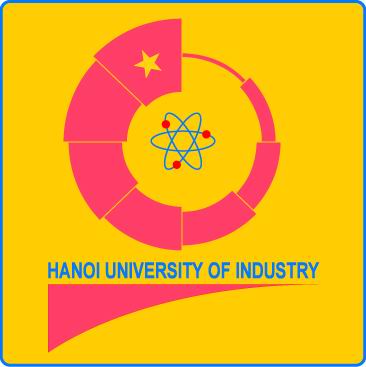
**NHẬN XÉT CỦA GIÁO VIÊN HƯỚNG DẪN**

**BỘ CÔNG THƯƠNG**

**TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP HÀ NỘI**



**ĐỀ TÀI NGHIÊN CỨU KHOA HỌC CẤP TRƯỜNG**

**NGHIÊN CỨU KỸ MỘT SỐ KĨ THUẬT PHÁT HIỆN**

**BẤT THƯỜNG TRONG GIÁM SÁT CAMERA**

Đơn vị thực hiện: Khoa công nghệ thông tin

Giáo viên hướng dẫn: Ths. **Ngô Đức Vĩnh**

Nhóm sinh viên: Nhóm BigData

1. Vũ Tiến Dũng - KTPM2K7
2. Vũ Tuấn Khắc - KTPM2K7
3. Dương Lê Phúc Nguyện-KTPM2K7

*--- Hà Nội, tháng 5 – 2015 ---*

Giáo viên hướng dẫn

**NHẬN XÉT CỦA GIÁO VIÊN PHẢN BIỆN**

` Giáo viên phản biện

**MỤC LỤC**

[LỜI CẢM ƠN 1](#_Toc420049242)

[DANH MỤC CÁC TỪ VIẾT TẮT 2](#_Toc420049243)

[DANH MỤC BẢNG BIỂU 3](#_Toc420049244)

[DANH MỤC HÌNH VẼ 3](#_Toc420049245)

[MỞ ĐẦU 6](#_Toc420049246)

[Chương I 7](#_Toc420049247)

[NỘI DUNG NGHIÊN CỨU 7](#_Toc420049248)

[1. Tổng quan về hệ thống giám sát bằng camera 8](#_Toc420049249)

[*1.1. Giới thiệu chung* 8](#_Toc420049250)

[*1.2. Ứng dụng* 11](#_Toc420049251)

[*1.3. Các bài toán cần giải quyết* 11](#_Toc420049252)

[2. Tổng quan về một hệ thống xử lý ảnh 12](#_Toc420049253)

[*2.1.* *Khái niệm cơ bản* 12](#_Toc420049254)

[*2.2.* *Cấu trúc của hệ thống xử lý ảnh* 13](#_Toc420049255)

[*2.3 Các vấn đề cơ bản trong xử lý ảnh* 14](#_Toc420049256)

[*2.4 Một số thuật toán xử lý ảnh thông dụng* 18](#_Toc420049257)

[3. Tổng quan về video 21](#_Toc420049258)

[*3.1.Cấu trúc của video* 21](#_Toc420049259)

[*3.2. Các đặc trưng của video* 22](#_Toc420049260)

[4. Một số kĩ thuật phát hiện chuyển động và theo dõi đa đối tượng 24](#_Toc420049261)

[*4.1. Các kĩ thuật phát hiện đối tượng chuyển động* 24](#_Toc420049262)

[*4.2. Một số hàm lọc nhiễu* 27](#_Toc420049263)

[*4.3. Các kĩ thuật theo dõi đa đối tượng chuyển động* 30](#_Toc420049264)

[4.4. Kĩ thuật khoanh vùng quan sát 34](#_Toc420049265)

[Chương II 36](#_Toc420049266)

[PHÁT HIỆN BẤT THƯỜNG TRONG GIÁM SÁT CAMERA 36](#_Toc420049267)

[1. Tổng quan 37](#_Toc420049268)

[*1.1. Phát biểu bài toán* 37](#_Toc420049269)

[*1.2. Mô hình của hệ thống* 37](#_Toc420049270)

[2. Các kĩ thuật triển khai cho bài toán phát hiện và theo vết đa đối tượng chuyển động 38](#_Toc420049271)

[*2.1. Thuật toán trừ ảnh và phát hiện đối tượng chuyển động* 38](#_Toc420049272)

[*2.3. Kĩ thuật theo vết đa đối tượng chuyển động – Thuật toán CamShift* 44](#_Toc420049273)

[3. Bài toán phát hiện té ngã 50](#_Toc420049274)

[*3.1. Đề xuất hướng tiếp cận dựa trên phân tích quỹ đạo chuyển động và phân lớp* 50](#_Toc420049275)

[*3.2. Phát hiện té ngã* 51](#_Toc420049276)

[*3.2. Học bán giám sát và kĩ thuật phân lớp SVM* 55](#_Toc420049277)

[4. Đánh giá các kĩ thuật áp dụng 60](#_Toc420049278)

[Chương III 61](#_Toc420049279)

[THỰC NGHIỆM 61](#_Toc420049280)

[1. Điều kiện thực nghiệm 62](#_Toc420049281)

[*1.1 Môi trường* 62](#_Toc420049282)

[*1.2 Thiết bị* 62](#_Toc420049283)

[*1.3 Cấu hình máy* 62](#_Toc420049284)

[*1.4 Khoảng cách thực nghiệm* 62](#_Toc420049285)

[2. Thực nghiệm 62](#_Toc420049286)

[*2.1. Phương pháp thực nghiệm* 62](#_Toc420049287)

[*2.2. Kết quả thực nghiệm* 63](#_Toc420049288)

[KẾT LUẬN & HƯỚNG PHÁT TRIỂN 65](#_Toc420049289)

[TÀI LIỆU THAM KHẢO 67](#_Toc420049290)

# LỜI CẢM ƠN

Nhiều năm trở lại đây, đi cùng với sự phát triển của kỷ nguyên công nghệ thông tin, kỹ thuật xử lý ảnh và thị giác máy tính đã có những vị trí nhất định trong sản xuất công nghiệp, quốc phòng, đời sống hàng ngày của chúng ta.

Sự tăng trưởng nhanh chóng của dân số kéo theo sự phức tạp của cuộc sống xã hội. Nhiều tệ nạn như trộm cắp, cướp giật thường xuyên xảy ra, cùng với đó là những hành vi phạm luật giao thông như vượt đèn đỏ, đi vào vùng cấm… dẫn đến những vụ tai nạn giao thông khảm khốc. Những ảnh hưởng tiêu cực ảnh hưởng đến nhiều khía cạnh của xã hội, đòi hỏi chúng ta phải hoạch định các chính sách, có những chế tài hợp lý, hoặc là xây dựng một hệ thống có thể cảnh báo những hành vi bất thường.

Dựa vào những nhận định trên, chúng tôi nhận thấy lĩnh vực phát hiện bất thường trong giám sát là một sân chơi mở dành cho những nghiên cứu có giá trị. Vì vậy, dưới sự chỉ dẫn của Ths Ngô Đức Vĩnh, chúng tôi quyết định lựa chọn đề tài “*Nghiên cứu một số kĩ thuật phát hiện bất thường trong giám sát camera*” làm đề tài nghiên cứu khoa học của mình, với hy vọng có thể đóng góp những đánh giá – nhận định có giá trị góp phần làm phong phú hơn nữa lĩnh vực nghiên cứu này.

Chúng tôi xin chân thành cảm ơn Thạc sĩ Ngô Đức Vĩnh – trưởng khoa Công nghệ thông tin trường ĐH Công nghiệp Hà Nội đã giúp chúng tôi định hướng đề tài, lựa chọn hướng nghiên cứu, đánh giá và động viên chúng tôi trong suốt thời gian thực hiện. Xin cảm ơn thầy Nguyễn Mạnh Cường, thầy Trần Thanh Hùng đã cho chúng tôi những lời khuyên bổ ích và trang bị thêm một số kiến thức về học máy và nhận dạng.

Sau cùng xin gửi lời cảm ơn chân thành tới gia đình, bạn bè đã động viên, đóng góp ý kiến và giúp đỡ trong quá trình học tập, nghiên cứu và hoàn thành đề tài nghiên cứu khoa học này.

# DANH MỤC CÁC TỪ VIẾT TẮT

|  |  |  |
| --- | --- | --- |
| **STT** | **TỪ VIẾT TẮT** | **ĐẦY ĐỦ** |
|  | PIXEL | PICTURE ELEMENT |
|  | RGB | RED -GREEN -BLUE |
|  | KF | KALMAN FILTER |
|  | EKF | EXTENDED KALMAN FILTER |
|  | MS | MEANSHIFT |
|  | CAMSHIFT | CONTINUOUSLY ADAPTIVE MEAN SHIFT |
|  | PCA | PRINCIPAL COMPONENT ANALYSIS |
|  | KNN | K - NEAREST NEIGHBORS. |
|  | OPENCV | OPEN COMPUTER VISION LIBRARY |
|  | CCTV | CLOSED CIRCUILT TELEVISION |
|  | KKT | KARUSH – KUHN – TUCKER |
|  | SVM | SUPPORT VECTOR MACHINE |

# DANH MỤC BẢNG BIỂU

[*Bảng I. 1 Xác suất ghép cặp đối tượng* 37](#_Toc419112211)

[*Bảng I. 2 Bảng kết quả ghép cặp* 37](#_Toc419112212)

# DANH MỤC HÌNH VẼ

[*Hình I. 1 Mô hình của một hệ thống giám sát thông minh* 9](#_Toc419994268)

[*Hình I. 2 Quá trình xử lý ảnh* 13](#_Toc419994269)

[*Hình I. 3 Các giai đoạn chính trong xử lý ảnh* 13](#_Toc419994270)

[*Hình I. 5 Biểu diễn ảnh với độ phân giải khác nhau* 14](#_Toc419994271)

[*Hình I. 6 Ví dụ về ảnh xám* 15](#_Toc419994272)

[*Hình I. 7 Ví dụ về ảnh nhị phân* 16](#_Toc419994273)

[*Hình I. 8 Ví dụ về ảnh màu* 16](#_Toc419994274)

[*Hình I. 9 Biểu diễn ảnh* 17](#_Toc419994275)

[*Hình I. 10 Ảnh với các hệ màu khác nhau* 17](file:///D:\1.HOC%20TAP\NCKH\BC-NCKH-Mot-so-ki-thuat-phat-hien-bat-thuong-trong-giam-sat-camera.docx#_Toc419994276)

[*Hình I. 11 Ví dụ về chuyển về ảnh xám bằng ma trận ba màu* 18](#_Toc419994277)

[*Hình I. 12 Ảnh thực tế sau khi chuyển đổi* 19](#_Toc419994278)

[*Hình I. 13 Ví dụ về khử nhiễu* 20](#_Toc419994279)

[*Hình I. 14 Hình ảnh sau khi khử nhiễu* 20](#_Toc419994280)

[*Hình I. 15 Ảnh sau khi giãn* 21](#_Toc419994281)

[*Hình I. 16 Cấu trúc phân đoạn video* 22](#_Toc419994282)

[*Hình I. 17 Các ảnh khác nhau nhưng có cùng biểu đồ* 22](#_Toc419994283)

[*Hình I. 18 Ví dụ về bài toán phát hiện khuôn mặt* 25](#_Toc419994284)

[*Hình I. 19 Ảnh sau khi dùng kĩ thuật trừ nền* 27](#_Toc419994285)

[*Hình I. 20 Biểu đồ mô tả hình thức của bộ lọc nói chung* 28](#_Toc419994286)

[*Hình I. 21 Ứng dụng lọc Kalman để lọc tín hiệu thu được từ toạ độ con trỏ của người dùng* 28](file:///D:\1.HOC%20TAP\NCKH\BC-NCKH-Mot-so-ki-thuat-phat-hien-bat-thuong-trong-giam-sat-camera.docx#_Toc419994287)

[*Hình I. 22 Bộ lọc Kalman bị che khuất* 29](#_Toc419994288)

[*Hình I. 23 Ví dụ sử dụng EKF của J. Hartikainen, A. Solin, S. Sarkka - Aalto University, August 2011* 30](#_Toc419994289)

[*Hình I. 24 Một số khó khăn khi theo vết đối tượng* 32](#_Toc419994290)

[*Hình II. 1 Mô hình hệ thống phát hiện bất thường* 37](#_Toc419994291)

[*Hình II. 5 Mô hình chung của kỹ thuật trừ nền* 38](#_Toc419994292)

[*Hình II. 6 Một ví dụ về kỹ thuật trừ ảnh* 38](#_Toc419994293)

[*Hình II. 7 Làm trơn ảnh sử dụng bộ lọc ảnh* 39](#_Toc419994294)

[*Hình II. 9 Lưu đồ thuật toán CamShift* 45](#_Toc419994295)

[*Hình II. 10 Mô hình kết hợp bộ lọc Kalman và thuật toán Camshift* 46](#_Toc419994296)

[*Hình II. 11 Kết quả mô phỏng ước lượng* 48](#_Toc419994297)

[*Hình II. 14 Khi ngã 2 chiều của hình chữ nhật bao quanh đối tượng thay đổi đột ngột* 51](#_Toc419994298)

[*Hình II. 15 Khi ngã góc hợp bởi trục đứng của đối tượng và nền thay đổi đột ngột* 51](#_Toc419994299)

[*Hình II. 16 Trọng tâm của đối tượng hạ thấp khi ngã* 52](#_Toc419994300)

[*Hình II. 17 Hình ảnh đối tượng ngã* 52](#_Toc419994301)

[*Hình II. 3 Mặt siêu phẳng tách các mẫu dương khỏi mẫu âm* 54](#_Toc419994302)

[*Hình III. 1 Các frame chạy thực nghiệm với thuật toán Camshift không dùng bộ lọc Kalman* 62](#_Toc419994303)

[*Hình III. 2 Các frame chạy thực nghiệm với thuật toán Camshift kết hợp bộ lọc Kalman* 62](#_Toc419994304)

# MỞ ĐẦU

Giám sát tự động trong môi trường công cộng là một lĩnh vực nghiên cứu được phát triển nhanh chóng. Nhiều CCTV cần được giám sát một cách tích cực, và phải tự động phát hiện các sự kiện bất thường cho các mục đích an ninh. Trong các môi trường công cộng như sân bay, cửa hàng, nhà ga… sự kiện bất thường là mối quan tâm của nhân viên an ninh khi họ có thể thấy được biểu hiện của mối nguy hiểm hoặc sự gây rối thông qua các cảnh báo.

Ngày nay, sự tăng trưởng nhanh chóng cùng với sự lão hóa nhanh của dân số làm cho cuộc sống ngày trở nên phức tạp hơn. Cụ thể, Việt Nam đứng thứ 13 trong các nước đông dân trên thế giới. Tổng điều tra dân số vào năm 2011 ghi nhận nước ta có[gần 86 triệu dân](http://doisong.vnexpress.net/tin-tuc/gia-dinh/viet-nam-co-gan-86-trieu-nguoi-2270912.html) và tính đến nay đã tăng trưởng lên hơn 90 triệu dân. Mỗi năm, ước tính có khoảng 1,5 – 2 triệu người trên 65 tuổi té ngã và xác suất té ngã đối với họ tăng lên cùng với tuổi tác. Hậu quả của việc té ngã đó là gãy xương hoặc nghiêm trọng hơn.

Các hiện tượng tiêu cực ảnh hưởng đến nhiều khía cạnh của xã hội, thách thức chúng ta phải hoạch định các chính sách, đưa ra các chế tài hợp lý, ví dụ xây dựng các nhà cung cấp chăm sóc y tế để đáp ứng nhu cầu của người cao tuổi. Và vấn đề chính chúng tôi quan tâm là làm thế nào để tạo ra sự an toàn cho người cao tuổi để họ được sống lâu hơn một cách độc lập.

