

**TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP HÀ NỘI**  
**KHOA CÔNG NGHỆ THÔNG TIN**

---



**BÁO CÁO**  
**NGHIÊN CỨU KHOA HỌC**

*Đề tài:*

**NGHIÊN CỨU VÀ XÂY DỰNG HỆ THỐNG  
KIỂM SOÁT VÀO RA VỚI SỰ TRỢ GIÚP CỦA CAMERA**

**Các sinh viên thực hiện:** TRỊNH XUÂN ĐỒNG - KHMT1K5  
DƯƠNG LÊ PHÚC NGUYỄN - KTPMK7  
VŨ TIẾN DŨNG - KTPMK7  
MAI XUÂN TIẾN - KTPMK7  
NGUYỄN TIẾN ĐẠT - KTPMK7

**Giảng viên hướng dẫn:** Ths. NGÔ ĐỨC VĨNH

*Hà Nội, 5/2014*

## LỜI NHẬN XÉT

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**Giáo viên hướng dẫn**

# LỜI NÓI ĐẦU

Ngày nay, với sự phát triển của công nghệ thông tin vô cùng mạnh mẽ đã xâm nhập vào hầu hết các mặt của đời sống xã hội trong mọi lĩnh vực như quản lý, an ninh, tư vấn, chống xâm nhập, các hệ thống tự động hóa... Trong đó các hệ thống về lĩnh vực an ninh là những hệ thống quan trọng và cần thiết trong cuộc sống. Có thể kể đến như hệ thống giám sát tự động, hệ thống phát hiện đối tượng đột nhập...

Thuật ngữ “Hệ thống giám sát bằng camera” đã trở nên phổ biến và quen thuộc với chúng ta, được ra đời từ những năm 1960 qua quá trình phát triển và hoàn thiện, hệ thống đã cung cấp và trợ giúp đắc lực cho con người trong việc thực hiện theo dõi, giám sát. Cùng với đó, sự tiến bộ vượt bậc của khoa học, ta có thể xây dựng hệ thống kiểm soát vào ra mà không cần con người phải trực tiếp theo dõi camera, điều này giúp cho con người giảm thiểu được thời gian và sức lực đồng thời vẫn cho một kết quả chính xác. Đối với bài toán đếm số người vào ra có 2 hướng tiếp cận để giải quyết đó là: dựa vào phần cứng (dùng đường tia laser, sóng từ...) và dựa vào kỹ thuật xử lý ảnh. Với kỹ thuật sử dụng xử lý ảnh không những giải quyết được bài toán giám sát mà còn áp dụng được cho rất nhiều lĩnh vực khác trong cuộc sống.

Chính vì vậy, sau một thời gian học tập và nghiên cứu nhóm đã quyết định chọn đề tài: “**Nghiên cứu và xây dựng hệ thống kiểm soát vào ra với sự trợ giúp của camera**”. Đầu tiên nhóm tác giả xin bày tỏ lòng kính trọng và biết ơn chân thành đến thầy giáo ThS.Ngô Đức Vĩnh – trưởng khoa Công nghệ thông tin là người trực tiếp hướng dẫn và tạo điều kiện cho nhóm được xây dựng, phát triển đề tài. Đồng thời nhóm cũng xin chân thành cảm ơn các thầy cô giáo đang giảng dạy tại khoa Công nghệ thông tin – Trường ĐH Công Nghiệp Hà Nội trong suốt những năm học vừa qua đã trang bị những kiến thức nền tảng thiết yếu phục vụ cho quá trình nghiên cứu và tìm hiểu đề tài.

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.

*Hà nội, 05/2014*

*Thay mặt cho nhóm tác giả - Chủ nhiệm đề tài*

**Trịnh Xuân Đồng**

# TÓM TẮT ĐỀ TÀI

Hệ thống đếm người vào,ra là một phần của hệ thống giám sát thông minh. Mặc dù hệ thống phát hiện và đếm số người vào ra đã thu hút được nhiều sự quan tâm, tuy nhiên cần phải có nhiều nghiên cứu hơn nữa để có thể giải quyết các vấn đề còn khó khăn trong lĩnh vực này. Trọng tâm của đề tài chính là phân loại các đối tượng và giám sát chúng trong thời gian thực. Đề tài sử dụng xử lý ảnh để đếm số người vào ra trong tòa nhà với sự trợ giúp của camera.

**Từ khóa:** Đếm số người vào ra, phát hiện đối tượng, theo vết đối tượng, theo dõi video, giám sát video, phân loại đối tượng.

**Abstract:** Object detection and counting systems is one of the system surveillance intelligent. Although object detection and counting systems are commercially available today, but there is need for further research to address the challenges of real world scenarios. The focus of this project the segmentation of groups of object into individuals and tracking them over time. This project using image processing to counting people at office with camera.

**Key word:** Counting people, Object Detection, Object Tracking, video tracking, video surveillance, object classification.

