp dụng một kỹ thuật được giới thiệu. Ở thí nghiệm cuối cùng, tất cả các kỹ thuật được sử dụng. Cụ thể như sau:

- Thí nghiệm thứ nhất: Không sử dụng MHI trong đo lường hình dáng cơ thể mà đã được mô tả trong phần 3.4.2
- Thí nghiệm thứ 2: Sử dụng phương thức tính hướng gradient thông thường được đề xuất tại [33] thay cho phương thức giới thiệu tại phần 1) của mục 3.3.1
- Không áp dụng thông tin chuyển động trọng tâm để hạn chế lỗi trong ước lượng độ lớn chuyển động sử dụng MHI, mô tả tại phần 2) của mục 3.3.1
- Áp dụng tất cả các kỹ thuật được đề xuất.

4.3. Kết quả và thảo luận

Tác giả sử dụng 4 tập dữ liệu thử nghiệm khác nhau, với 4 môi trường khác biệt rõ rệt: Góc camera và chiều cao đặt camera là khác nhau; khoảng cách từ camera đến vùng phát hiện ngã khác nhau; điều kiện chiếu sáng khác nhau; v.v.. Vì vậy, với mỗi bộ dữ liệu cần đặt các ngưỡng giá trị cho các đặc trưng khác nhau. Các ngưỡng này được xác định thông qua thực nghiệm để đạt được kết quả tốt nhất.

Bảng 4.2 dưới đây trình bày các kết quả thí nghiệm đối với từng tập dữ liệu. Các dòng từ 1 đến 4 tương ứng với các thí nghiệm 1 đến thí nghiệm 4 đã được mô tả ở phần trước của chương này.

Bảng 4.2. Kết quả thực nghiệm

Datasets	Thí nghiệm	Recall(%)	Precision(%)	F-measure
Lecture-room	1	92.86	81.25	0.867
	2	92.86	92.86	0.929
	3	100	82.35	0.903
	4	**100**	**93.33**	**0.965**
Coffee-room	1	88.71	94.83	0.917
	2	82.26	100	0.903
	3	70.97	100	0.83
	4	**90.32**	**94.92**	**0.926**
Home	1	87.88	93.55	0.906
	2	87.88	100	0.935
	3	72.73	96	0.828
	4	**93.94**	**96.88**	**0.954**

Office	1	82.35	77.78	0.8
	2	64.71	73.33	0.688
	3	52.94	52.94	0.529
	4	**82.35**	**87.50**	**0.848**

Từ bảng kết quả có thể nhận thấy, ở thí nghiệm cuối cùng khi áp dụng tất cả các kỹ thuật đã được đề xuất, tỉ lệ khôi phục (Recall) là cao hơn hẳn. Tuy nhiên tỉ lệ chính xác trong hai trường hợp là thấp hơn chút ít. Cụ thể là thí nghiệm 4 so với thí nghiệm 2 và 3 cho tập dữ liệu Coffee-room; thí nghiệm 4 so với thí nghiệm 2 đối với tập dữ liệu Home. Tuy nhiên trong ngữ cảnh bài toán phát hiện ngã, tỉ lệ khôi phục thường là quan trọng hơn nếu tỉ lệ chính xác là không quá khác biệt. Vì vậy để đánh giá hiệu quả của giải thuật, cần căn cứ vào hệ số điều hòa F-measure. Hệ số này thể hiện mối tương quan giữa tỉ lệ khôi phục và tỉ lệ chính xác của kết quả. Trong các thí nghiệm, thí nghiệm thứ 4 luôn cho giá trị F-measure cao hơn hẳn các thí nghiệm khác. Ngoài ra F-measure cũng đạt giá trị rất cao, lần lượt là 0.965 cho tập Lecture-room; 0.926 cho tập Coffee-room; 0.954 cho tập Home và 0.848 đối với tập Office.

Ngoài ra có thể thấy kết quả thí nghiệm đối với tập Lecture-room là cao nhất trong các tập dữ liệu thử nghiệm, còn tập Office cho kết quả thấp nhất. Điều này là do ở tập Lecture-room, camera được đặt ở vị trí phù hợp khi đủ xa khu vực người di chuyển, dẫn đến tránh được hiện tượng tự động đo sáng lại của camera. Cộng với việc môi trường ở tập này không có chứa các nguồn sáng phức tạp như cửa sổ, khiến cho giải thuật phát hiện chuyển động không gặp nhiều khó khăn. Ngược lại, môi trường trong tập Office có cửa sổ nên khi người di chuyển che khuất một phần cửa sổ sẽ làm thay đổi đột ngột điều kiện chiếu sáng của khung cảnh. Camera cũng được đặt thấp gần như ngang người. Khi người di chuyển lại gần camera khiến camera đo sáng lại, gây ra sự thay đổi độ sáng trên khắp khung hình. Những điều này gây rất nhiều khó khăn cho giải thuật trừ nền được áp dụng.