Một điều rõ ràng là chúng ta có thể nhận thấy là ngày nay có rất nhiều đóng góp cho sự phát triển một hệ thống giám sát thông minh, công nghệ cảm biến và mạng lưới camera phát triển nhanh chóng, góp phần không nhỏ vào sự phát triển chăm sóc y tế. Có rất nhiều phương pháp để phát hiện té ngã, ở trong bài báo cáo,với cách tiếp cận dựa vào quỹ đạo chuyển động của đối tượng, chúng tôi đã phát hiện được té ngã với tỉ lệ cao.

# Chương I

# NỘI DUNG NGHIÊN CỨU

**Các nội dung chính:**

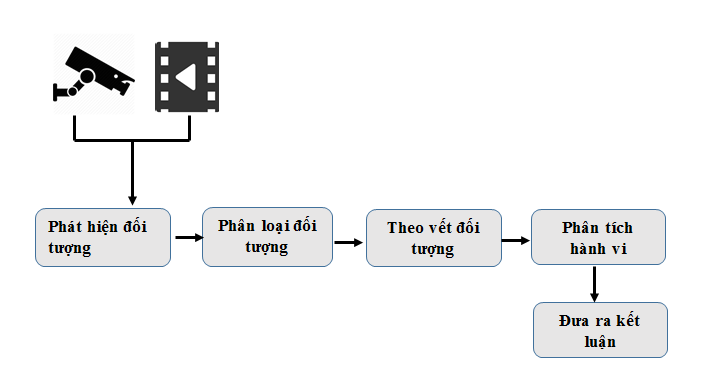
* Hệ thống giám sát thông minh bằng camera.
* Các vấn đề trong xử lý ảnh.
* Tổng quan về video.
* Kĩ thuật phát hiện và theo dõi đa đối tượng. chuyển động

*Trong chương này chúng tôi sẽ giới thiệu về hệ thống giám sát thông minh bằng camera. Đưa ra các khái niệm cơ bản về xử lý ảnh và video, cũng như các vấn đề cần giải quyết trong một hệ thống giám sát camera, qua đó tiến tới đề cập tới các thuật toán liên quan trong bài toán giám sát camera. Từ đó mô hình và giải quyết bài toán “phát hiện bất thường trong giám sát camera”.*

## 1. Tổng quan về hệ thống giám sát bằng camera

### *1.1. Giới thiệu chung*

Giám sát tự động là một hướng được nghiên cứu và phát triển trong lĩnh vực nhận dạng và xử lý ảnh và tạo cách tiếp cận cho phần mềm thiết kế chuyên dụng cho các thiết bị giám sát tự động trong nhiều năm gần đây. Việc phát hiện ra các đối tượng chuyển động trong camera nhờ các kỹ thuật xử lý ảnh đã đoán nhận một số hành vi của đối tượng là một việc làm có ý nghĩa khoa học và thực tiễn. Hệ thống giám sát video từ lâu đã được sử dụng để giám sát an ninh ở các khu vực nhạy cảm. Hệ thống giám sát video đã trải qua ba thế hệ được gọi là 1GSS, 2GSS và 3GSS. Các hệ thống giám sát thế hệ đầu tiên (1GSS, 1960-1980) được dựa trên hệ thống truyền, thu thập và xử lý hình ảnh tương tự. Chúng truyền các hình ảnh thu được từ các camera ở các góc nhìn khác nhau về một trung tâm điều khiển. Chúng có các hạn chế lớn như đòi hỏi băng thông cao, khó lưu trữ và phục hồi các sự kiện do số lượng băng video rất lớn và khó phát hiện các sự kiện tức thời mà phải phụ thuộc vào sự tập trung hạn chế của người điều hành. Các hệ thống giám sát thế hệ tiếp theo (2GSS, 1980-2000) là hybrid. Chúng sử dụng cả hai hệ thống tương tự và kỹ thuật số phụ để giải quyết một số nhược điểm của thế hệ trước đó. Chúng đã sử dụng những tiến bộ mới nhất trong phương pháp xử lý video kỹ thuật số nhằm hỗ trợ cho các nhà điều hành lọc ra những sự kiện giả mạo. Hầu hết các công việc trong 2GSS được tập trung vào việc phát hiện sự kiện trong thời gian thực. Hệ thống giám sát thế hệ thứ ba (3GSS, 2000) cung cấp cho hệ thống kỹ thuật số điểm tới điểm. Hình ảnh thu thập và xử lý ở các cảm biến, thông tin được truyền thông qua các mạng di động băng thông rộng và cố định không đồng nhất và hình ảnh được lưu trữ tại các máy chủ trung tâm. Không giống như các thế hệ trước, trong 3GSS một số phần của quá trình xử lý hình ảnh phân chia cho các cảm biến bằng cách sử dụng các máy ảnh thông minh có thể số hoá và nén các tín hiệu hình ảnh tương tự lại và thực hiện phân tích hình 28 ảnh bằng các thuật toán như chuyển động và nhận diện khuôn mặt với sự giúp đỡ của trực tiếp từ các bộ phận máy tính. Một hệ thống giám sát video "thông minh" đòi hỏi phải nhanh chóng, đáng tin cậy và các thuật toán mạnh mẽ cho việc phát hiện đối tượng chuyển động, phân loại, theo dõi và phân tích hành vi. Bắt đầu từ 2GSS, một số lượng đáng kể của nghiên cứu đã được cống hiến cho sự phát triển của các thuật toán thông minh. Về mô hình của hệ thống, như chúng ta biết kết quả thu nhận từ các camera giám sát hoặc webcam là các frame ảnh, kết quả nghiên cứu chính của bài báo cáo ở đây là việc phát hiện đối tượng chuyển động trong các frame ảnh đó. Frame ảnh thu nhận được từ các camera hoặc webcam sẽ được xử lý qua các công đoạn sau: Phát hiện đối tượng chuyển động, đánh dấu các đối tượng vừa phát hiện, phân loại chúng được tiến hành xử lý và được kết quả là đối tượng đang cần theo vết ở vị trí nào, để tiến hành đánh dấu (tô màu, kẻ khung) và từ đó liên tục bám sát đối tượng theo một ngưỡng nhất định.

****

*Hình I. 1 Mô hình của một hệ thống giám sát thông minh*

Mỗi hệ thống giám sát có các mục đích khác nhau nhưng đa phần chúng đều có chung sơ đồ hệ thống gồm 5 phần chính như trên. Mỗi thành phần đảm nhiệm chức năng như chính tên của nó. Các thành phần sau thì sẽ lấy dữ liệu từ thành phần trước để xử lý. Nghĩa là module sau sẽ lấy dữ liệu từ module trước.

* **Module phát hiện đối tượng:**

- Dữ liệu vào: Chuỗi ảnh, video, camera...

- Dữ liệu ra: Danh sách các đối tượng sau khi phát hiện chuyển động

- Vai trò: Cung cấp dữ liệu cho module Phân loại đối tượng , phát hiện được đối tượng trong ảnh hoặc video có chuyển động hay không

- Phương pháp sử dụng: Trừ nền, lọc Kalman, Optical flow.

* **Module phân loại đối tượng:**

- Dữ liệu vào: Danh sách các đối tượng trong ảnh, video

- Dữ liệu ra: Phân loại đối tượng đó là gì: Con người, xe cộ, con vật...

- Vai trò: Giúp nhận dạng đúng đối tượng đó là gì, căn cứ vào đó kết hợp với mục đích của hệ thống mà đưa ra cách làm tiếp theo.

- Phương pháp sử dụng: Mạng nơ ron, KNN, K-Means.

* **Module theo vết đối tượng:**

- Dữ liệu vào: Danh sách các vùng chứa đối tượng chuyển động

- Dữ liệu ra: Danh sách các đối tượng kèm theo hướng chuyển động của chúng  
- Vai trò: Giám sát sự chuyển động của đối tượng từ đó ghi lại quá trình hoạt động và cung cấp dữ liệu cho module sau để xem đối tượng đó đang làm gì.

- Phương pháp sử dụng: Mean Shift, Camshift, Optical flow, Kalman..

* **Module phân tích hành vi:**

- Dữ liệu vào: Danh sách các đối tượng kèm theo thông tin của đối tượng như hướng chuyển động , hình dáng, kích thước, màu sắc.

- Dữ liệu ra: Danh sách các đối tượng kèm theo hành vi được định nghĩa từ trước.  
- Vai trò: Module sẽ dựa vào thông tin của đối tượng, từ đó kết luận và phân tích xem đối tượng đó đang làm gì.

- Phương pháp sử dụng: Các phương pháp liên quan đến hình học, luồng quang học, mô phỏng.

* **Module đưa ra kết luận:**

- Phụ thuộc vào từng hệ thống cụ thế mà module sẽ có vai trò và đầu ra cụ thể. Ví dụ trong hệ thống phát hiện bất thường, đầu ra là kết luận đối tượng đó ở thời điểm quan sát bình thường hay bất thường.

### *1.2. Ứng dụng*

* An ninh công cộng và thương mại:

- Giám sát ngân hàng, siêu thị, sân bay, bảo tàng, nhà ga, nhà riêng và bãi đỗ xe, phát hiện và phòng chống tội phạm

- Theo dõi đường cao tốc và đường sắt để phát hiện tai nạn

- Giám sát tài sản và rừng để phát hiện cháy

- Quan sát các hoạt động của người cao tuổi và tàn tật để cảnh báo sớm và đánh giá hiệu quả của các điều trị y tế.

* Khai thác dữ liệu video:
* Đo lưu lượng giao thông, tắc nghẽn giao thông dành cho người đi bộ
* Lập nhân khẩu học của người tiêu dùng tại các trung tâm mua sắm và vui chơi giải trí, công viên
* Trích xuất các số liệu thống kê từ các hoạt động thể thao
* Thực thi pháp luật:

- Phát hiện vượt đèn đỏ và lấn làn đường

### *1.3. Các bài toán cần giải quyết*

Một hệ thống giám sát thông minh bằng hình ảnh là một tập hợp các bài toán nhỏ. Nhìn một cách tổng quan:

- Đầu vào của hệ thống sẽ là hình ảnh thu được tại các điểm quan sát

- Đầu ra của hệ thống là các thông tin về chuyển động, hành vi của đối tượng được giám sát. Việc xử lý của hệ thống giám sát thông minh bằng hình ảnh tóm lại có thể hiểu là việc phân tích và xử lý hình ảnh video qua việc giải quyết các bài toán sau:

**Bài toán 1**: Phát hiện các đối tượng chuyển động là bước cơ bản đầu tiên trong bài toán phân tích hình ảnh video, công việc này khái quát lại đó là việc tách các đối tượng chuyển động trong các hình ảnh nền của đối tượng đó. Phương pháp thường được sử dụng trong bài toán này đó là: Phương pháp trừ ảnh nền, các phương pháp dựa trên thống kê, phương pháp dựa trên sự chênh lệch tạm thời và các phương pháp phân luồng thị giác.

**Bài toán 2**: Phân lớp đối tượng là công việc phân loại ra các lớp đối tượng đã được tìm ra theo các lớp đã được định nghĩa từ trước như: Lớp người, lớp phương tiện, lớp động vật... Đây là bước cần thiết để tiếp tục phân tích các hoạt động của chúng. Hiện tại có hai hướng tiếp cận chính để giải quyết bài toán này đó là: Hướng tiếp cận dựa trên hình dáng của các vết và hướng tiếp cận dựa trên chuyển động của đối tượng. Hướng tiếp cận dựa trên hình dáng của đối tượng là hoàn toàn phụ thuộc vào hình dáng, tính chất 2D của các vết tìm được, trong khi đó hướng tiếp cận dựa trên chuyển động của đối tượng dựa trên các tính chất chuyển động của đối tượng theo thời gian.

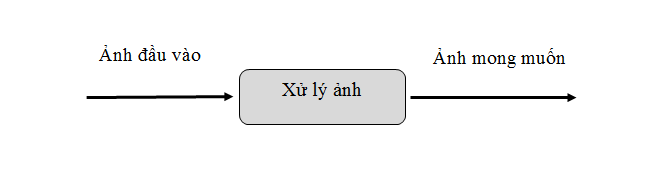
**Bài toán 3**: Theo dõi đối tượng đó là công việc đưa ra một chuỗi các hành vi của đối tượng chuyển động trong một khoảng thời gian xác định dựa vào các khung hình thu được từ video. Thủ tục này đưa ra các thông tin về đối tượng được theo dõi như đường đi của đối tượng, tốc độ hay hướng chuyển động của đối tượng. Từ đó có thể dự đoán được hành động của các đối tượng và mô tả được hành động của chúng.

Các bài toán này không những được nghiên cứu và áp dụng cho các hệ thống giám sát mà còn được áp dụng vào các bài toán trong lĩnh vực khác như: Thực tại ảo, nén hình ảnh, biên tập video và cơ sở dữ liệu đa phương tiện, là các hướng tiếp cận phát triển công nghệ đa phương tiện trong tương lai.

## 2. Tổng quan về một hệ thống xử lý ảnh

### *2.1. Khái niệm cơ bản*

Quá trình xử lý ảnh được xem như là quá trình thao tác với ảnh đầu vào nhằm cho ra kết quả như mong muốn .



*Hình I. 2 Quá trình xử lý ảnh*

### *2.2. Cấu trúc của hệ thống xử lý ảnh*

Xử lý ảnh là một ngành khoa học không còn quá mới mẻ song còn có nhiều vấn đề còn tồn tại nếu chúng ta muốn triển khai ở quy mô công nghiệp. Để có thể hình dung một hệ thống xử lý ảnh chuyên dụng hay một hệ thống xử lý ảnh dùng trong nghiên cứu, đào tạo, chúng ta sẽ xem xét cụ thể các giai đoạn trong một hệ thống xử lý ảnh:



*Hình I. 3 Các giai đoạn chính trong xử lý ảnh*

Trước hết là quá trình thu nhận ảnh. Ảnh có thể thu nhận qua camera hoặc qua các bộ sensor cảm ứng. Sau đó ảnh sẽ được số hóa, đây là quá trình biến đổi tín hiệu tương tự sang tín hiêu rời rạc giúp cho máy tính có thể dễ dàng biểu diễn và lưu trữ, bên cạnh đó, đây cũng là giai đoạn quan trọng để lấy đầu vào cho các giai đoạn sau. Quá trình phân tích ảnh thực ra là bao gồm nhiều công đoạn nhỏ. Trước hết ảnh sẽ được tăng cường để nâng cao chất lượng ảnh. Do những nguyên yếu tố khác nhau như thiết bị thu nhận ảnh, môi trường thu nhận ảnh... nên ảnh có thể bị suy biến. Do vậy cần tăng cường và khôi phục lại ảnh để làm nổi bật một số đặc tính của ảnh, hay làm cho ảnh gần giống nhất với trạng thái gốc. Giai đoạn tiếp theo là là phát hiện các đặc tính như phát hiện biên của ảnh, phân vùng ảnh, trích chọn các đặc tính... Cuối cùng, tùy vào mục đính của ứng dụng, sẽ là giai đoạn nhận dạng, phân lớp hay trích rút thông tin.

### *2.3 Các vấn đề cơ bản trong xử lý ảnh*

Như đã đề cập trong phần trên, chúng ta thấy được một cách khái quát các vấn đề chính trong xử lý ảnh. Để hiểu chi tiết hơn, trước tiên ta phải xem xét hai khái niệm (thuật ngữ) thường được sử dụng trong xử lý ảnh đó là Pixel (phần tử ảnh) và gray level (mức xám), dưới đây là những vấn đề chính.

#### **2.3.1 Một số khái niệm chính.**

**- Pixel(Picture Element)**: Ảnh trong thực tế là một ảnh liên tục về không gian và giá trị độ sáng. Để có thể xử lý ảnh bằng máy tính cần thiết phải số hóa ảnh. Trong quá trình số hóa, người ta biến đổi tín hiệu liên tục sang tín hiệu rời rạc thông qua quá trình lấy mẫu (rời rạc hóa về không gian) và lượng hóa thành phần giá trị mà về nguyên tắc bằng mắt thường không phân biệt được hai điểm kề nhau. Trong quá trình này, người ta sử dụng khái niệm Picture Element mà ta quen gọi tắt là Pixel – Phần tử ảnh. Như vậy một ảnh là một tập hợp các picel. Mỗi pixel gồm một cặp tọa độ x,y và màu.