# MỤC LỤC

<b>LỜI NHẬN XÉT .....</b>	<b>2</b>
<b>LỜI NÓI ĐẦU .....</b>	<b>3</b>
<b>TÓM TẮT ĐỀ TÀI .....</b>	<b>4</b>
<b>MỤC LỤC .....</b>	<b>5</b>
<b>DANH MỤC CÁC HÌNH .....</b>	<b>8</b>
<b>DANH SÁCH BẢNG BIỂU.....</b>	<b>10</b>
<b>DANH SÁCH CÁC TỪ VIẾT TẮT .....</b>	<b>11</b>
<b>MỞ ĐẦU .....</b>	<b>12</b>
<b>CHƯƠNG 1. TỔNG QUAN VỀ CÁC VẤN ĐỀ NGHIÊN CỨU .....</b>	<b>15</b>
<b>1.1 Tổng quan về một hệ thống xử lý ảnh.....</b>	<b>15</b>
1.1.1 Khái niệm cơ bản. ....	15
1.1.2 Cấu trúc của hệ thống xử lý ảnh. ....	15
1.1.3 Các vấn đề cơ bản trong xử lý ảnh.....	17
1.1.4 Một số thuật toán xử lý ảnh thông dụng. ....	20
<b>1.2 Tổng quan về video. ....</b>	<b>23</b>
1.2.1 Cấu trúc của video.....	23
1.2.2 Thu nhận, lưu trữ video.....	23
1.2.3 Một số dạng chuẩn nén. ....	23
1.2.4 Các đặc trưng của video.....	25
<b>CHƯƠNG 2. HỆ THỐNG GIÁM SÁT BẰNG CAMERA .....</b>	<b>27</b>
<b>2.1 Tổng quan hệ thống giám sát bằng camera.....</b>	<b>27</b>
2.1.1 Giới thiệu chung.....	27
2.1.2 Ứng dụng của hệ thống giám sát.....	29
2.1.3 Các bài toán cần phải giải quyết. ....	30
<b>2.2 Các phương pháp phát hiện đối tượng chuyển động.....</b>	<b>31</b>
2.2.1: Tổng quan về các phương pháp phát hiện đối tượng chuyển động. ....	31
2.2.2 Phương pháp trừ nền.....	32
2.2.3 Phương pháp thống kê.....	33
2.2.4 Phương pháp luồng quang học.....	34

2.2.5 Phương pháp khác biệt thời gian.....	35
<b>2.3 Các thuật toán theo vết đối tượng chuyển động.....</b>	<b>36</b>
2.3.1 Tổng quan về theo vết đối tượng. ....	36
2.3.2 Thuật toán Meanshift. ....	38
2.3.3 Thuật toán Camshift.....	39
2.3.4 Thuật toán lọc Kalman. ....	40
2.3.5 Theo vết đối tượng bằng so khớp mẫu.....	42
<b>2.4 Các kỹ thuật đường ranh giới ảo.....</b>	<b>43</b>
2.4.1 Kỹ thuật ranh giới ảo một đường.....	44
2.4.2 Kỹ thuật ranh giới ảo hai đường. ....	45
2.4.3 Kỹ thuật ranh giới ảo ba đường.....	46
<b>2.5 Các thuật toán nhận dạng đối tượng.....</b>	<b>46</b>
2.5.1 Tổng quan về bài toán nhận dạng. ....	46
2.5.2 Nhận dạng đối tượng bằng thuật toán K-Means. ....	48
2.5.3 Nhận dạng đối tượng bằng thuật toán KNN. ....	50
<b>CHƯƠNG 3. BÀI TOÁN KIỂM SOÁT VÀO RA BẰNG CAMERA .....</b>	<b>52</b>
<b>3.1 Phát biểu bài toán.....</b>	<b>52</b>
3.1.1 Giới thiệu chung.....	52
3.1.2 Mô hình của hệ thống.....	52
<b>3.2: Các kỹ thuật triển khai.....</b>	<b>53</b>
3.2.1 Thuật toán trừ nền. ....	53
3.2.2 Thuật toán phát hiện các khối. ....	55
3.2.3 Thuật toán xử lý các vùng đối tượng nhỏ. ....	57
3.2.4 Thuật toán theo vết đối tượng. ....	57
3.2.5 Thuật toán ranh giới ảo hai đường. ....	59
3.2.6 Thuật toán KNN.....	60
<b>3.3 Đề xuất, cải tiến. ....</b>	<b>62</b>
3.3.1 Cải tiến vùng quan sát của camera.....	62
3.3.2 Cải tiến phát hiện các đối tượng dính liền nhau.....	63
<b>3.4 Điều kiện thực nghiệm.....</b>	<b>65</b>

3.4.1 Môi trường thực nghiệm. ....	65
3.4.2 Thiết bị thực nghiệm. ....	65
3.4.3 Khoảng cách thực nghiệm.....	65
<b>3.5: Chương trình demo. ....</b>	<b>65</b>
3.5.1 Các bước thực hiện của chương trình. ....	65
3.5.2 Mô hình các lớp quan trọng được cài đặt.....	66
3.5.3 Kết quả thực nghiệm và hình ảnh chương trình.....	67
<b>3.6 Đánh giá các kỹ thuật áp dụng. ....</b>	<b>69</b>
3.6.1 Kết quả đạt được. ....	69
3.6.2 Tồn tại. ....	69
<b>KẾT LUẬN .....</b>	<b>70</b>
<b>TÀI LIỆU THAM KHẢO .....</b>	<b>71</b>
<b>BẢNG ĐỐI CHIẾU THUẬT NGỮ VIỆT – ANH .....</b>	<b>73</b>