Tóm lại, các trường hợp phát hiện lỗi là do nhiễu tạo ra trong quá trình phân tách vùng chuyển động vì các lý do về thay đổi điều kiện chiếu sáng đột ngột. Quan sát trong quá trình thí nghiệm, một số lỗi xuất hiện vì có các hoạt động thường ngày có đặc điểm rất giống với hành động ngã, ví dụ như ngồi xuống dứt khoát; chủ động nằm với tốc độ nhanh.

Thời gian xử lý

Với đặc điểm của bài toán phát hiện hành động thời gian thực, yêu cầu tính toán của hệ thống phải đảm bảo năng lực xử lý tối thiểu là 10fps. Với hệ thống thực hiện thí nghiệm đã mô tả ở phần trước, tốc độ xử lý của giải thuật là xấp xỉ 90fps, đáp ứng rất tốt yêu cầu hoạt động thời gian thực của bài toán.

CHƯƠNG 5.
KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Qua quá trình thực hiện luận văn, tác giả đã tiến hành tìm hiểu lý thuyết tổng quan về lĩnh vực xử lý ảnh số và thị giác máy, có cơ hội tiếp cận với một số giải thuật học máy. Tác giả cũng đi sâu vào tìm hiểu các giải thuật thị giác máy quan trọng như giải thuật trừ nền, giải thuật trích chọn một số đặc trưng quan trọng như góc, điểm bất biến, đặc trưng SIFT, SUFT, v.v.., có hiểu biết cơ bản về các dạng bài toán thuộc ngành thị giác máy, qua đó làm nền tảng cho quá trình học tập nghiên cứu tiếp theo của bản thân trong lĩnh vực này.

Về mặt thực tiễn, luận văn này đã giới thiệu một phương thức tự động phát hiện ngã dựa trên việc kết hợp các đặc trưng chuyển động như hướng và độ lớn, với các đặc trưng về sự thay đổi hình dáng cơ thể. Luận văn cũng đã đề xuất một số cải tiến trong giải thuật MHI, đề xuất sử dụng đặc điểm chuyển động trọng tâm cơ thể để cải thiện kết quả đo lường chuyển động dựa trên MHI. Điểm mấu chốt của phương thức được đề xuất là là thông qua phân tích đặc điểm thực tế của quá trình ngã dựa trên quan sát kỹ lưỡng, từ đó có thể khai thác hợp lý các đặc trưng đã trích rút được. Việc đưa ra dự đoán về việc ngã không dựa trên giá trị của các đặc trưng trong cùng một thời điểm, mà dựa trên quan sát giá trị các đặc trưng này trên toàn bộ khoảng thời gian tương ứng với hành động ngã, từ khi bắt đầu xuất hiện chuyển động nhanh bất thường đến khi không xuất hiện chuyển động sau ngã.

Các ngưỡng được xác định thủ công dựa trên suy luận từ các đặc điểm của việc ngã và quá trình quan sát các video thử nghiệm. Với việc lựa chọn tập dữ liệu thực nghiệm với nhiều môi trường khác nhau, điều kiện ánh sáng khác nhau, vị trí và góc độ camera được đặt khác nhau, kịch bản ngã phong phú và được xen giữa bởi các hoạt động thông thường hằng ngày, kết quả đạt được của luận văn là hết sức khả quan.

Các trường hợp nhận diện sai lầm chủ yếu là do nhiễu, do thay đổi ánh sáng đột ngột hoặc do người di chuyển quá gần ống kính camera, khiến kích hoạt chức năng tự động đo sáng của camera, ảnh hưởng đến giải thuật phân tách vùng chuyển động. Một số trường hợp nhận diện nhầm các hành động như nằm, ngồi dứt khoát. Để giải quyết các vấn đề trên, tác giả dự kiến tìm hiểu các giải pháp trừ nền phù hợp hơn nữa, giúp loại trừ trường hợp camera điều chỉnh độ sáng, bổ xung các kỹ thuật phát hiện vùng đầu người (head detection) và kỹ thuật giới hạn vùng quan tâm (inactivity zone) trong các nghiên cứu tiếp theo. Ngoài ra, để mở rộng phạm vi của bài toán trong trường hợp bối cảnh có nhiều hơn một người, tác giả dự định tìm hiểu các kỹ thuật theo vết đối tượng (object tracking) cho việc cải tiến phương thức đã đề xuất.