*Hình I. 5 Biểu diễn ảnh với độ phân giải khác nhau*

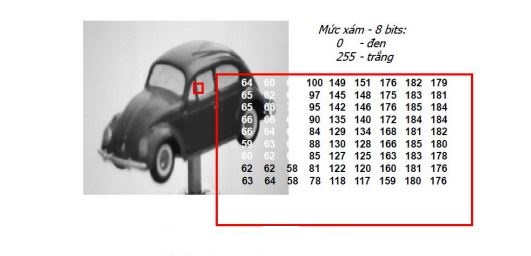
**- Độ phân giải của ảnh**: Là mật độ điểm ảnh được ấn định trên một ảnh số được hiển thị.

+ Theo định nghĩa , khoảng cách giữa các điểm ảnh phải được chọn sao cho mắt người vẫn cảm nhận được sự liên tục của ảnh. Việc lựa chọn khoảng cách thích hợp tạo nên mật độ phân bổ, đó chính là độ phân giải và được phân bố theo hai trục x,y trong không gian hai chiều.

+ Ví dụ: Độ phân giải trên màn hình CGA (Color Graphic Adaptor) là một lưới điểm theo chiều ngang màn hình: 320 điểm theo chiều dọc \* 200 điểm ảnh (320\*200). Rõ ràng , cùng màn hình CGA 12” ta nhận thấy mịn hơn màn hình CGA 17” (320\*200) . Lý do là tuy cùng một độ phân giải nhưng kích thước càng nhỏ thì mật độ phân bổ càng nhiều, dẫn đến màn hình sẽ cho hình ảnh mịn hơn.

- **Mức xám** của điểm ảnh: Là cường độ sáng của nó được gán bằng giá trị số tại điểm đó. Các thang giá trị mức xám thông thường: 16, 32, 64, 128, 256 (Mức 256 là phổ dụng do máy tính dùng 1byte = 8bit để biểu diễn 1 điểm ảnh)

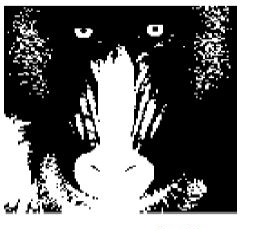
- **Ảnh xám**: Ảnh được biểu diễn bởi 1 ma trận các điểm ảnh và có mức xám từ 0 đến 255. Ảnh xám được sử dụng nhiều trong xử lý ảnh vì những tính chất ưu việt trong việc sử dụng duy nhất một ma trận để biểu diễn.



*Hình I. 6 Ví dụ về ảnh xám*

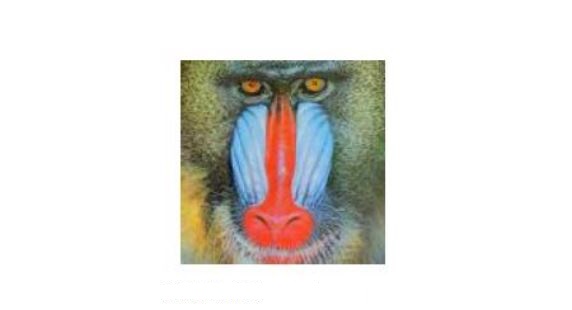
- **Ảnh đen trắng**: Là ảnh chỉ chứa hai màu đen trắng tương ứng với mức xám là 0 hoặc 255

- **Ảnh nhị phân**: Là ảnh đen trắng nhưng giá trị của điểm ảnh sẽ là 0,1.



*Hình I. 7 Ví dụ về ảnh nhị phân*

- **Ảnh màu**: Trong khuôn khổ lý thuyết ba màu (RED, BLUE, GREEN) để tạo nên thế giới màu sinh động. Máy tính sử dụng 3 byte để biểu diễn 1 điểm ảnh màu, suy ra sẽ có khoảng 16,7 triệu màu.

****

*Hình I. 8 Ví dụ về ảnh màu*

#### 2.3.2. Thu nhận ảnh

Để có thể xử lý ảnh trên máy tính cũng như lưu trữ, ảnh cần được thu nhận bởi các thiết bị chuyên dụng như: camera cộng với bộ chuyển đổi tương tự số AD (Analog to Digital) hoặc máy quét chuyên dụng.

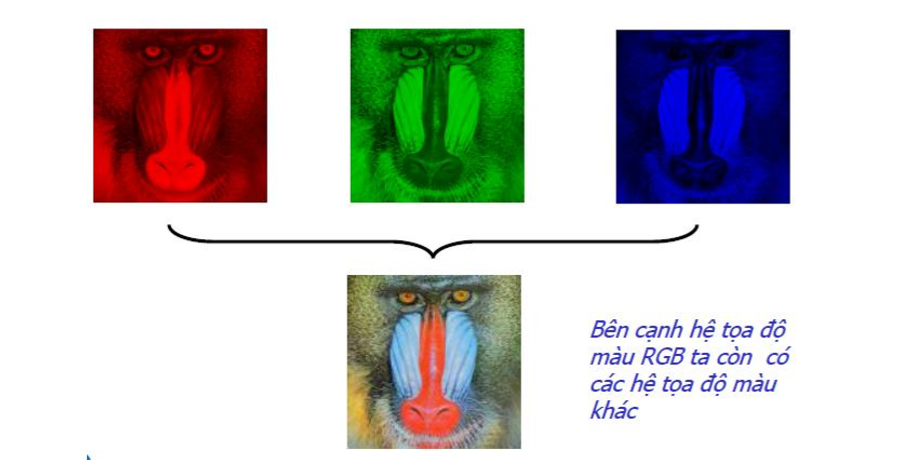
##### **2.3.2.1. Biểu diễn ảnh**

Sau quá trình số hóa sẽ thu được một ma trận các điểm ảnh tương ứng với ảnh cần xét, mỗi phần tử của ma trận tương ứng với điểm ảnh. Các điểm này thường được đặc trưng bởi hệ tọa độ màu RGB.



*Hình I. 9 Biểu diễn ảnh*

Tùy theo mục đích sử dụng mà ta ảnh có thể được lưu trữ dưới dạng tổ hợp của 3 ma trận (đối với ảnh màu), hay lưu dưới dạng 1 ma trận (đối với ảnh xám hoặc ảnh nhị phân hay ảnh đen trắng).



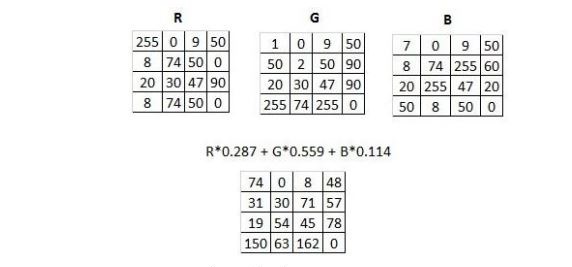
*Hình I. 10 Ảnh với các hệ màu khác nhau*

### *2.4 Một số thuật toán xử lý ảnh thông dụng*

#### 2.4.1 Thuật toán chuyển ảnh màu về ảnh xám

Như đã biết, ảnh đầu vào của hệ thống kiểm soát vào ra thường là ảnh màu. Được biểu diễn bởi 3 ma trận màu Red – Blue – Green (R –G –B). Vấn đề đặt ra là trong quá trình phát hiện đối tượng, ánh sáng trong vùng quan sát có thể bị thay đổi một chút (Ví dụ khi có đám mây đi ngang qua thì bầu trời sẽ tối hơn). Bên cạnh đó, việc xử lý trên 3 ma trận để phát hiện ra đối tượng chuyển động trong thời gian thực rất khó khăn và đôi khi lại thừa thông tin. Chính vì vậy ta phải biến đổi ảnh màu về ảnh xám, được biểu diễn bằng một ma trận duy nhất với giá trị của ma trận từ 0-255 (0- Màu đen, 255- Màu trắng). Điều này sẽ khắc phục được những yếu điểm trong việc xử lý ảnh màu như đã nêu trên. Đây cũng là cách mà phần lớn các hệ thống xử lý ảnh cần phải làm trước khi tiến hành các bước về sau. Để chuyển từ ảnh màu về ảnh xám ta sẽ sử dụng 3 ma trận R-G-B với công thức sau: Ảnh xám I=R\*0.287 + G\*0.559 + B\*0.114.

Hình bên dưới là ví dụ về việc sử dụng công thức trên.

****

*Hình I. 11 Ví dụ về chuyển về ảnh xám bằng ma trận ba màu*

*Hình I. 12 Ảnh thực tế sau khi chuyển đổi*

#### 2.4.2 Thuật toán khử nhiễu.

Nhiễu ảnh là hiện tượng một hoặc một nhóm điểm ảnh có giá trị cao bất thường so với các điểm ảnh xung quanh nó. Các kết quả đầu ra của thuật toán phát hiện đối tượng chuyển động vừa được đề cập ở trên thường có chứa các nhiễu, do đó cần phải xử lý để loại bỏ chúng trước khi chuyển kết quả sang các bước sau. Có nhiều yếu tố khác nhau gây ra nhiễu như:

- Nhiễu máy quay: đây là nhiễu do các bộ phận thu nhận ảnh của camera.

- Nhiễu phản xạ: khi một nguồn sáng, như ánh sáng mặt trời di chuyển sẽ làm cho một số phần trong cảnh nền phản chiếu ánh sáng. Hiện tượng này làm cho thuật toán phát hiện foreground thất bại hoặc phát hiện nhầm khu vực phản xạ là foreground.

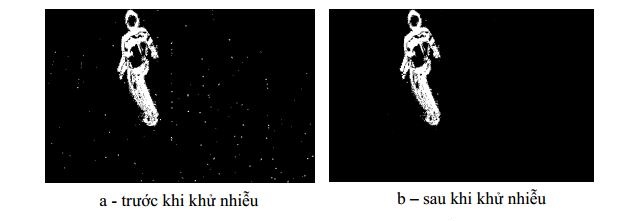
- Nhiễu đối tượng cùng màu nền: Một số phần của đối tượng có thể cùng màu sắc với ảnh nền tham chiếu.

- Bóng và thay đổi ánh sáng đột ngột: Hầu hết các thuật toán phát hiện đều phát hiện bóng của các đối tượng được phát hiện là foreground. Ngoài ra, thay đổi ánh sáng đột ngột (ví dụ: như đèn trong căn phòng được bật lên) làm cho các thuật toán không phát hiện được foreground chính xác. Có nhiều thuật toán để khử nhiễu. Tuy nhiên trong đề tài này, ảnh được khử nhiễu là ảnh đen trắng (0-255) nên nhóm tác giả xin trình bày 1 thuật toán đơn giản

****

*Hình I. 13 Ví dụ về khử nhiễu*

Dựa vào tính chất của nhiễu là có giá trị cao hơn so với các điểm lân cận. Chính vì vậy đối với ảnh A cần khử nhiểu, ta chuyển ảnh đó về ảnh xám và tìm xem có giá trị xám nào cao hơn bất thường không. Cụ thể là ta sẽ loại bỏ những điểm ảnh nào có giá trị bằng 255 (màu trắng) mà xung quanh nó là 0 (màu đen). Hình bên dưới chính là kết quả của quá trình khử nhiễu.

*Hình I. 14 Hình ảnh sau khi khử nhiễu*

Đối với ảnh có vùng nhiễu lớn (tồn tại những vùng ảnh màu trắng với kích thước rất nhỏ) thì thuật toán đưa ra phía trên chưa thực sự phát huy được hết hiệu quả. Tuy nhiên trong đề tài này, việc các vùng nhiễu lớn có thể được loại bỏ bằng cách giãn ảnh, sau đó xóa bỏ những vùng ảnh nhỏ nếu vùng ảnh đó có kích thước nhỏ hơn một ngưỡng cho phép nào đó.  
2.1.4.3 Thuật toán giãn ảnh.

Giãn ảnh là khái niệm thuộc về phép biến đổi hình thái học trên ảnh, được áp dụng cho ma trận điểm ảnh foreground để loại bỏ nhiễu gây ra bởi ba mục đầu tiên được liệt ở trên. Mục tiêu của việc áp dụng các phép hình thái này là để loại bỏ các điểm ảnh nhiễu foreground, những điểm ảnh thực tế không thuộc về foreground, và để loại bỏ các điểm ảnh gần foreground, và các điểm ảnh bên trong vùng các đối tượng Giãn nở là sự mở rộng biên foreground lên 1 đơn vị điểm ảnh. Giãn ảnh nhằm phục vụ cho thuật toán loang tìm các khối liên thông.

****

*Hình I. 15 Ảnh sau khi giãn*

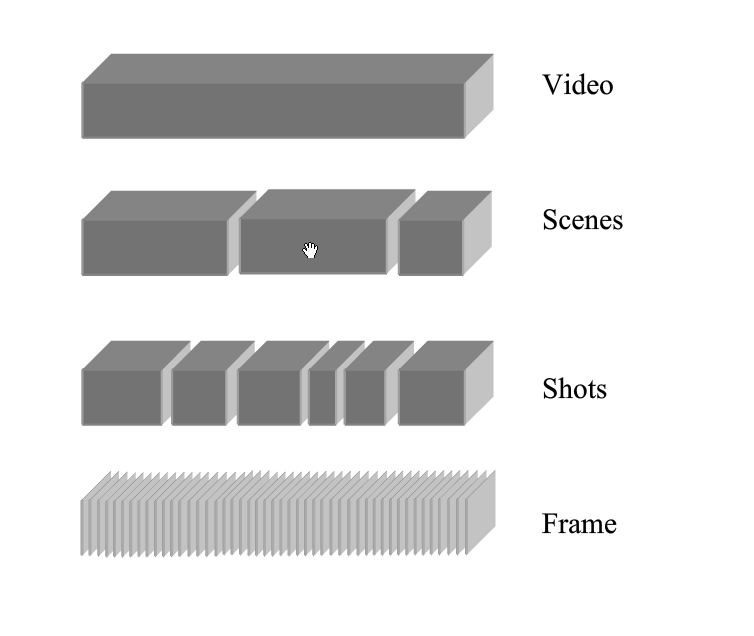
## 3. Tổng quan về video

### *3.1.Cấu trúc của video*

Video được cấu thành từ một tập liên tiếp các khung hình – frame, ghi nhận lại các hình ảnh quan sát được của các sự kiện xảy ra trong một khoảng thời gian nhất định. Có 2 dạng tín hiệu video thông dụng là tín hiệu analog và digital. Đề tài lấy dữ liệu hình ảnh từ camera nên chúng ta chỉ xét đến video được lưu trữ dưới dạng digital chứ không xét đến âm thanh.

Mô hình cấu trúc của video được mô hình như hình dưới đây:

* Frame: là thành phần cơ bản trong chuỗi video. Mỗi frame tương ứng với một hình ảnh trong thế giới thực tại một thời điểm.
* Shot là một chuỗi các frame liên tiếp được camera ghi nhận không có sự ngắt quãng nào xảy ra. Shot là đơn vị cơ bản để xây dựng phân tích nội dung video.
* Các shot liên tiếp nhau được kết hợp để tạo thành một cảnh (Sceners).
* Tất cả các scenes được tập hợp thành một video.

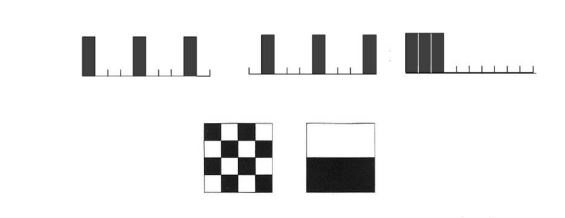
  
 *Hình I. 16 Cấu trúc phân đoạn video*

Khi video được chiếu, các frame lần lượt được hiển thị ở tốc độ nhất định. Tốc độ thường thấy là 24 frame / giây. Với càng nhiều frame trên 1 giây thì video sẽ được thể hiện càng mịn và chuyển động linh hoạt.