# DANH MỤC CÁC HÌNH

Hình 1.1	Quá trình xử lý ảnh .....	15
Hình 1.2:	Các giai đoạn chính trong xử lý ảnh [1] .....	15
Hình 1.3:	Các thành phần chính của hệ thống xử lý ảnh [1].....	16
Hình 1.4:	Biểu diễn ảnh với độ phân giải khác nhau .....	17
Hình 1.5:	Ví dụ về ảnh xám .....	18
Hình 1.6:	Ví dụ về ảnh nhị phân .....	18
Hình 1.7:	Ví dụ về ảnh màu .....	19
Hình 1.8:	Hệ toạ độ màu RGB .....	19
Hình 1.9:	Ví dụ về biểu diễn ảnh màu (hệ màu RGB).....	20
Hình 1.10a:	Ví dụ về chuyển về ảnh xám sử dụng ba ma trận màu .....	20
Hình 1.10b:	Ảnh thực tế khi chuyển đổi ảnh màu về ảnh xám.....	21
Hình 1.11a:	Ví dụ về khử nhiễu .....	21
Hình 1.11b:	Hình ảnh sau khi khử nhiễu .....	22
Hình 1.12:	Ví dụ về ảnh sau khi giãn.....	22
Hình 1.13:	Cấu trúc phân đoạn video .....	23
Hình 1.14:	Một số hình ảnh của định dạng video .....	24
Hình 1.15:	Các hình ảnh khác nhau nhưng có cùng biểu đồ .....	25
Hình 2.1:	Mô hình của hệ thống giám sát .....	28
Hình 2.2:	Mô hình hoạt động của phương pháp trừ nền .....	33
Hình 2.3:	Phương pháp luồng quang học phát hiện chuyển động của ô tô .....	35
Hình 2.4:	Ví dụ phương pháp khác biệt thời gian.....	35
Hình 2.5:	Các khó khăn trong bài toán theo vết đối tượng chuyển động [14].....	37
Hình 2.7:	Mô hình tổng quát của một ước lượng dùng bộ lọc Kalman .....	41
Hình 2.8:	Mô hình ranh giới ảo 1 đường .....	44
Hình 2.9:	Mô hình ranh giới ảo 2 đường .....	45
Hình 2.10:	Mô hình ranh giới ảo 3 đường .....	46
Hình 2.11:	Mô tả các điểm trên mặt phẳng tọa độ.....	49
Hình 3.1:	Mô hình hệ thống giám sát vào ra bằng camera .....	53
Hình 3.2:	Sơ đồ hệ thống phát hiện đối tượng chuyển động .....	54



Hình 3.3: Phát hiện ảnh Foreground .....	55
Hình 3.4: Các phép liên thông.....	56
Hình 3.5: Ví dụ sự khác biệt các phép liên thông .....	56
Hình 3.6: Phép bóc tách các block .....	56
Hình 3.7: Sơ đồ tổng quát thuật toán theo vết đối tượng .....	57
Hình 3.8: Ví dụ về kết nối các đối tượng .....	58
Hình 3.9: Áp dụng ranh giới ảo hai đường cho bài toán.....	59
Hình 3.10: Tập ảnh huấn luyện chứa đầu người .....	61
Hình 3.11: Tập ảnh huấn luyện không chứa đầu người .....	61
Hình 3.12: Phát hiện vùng đầu người bằng thuật toán KNN .....	61
Hình 3.13: Sơ đồ cải tiến vùng quan sát .....	62
Hình 3.14: Vùng quan sát sau cải tiến .....	63
Hình 3.15: Sơ đồ tóm tắt quy trình phát hiện đầu người .....	63
Hình 3.16: Cải tiến khu vực chứa đầu người .....	64
Hình 3.17: Tiền xử lý ảnh trước khi phát hiện đầu người.....	64
Hình 3.18: Các lớp chính được cài đặt trong hệ thống .....	66
Hình 3.19: Form giới thiệu về hệ thống .....	68
Hình 3.20: Form đếm người vào/ra.....	68

## DANH SÁCH BẢNG BIỂU

Bảng 2.1: Bảng phân loại các thuật toán phát hiện đối tượng .....	31
Bảng 2.2 : Một số thuật toán nhận dạng, phân lớp đối tượng .....	47
Bảng 2.3: Danh sách các tọa độ điểm .....	48
Bảng 2.4: Biểu diễn văn bản theo vector .....	51
Bảng 3.1: Kết quả thực nghiệm của hệ thống .....	67

## DANH SÁCH CÁC TỪ VIẾT TẮT

STT	KÝ HIỆU	DIỄN GIẢI
1.	PIXCEL	PICTURE ELEMENT
2.	CGA	COLOR GRAPHIC ADAPTOR
3.	AD	ANALOG TO DIGITAL
4.	RGB	RED - GREEN -BLUE
5.	3GP	3GPP MULTIMEDIA FILE
6.	ASF	ADVANCED SYSTEMS FORMAT FILE
7.	AVI	AUDIO VIDEO INTERLEAVE FILE
8.	MOV	APPLE QUICKTIME MOVIE FILE
9.	MP4	MPEG-4 VIDEO FILE
10.	RM	REAL MEDIA FILE
11.	VOB	DVD VIDEO OBJECT FILE
12.	WMV	WINDOWS MEDIA VIDEO FILE
13.	KNN	K NEAREST NEIGHBOR
14.	PCA	PRINCIPAL COMPONENT ANALYSIS
15.	IIR	INFINITE IMPULSE RESPONSE
16.	ANN	ARTIFICIAL NEURAL NETWORK
17.	CAMSHIFT	CONTINUOUSLY ADAPTIVE MEAN SHIFT