Kết quả nghiên cứu và kỹ thuật được đề xuất đồng thời cũng được trình bày trong bài báo [51] gửi hội thảo quốc tế SoICT và đã được chấp nhận.

Danh mục công trình khoa học của tác giả liên quan đến luận văn

1. Viet Anh Nguyen, Thanh Ha Le and Thuy Thi Nguyen. Single camera based Fall detection using Motion and Human shape Features. In *The Seventh International Symposium on Information and Communication Technology (SoICT 2016)*, đã được chấp nhận đăng trong kỷ yếu và trình bày tại hội thảo.

TÀI LIỆU THAM KHẢO

Tài liệu tiếng Việt

[1] Bộ Y tế, "Ngày quốc tế người cao tuổi (IDOP) 2015: Già hóa dân số do nâng cao chất lượng cuộc sống," 1 tháng 10 2015. Bản điện tử: http://moh.gov.vn:8086/news/Pages/ChuongTrinhMucTieuQuocGiaYTe.aspx?ItemID=4110. [Truy cập 3 tháng 5, 2016]

[2] Quỹ Dân số Liên hợp quốc, "Già hóa dân số và người cao tuổi ở Việt Nam: Thực trạng, dự báo và một số khuyến nghị chính sách," Hà Nội, 2011.

Tài liệu tiếng Anh

[3] A. Bourke, J. O'Brien, G. Lyons: Evaluation of a threshold-based tri-axial accelerometer fall detection algorithm. *Gait Posture* 2007, 26:194–199.

[4] A. F. Bobick and J. W. Davis. The recognition of human movement using temporal templates. *IEEE transactions on pattern analysis and machine intelligence*, 23(3):257– 267, March 2001.

[5] A. H. Nasution and S. Emmanuel. Intelligent video surveillance for monitoring elderly in home environments. *In Multimedia Signal Processing, 2007. MMSP 2007. IEEE 9th Workshop* on pages 203–206. IEEE, 2007.

[6] A. Leone, G. Diraco, and P. Siciliano. Detecting falls with 3d range camera in ambient assisted living applications: A preliminary study. *Medical Engineering & Physics*, 33:770–781, July 2011.

[7] A. W. Fitzgibbon, M. Pilu, and R. B. Fisher. Direct least squares fitting of ellipses. *Technical Report DAIRP-794*, January 1996.

[8] A. Zweng, S. Zambanini, and M. Kampel. Introducing a statistical behavior model into camera-based fall detection. *In International conference on Advances in visual computing,* pages 163–172, 2010.

[9] B. Mirmahboub, S. Samavi, N. Karimi, and S. Shirani. Automatic monocular system for human fall detection based on variations in silhouette area. *IEEE Transactions on Biomedical Engineering*, 60:427–436, February 2013.

[10] C. Doukas, I. Maglogiannis, F. Tragkas, D. Liapis, G. Yovanof: Patient Fall Detection using Support Vector Machines. *Int Fed Inf Process* 2007, 247:147–156.

[11] C. F. Lai, S. Y. Chang, H. C. Chao, Y. M. Huang: Detection of cognitive injured body region using multiple triaxial accelerometers for elderly falling. *IEEE Sensors* J 2011, 11:763–770.

[12] C. Rougier, E. Auvinet, J. Rousseau, M. Mignotte, and J. Meunier. Fall detection from depth map video sequences. *In International Conference on Smart Homes and Health Telematics*, pages 121–128, June 2011.

[13] C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau. Monocular 3d head tracking to detect falls of elderly people. In *Engineering in Medicine and Biology Society*, 2006. EMBS'06. 28th Annual International Conference of the IEEE, pages 6384–6387. IEEE, 2006.

[14] C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau. Robust video surveillance for fall detection based on human shape deformation. *IEEE Transactions on Circuits and Systems for Video Technology*, 21:611–622, May 2011.

[15] C. Stauffer and W.E.L. Grimson. Adaptive background mixture models for real-time tracking. *International conference on Computer Vision and Pattern Recognition*, 2, 1999.

[16] D. Anderson, R. H. Luke, J. M. Keller, M. Skubic, M. J. Rantz, and M. A. Aud. Modeling human activity from voxel person using fuzzy logic. *Fuzzy Systems, IEEE Transactions* on, 17(1):39–49, 2009

[17] D. Anderson, R. Luke, J. Keller, M. Skubic, M. Rantz, and M. Aud. Linguistic summarization of video for fall detection using voxel person and fuzzy logic. *Computer Vision and Image Understanding*, 113(1):80–89, Jan 2009.