### *3.2. Các đặc trưng của video*

#### 3.3.1 Màu sắc (color).

Màu là một thuộc tính qua trọng của ảnh. Biểu đồ màu biểu diễn sự phân bố màu, là một đặc trưng phổ biến nhất hiện nay. Biểu đồ màu không phụ thuộc vào độ quay hay dịch chuyển ảnh cũng như chiều nhìn ảnh. Có một vấn đề với biểu đồ màu là nó không biểu diễn thông tin về không gian phân bố các điểm ảnh. Do đó hai ảnh có cũng biều đồ màu nhưng nội dung có thể sẽ rất khác nhau.



*Hình I. 17 Các ảnh khác nhau nhưng có cùng biểu đồ*

#### 3.3.2 Kết cấu (Texture).

Là một đặc trưng quan trọng của bề mặt, nơi xảy ra việc lặp lại mẫu cơ bản. Có hai dạng biểu diễn texture phổ biến: Biểu diễn dạng ma trận đồng thời và biểu diễn Tamura. Ma trận đồng thời mô tả hướng và khoảng cách giữa các điểm ảnh, nhờ đó các thống kê có thể được trích chọn. Biểu diễn Tamura được thúc đẩy nhờ các nghiên cứu về tâm lý trong việc thu nhận trực giác của con người và nó bao gồm các đại lượng đo tính thô, độ tương phản, hướng, tính trơn, tính cân đối và độ ráp. Các đặc trưng Tamura rất hấp dẫn trong việc hiểu nội dung ảnh vì nó biểu diễn trực quan. Ngoài ra còn có một số các dạng biểu diễn khác như trường ngẫu nhiên Markov, biến đổi Gabor, biến đổi gợn sóng...

#### 3.3.3 Hình dạng (Shape)

Các đặc trưng hình dáng có thể được biễu diễn sử dụng phân tích hình dáng truyền thống như bất biến Momen , mô tả Fourier, mô hình học tự động quay lui và các thuộc tính hình học. Các đặc trưng này có thể được chia thành đặc trưng toàn cục và đặc trưng cục bộ. Đặc trưng toàn cục là đặc trưng thuộc tính thu được từ toàn bộ hình dáng ảnh, chẳng hạn như chu vi, tính tròn, momen trung tâm, hướng trục chính... Đặc trưng cục bộ là đặc trưng thu được từ việc thao tác với một phần của ảnh, không phụ thuộc vào toàn bộ ảnh.

#### **3.3.4 Chuyển động (Motion)**

Chuyển động là một thuộc tính quan trọng của video. Thông tin về chuyển động có thể được sinh ra bằng các kỹ thuật ghép khối hoặc luồng ánh sáng. Các đặc trưng chuyển động như Momen của trường chuyển động, biểu đồ chuyển động hoặc là các tham số chuyển động toàn cục có thể dược trích chọn từ các vector chuyển động. Các đặc trưng mức cao phản ánh di chuyển của camera như quét, nghiêng, phóng to, thu nhỏ.... cũng có thể được trích chọn.

## 4. Một số kĩ thuật phát hiện chuyển động và theo dõi đa đối tượng

### *4.1. Các kĩ thuật phát hiện đối tượng chuyển động*

#### 4.1.1. Tổng quan

Mục tiêu của bài toán là xác định các đối tượng chuyển động xuất hiện trong frame hiện tại. Kỹ thuật phát hiện đối tượng chuyển động được sử dụng rộng rãi trong thực tế, là bước đầu tiên của các hệ thống theo dõi đối tượng chuyển động, hệ thống giám sát, …

Bài toán được đề xuất và được rất nhiều nhà khoa học dày công nghiên cứu từ giai đoạn đầu tiên của các hệ thống giám sát 1960-1980. Từ đó đến nay, đã có rất nhiều nghiên cứu có giá trị thực tiễn và vẫn đang tiếp tục được phát triển theo chiều sâu.

Với kỹ thuật phát hiện đối tượng chuyển động, người ta chia thành 2 phân nhánh chính:

* Nền cố định
* Nền thay đổi

Với bài toán nền cố định, ta có thể áp dụng kỹ thuật sai chênh lệch tạm thời hoặc thuật toán trừ nền. Tuy nhiên, với bài toán có nền thay đổi – chúng ta phải áp dụng rất nhiều kỹ thuật phức tạp. Nhánh nghiên cứu với nền thay đổi đòi hỏi sự thách thức lớn hơn.

Các bài toán với nền cố định có thể giải quyết bằng một số kỹ thuật cơ bản như sau:

* Thuật toán trừ ảnh: Trừ ảnh của thời điểm **ngay trước đó**, trừ ảnh với các điểm ảnh có **giá trị trung vị của K thời điểm trước** đó, trừ ảnh với các gia trị điểm ảnh là **trung bình của K thời điểm trước**…
* Dựa trên phân bố Gaussian
* Dựa trên ước lượng MeanShift
* Dựa trên ước lượng và lan truyền
* Dựa vào nền riêng (eigen backgrounds)

Ngoài ra, có các vấn đề đi song song cùng với bài toán phát hiện đối tượng chuyển động là kỹ thuật loại bỏ bóng (shadow) khỏi đối tượng. Đây cũng là loạt bài toán rất thú vị và rất hấp dẫn đối với các phòng nghiên cứu của một số trường ĐH trên thế giới.

Trong phạm vi bài báo cáo này, chúng tôi sẽ trình bày thuật toán trừ nền với 2 kỹ thuật chủ đạo là trừ nền trước và trừ nền trung bình với ngưỡng.

#### 4.2.2. Cách tiếp cận dựa trên đặc trưng của ảnh và các bộ phân lớp

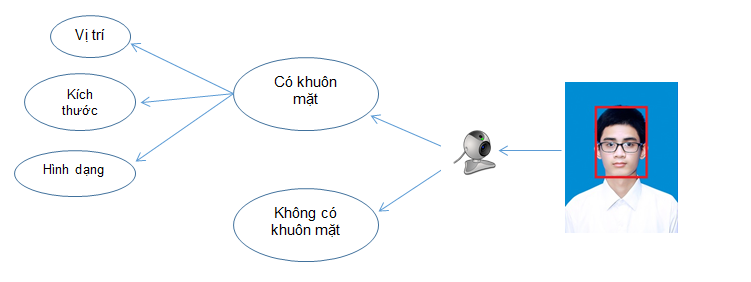
Để giải quyết bài toán phát hiện đối tượng chuyển động, ta có thể quy về bài toán phát hiện đối tượng trong ảnh. Với phương pháp này, ta dễ dàng áp dụng các kỹ thuật phân tích đặc trưng, phân lớp và nhận dạng đối tượng trên ảnh tĩnh để áp dụng cho bài toán phát hiện đối tượng chuyển động.

Tuy nhiên, với cách giải quyết này – đặc trưng của bài toán không được làm rõ.

Một số kỹ thuật có thể được áp dụng như:

* Sử dụng đặc trưng haarlike
* Sử dụng kỹ thuật HOG (Histogram Orriented Gradient)
* Sử dụng các kỹ thuật xử lý ảnh để phân tích đặc trưng cho một số đối tượng quan tâm riêng biệt (chẳng hạn như phát hiện khuôn mặt chuyển động, ta sẽ xây dựng kỹ thuật phân tích đặc trưng riêng cho khuôn mặt.)

Dưới đây là ví dụ mô phỏng cho kỹ thuật phát hiện khuôn mặt sử dụng đặc trưng haarlike



*Hình I. 18 Ví dụ về bài toán phát hiện khuôn mặt*

*Tuy nhiên, việc áp dụng cách tiếp cận trên – chỉ phù hợp cho từng bài toán cụ thể. Chứ không phù hợp cho đa số bài toán phân đoạn chuyển động. Vì vậy, chúng tôi quyết định không lựa chọn kỹ thuật trên cho bài báo cáo của mình.*

#### 4.2.3. Cách tiếp cận dựa trên thuật toán trừ nền

**+ Tổng quan kỹ thuật trừ nền**

Trừ nền là một kỹ thuật thường được sử dụng cho việc phân đoạn chuyển động. Mục đích của phương pháp là để phát hiện các khu vực di chuyển bằng cách trừ từng điểm ảnh của hình ảnh hiện tại với một hình nền tham chiếu được tạo ra bằng cách trung bình hình ảnh theo thời gian trong một khoảng thời gian. Các điểm ảnh có sự khác biệt lớn hơn một ngưỡng được phân loại thành foreground. Sau khi tạo ra một ma trận điểm ảnh foreground, một số phương pháp hình thái trong xử lý ảnh chẳng hạn như phép giãn nở, phép xói mòn và phép đóng ảnh được thực hiện để làm giảm ảnh hưởng của nhiễu và cải thiện các khu vực được phát hiện. Các nền tham chiếu được cập nhật mới theo thời gian để thích ứng với các đoạn cảnh chuyển động.

Có nhiều phương pháp khác nhau để thực hiện trừ nền. Heikkila và Silven sử dụng phương pháp đơn giản của phương pháp này đó là một điểm ảnh tại vị trí (x, y) trong hình ảnh hiện tại It thuộc về foreground nếu: |It(x,y) – Bt(x-y)| > τ được thỏa mãn trong đó τ là một ngưỡng được xác định trước. Hình nền Bt được cập nhật bằng cách sử dụng bộ lọc Infinite Impulse Response (IIR) như sau:

Bt+1= αIt + (1- α)Bt

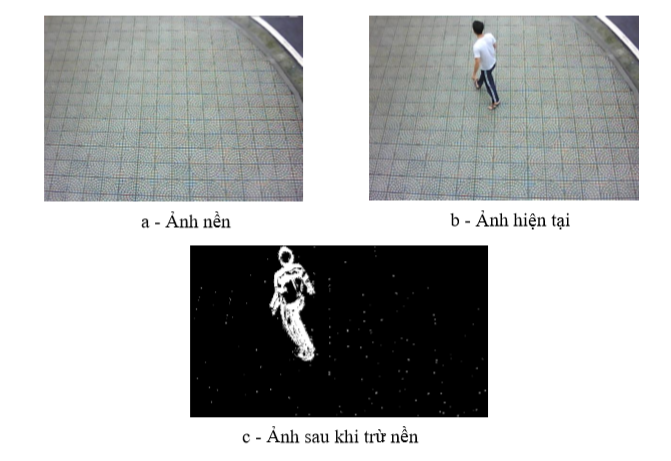
Sau đó thực hiện tạo ma trận các điểm ảnh foreground bằng các phép hình thái và loại bỏ các vùng có kích thước nhỏ. Ta có thể hình dung phương pháp này như hình ảnh bên dưới.

Từ trái qua phải:

- Ảnh nền

- Ảnh chứa đối tượng

- Khoanh vùng đối tượng sau khi trừ nền



*Hình I. 19 Ảnh sau khi dùng kĩ thuật trừ nền*

Vì kỹ thuật này khá đơn giản, dễ cài đặt và có thể thực hiện trong thời gian thực nên được sử dụng rất phổ biến. Tuy nhiên nó cũng có những hạn chế nhất định như màu của đối tượng trùng với màu nền hay đối tượng bất động…

### *4.2. Một số hàm lọc nhiễu*

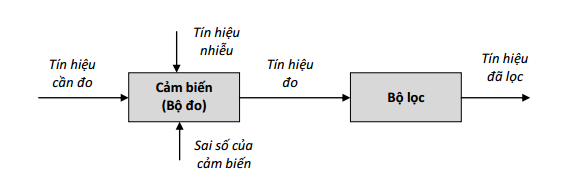
#### 4.2.1 Tổng quan

Trong quá trình xử lý của một số kĩ thuật, dữ liệu trả về có thể chứa nhiều nhiễu, có thể chứa nhiễu môi trường, của nguồn hay quá trình xử lý. Vấn đề đặt ra ở đây là làm thế nào để loại bỏ nhiễu để có giá trị gần đúng nhất với tín hiệu chuẩn. Các bộ lọc được giới thiệu ở đây nhằm giải quyết các vấn đề đó.

#### 4.2.2 Bộ lọc Kalman

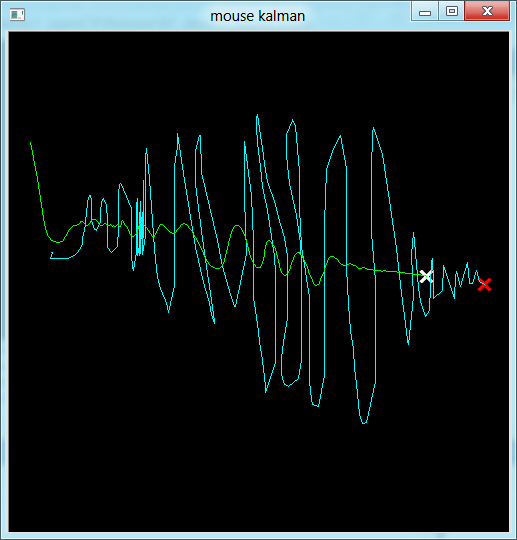
Năm 1960, R.E. Kalman công bố bài báo nổi tiếng về một giải pháp đệ quy để giải quyết bài toán lọc thông tin rời rạc tuyến tính (Discrete data linear filtering). Bài báo có tựa đề “A New Approach to Linear Filtering and Prediction Problems”. Khoảng 55 năm trôi qua, bộ lọc Kalman đã trở nên phổ biến. Nó xuất hiện trong rất nhiều ứng dụng và bởi vì vấn đề lọc mà nó giải quyết là một vấn đề cơ bản trong rất nhiều lĩnh vực, nên nó vẫn luôn còn được ứng dụng nhiều hiện nay. Một cách tổng quát, bộ lọc Kalman là một tập hợp các phương trình toán học giúp tối ưu ước lượng trạng thái của một hệ (theo nghĩa giá trị ước đoán có sai số bình phương trung bình nhỏ nhất – MMSE) dựa trên mô hình hệ thống (system model), giá trị đo (measurement value) và các hiểu biết về nhiễu (của hệ thống lẫn phép đo).

Biểu đồ sau đây là một mô tả hình thức của một bộ lọc nói chung.



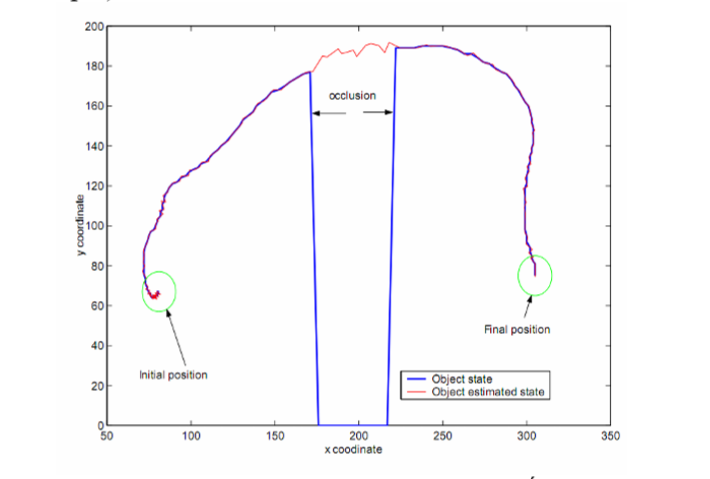
*Hình I. 20 Biểu đồ mô tả hình thức của bộ lọc nói chung*

***Hình I. 21 Ứng dụng lọc Kalman để lọc tín hiệu thu được từ toạ độ con trỏ của người dùng***



#### 4.2.3. Bộ lọc Kalman mở rộng (Extended Kalman Filter - EKF)

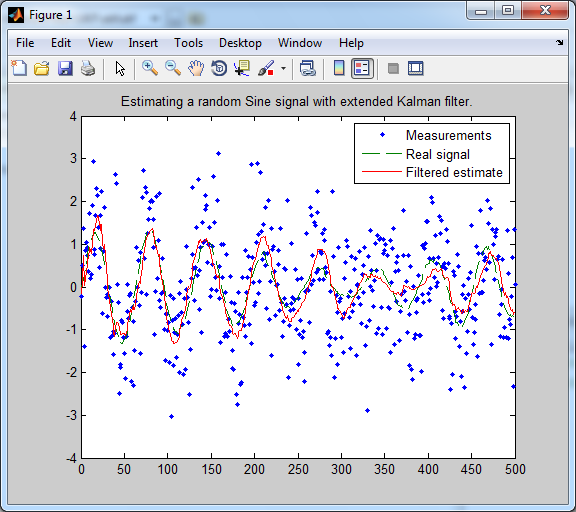
Một lợi thế của bộ lọc Kalman cho việc bám bắt đối tượng là nó có thể ước lượng được vị trí đối tượng ngay cả khi bị che khuất nhỏ. Cách để thực hiện việc này là xem xét hai giai đoạn của bộ lọc: dự đoán và hiệu chỉnh. Có nghĩa là nếu việc giới hạn vị trí của đối tượng là không trong một chuỗi của trạng thái dự đoán của bộ lọc (trong thời điểm k), ta có thể coi rằng đối tượng bị che khuất bởi một vài đối tượng khác, do đó ta sẽ không thể dùng hiệu chỉnh đo và sẽ chỉ lấy giá trị lọc tiên nghiệm vị trí của đối tượng. *Hình I. 22* chỉ ra hoạt động của bộ lọc khi đối tượng bị che khuất.