# MỞ ĐẦU

## 1. Lý do chọn đề tài

Ra đời từ những năm 1960 qua quá trình hoàn thiện và phát triển, ngày nay một hệ thống giám sát thông minh, tự động là một trong những hệ thống trợ giúp đắc lực nhất cho con người thực hiện theo dõi, giám sát. Từ các hình ảnh thu được từ những nơi được quan sát, ta có thể phát hiện được chuyển động của các đối tượng trong các khung hình, có thể xác định được đối tượng chuyển động là người hay xe cộ, đồ vật... Minh họa rõ hơn, với một bài toán giám sát giao thông, hệ thống giám sát thông minh có thể cho chúng ta biết được số lượng phương tiện lưu thông qua đoạn đường được theo dõi, đưa ra thông tin về tốc độ chuyển động, đường đi của đối tượng đang được theo dõi. Với một bài toán phát hiện, dự đoán hệ thống giám sát có thể phát hiện một đám cháy, tự động cảnh báo cháy ở nơi được quan sát và theo dõi. Hay với một hệ thống kiểm soát lượng người vào ra trong tòa nhà có thể cung cấp cho chúng ta số lượng các xe máy, xe ô tô, con người vào ra trong tòa nhà...

Hiện nay trên thế giới các hệ thống giám sát thông minh bằng hình ảnh đã được phát triển và đã được chứng minh được hiệu quả nhất định trên một số lĩnh vực như giám sát hoạt động của con người, giám sát giao thông... Hay ở Việt Nam, một số hệ thống được phát triển bởi các công ty phần mềm chuyên biệt về hệ thống giám sát như “Hệ thống giám sát vào/ra – Công ty TNT Việt Nam”, “Hệ thống giám sát vào/ra – Công ty SunMedia”, “Hệ thống giám sát vào/ra – Công ty SRP”,...

Tuy nhiên, các hệ thống nói trên vẫn gặp phải một số tồn tại khó khăn như hiệu quả của việc quan sát luôn phụ thuộc vào điều kiện môi trường quan sát, kiểu chuyển động của đối tượng hay một số lý do khách quan khác. Chính vì thế đây vẫn còn là một lĩnh vực mới, chưa thực sự hoàn thiện và cần phải nhiều công sức nghiên cứu, phát triển hơn nữa nhằm đưa ra được một hệ thống thông minh, tối ưu, chính xác và hoàn thiện nhất.

Xuất phát từ những nhu cầu cũng và lý do trên, nhóm tác giả quyết định chọn đề tài **“Nghiên cứu và xây dựng hệ thống kiểm soát vào ra với sự trợ giúp của camera”**. Với mong muốn được tìm hiểu thêm kiến thức và xây dựng được một hệ thống kiểm soát trong lĩnh vực xử lý ảnh và ứng dụng. Bên cạnh đó sẽ đề xuất một số giải pháp cải tiến những khó khăn tồn tại nhằm góp phần hoàn thiện hệ thống hơn nữa.

## 2. Mục tiêu nghiên cứu

- Tìm hiểu về hệ thống giám sát tự động.
- Tìm hiểu về các kỹ thuật theo dõi đối tượng chuyển động.
- Tìm hiểu về kỹ thuật đường ranh giới ảo áp dụng vào đề tài.
- Tìm hiểu về các thuật toán liên quan đến lĩnh vực xử lý ảnh.

- Nghiên cứu các giải thuật, thuật toán trí tuệ nhân tạo, hệ chuyên gia.
- Tìm hiểu các kỹ thuật nhận dạng đối tượng chuyển động.
- Đề xuất phương pháp cải tiến
- Đánh giá hiệu quả của phương pháp cải tiến.
- Vận dụng xây dựng ứng dụng đếm số người ra vào trong một tòa nhà dưới sự hỗ trợ của camera.

### **3. Phương pháp nghiên cứu**

- Nghiên cứu dựa trên các bài báo, tạp chí khoa học về giám sát tự động, phát hiện đối tượng chuyển động, ranh giới ảo từ đó xây dựng góc nhìn tổng quan về lĩnh vực đang nghiên cứu.
- Từng bước nghiên cứu, đánh giá và đề xuất các kỹ thuật cải thiện cho nội dung nghiên cứu.
- Tham khảo các mã nguồn mở về xử lý ảnh, phát hiện chuyển động và theo dõi chuyển động trên các website như [codeproject.com](http://codeproject.com), [codeplex.com](http://codeplex.com), [sourceforge.net](http://sourceforge.net), [github.com](http://github.com), [emgu.com](http://emgu.com), [opencv.com](http://opencv.com), [msdn.com](http://msdn.com)... Nhằm mục đích tìm hiểu quá trình cài đặt các thuật toán xử lý ảnh cơ bản và nâng cao trên ngôn ngữ lập trình C#.
- Tìm hiểu kiến thức về xử lý ảnh thông qua các giáo trình Xử lý ảnh, các bài viết về xử lý ảnh trên các diễn đàn – song song với việc tham khảo các tài liệu do giáo viên hướng dẫn cung cấp.
- Tìm hiểu, cài đặt các thư viện mã nguồn mở về xử lý ảnh hỗ trợ cho ngôn ngữ C#, qua đó giảm quá thời gian trình xây dựng hệ thống.
- Thực nghiệm dựa trên xây dựng ứng dụng và đánh giá các kết quả nghiên cứu.

### **4. Đối tượng nghiên cứu**

- Nguyên lý của hệ thống giám sát.
- Kỹ thuật phát hiện đối tượng chuyển động trong video.
- Kỹ thuật theo dõi đối tượng chuyển động.
- Đường ranh giới ảo.
- Cải tiến các kỹ thuật ranh giới ảo và miền giám sát.