[18] E. Auvinet, F. Multon, A. St-Arnaud, J. Rousseau, and J. Meunier. Fall detection using body volume reconstruction and vertical repartition analysis. In *International conference on Image and signal processing*, pages 376–383, 2010.

[19] E. Auvinet, F. Multon, A. St-Arnaud, J. Rousseau, and J. Meunier. Fall detection with multiple cameras: An occlusion-resistant method based on 3-d silhouette vertical distribution. *IEEE Transactions on Information Technology in Biomedicine*, 15:290–300, March 2011.

[20] E. Auvinet, L. Reveret, A. St-Arnaud, J. Rousseau, and J. Meunier. Fall detection using multiple cameras. In *Engineering in Medicine and Biology Society,* 2008. EMBS 2008. 30th Annual International Conference of the IEEE, pages 2554–2557. IEEE, 2008.

[21] F. Bagalà, C. Becker, A. Cappello, L. Chiari, K. Aminian and H. Jeffrey, "Evaluation of Accelerometer-Based Fall Detection Algorithms on Real-World Falls," vol. 7, no. 5, May 2012.

[22] F. Bianchi, S. J. Redmond, M. R. Narayanan, S. Cerutti, N. H. Lovell: Barometric pressure and triaxial accelerometry-based falls event detection. *IEEE Trans Neural Syst Rehabil* Eng 2010, 18:619–627.

[23] F. Sposaro, G. Tyson: iFall: an Android application for fall monitoring and response. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Minneapolis: Institute of Electrical and Electronics Engineers; 2009:6119–6122. doi:10.1109/IEMBS.2009.5334912.

[24] G. Diraco, A. Leone, and P. Siciliano. An active vision system for fall detection and posture recognition in elderly healthcare. In *Design, Automation & Test in Europe Conference & Exhibition*, pages 1536–1541, March 2010.

[25] H. Foroughi, A. Naseri, A. Saberi, and H. S. Yazdi. An eigenspace-based approach for human fall detection using integrated time motion image and neural network. In *Signal Processing*, 2008. ICSP 2008. 9th International Conference on, pages 1499–1503. IEEE, 2008.

[26] H. Kerdegari, K. Samsudin, A. R. Ramli, S. Mokaram: Evaluation of fall detection classification approaches. In *Proceedings of the 4th International Conference on Intelligent and Advanced Systems*. Kuala Lumpur: Institute of Electrical and Electronics Engineers; 2012:131–136. doi:10.1109/ICIAS.2012.6306174.

[27] I. C. Lopes, B. Vaidya, J. Rodrigues: Towards an autonomous fall detection and alerting system on a mobile and pervasive environment. *Telecommun Syst* 2011, 48:1–12.

[28] I. Charfi, J. Miteran, J. Dubois, M. Atri, and R. Tourki. Definition and performance evaluation of a robust svm based fall detection solution. *SITIS'12*, page 218–224, 2012.

[29] J. Chen, K. Kwong, D. Chang, J. Luk and R. Bajcsy, "Wearable Sensors for Reliable Fall Detection," in *Proceedings of the IEEE Engineering in Medicine and Biology 27th Annual Conference*.

[30] J. Cheng, X. Chen, M. Shen: A framework for daily activity monitoring and fall detection based on surface electromyography and accelerometer signals. *IEEE J Biomed and Health Inform* 2013, 17(1):38–45.

[31] J. Dai, X. Bai, Z. Yang, Z. Shen, D. Xuan: Mobile phone-based pervasive fall detection. *Pers Ubiquitous Comput* 2010, 14:633–643.

[32] J. Tao, M. Turjo, M. Wong, M. Wang, and Y. Tan. Fall incidents detection for intelligent video surveillance. In *IEEE Conference on Information, Communications and Signal Processing*, pages 1590–1594, 2005.

[33] J. W. Davis. Recognizing movement using motion histograms. Technical report, 1999.

[34] K. Kim, T.H. Chalidabhongse, D. Harwood, and L. Davis. Real-time foreground-background segmentation using codebook model. *Real-Time Imaging*, 11:172–185, 2005

[35] K. Tra and T. Pham. Human fall detection based on adaptive background mixture model and hmm. In *International Conference on Advanced Technologies for Communications*, pages 95–100, Oct 2013.

[36] M. Kangas, A. Konttila, P. Lindgren, I. Winblad, T. Jms: Comparison of low-complexity fall detection algorithms for body attached accelerometers. *Gait Posture* 2008, 28:285–291.