*Hình I. 22 Bộ lọc Kalman bị che khuất*

Bộ lọc Kalman thông thường chỉ đáp ứng được các hệ thống phi tuyến có giả định nhiễu Gauss. Trên thực tế, có rất ít hệ thống có các điều kiện lý tưởng cho Kalman Filter hoạt động tốt. Vì vậy, Kalman Filter thường ít được triển khai trong các hệ thống thực tế. Tuy nhiên, không phải vì vậy mà ta kết luận Kalman Filter kém hơn so với các bộ lọc phi tuyến khác. Trên thực tế, có các hệ thống khi triển khai Kalman Filter cho ta hiệu quả tốt hơn hẳn, bởi đặc tính của hệ thống đó là trạng thái biến đổi tuyến tính.

Jazwinski, 1996; Maybeck, 1982; Bar-Shalom, 2001; Grewal and Andrews, 2001; Sarkka, 2006, đã có những nghiên cứu để mở rộng bộ lọc Kalman tuyến tính, từ đó đề xuất một bộ lọc có tên Lọc Kalman mở rộng (Extended Kalman filter - EKF).



*Hình I. 23 Ví dụ sử dụng EKF của J. Hartikainen, A. Solin, S. Sarkka - Aalto University, August 2011*

Trong ví dụ của cuốn sách Optimal Filtering with Kalman Filter and Smoothers của J. Hartikainen, A. Solin, S. Sarkka – trường ĐH Aalto, họ sử dụng EKF để ước lượng việc đo đạc một hệ thống có trạng thái biến đổi theo dạng sin – nghĩa là trạng thái biến đổi phi tu­­­yến.

#### 4.2.4. Đánh giá

Hiện nay có một số bộ lọc ra đời nhằm cải tiến và nâng cao chất lượng lọc trong một số bài toán có độ phức tạp cao, ví dụ như Unscented Kalman, Multiple Model Estimator. Tuy nhiên trong khuôn khổ đề tài với camera đặt cố định thì lọc Kalman giải quyết rất tốt vấn đề còn tồn tại. Vì vậy chúng tôi chỉ triển khai kĩ thuật lọc Kalman trong bài toán này.

### *4.3. Các kĩ thuật theo dõi đa đối tượng chuyển động*

##### **4.3.1. Tổng quan**

Theo vết đối tượng (Object Tracking) là bài toán quan trọng trong lĩnh vực thị giác máy tính. Do nhu cầu phân tích video ngày càng cao cũng với khả năng tính toán của máy tính cũng được nâng cao, các thuật toán theo vết đối tượng cũng thu hút được rất nhiều sự quan tâm. Trong lĩnh vực phân tích nội dung video có ba bước quan trọng: Phát hiện đối tượng cần quan tâm, theo dõi đối tượng đó qua các frame, phân tích vết di chuyển của đối tượng để nhận dạng hành vi của đối tượng mà cụ thể ở đây là phát hiện các hành vi bất thường. Chính vì vậy, mà khâu theo vết đối tượng vô cùng quan trọng và cần sự chuẩn xác cao.

Theo vết đối tượng có thể được định nghĩa là bài toán ước lượng vết di chuyển của đối tượng trong mặt phẳng ảnh khi đối tượng chuyển động. Hay nói cách khác, bộ theo vết đối tượng (tracker) phải gán nhãn nhất quán cho các đối tượng được quan tâm trong video. Ngoài ra còn tùy thuộc vào lĩnh vực cụ thể mà tracker cung cấp các thông tin khác của đối tượng như hình dáng, diện tích, hướng chuyển động. Bài toán theo vết đối tượng rất phức tạp và khó khăn do các nguyên nhân chính sau:

- Mất mát thông tin do chiếu từ không gian 3D xuống mặt phẳng 2D

- Nhiễu

- Chuyển động phức tạp của đối tượng

- Đối tượng bị che khuất và che khuất lẫn nhau

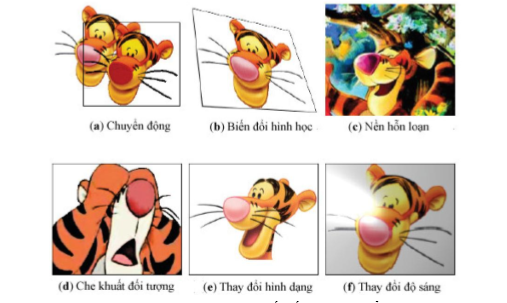
- Sự thay đổi về ánh sáng, yếu tố ngoại cảnh

- Các đối tượng dính liền vào nhau

- Đối tượng là người thì có thể bị che khuất bởi kính, tóc, khẩu trang, mũ, nón..

- Màu của đối tượng trùng với màu nền.

- Yêu cầu xử lý trong thời gian thực....



*Hình I. 24 Một số khó khăn khi theo vết đối tượng*

Có rất nhiều thuật toán theo vết đối tượng đã được đề xuất. Mỗi thuật toán dựa trên những giả thiết và ứng dụng khác nhau. Các thuật toán có thể được phân loại dựa trên cách mô hình hóa đối tượng được quan tâm, các loại đặc trưng dùng để phân biệt đối tượng đó. Như trong [] các phương pháp theo vết đối tượng được phân loại thành ba nhóm chính:

- Point tracking : Mối đối tượng được quan tâm được biểu diễn bằng một tập các điểm. Các điểm này có thể được phát hiện trong mối Frame bằng một bộ phát hiện điểm (point detector), chẳng hạn như [],[] và được liên kết dựa trên các ràng buộc về chuyển động , vị trí. Các phương pháp tiêu biểu cho nhóm này bao gồm Kalman filter[], Greedly Optimal Asignment .

- Kernel tracking: Các phương pháp theo vết trong nhóm này tính toán chuyển động của các đối tượng để theo vết qua các frame. Mô hình của đối tượng có thể được biểu diễn dưới dạng mẫu – template, hoặc mô hình mật độ - density based model[], KLT[], SVT[].

- Silhouette tracking: Là phương pháp trong đó đối tượng được theo vết bằng cách ước lượng vùng đối tượng (Object region) trong mỗi frame bằng các phương pháp như so khớp hình dáng (shape matching)[], hoặc contour evalution [].

Ngoài ra các phương pháp theo vết đối tượng còn được phân loại dựa vào số đối tượng được theo vết:

- Theo vết đơn đối tượng (single object tracking)

- Theo vết đa đối tượng (multiple object tracking)

*Trong đề tài này, như đã nói ở trên, nhóm tác giả mong muốn nâng cao độ chính xác ở bước theo vết đối tượng chuyển động, qua đó có thể đưa ra thông tin chính xác của đối tượng đang có hành vi bất thường, vì vậy chúng tôi xin phép được trình bày một số thuật toán theo vết đối tượng cơ bản được ứng dụng nhiều trong các hệ thống. Từ đó hình thành ý tưởng và đề xuất kĩ thuật cụ thể áp dụng cho module theo vết đối tượng chuyển động.*

##### **4.3.2. Thuật toán Meanshift**

Dorin Comaniciu đã giới thiệu phương pháp theo vết màu Meanshift []. Đây là một phương pháp theo vết tối ưu hóa tối thiểu cục bộ. Mỗi vị trí xi trong miền ứng viên của theo vết sẽ tương ứng với một trọng số wi.

Với b(xi) là giá tị màu tại xi và qu là giá trị màu tại u của mô hình đích, và pu(y0) là giá trị màu tại u của mô hình ứng viên. Vị trí mới của đối tượng là vị trí mà khoảng cách nhỏ nhất tới mô hình đích và mô hình ứng viên được tính bởi:

Với g(x)= - k’(x) và k(x) là một kernel function.

Quá trình được lặp lại cho đến khi không có sự thay đổi trong vị trí mới.

Các bước tính cơ bản trong phương pháp theo vết Meanshift:

Cho trước mô hình đích {qu}u = 1….mvà vị trí của nó trong khung hình trước y0.

**Bước 1:** Khởi tạo vị trí của đích trong khung hiện tại với y0, tính mô hình ứng viên {pu(y0)}u=1…m và tính p[p(y0),q] =

**Bước 2:** Tính được các trọng số wi

**Bước 3:** Tính vị trí đích ứng viên kế tiếp theo

**Bước 4:** Tính mô hình mới {pu(yi)}u=1…m và tính được p[p(y0),q] =

**Bước 5:** Trong khi p[p(y1),q] < p[p(y0),q], thực hiện y1<- ½ (y1+y0)

**Bước 6:** Nếu |y1-y0| < ε thì dừng, nếu không y0< - y1, tiếp tục bước 2.

*Ưu điểm của thuật toán Mean-shift*

Meanshift là một trong phương pháp đơn giản và hiệu quả cho theo vết thời gian thực. Nó tỏ ra rất hiệu quả trong những tình huống mà sự hình dạng của đối tượng thay đổi do máy quay chuyển động, có sự nhập nhằng, ảnh nền lộn xộn hay mục tiêu thay đổi tỉ lệ. Tuy nhiên nó lại mất tác dụng khi màu nền và màu đối tượng giống nhau, theo đánh giá của chúng tôi thì nó chỉ tối ưu cục bộ chứ không toàn cục.

##### **4.3.3. Camshift – Một cải tiến của MeanShift**

Thuật toán Camshift được phát triển trên cơ sở thuật toán MeanShift. Camshift viết tắt của “Continuously Adaptive Meanshift”. Nó bao gồm thuật toán cơ sở Meanshift với sự thay đổi thích ứng kích cỡ vùng biên của đối tượng. Các bước thực thi thuật toán như sau:

Bước 1. Kích chọn kích thước khung bám

Bước 2. Chọn vị trí khởi tạo của khung bám (cửa sổ bám)

Bước 3. Tính giá trị trung bình bên trong khung bám

Bước 4. Đặt tâm khung bám ở vị trí trung bình đã tính trong bước 3

Bước 5. Lặp lại bước 3 và bước 4 cho đến khi hội tụ (hoặc cho đên khi độ dịch chuyển vị trí tâm nhỏ hơn một ngưỡng đặt trước).

Chi tiết về thuật toán chúng tôi xin trình bày ở chương 2.

## 4.4. Kĩ thuật khoanh vùng quan sát

Kĩ thuật khoanh vùng quan sát, hay có thể hiểu là xác định các vùng quan tâm khác nhau trong mặt phẳng ảnh, từ đó chúng ta có thể khoanh vùng hoặc loại bỏ những khu vực không cần quan tâm như cây cối… Kĩ thuật được thực hiện đơn giản bằng cách tự cấu hình các vùng quan tâm tại những địa điểm khác nhau, việc cấu hình được thực hiện thủ công. Một ưu điểm của kĩ thuật này là có thể cấu hình thủ công vùng quan tâm tùy từng địa điểm quan sát, hay là loại bỏ những vùng không cần xử lý nhằm giảm khối lượng xử lý cho chương trình.

## 

## Chương II

# PHÁT HIỆN BẤT THƯỜNG TRONG GIÁM SÁT CAMERA

**Nội dung chính**

* Theo vết đối tượng dùng thuật toán CamShift và bộ lọc Kalman
* Phân tích quỹ đạo chuyển động của đối tượng
* Phát hiện bất té ngã dùng kĩ thuật phân lớp SVM

*Trong chương này chúng tôi sẽ đi sâu vào phân tích các khía cạnh của bài toán. Trong mục 2 của chương chúng tôi đề xuất các kĩ thuật triển khai cho bài toán phát hiện và theo vết đa đối tượng chuyển động. Tiếp đến mục 3 là đề xuất hướng tiếp cận và kĩ thuật giải quyết bài toán phát hiện bất thường trong giám sát camera.*

## 1. Tổng quan

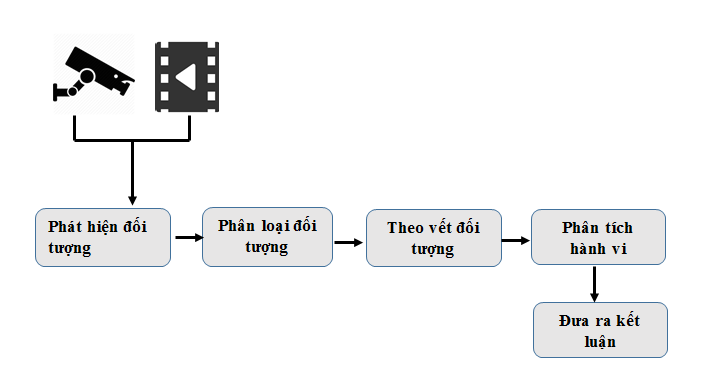
### *1.1. Phát biểu bài toán*

CCTV cung cấp hình ảnh cho nhân viên an ninh để ngăn chặn hoặc giảm thiểu những vấn đề bất thường như gây rối, đánh nhau hay té ngã. Ở những nơi công cộng như sân bay, trường học, viện dưỡng lão, nhân viên an ninh phải thực sự nhạy bén với các sự kiện bất thường hay các hành vi đáng ngờ. Thực sự phân biệt các sự kiện như vậy bằng mắt thường là không dễ dàng, đòi hỏi phải liên tục tập trung chú ý. Họ không chỉ phải giám sát trực tiếp tại thời điểm xảy ra bất thường và còn phải xem lại các video được lưu khi phải điều tra một vụ án nào đó. Vậy làm thế nào để tự động phát hiện ra những sự kiện bất thường?

Ngày nay, có nhiều đóng góp được thực hiện để giải quyết các vấn đề này. Khái niệm thế nào là bất thường rất vô tận, vậy nên điều quan trọng là chúng ta phải mô hình và giới hạn những vấn đề mang tính cấp thiết đối với cuộc sống của xã hội hiện nay. Trong bài báo cáo chúng tôi hướng đến đối tượng chính đó là “phát hiện té ngã” bởi vì lợi ích mà nó mang lại cho cuộc sống rất lớn như phát hiện người cao tuổi té ngã nhưng không có khả năng tự đứng dậy.