### **5. Kết quả nghiên cứu**

- Hoàn thiện quá trình nghiên cứu, đánh giá về các kỹ thuật nhận dạng và theo dõi đối tượng chuyển động, ranh giới ảo trong các nghiên cứu trước đó.
- Nắm vững được mô hình và cách thức hoạt động của hệ thống kiểm soát vào ra trong tòa nhà.
- Đề xuất một kỹ thuật cải tiến.
- Xây dựng thử nghiệm đếm người vào ra tòa nhà và đánh giá kết quả thực nghiệm.

- Đánh giá hiệu quả của đề xuất.

## **6. Ý nghĩa khoa học và thực tiễn của đề tài**

### **- Về mặt lý thuyết**

- + Ứng dụng thành công công nghệ xử lý ảnh vào trong thực tế
- + Tạo tiền đề cho những nghiên cứu sau này.

### **- Về mặt thực tiễn**

- + Cung cấp các thuật toán hỗ trợ cho việc phát hiện và kiểm soát đối tượng chuyển động trong video.
- + Giảm giá thành khi triển khai so với hệ thống khác không dùng xử lý ảnh.
- + Thống kê và kiểm soát được lượng người vào, ra trong các tòa nhà, siêu thị. Từ đó có những chiến lược thích hợp trong việc phát triển kinh doanh tương lai.
- + Tạo tiền đề cho các nghiên cứu và ứng dụng sau này.

## **7. Cấu trúc của báo cáo**

- **Chương 1:** Tổng quan về các vấn đề nghiên cứu.
- **Chương 2:** Hệ thống giám sát bằng camera.
- **Chương 3:** Bài toán kiểm soát vào ra bằng camera.

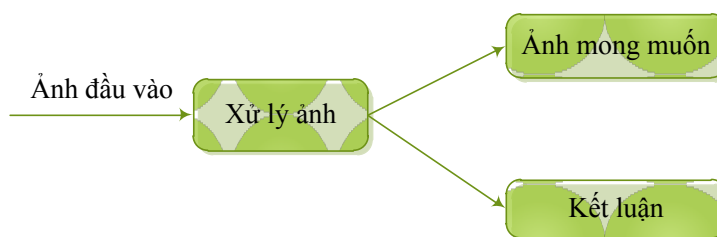
# CHƯƠNG 1. TỔNG QUAN VỀ CÁC VẤN ĐỀ NGHIÊN CỨU

**Tóm tắt nội dung chính:** Trong chương này chúng ta sẽ nghiên cứu một số vấn đề cơ bản liên quan đến xử lý ảnh và video. Các vấn đề được nêu ra trong chương là những khái niệm, thuật toán, cách thức lưu trữ, xử lý ảnh, video. Thông qua đó, tạo tiền đề cho việc phát triển, xử lý các vấn đề khó khăn liên quan đến đề tài đang xây dựng.

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

### 1.1.1 Khái niệm cơ bản.

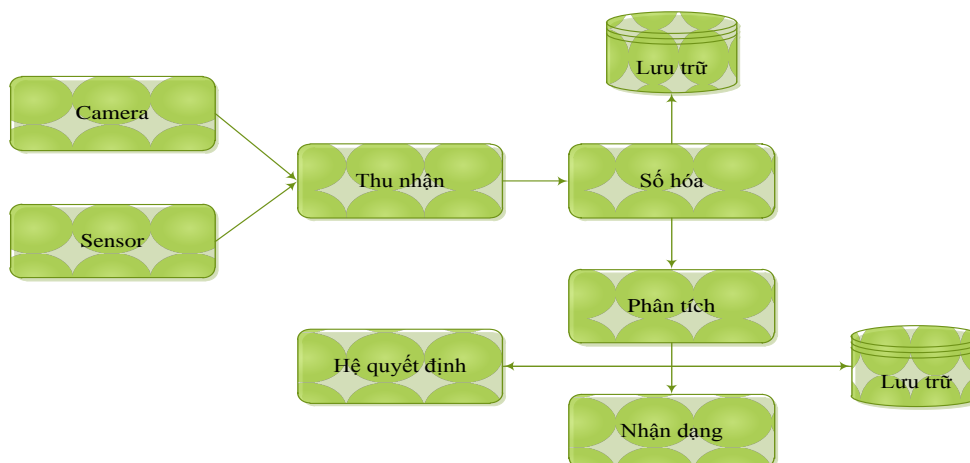
Quá trình xử lý ảnh được xem như là quá trình thao tác với ảnh đầu vào nhằm cho ra kết quả như mong muốn. Kết quả đầu ra của quá trình này có thể là ảnh tốt hơn hoặc một kết luận nào đó. Kết luận ở đây có thể là việc rút trích thông tin, phân loại, nhận dạng ảnh[1].