[37] M. Kangas, I. Vikman, J. Wiklander, P. Lindgren, L. Nyberg, T. Jämsä: Sensitivity and specificity of fall detection in people aged 40 years and over. *Gait Posture* 2009, 29:571–574.

[38] M. Kepski and B. Kwolek. Fall detection using ceiling-mounted 3d depth camera. In *International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, pages 640–647, 2014.

[39] M. V. Albert, K. Kording, M. Herrmann, A. Jayaraman: Fall classification by machine learning using mobile phones. *PLoS One* 2012, 7:e36556.

[40] M. Yuwono, B. Moulton, S. Su, B. Celler, H. Nguyen: Unsupervised machine-learning method for improving the performance of ambulatory fall-detection systems. *Biomed Eng Online* 2012, 11:1–11.

[41] N. Thome, S. Miguet, and S. Ambellouis. A real-time, multiview fall detection system: A lhmm-based approach. Circuits and Systems for Video Technology, *IEEE Transactions* on, 18(11):1522–1532, 2008.

[42] R. Cucchiara, A. Prati, and R. Vezzani. A multi-camera vision system for fall detection and alarm generation. *Expert Systems*, 24(5):334–345, Nov 2007.

[43] R. M. Haralick and L. G. Shapiro. *Computer and Robot Vision*, volume 1. Addition Wesley, 1993.

[44] R. Y. W. Lee, A. J. Carlisle: Detection of falls using accelerometers and mobile phone technology. *Age Ageing* 2011, 0:1–7.

[45] S. Abbate, M. Avvenuti, F. Bonatesta, G. Cola, P. Corsini, A. Vecchio: A smartphone-based fall detection system. *Pervasive Mob Comput* 2012, 8:883–899.

[46] S. H. Fang, Y. C. Liang, K. M. Chiu: Developing a mobile phone-based fall detection system on android platform. In *Proceedings of the Conference on Computing, Communications and Applications*. Hong Kong: Institute of Electrical and Electronics Engineers; 2012:143–146. doi:10.1109/ComComAp.2012.6154019.

[47] S. Miaou, P. Sung, and C. Huang. A customized human fall detection system using omni-camera images and personal information. In *Transdisciplinary Conference on Distributed Diagnosis and Home Healthcare*, pages 39–42, Apr 2006.

[48] S. Shan, T. Yuan: A wearable pre-impact fall detector using feature selection and support vector machine. In *Proceedings of the IEEE 10th International Conference on Signal Processing*. Beijing: Institute of Electrical and Electronics Engineers; 2010:1686–1689. doi:10.1109/ICOSP.2010.5656840.

[49] T. Zhang, J. Wang, L. Xu, P. Liu: Fall detection by wearable sensor and one-class SVM algorithm. In *Lecture Notes in Control and Information Science*, Volume 345. Edited by Huang DS, Li K, Irwin GW. Berlin Heidelberg: Springer; 2006:858–863.

[50] U. Lindemann, A. Hock, M. Stuber, W. Keck and C. Becker, "Evaluation of a fall detector based on accelerometers: A pilot study," *Medical and Biological Engineering and Computing*, vol. 43, no. 5, pp. 548-551, 2005.

[51] Viet Anh Nguyen, Thanh Ha Le and Thuy Thi Nguyen. Single camera based Fall detection using Motion and Human shape Features. In *The Seventh International Symposium on Information and Communication Technology (SoICT 2016)*, accepted.

[52] W. Feng, R. Liu, and M. Zhu. Fall detection for elderly person care in a vision-based home surveillance environment using a monocular camera. *Signal, Image and Video Processing*, 8(6):1129–1138, 2014.

[53] World Health Organization, WHO Global Report on Falls Prevention in Older Age, 2007.

[54] Y. T. Liao, C.-L. Huang, and S.-C. Hsu. Slip and fall event detection using bayesian belief network. *Pattern recognition,* 45(1):24–32, 2012.

[55] Yannick Benezeth, Pierre-Marc Jodoin, Bruno Emile, Hélène Laurent, Christophe Rosenberger. Comparative study of background subtraction algorithms. *Journal of Electronic Imaging*, Society of Photo-optical Instrumentation Engineers, 2010, 19, <10.1117/1.3456695>. <inria-00545478>

[56] Z. Zhang, E. Becker, R. Arora, and V. Athitsos. Experiments with computer vision methods for fall detection. In *International Conference on Pervasive Technologies Related to Assistive Environments*, pages 25:1–25:4, 2010.