### *1.2. Mô hình của hệ thống*

Là một trong những loại thuộc hệ thống giám sát thông minh, hệ thống phát hiện bất thường cũng có 5 thành phần chính đó là: Phát hiện đối tượng, phân loại đối tượng, theo vết đối tượng, phân tích hành vi và kết luận. Với dữ liệu vào, ra, vai trò và các phương pháp thực hiện giống như hệ thống giám sát thông minh tổng quát đã trình bày phía trên. Tuy nhiên, trong đề tài này, do điều kiện thực nghiệm cũng như thời gian xây dựng chương trình có hạn nên nhóm tác giả xin phép module “phân loại đối tượng” sẽ được bỏ qua vì trong điều kiện vị trí thực nghiệm là trong trường học, hoặc viện dưỡng lão, bệnh viện sẽ chỉ có con người vào/ra nên việc phân loại đối tượng sẽ không cần thiết. Hình bên dưới sẽ là mô tả khái quát hệ thống



*Hình II. 1 Mô hình hệ thống phát hiện bất thường*

Trong năm modul trên thì modul phát hiện đối tượng chuyển động và theo vết đối tượng là hai thành phần cực kỳ quan trọng của hệ thống, chúng sẽ ảnh hưởng tới hiệu kết quả cũng như tốc độ của hệ thống. Đối với đặc thù của hệ thống là xử lý trong thời gian thực nên việc xử lý chính xác là cực kì khó khăn, vậy nên chúng tôi đề xuất các kĩ thuật triển khai sau đây:

## 2. Các kĩ thuật triển khai cho bài toán phát hiện và theo vết đa đối tượng chuyển động

### *2.1. Thuật toán trừ ảnh và phát hiện đối tượng chuyển động*

Kỹ thuật trừ ảnh là kỹ thuật sơ khai nhất được áp dụng trong quá trình phát hiện chuyển động trong video hoặc chuỗi hình ảnh. Ưu điểm của kỹ thuật trên là dễ triển khai và độ phức tạp tối thiểu. Tuy nhiên, nhược điểm của thuật toán này thể hiện ở việc chỉ hoạt động hiệu quả trong trường hợp có sự thay đổi rõ rệt trên từng vùng ảnh đến mức mắt thường có thể quan sát được.

**Các bước tổng quát của kỹ thuật trừ ảnh**

Cập nhật nền

Ảnh hiện tại

Ảnh nền

Trừ ảnh

Làm trơn và khoanh vùng

đối tượng

*Hình II. 5 Mô hình chung của kỹ thuật trừ nền*

**Bước 1:** Trừ ảnh

Trong bước này, ta thực hiện việc duyệt qua từng pixel ảnh của 2 ảnh đưa vào. Ta so sánh trị tuyệt đối giữa giá trị điểm ảnh này với một ngưỡng α cho trước nào đó. Nếu trị tuyệt đối sai biệt lớn hơn ngưỡng thì ta gán màu cho ảnh sai biệt tại điểm đang xét là trắng, ngược lại là màu đen.



*Hình II. 6 Một ví dụ về kỹ thuật trừ ảnh*

**Bước 2:**  Làm trơn ảnh

Trên thực tế, việc thu nhận hình ảnh từ các thiết bị ghi hình có thể gặp sai sót hoặc bị ảnh hưởng trong quá trình nén và giải nén video. Nên sẽ có rất nhiều điểm nhiễu, ta thường gọi là nhiễu sạn. Ngay cả sau khi loại bỏ các điểm ảnh nhiễu, vẫn còn sót một số khu vực nhỏ do việc phân khúc đối tượng không chính xác.

Để loại bỏ nhiễu, mỗi khung hình, cần tính toán kích thước trung bình của các khối (y). Khối nào có kích thước nhỏ hơn (α) lần của kích thước trung bình (Kích thước khối < α\* γ) được xóa bỏ khỏi ma trận điểm ảnh foreground. Ngoài ra, do lỗi phân khúc, một số bộ phần của đối tượng bị tách rời ra khỏi phần chính. Để sửa lỗi này, các khối sẽ được nhập lại với nhau và nhãn của khối sẽ được điều chỉnh lại.

|  |  |
| --- | --- |
| 1. ảnh trước khi khử nhiễu | 1. ảnh sau khi khử nhiễu |

*Hình II. 7 Làm trơn ảnh sử dụng bộ lọc ảnh*

**Bước 3:** Khoanh vùng và gán nhãn đối tượng

Đây là khâu cuối giúp ta khoanh vùng để xác định được đối tượng chuyển động trong video. Bài toán khoanh vùng và gán nhãn – tương ứng với bài toán loang. Tuy nhiên, nếu gọi lời gọi đệ quy cho kỹ thuật loang thông thường, độ phức tạp của bài toán sẽ cao và không ổn định. Vì vậy, chúng tôi nghiên cứu và lựa chọn một kỹ thuật khoanh vùng và gán nhãn khác hiệu quả hơn có tên gọi là Labeling.

Thuật toán khoanh vùng đối tượng nhanh – độ phức tạp O(n2)

|  |
| --- |
| Input: Ảnh sai biệt (I\_diff)  Output: Danh sách các đối tượng (listObject)  L = 1 *// Khởi tạo số thứ tự nhãn*  Duyệt qua từng pixel của ảnh I\_diff  if pixel **X** không phải là nền  if lân cận **A,B,C & D** đã được gán nhãn( = 0)  gán nhãn pixel **X** là L  L = L+1  else  num = MIN(nhãn của các lân cận **A,B,C & D)** điều kiện #0  gán nhãn pixel **X** và pixels **A, B, C & D** if không phải là nền với giá trị num  end if  end if  done  Duyệt qua từng pixel(x, y) của ảnh I\_diff  Index = nhãn của (x, y)  if(Index>0) cập nhật tọa độ, kích thước của **listObject[Index]**  done |

**Bước 4:** Cập nhật ảnh nền

*Với kỹ thuật trừ nền, chúng ta thường có 3 cách triển khai:*

* Ảnh nền là ảnh trước đó
* Ảnh nền là ảnh trung bình tại K thời điểm trước đó
* Ảnh nền là ảnh với từng điểm ảnh là trung vị của các điểm ảnh tương ứng trong K ảnh trước đó

**+ Phát hiện đối tượng chuyển động dựa trên việc tính ảnh sai biệt giữa 2 frame liên tiếp**

Trong kỹ thuật này, ta duyệt qua từng điểm ảnh và lấy trị tuyệt đối của điểm ảnh hiện tại trên 2 ảnh previous và current. Nếu độ chênh lệch trên điểm (x, y) đó lớn hơn ngưỡng **motionThreshold** thì ta gán màu của ảnh sai biệt là trắng. Ngược lại gán màu đen cho ảnh sai biệt.

|  |
| --- |
| **Input :** Ảnh frame trước (Iprevious) & Ảnh hiện tại (Icurrent)  **Output :** Ảnh sai biệt (Idiff)  **BEGIN**  for pixel (x, y) in Icurrent  if ( ***| get(Icurrent, x, y) – get(Icurrent, x, y) | > motionThreshold*** )  Set(***Idiff***, x, y, white) ;  else  Set(***Idiff***, x, y, black)  **END.** |

**+ Phát hiện đối tượng chuyển động dựa vào kỹ thuật ảnh có giá trị điểm ảnh là trung bình của chuỗi K ảnh liên tiếp ngay sau đó**

Với kỹ thuật trừ nền trung bình, ta luôn lưu lại tập các ảnh tại K thời điểm liên tiếp trước đó. Trước tiên, ta sẽ tính ảnh trung bình (avg) từ tập các ảnh tại K thời điểm ngay trước đó. Để tính ảnh sai biệt, ta duyệt qua từng điểm ảnh và lấy trị tuyệt đối của điểm ảnh hiện tại trên 2 ảnh avg và current. Nếu độ chênh lệch trên điểm (x, y) đó lớn hơn ngưỡng **motionThreshold** thì ta gán màu của ảnh sai biệt là trắng. Ngược lại gán màu đen cho ảnh sai biệt. Kết thúc bằng việc cập nhật tập K ảnh bằng cách bớt đi ảnh đầu tiên và thêm ảnh hiện tại vào cuối danh sách.

|  |
| --- |
| **Input :** tập gồm K trước đó (listImages) & Ảnh hiện tại (current)  **Output :** Ảnh sai biệt (diff)  **BEGIN**  avg <= getAverageImage(listImages)  for pixel (x, y) in current  if ( ***| get(current, x, y) – get(avg, x, y) | > motionThreshold*** )  Set(diff, x, y, white) ;  else  Set(diff, x, y, black)  listImage <= updateListImages(current)  **END.** |

Chúng tôi đề xuất sử dụng K = 7, nghĩa là sử dụng chuỗi 7 hình ảnh liên tiếp trước đó. Việc lựa chọn tham số K có ảnh hưởng trực tiếp đến hiệu quả thuật toán. Dựa trên kinh nghiệm cá nhân khi trải qua một số thực nghiệm kỹ thuật trừ ảnh áp dụng trong theo dõi đối tượng chuyển động, … chúng tôi thấy K = 7 là lựa chọn tốt để cân bằng cả yếu tố độ chính xác và tốc độ thực thi.

Hình dưới đây mô tả các bước làm việc của thuật toán trừ ảnh đề xuất.

Cập nhật nền

Ảnh hiện tại

Ảnh

Trung bình

Trừ ảnh

Làm trơn và khoanh vùng

đối tượng

*Thuật toán khoanh vùng nhanh*

*Chuỗi K ảnh nền*

### *2.3. Kĩ thuật theo vết đa đối tượng chuyển động – Thuật toán CamShift*

#### Trong vấn đề theo vết đối tượng, điều chúng ta cần làm là chính xác hóa đối tượng, hay nói cách khác là tìm hình ảnh của đối tượng đã phát hiện được ở thời điểm trước xem nó có xuất hiện trong thời điểm hiện tại hay không. Để làm được điều này, chúng ta có rất nhiều cách, ví dụ như đối sánh ảnh, cụ thể ở đây là đối sánh qua đặc trưng màu sắc, thuật toán Camshift sẽ thực hiện việc này. Tại thời điểm t, sau khi thực hiện kĩ thuật trừ ảnh xong, chúng ta sẽ thu được một khối hình chữ nhật đại diện cho đối tượng di chuyển, sau đó Camshift sẽ phân tích là lấy Histogram của vùng ảnh đó, tại thời điểm tiếp theo, nó duyệt tất cả các vùng trong không gian ảnh, sau đó so sánh xem có đối tượng này có Histogram tương đồng hay không, nếu đúng thì đó là đối tượng tại thời điểm t-1.

#### 2.3.1. Thuật toán CamShift

Thuật toán CamShift được phát triển trên cơ sở thuật toán MeanShift. CamShift viết tắt của “Continuously Adaptive MeanShift”. Nó bao gồm thuật toán cơ sở Meanshift với các bước thay đổi linh hoạt kích cỡ của vùng bao hình chữ nhật bao quanh đối tượng.

Các bước tiến hành thuật toán CamShift:

B1. Kích chọn kích thước khung bám. (vùng biên của đối tượng)

B2. Chọn vị trí khởi tạo của khung bám

B3. Tính giá trị trung bình bên trong khung bám

B4. Đặt tâm khung bám ở vị trí trung bình đã tính trong bước 3

B5. Lặp lại bước 3 và bước 4 cho đến khi hội tụ (hoặc cho đến khi độ dịch chuyển vị trí tâm nhỏ hơn một ngưỡng đặt trước)

Chứng minh hội tụ:

Giả sử trong không gian Euclid có một phân bố f, ta sẽ chứng minh từng bước:

B1. Chọn khung bám W có kích thước là s

B2. Tâm của khung bám ở điểm P­k

B3. Tính giá trị trung bình bên trong khung bám

Sự thay đổi là gradien của f(pk)

B4. Tâm khung bám:

B5. Lặp lại bước 3 và bước 4 cho đến khi hội tụ hoặc ( ) 0 fp do đó thuật toán Camshift là hội tụ Đối với ảnh 2D rời rạc thì vị trí tâm của khung bám được tính như sau:

Tìm Moment bậc không

sau đó tính moment bậc 1 cho x và y:

và vị trí tâm của khung bám là:

;

ở đây I(x,y) là giá trị điểm ảnh (xác suất phân bố) tại vị trí (x,y) và x, y nằm trong khung bám.

**Lưu đồ của thuật toán Camshift**



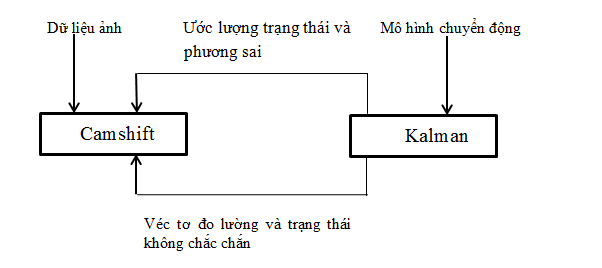
*Hình II. 9 Lưu đồ thuật toán CamShift*

Đầu tiên là một Histogram được tạo ra, Histogram này chứa các thuộc tính liên quan đến màu sắc và tiếp theo tâm và kích cỡ của mục vùng biên của đối tượng được tính toán để theo dõi mục tiêu khi hình dạng và kích cỡ của nó thay đổi. Tính xác suất phân bố Histogram nhận được và dịch chuyển đến vị trí mới với mỗi khung hình vừa nhận được từ video. CamShift sẽ dịch chuyển đến vị trí mà nó ước lượng trên đối tượng nó mà nó tập trung nhiều điểm sáng gần giống nhất trong bức ảnh nó tìm vị trí mới bắt đầu từ vị trí trước đó và tính toán giá trị trọng tâm vừa tìm được.

#### 2.3.2. Kết hợp bộ lọc Kalman với thuật toán CamShift

Trong thuật toán bám đối tượng trong video, Camshift được giới thiệu ở phần trên có thể thấy đối tượng được bám tốt khi nền ảnh không ảnh hưởng đến đối tượng, như lẫn màu hay che khuất đối tượng. Khi đối tượng bám bị che khuất hay lẫn màu với nền ảnh, sẽ gây ra hiện tượng mất vết. Hơn nữa một nhược điểm nữa của thuật toán Camshift là nó phải quét tất cả khu vực trên không gian ảnh để tìm đối tượng tương đương, vì vậy thuật toán chạy rất chậm. Để khắc phục nhược điểm này và nâng cao chất lượng theo vết, bộ lọc Kalman được sử dụng .

Sau khi tiến hành thuật toán Camshift ta sẽ có được trạng thái hệ thống xk, yk, vị trí x và y của đối tượng tại thời điểm k. Sự đánh giá được đề cập phía trên cho ta thấy hạn chế của Camshift, vậy nên ta có thể sử dụng bộ lọc Kalman để tạo ra phương pháp giới hạn vị trí của đối tượng hiệu quả hơn, điều đó có nghĩa là thay vì tìm kiếm đối tượng trong toàn bộ không gian ảnh ta sẽ phải tìm những khu vực tin cậy do bộ lọc Kalman gợi ý.



*Hình II. 10 Mô hình kết hợp bộ lọc Kalman và thuật toán Camshift*

Các bước sử dụng bộ lọc Kalman cho việc bám đối tượng là:

*Bước 1: Khởi tạo (k=0)*

Trong bước này nó sẽ tìm kiếm đối tượng trong toàn bộ bức ảnh do chúng ta không biết vị trí trước của bức ảnh. Theo cách này ta nhận được x0. Tương tự ta có thể đánh giá một lượng sai phương lớn ban đầu (P0).

*Bước 2: Dự đoán( k > 0)*

Trong giai đoạn này sử dụng bộ lọc Kalman ta dự đoán vị trí tương đối của đối tượng, như vậy vị trí xk- được coi như trung tâm tìm kiếm để tìm kiếm đối tượng.

*Bước 3: Hiệu chỉnh (k>0)*

Trong phần này ta đặt đối tượng (là một chuỗi điểm được dự đoán trong trạng thái tiên nghiệm xk và ta sử dụng vị trí thực của nó (đo đạc) để thực hiện việc hiệu chỉnhtrạng thái bằng cách sử dụng bộ lọc Kalman tìm kiếmxk-

Bước 2 và 3 được thực hiện trong khi việc theo vết đối tượng vẫn đang tiến hành.

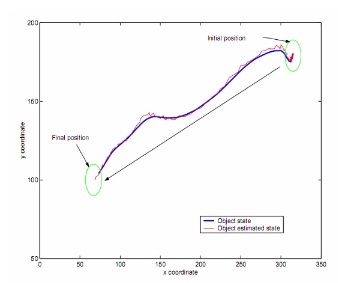
Để minh họa những kết quả của việc dùng bộ lọc Kalman trong việc theo vết đối tượng, chúng tôi lựa chọn cho bám theo một quả bóng và xem xét các trường hợp sau:

1. Trong bài thử nghiệm này nhóm tác giả tham khảo một chương trình thực hiện theo vết một quả bóng, quả bóng sẽ di chuyển tương đương với một đường thẳng tuyến tính,được miêu tả bởi hệ thống các phương trình sau:



Trong sơ đồ trên là kết quả mô phỏng cho thấy vị trí ước lượng bám rất sát với vị trí thực của đối tượng bám





.