Hình 1.1 Quá trình xử lý ảnh

### 1.1.2 Cấu trúc của hệ thống xử lý ảnh.

Xử lý ảnh là một ngành khoa học còn tương đối mới mẻ so với nhiều ngành khoa học khác, nhất là trên quy mô công nghiệp, song trong xử lý ảnh đã bắt đầu xuất hiện những máy tính chuyên dụng. 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 1.2):



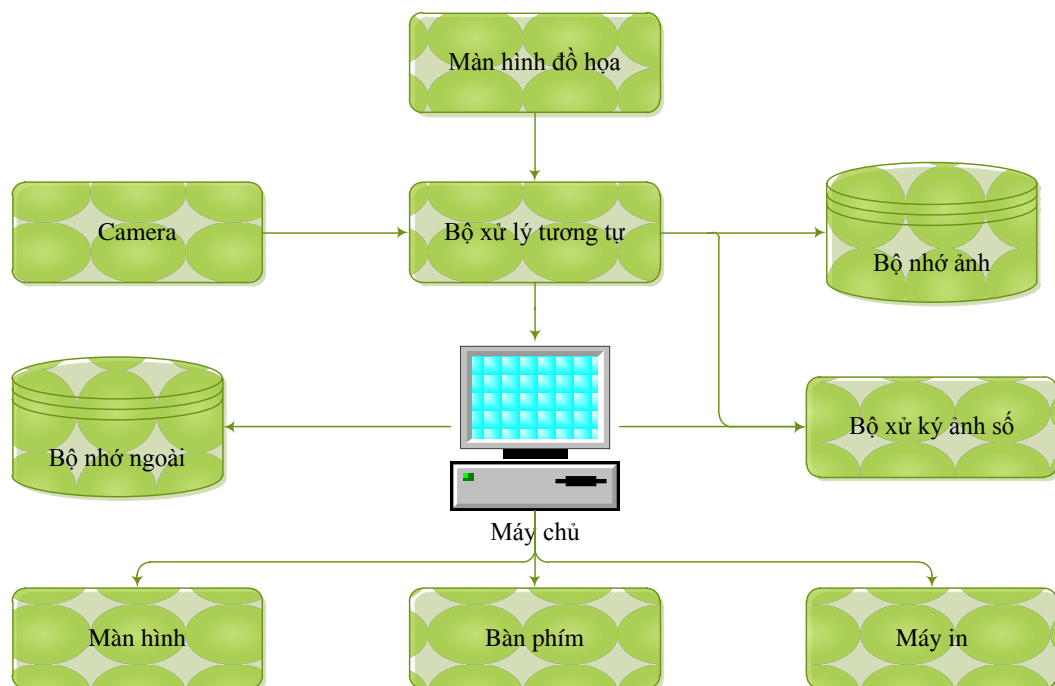
Hình 1.2: Các giai đoạn chính trong xử lý ảnh [1]

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 hiệu 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 nhân khác nhau: Do thiết bị thu nhận ảnh, do 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 phát hiện các đặc tính như phát hiện biên của ảnh, phân vùng ảnh, trích chọn các đặc tính...

Cuối cùng, tùy vào mục đích của ứng dụng, sẽ là giai đoạn nhận dạng, phân lớp hay các quyết định khác.

Với các giai đoạn trên, ta có thể thấy rằng một hệ thống xử lý ảnh cần phải có những thành phần tối thiểu như sau (Hình 1.3):



*Hình 1.3: Các thành phần chính của hệ thống xử lý ảnh [1].*

- Bộ phận camera: Để thu nhận ảnh từ bên ngoài vào hệ thống.
- Bộ xử lý tương tự: Bao gồm các chức năng sau:
  - + Chọn camera thích hợp nếu có nhiều camera.
  - + Chọn màn hình hiển thị tín hiệu.
  - + Thu nhận tín hiệu bởi camera thu nhận. Thực hiện lấy mẫu và số hóa.
  - + Tiền xử lý ảnh khi thu nhận.



- Bộ xử lý ảnh số: Gồm nhiều bộ xử lý chuyên dụng: Xử lý lọc, trích chọn đường bao, nhị phân hóa ảnh...
- Máy chủ: Đóng vai trò điều khiển các thành phần ở trên.
- Bộ nhớ ngoài: Dùng để lưu trữ lại ảnh sử dụng cho các mục đích sau này.

### 1.1.3 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.

#### 1.1.3.1 Một số khái niệm chính.

- **Pixel(Picture Element)** - Điểm ảnh: 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 pixel. Mỗi pixel gồm một cặp tọa độ x,y và màu.



*Hình 1.4: 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 bố, đó chính là độ phân giải và được phân bố theo hai trục x,y trong không gian hai chiều.

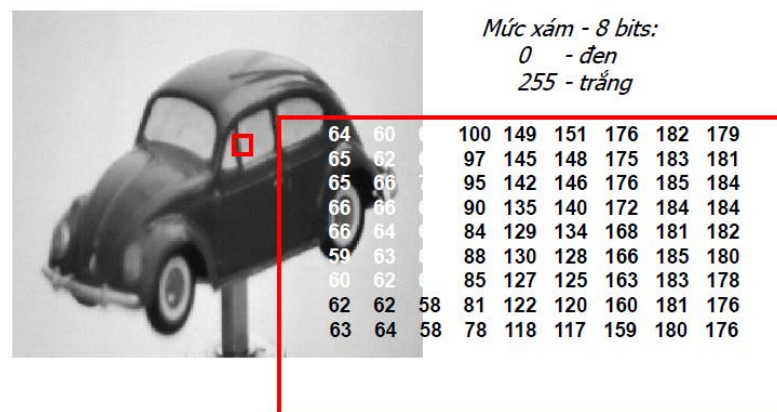
+ Ví dụ: Độ phân giải trên màn hình CGA (Color Graphic Adaptor) là một lưới điểm theo chiều ngang màn hình: 320 điểm theo chiều dọc \* 200 điểm ảnh

(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 bố 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 1.5: 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 1.6: 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.



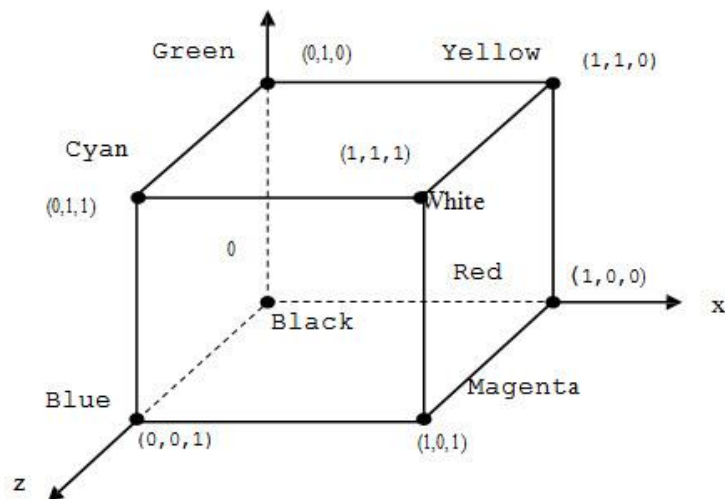
Hình 1.7: Ví dụ về ảnh màu

### 1.1.3.2 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.

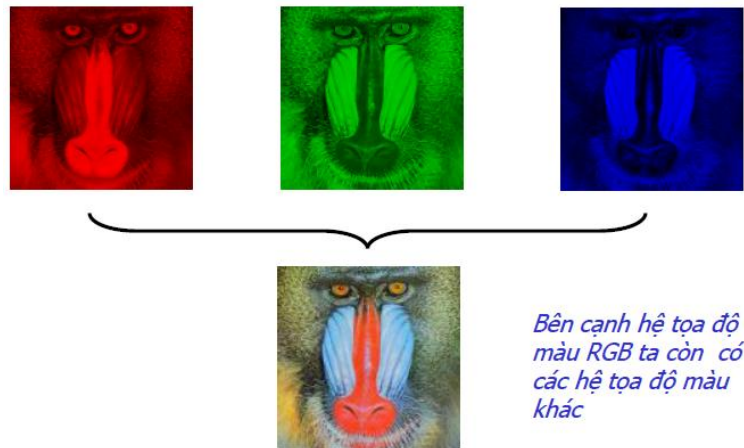
### 1.1.3.3 Biểu diễn ảnh

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 1.8):



Hình 1.8: Hệ tọa độ màu RGB

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 1.9)



Hình 1.9: Ví dụ về biểu diễn ảnh màu (hệ màu RGB)

#### 1.1.4 Một số thuật toán xử lý ảnh thông dụng.

##### 1.1.4.1 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.

R				G				B			
255	0	9	50	1	0	9	50	7	0	9	50
8	74	50	0	50	2	50	90	8	74	255	60
20	30	47	90	20	30	47	90	20	255	47	20
8	74	50	0	255	74	255	0	50	8	50	0

$$R * 0.287 + G * 0.559 + B * 0.114$$

74	0	8	48
31	30	71	57
19	54	45	78
150	63	162	0

Hình 1.10a: Ví dụ về chuyển về ảnh xám sử dụng ba ma trận màu





Ảnh màu



Ảnh xám

Hình 1.10b: Ảnh thực tế khi chuyển đổi ảnh màu về ảnh xám

#### 1.1.4.2 Thuật toán khử nhiễu.

Nhiều ảnh là hiện tượng một 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 đáp ứng trong thời gian thực để giải quyết vấn đề này.

0	0	0
0	255	0
0	0	0

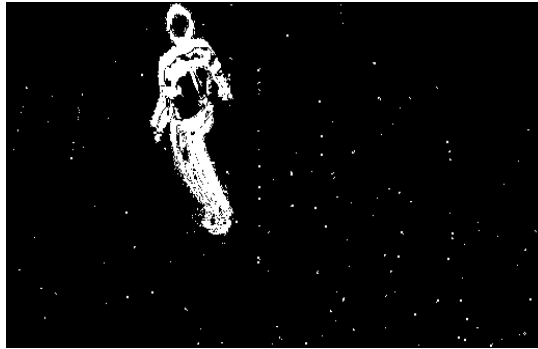
Trước khi khử nhiễu

0	0	0
0	0	0
0	0	0

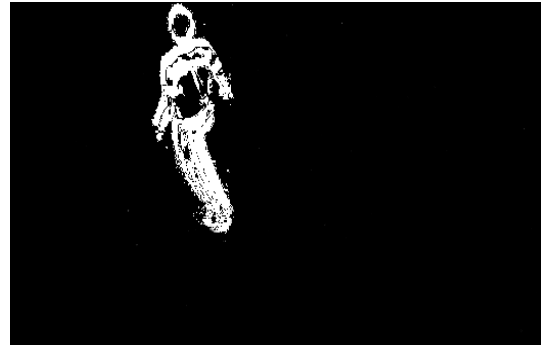
Sau khi khử nhiễu

Hình 1.11a: 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 nhiễu, 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.



a - trước khi khử nhiễu



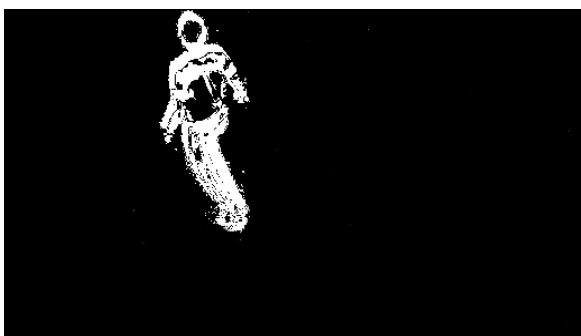
b – sau khi khử nhiễu

*Hình 1.11b: 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.