*Hình II. 11 Kết quả mô phỏng ước lượng*

#### 2.3.3. Đánh giá

*Việc kết hợp thuật toán CamShift và bộ lọc Kalman cho chúng ta thấy hiệu quả rõ rệt, một lợi thế của bộ lọc Kalman cho việc bám bắt đối tượng là nó có thể ước lượng được vị trí đối tượng ngay cả khi bị che khuất nhỏ. Tuy nhiên, nhược điểm của sự kết hợp này là khối lượng tính toán lớn, do vậy thời gian xử lý lâu hơn, nhưng điều này có thể khắc phục bằng cách tăng tốc độ CPU của máy tính hoặc nhúng chương trình vào các phần cứng chuyên dụng.*

*Một hạn chế khác là khi đối tượng bị che khuất với thời gian đủ lâu (trên 3s), hoặc khi có đối tượng khác chuyển động cùng chiều che khuất dần cũng dẫn đến hiện tượng mất bám. Điều này cũng mở ra hướng phát triển tiếp theo là kết hợp thêm thuật toán* ***Correlation-base template matching*** *sẽ có thể khắc phục được hạn chế này.*

## Bài toán phát hiện té ngã

### *3.1. Đề xuất hướng tiếp cận dựa trên phân tích quỹ đạo chuyển động và phân lớp*

Khái niệm “bất thường” trong giám sát camera là một sự mô hình đa dạng, chúng ta có thể hiểu theo nhiều phương diện khác nhau. Ví dụ như : tốc độ bất thường, vị trí bất thường, hành vi bất thường, chiều cao bất thường… Vậy điều quan trọng của bài toán là phải mô hình hóa một cách chính xác khái niệm bất thường. Trong bài báo cáo này, chúng tôi chọn các tiếp cận phát hiện bất thường dựa vào quỹ đạo chuyển động của đối tượng. Khi theo dõi đối tượng chuyển động chúng ta có thể thu thập các thông tin như độ cao, hình dáng, trọng tâm, vị trí và hướng di chuyển dựa vào quỹ đạo chuyển động. Qua đó chúng ta có thể gom nhóm và phân loại chúng.

Bản chất của phương pháp này là sự phân loại các đối tượng được biểu diễn theo mô hình là các trạng thái chuyển động dựa trên các đặc trưng và gán chúng vào một lớp (gắn nhãn) dựa trên những tập luật hoặc các mẫu chuẩn. Có rất nhiều thuật toán được đề sử dụng như K-Means, K-NN, mạng nơ ron, Bayesian Networks, SVM, PCA.

Sau đây là tóm tắt vài thuật toán tiêu biểu:

|  |  |  |
| --- | --- | --- |
| STT | Tên thuật toán | Tóm tắt |
|  | Mạng nơ ron | Áp dụng nguyên lý của nơ ron thần kinh. |
|  | K-Means | Thuật toán phân cụm, sử dụng khoảng cách ngắn nhất đến mẫu đã huấn luyện. |
|  | K-NN | Sử dụng K khoảng cách ngắn nhất đến dữ liệu huấn luyện để phân lớp. |
|  | Bayesian Networks | Sử dụng nguyên lý xác suất. |
|  | PCA | Kĩ thuật phân tích và trích rút thành phần đặc trưng của đối tượng. |
|  | SVM | Support vector machine – Kĩ thuật phân lớp dựa trên lý thuyết thống kê. |

*Trong bài báo cáo chúng tôi đề xuất phương pháp phân tích và thống kê đặc trưng của đối tượng và dùng thuật kĩ thuật phân lớp SVM. Sau khi phân tích quỹ đạo cũng như các thành phần đặc trưng của đối tượng chúng tôi tiến hành phân lớp cho đối tượng. Ở mức đơn giản nhất, chúng tôi phân thành 2 lớp:*

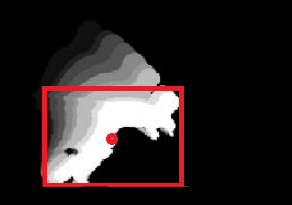
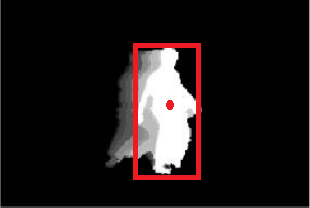
* *Lớp đối tượng ngã*
* *Lớp đối tượng không ngã*

### *3.2. Phát hiện té ngã*

#### a. Mô hình đối tượng bằng hình chữ nhật bao quanh

Trong tất cả các kĩ thuật phát hiện ngã, phần đầu tiên chính là dựa trên sự thu nhận tín hiệu từ thị giác trực tiếp, ví dụ sự thay đổi đột ngột kích thước của đối tượng. Dựa vào tỉ lệ của đường viền của đối tượng là một kĩ thuật đơn giản để phát hiện ngã. Đường viền bao quanh đối tượng tương ứng với một hình chữ nhật được tạo ra khi đốitượng chuyển động và nó cho ta biết vị trí hay hình dáng của đối tượng. Tỉ lệ giữa hai chiều trong hướng x và y của hình chữ nhật bao quanh đối tượng cho chúng ta biết trạng thái của đối tượng đang chuyển động, khi một người ngã, hình chữ nhật bao quanh đối tượng có sự thay đổi đột ngột theo hướng x và y của trục tọa độ.

* Theo hướng dọc và ngang: trong suốt quá trình ngã hướng dọc sẽ ít hơn ban đầu, trong khi đó hướng ngang có thể tăng lên đáng kể.



*Hình II. 14 Khi ngã 2 chiều của hình chữ nhật bao quanh đối tượng thay đổi đột ngột*

* Theo góc ngã: là góc giữa trục của đối tượng và nền, khi người ngã, góc này sẽ có sự thay đổi đột ngột, cụ thể là giảm dần.



*Hình II. 15 Khi ngã góc hợp bởi trục đứng của đối tượng và nền thay đổi đột ngột*

* Theo trọng tâm: khi ngã, trọng tâm của đối tượng hạ xuống một cách đột ngột.



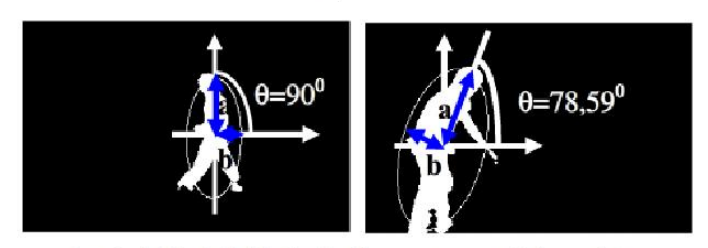
*Hình II. 16 Trọng tâm của đối tượng hạ thấp khi ngã*



*Hình II. 17 Hình ảnh đối tượng ngã*

Từ những phân tích trên chúng tôi nhận thấy rằng các đặc trưng: chiều rộng, chiều cao của hình chữ nhật bao quanh đối tượng, trọng tâm của đối tượng có thể biểu hiện được việc đối tượng có ngã hay không.

#### b.Mô hình đối tượng bằng hình 2D elip



Để xây dựng mô hình 2D elip cho đối tượng, cần tính được các thông số sau:

* Tọa độ trọng tâm elip
* Góc lệch elip so với phương ngang
* Độ dài 2 bán trục a, b của elip

**Xác định tâm Elip:**

+ Hoành độ (tung độ) là trung bình cộng của các hoành độ, tung độ của các điểm ảnh trắng.

Với : - i = (1…Height), j = (1…Width), Height là chiều cao, Width là chiều rộng của khung hình.

* P(i,j) là giá trị nhị phân tại điểm (i,j), P(i, j) = 0 nếu điểm ảnh màu đen, P(i, j ) = 1 nếu điểm ảnh màu trắng.

**Xác định góc quay Elip:**

*tan2Ө* =

Ta có thể dời trục về tâm elip, đồng thời áp dụng công thức:

*Ө = arctan()*

Với : - i = (1…Height), j = (1…Width), Height là chiều cao, Width là chiều rộng của khung hình.

* x = i – Ox, y = j – Oy (Ox, Oy là tọa độ trọng tâm của elip)

**Xác định độ dài 2 bán trục:**

Gọi d1, d2 lần lượt là khoảng cách từ O đến trung điểm nửa trên trục dài O1, và trục ngắn O2:

Tọa độ O1 có hoành độ (tung độ) bằng trung bình cộng của hoành độ (tung độ) của các điểm trắng thỏa mãn: Tung độ lớn hơn tung độ O, hiệu của góc tạo bởi đường thẳng xuất phát từ trọng tâm O đến nó so với phương ngang và góc *Ө* không vượt quá ∆ *Ө*.

Tọa độ O2 có hoành độ (tung độ) bằng trung bình cộng của hoành độ (tung độ) của các điểm trắng thỏa mãn: Tung độ lớn hơn tung độ O, hiệu của góc tạo bởi đường thẳng xuất phát từ trọng tâm O đến nó so với phương ngang và góc *(Ө+)* không vượt quá ∆ *Ө*.

Độ lớn của bán trục dài a và bán trục ngắn b được xác định như sau:

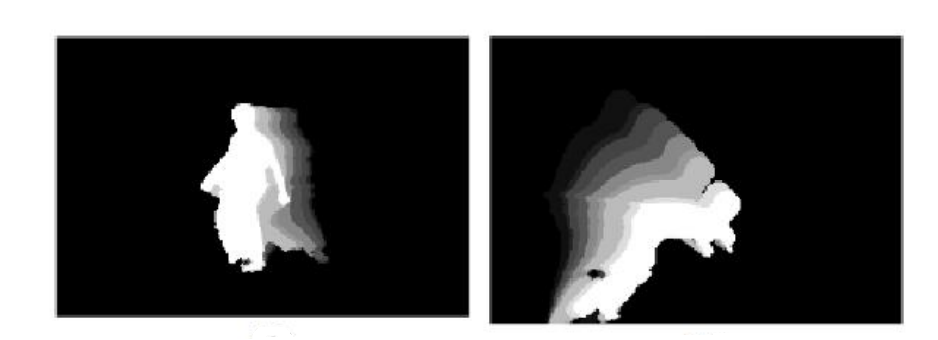
**Độ lệch tâm tức thời của elip:**

Độ lệch tâm tức thời của elip chính là tâm sai của elip tại khung hình đang xét:

Với a, b là độ dài 2 bán trục của elip.

**Tốc độ chuyển động của đối tượng:**

Sau khi chuyển ảnh màu về ảnh xám để phân tích, chúng tôi nhận thấy rằng, với mỗi ảnh xám, mỗi điểm ảnh có giá trị trong đoạn [0, 255], trong đó 255 là giá trị của điểm trắng nhất, 0 là giá trị của điểm đen nhất, “White pixel” là số điểm trắng có giá trị 255, “Gray pixel” là số điểm xám có giá trị nằm trọng đoạn [0, 255]. Tốc độ chuyển động được tính bằng công thức:



*Minh họa về tốc độ chuyển động của đối tượng*

1. Khi đối tượng chuyển động chậm
2. Khi đối tượng chuyển động nhanh (bị ngã đột ngột)

*Dựa trên những phân tích trên chúng tôi lựa chọn phương pháp phân lớp bằng thuật toán SVM, do việc biểu diễn đối tượng bằng mô hình 2D elip khá khó khăn, trong thời gian gấp rút chúng tôi mới chỉ đang trong giai đoạn phân tích, vì thế chúng tôi trích rút 3 đặc trưng chính của đối tượng như sau: chiều cao, chiều rộng hình chữ nhật bao quanh đối tượng và tâm của đối tượng.*

### *3.2. Học bán giám sát và kĩ thuật phân lớp SVM*

Học bán giám sát là một mô hình học máy liên quan đến việc tận dụng các dữ liệu chưa gán nhãn để xây dựng các bộ phân lớp tốt hơn . Học bán giám sát là việc học trên cả dữ liệu đã và chưa được gán nhãn. Từ một số luợng lớn các dữ liệu chưa được gán nhãn, và một luợng nhỏ dữ liệu đã được gán nhãn ban đầu để xây dựng một bộ phân lớp thậm chí là tốt hơn. Trong quá trình học như thế, phương pháp sẽ tận dụng được những thông tin phong phú của dữ liệu chưa gán nhãn mà chỉ yêu cầu một số luợng nhỏ các dữ liệu đã gán nhãn. Như tổng hợp của X.J. Zhu, 2008, học bán giám sát nhận được sự quan tâm đặc biệt của cộng đồng nghiên cứu về học máy. Theo X.J. Zhu, 2008, X. Zhu và A. B. Goldberg, 2009, học bán giám sát có thể chia thành:

• Học giám sát (tất cả các dữ liệu đều đã được gán nhãn) cộng thêm dữ liệu chưa gán nhãn.

• Học không giám sát (tất cả các dữ liệu đều chưa được gán nhãn) cộng thêm dữ liệu gãn nhãn.

Nhiệm vụ chính của học bán giám sát chính là mở rộng tập các dữ liệu gán nhãn ban đầu . Việc học bán giám sát sẽ hữu ích khi chúng ta có nhiều dữ liệu chưa được gán nhãn hơn dữ liệu gãn nhãn. Trong thực tế, tồn tại nhiều dữ liệu không gán nhãn hơn là dữ liệu gán nhãn; bởi vì dữ liệu không gãn nhãn có thể dễ dàng được thu thập với chi phí thấp, nhưng công việc gãn nhãn cho các dữ liệu lại tốn rất nhiều thời gian, công sức và tiền bạc. Chính vì khả năng tận dụng cả dữ liệu đã gán nhãn để đạt được kết quả cao hơn học giám sát nên học bán giám sát mang giá trị thực tiễn cao.

#### 3.2.1. Kĩ thuật phân lớp Support Vector Machine (SVM)

Kĩ thuật phân lớp SVM dựa trên nguyên lý tối thiểu hóa rủi ro cấu trúc (Structural Risk Minimization) trong lý thuyết thống kê (T. Joachims, 1997). Xét bài toán phân lớp văn bản thành các lớp mẫu dương và mẫu âm:

Cho một tập huấn luyện các cặp (xi, yi), i = 1, …, l; trong đó *xi* ∈ Rn là không gian vector đặc trưng n chiều; *yi* ∈ {-1, 1}, các mẫu dương là các mẫu *xi* thuộc lĩnh vực quan tâm và được gán nhãn *yi* = 1 và các mẫu âm là các mẫu *xi* không thuộc lĩnh vực quan tâm và được gán nhãn yi = −1. Bài toán đặt ra là khi đưa ra một vector đặc trưng x mới, cần dự đoán được y sao cho khả năng lỗi xảy ra là tối thiểu.

* **Trường hợp khả tách tuyến tính**

Trong trường hợp này, bộ phân lớp SVM là mặt siêu phẳng phân tách các mẫu dương khỏi các mẫu âm với lề cực đại, được xác định bằng khoảng cách giữa các mẫu dương và các mẫu âm gần mặt siêu phẳng lề tối ưu nhất . Các mặt siêu phẳng trong không gian đối tượng có phương trình là wTx + b = 0, trong đó w là vector pháp tuyến, b là tham số mô hình phân lớp. Khi thay đổi w và b, hướng và khoảng cách từ gốc tọa độ đến mặt siêu phẳng thay đổi.



*Hình II. 3 Mặt siêu phẳng tách các mẫu dương khỏi mẫu âm*

Hàm tuyến tính phân biệt hai lớp như sau:

Trong đó:

* là vector trọng số hay vector chuẩn của siêu phẳng phân cách, T là kí hiệu chuyển vị.
* là độ lệch
* là vector đặc trưng, làm hàm ánh xạ từ không gian đầu vào sang không gian đặc trưng.

Tập dữ liệu đầu vào gồm N mẫu vector {x1, x2,...,xN}, với các giá trị nhãn tương ứng là {t1,…,tN} trong đó .

Nếu *f(x) = +1* thì *x* thuộc về lớp dương và ngược lại, nếu *f(x) = −1* thì *x* thuộc về lớp âm. Mục tiêu của phương kĩ thuật SVM là ước lượng vector trọng số w và hệ số b để cực đại hóa lề giữa các lớp dữ liệu dương và âm. Các giá trị khác nhau của lề cho ta các họ mặt siêu phẳng khác nhau, và lề càng lớn thì lỗi tổng quát hóa của bộ phân lớp càng giảm. Tập dữ liệu huấn luyện là khả tách tuyến tính, ta có các ràng buộc sau:

wTxi + b ≥ +1 nếu yi = +1

wTxi + b ≤ −1 nếu yi = −1

Hai mặt siêu phẳng có phương trình là wT x + b = ±1 được gọi là các mặt siêu phẳng hỗ trợ (các đường nét đứt trên *Hình II.* ***3*** )

Để xây dựng một mặt siêu phẳng lề tối ưu thì:

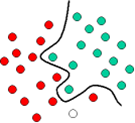
- Vector *w* sẽ được tính: *w =*

- Tham số *b* được xác định sử dụng điều kiện KKT như sau:

*a*i [ yi (wT xi + b) – 1] = 0

Các mẫu *xi* tương ứng với *αi> 0* là những mẫu nằm gần mặt siêu phẳng quyết định nhất và được gọi là các vector hỗ trợ. Những vector hỗ trợ là những thành phần quan trọng nhất của tập dữ liệu huấn luyện. Bởi vì nếu chỉ có các vector hỗ trợ, ta vẫn có thể xây dựng mặt siêu phẳng lề tối ưu như khi có một tập dữ liệu huấn luyện đầy đủ.

* **Trường hợp không khả tách tuyến tính**

****

Với dữ liệu huấn luyện không khả tách tuyến tính thì ta có thể giải quyết theo hai cách. Cách thứ nhất sử dụng một mặt siêu phẳng lề mềm, nghĩa là cho phép một số mẫu huấn luyện nằm về phía sai của mặt siêu phẳng phân tách hoặc vẫn ở vị trí đúng nhưng rơi vào vùng giữa mặt siêu phẳng phân tách và mặt siêu phẳng hỗ trợ tương ứng. Trong trường hợp này, các hệ số Lagrange của bài toán quy hoạch toàn phương có thêm một cận trên C dương - tham số do người sử dụng lựa chọn. Tham số này tương ứng với giá trị phạt đối với các mẫu bị phân loại sai. Cách thứ hai sử dụng một ánh xạ phi tuyến Φ để ánh xạ các điểm dữ liệu đầu vào sang một không gian mới có số chiều cao hơn. Trong không gian này, các điểm dữ liệu trở thành khả tách tuyến tính, hoặc có thể phân tách với ít lỗi hơn so với trường hợp sử dụng không gian ban đầu. Một mặt quyết định tuyến tính trong không gian mới sẽ tương ứng với một mặt quyết định phi tuyến trong không gian ban đầu. Với k là một hàm nhân thoản mãn:

*k(xi, xj) = Φ(xi)T. Φ(xj)*

Nếu chọn một hàm nhân phù hợp, ta có thể xây dựng được nhiều bộ phân loại khác nhau.

Có một số hàm nhân cơ bản sau đây:

- Hàm nhân đa thức: k(xi, xj) = ( *gamma*.xiTxj + *coef0*)*degree*

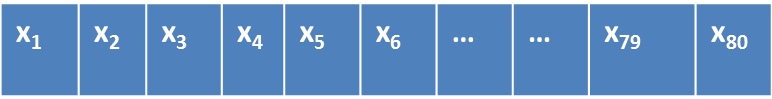
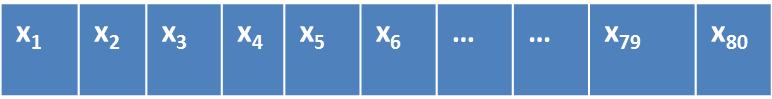
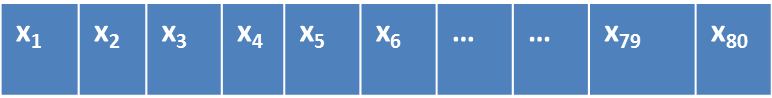
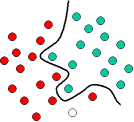
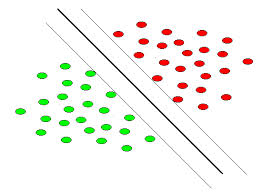
- Hàm vòng RBF (Radial Basic Function): k(xi, xj) = exp (-*gamma*.|xi-xj|2 )

- Hàm chữ S Sigmoid: k(xi, xj) = tan(*gamma*.xiTxj + *coef0*)

trong đó *gamma*, *coef0* và *degree* là các tham số nhân.

#### 3.2.2. Áp dụng thuật kĩ thuật phân lớp SVM giải quyết bài toán phát hiện ngã

Đầu tiên chúng tôi thu thập dữ liệu huấn luyện gồm có các đặc trưng: chiều dài, chiều rộng của hình chữ nhật bao quanh đối tượng, tọa độ tâm của đối tượng trong một thời điểm. Dữ liệu huấn luyện được biểu diễn dưới dạng 1 vector (1 x N, N là số phần tử trong vector gồm 20 frame ảnh liên tiếp, mỗi frame trích rút 4 đặc trưng của đối tượng). Ở đây, chúng tôi định nghĩa 2 lớp: ngã và không ngã.



Y = WTx + b

Lưu đồ thực hiện kĩ thuật phân lớp SVM trong bài toán phát hiện ngã như sau:



Sau khi đưa dữ liệu vào bộ phân lớp SVM, chúng ta sẽ thu được phương trình của siêu phẳng phân tách 2 lớp dữ liệu. Dựa vào đó chúng ta phân lớp cho dữ liệu mới được sinh ra.

## Đánh giá các kĩ thuật áp dụng

Việc sử dụng bộ lọc Kalman làm cho thuật toán Camshift chạy nhanh hơn, tỉ lệ chính xác cao hơn

Khi sử dụng kĩ thuật phân lớp SVM để phát hiện té ngã chúng tôi nhận thấy:

- Kĩ thuật phát huy tác dụng khi khai thác đúng các đặc trưng của đối tượng biểu hiện cho việc té ngã.

- Chương trình chạy nhanh vì quá trình huấn luyện dữ liệu là offline, trong quá trình xử lý online chương trình phải xử lý rất ít dữ liệu, so với thuật toán phân cụm K-NN tốc độ của kĩ thuật SVM nhanh hơn đáng kể.

- Bộ dữ liệu huấn luyện phải chuẩn, nếu sơ sài sẽ gây ra sai số rất lớn.

# 

# Chương III

# THỰC NGHIỆM

**Nội dung chính**

* Điều kiện thực nghiệm
* Kết quả thực nghiệm

## 1. Điều kiện thực nghiệm

### *1.1 Môi trường*

* Các môi trường thông thường như: *sân trường, đường sá,…* với các yếu tố sau:
  + Màu sắc nền và đối tượng có độ tương phản nhất định
  + Các đối tượng di chuyển với mật độ và khoảng cách tương ứng ở mức trung bình
* Yếu tố màu sắc: Các vùng màu sắc có độ tương phản nhất định. Trong trường hợp màu sắc của nền và đối tượng đồng nhất, thì chương trình hoạt động rất kém hiệu quả.

### *1.2 Thiết bị*

* Máy tính
* Thiết bị ghi hình hoặc file Video:
  + File video: Kích thước lớn hơn 600 x 480 pixel

### *1.3 Cấu hình máy*

* CPU tối thiểu: 2.0 GHz
* RAM tối thiểu: 2 GB

### *1.4 Khoảng cách thực nghiệm*

* Camera đặt trước đối tượng với khoảng cách từ 4m - 30m..

## 2. Thực nghiệm

### *2.1. Phương pháp thực nghiệm*

Để thực nghiệm và đánh giá cách tiếp cận dựa trên quỹ đạo chuyển động và sử dụng kĩ thuật phân lớp SVM, chúng tôi sử dụng 2 tập video với độ dài tầm 15s/1 video, tập thứ nhất gồm các video được dàn dựng với các hành động ngã thật, tập video thứ 2 là những cảnh quay các hoạt động thường ngày như ngồi, nằm có khả năng nhầm lẫn với té ngã. Chúng tôi sử dụng các thông số sau để đánh giá độ chính xác của chương trình: Tỉ lệ phát hiện (RC), Độ tin cậy (PR), Độ chính xác (Acc).

Tính bằng các công thức:

RC = , PR = , Acc =

Trong đó:

* True positives – TP là tổng số hành động ngã phân loại đúng.
* False positives – FP là tổng số hành động không ngã phân loại sai.
* True negatives – TN là tổng số hành động không ngã phân loại đúng.
* False negatives – FN là tổng số hành động ngã bị phân loại sai.

Trường hợp có xảy ra té ngã mà hệ thống không phát hiện và không cảnh báo là sai tiêu cực. Ngược lại, nếu đối tượng hoạt động bình thường mà hệ thống lại cảnh báo té là sai tích cực

Có nhiều loại ngã khác nhau như: ngã ngang, ngã trực diện, ngã chéo. Nhưng do thời gian hạn chế nên chúng tôi chỉ mới xây dựng được các cảnh ngã ngang so với camera do vậy có một số trường hợp sẽ bị nhận dạng sai.

Các video xuất phát từ việc tự quay và tổng hợp tại các địa điểm:

* Sân thư viện khu A trường đại học Công nghiệp Hà Nội (quay từ tầng 3 – Nhà A11)
* Sân nhà A12 khu A trường đại học Công Nghiệp Hà Nội (Quay từ tầng 3)

### *2.2. Kết quả thực nghiệm*

Bảng kết quả thực nghiệm.

Từ bảng thực nghiệm cho thấy, tỉ lệ phát hiện ngã chính xác cao ở các video không có nhiễu là các hoạt động thường ngày, vì khâu trích rút đặc trưng chưa chính xác nên tỉ lệ phát hiện nhầm các hoạt động như ngồi, nằm tương đối nhiều.

Dưới đây là một số hình ảnh của kết quả thực nghiệm sự kết hợp bộ lọc Kalman và thuật toán Camshift. Chúng ta có thể nhận thấy nếu không dùng bộ lọc Kalman, khi bám bắt đối tượng, vùng biên bao quanh bị loang to và không tập trung vào đối tượng, khi sử dụng bộ lọc Kalman thì độ chính xác cao hơn.



*Hình III. 1 Các frame chạy thực nghiệm với thuật toán Camshift không dùng bộ lọc Kalman*



*Hình III. 2 Các frame chạy thực nghiệm với thuật toán Camshift kết hợp bộ lọc Kalman*

# KẾT LUẬN & HƯỚNG PHÁT TRIỂN

Sau thời gian thực hiện đề tài, chúng tôi có những kết luấn sau:

* Mô hình được đối tượng nghiên cứu, đưa ra được phương pháp và hướng tiếp cận cho bài toán
* Kĩ thuật phân lớp SVM phát hiện té ngã với tỉ lệ cao với các tập dữ liệu chuẩn, những tập dữ liệu có nhiễu là các hoạt động thường ngày tỉ lệ chính xác còn thấp, nguyên nhân chủ yếu là do khâu trích rút đặc trưng của đối tượng chưa chuẩn xác.
* Bộ dữ liệu huấn luyện còn ít dẫn đến một số trường hợp phát hiện sai.
* Trong điều kiện ánh sáng tương phản mạnh như trời nắng thì hệ thống làm việc kém hiệu quả.
* Ngoài ứng dụng vào lĩnh vực đang nghiên cứu, một số cải tiến như kết hợp bộ lọc Kalman và Camshift có thể giải quyết tốt những vấn đề khác như chương trình phát hiện bàn tay ứng dụng vào mô hình điều khiển từ xa.

**Hướng phát triển**

* Xem xét thời gian xử lý của hệ thống, hướng đến việc tích hợp chương trình vào phần cứng chuyên dụng.
* Thu thập thêm dữ liệu huấn luyện
* Nghiên cứu khâu phân loại đối tượng để không phải lẫn với các đối tượng khác loại.
* Xây dựng mô hình hóa cơ thể người 2D elip hoặc mô hình hóa cơ thể người 3D để tạo chiều sâu cho đối tượng, trích rút các thông tin đặc trưng chi tiết hơn.
* Tiếp cận kĩ thuật tracking muticamera để lấy được đầy đủ thông tin của đối tượng đang theo dõi.
* Phát triển thêm khâu nhận dạng đối tượng: phân tách rõ rệt hơn các hoạt động thường ngày như nằm, ngồi hay là té ngã.

# TÀI LIỆU THAM KHẢO

[]. A. Colombari, A. Fusiello, V. Murino, “*Segmentation and Tracking of Multiple Video Object*”, Dipartimento di Informatica, University of Verona, 2008

[]. A.S. Ogale, C. Fermuller and Y. Aloimonos, “*Motion Segmentation Using Occlusions*”, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.27, No.6, 2005, pp.988-992.

[]. A. Yilmaz, Object Tracking: A Survey, ACM ComputingSurveys, 2006.

[]C. Harris and M. Stephens, A combined corner and edge detector, Alvey Vision Conference, 1988

[] D. G. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision, 2004

[] T. J. Broida and R. Chellappa, Estimation of object motion parameters from noisy images, IEEE Trans. on Pattern Analysis and Machine Intelligence,1986.

[]C. Veenman, M. Reinders, and E. Backer, Resolving motion correspondence

for densely moving points, IEEE Transaction. on Pattern. and Analysis,2001.

[]J. Shi and C. Tomasi, Good features to track, IEEE Conference on ComputerVision and Pattern Recognition., 1994.

[]D. Comaniciu, V. Ramesh, and P. Meer, Kernel-based object tracking, IEEE Transactions on Pattern Analysis an Machine Intelligence, 2003.

[]S. Avidan, Support vector tracking, IEEE Transactions on Pattern Analysisand

Machine Intelligence, 2004.

[]J. Kang, I. Cohen, and G. Medioni, Object reacquisition using invariant

appearance model, International Conference on Pattern Recongnition, 2004.

[] Yaakov Bar-Shalom, X.-Rong Li, Thiagalingam Kiruba, “*Estimation with Applications to Tracking & Navigation*”, A Wiley-Interscience Publication, Canada, 2001

[]M. Isard and A. Blake, Condensation -Conditional Density Propagation for Visual Tracking, International Journal of Computer Vision (IJCV), vol. 29, 1998, pp. 5-28

[] Donald B. Reid, “*An Algorithm for Tracking Multiple Targets*”, IEE Transaction on automatic control, 1979

[] <http://en.wikipedia.org/wiki/Kalman_filter>

[]Trần Quốc Đính - Nghiên cứu bộ lọc Kalman áp dụng cho bài toán cảm biến – TP Hồ Chí Minh tháng 8 năm 2008

[] Yaron Ukrainitz & Bernard Sarel - Mean Shift Theory and Applications

[]<http://en.wikipedia.org/wiki/Mean_shift>

[]N.M Tiến – P.X Minh - Hà Th.Kim Xuyên - (2011)- Hội nghị toàn quốc về Điều khiển và Tự động hoá -VCCA-2011

[] 2010\_Probabilistic framework for multi-target tracking using multi-camera applied to fall detection\_Thesis

[]Yingying Zhu, Nandita M. Nayak, and Amit K. Roy-Chowdhury - Context-Aware Activity Recognition and Anomaly Detection in Video – 2013

[] Subashree, Sobitha Ahila - Video Mining And Anomaly Detection\_Survey - 2014

(2011). A Method using Kalman Filter combining with Image Tracking Camshift Algorithm to bring higher tracking Quality in automatically searching and tracking target Robot System Hội nghị toàn quốc về điều khiển và tự động hóa.