#### **1.1.4.3 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.



a - Ảnh trước khi giãn



b - Ảnh sau khi giãn

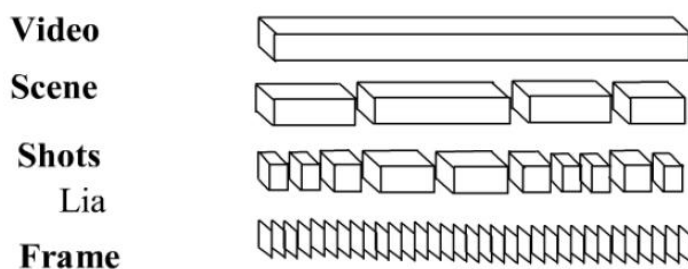
*Hình 1.12: Ví dụ về ảnh sau khi giãn*

## 1.2 Tổng quan về video.

### 1.2.1 Cấu trúc của video.

Video là tập hợp các khung hình, mỗi khung hình là một ảnh. Shot (lia) là đơn vị cơ sở của video. 1 lia là một đơn vị vật lý của dòng video, gồm các chuỗi các khung hình liên tiếp, không thể chia nhỏ hơn, ứng với một thao tác camera đơn.

Scene (cảnh) là đơn vị logic của dòng video, một cảnh gồm các lia liên quan về không gian và thời gian, cùng miêu tả một nội dung ngữ nghĩa hoặc một tình tiết.



Hình 1.13: Cấu trúc phân đoạn video

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

### 1.2.2 Thu nhận, lưu trữ video.

Video được thu nhận chủ yếu qua các thiết bị như camera. Camera sẽ thu thập âm thanh và hình ảnh để được video. Nghĩa là video= âm thanh + hình ảnh. Sau đó, video sẽ được nén hoặc định dạng tùy theo mục đích sử dụng. Việc lưu trữ video phụ thuộc khá nhiều vào định dạng của video. Với cùng 1 video nhưng định dạng khác nhau sẽ có kích thước khác nhau.

### 1.2.3 Một số dạng chuẩn nén.

#### 1.2.3.1 Chuẩn .AVI.

“Audio Video Interleave” [2] - được viết tắt là AVI, là một đa phương tiện định dạng container của Microsoft được giới thiệu vào tháng 11 năm 1992. AVI là tập tin có thể chứa cả âm thanh và video dữ liệu trong một container cho phép đồng bộ tập tin âm thanh-với-video. Cũng giống như các định dạng video DVD, AVI file hỗ trợ nhiều tuyến âm thanh và video, mặc dù các tính năng này ít khi được sử dụng. Hầu hết các tập tin AVI cũng sử dụng định dạng mở rộng tập tin được phát triển bởi Matrox nhóm OpenDML trong tháng hai 1996. Những tập tin này được hỗ trợ bởi Microsoft, và không chính thức gọi là "AVI 2.0"

#### 1.2.3.2 Một số dạng chuẩn khác.



*Hình 1.14: Một số hình ảnh của định dạng video*

Có rất nhiều định dạng video phổ biến ngày nay. Một số định dạng phổ biến như 3GP, AVI, MPEG, RM và WMV. Bên cạnh các định dạng phổ biến đó, còn có một số định dạng ít phổ biến hơn và có thể gây nhầm lẫn. Sau đây là 7 định dạng video mà chúng ta hay sử dụng:

- **3GP – 3GPP MULTIMEDIA FILE** - Định dạng 3GP là định dạng video được phát triển bởi dự án 3rd Generation Partnership. 3GPP và 3GPP2 là các tiêu chuẩn được sử dụng cho mục đích sáng tạo, phân phối và phát lại các tập tin đa phương tiện trên mạng không dây 3G tốc độ cao. Định dạng này được sử dụng phổ biến trên điện thoại có hỗ trợ quay phim. 3GP dựa trên chuẩn MPEG-4 - vốn bắt nguồn từ định dạng QuickTime của Apple.

- **ASF – ADVANCED SYSTEMS FORMAT FILE** - ASF hay Advanced Systems Format File là định dạng được Microsoft phát triển. Định dạng này được dùng để truyền tải các tập tin đa phương tiện chứa văn bản, đồ họa, âm thanh, video và hoạt họa. File ASF chủ yếu là tập tin Windows Media Audio và Windows Media video. Tập tin ASF chỉ đặc tả cấu trúc của file audio hay âm thanh được truyền tải mà không chỉ rõ phương pháp mã hóa.

- **MOV – APPLE QUICKTIME MOVIE FILE** là định dạng được Apple phát triển. Đây là một định dạng đa phương tiện phổ biến, thường được dùng trên Internet do ưu điểm tiết kiệm dung lượng của nó.

- **MP4 – MPEG-4 VIDEO FILE** - Định dạng MPEG-4 là một tiêu chuẩn được phát triển bởi Moving Picture Experts Group, được sử dụng trên rất nhiều điện thoại và các thiết bị chơi video (còn gọi là thiết bị chơi MP4). Định dạng MPEG-4 sử dụng một định dạng nén riêng biệt cho video và audio. Video được nén với kiểu mã hóa video MPEG-4 còn âm thanh thì được nén theo kiểu mã hóa AAC.

- **RM – REAL MEDIA FILE** - Real Media là định dạng được phát triển bởi RealNetworks. Real Media chứa cả thông tin về video (RealVideo) và âm thanh



(RealAudio) và thường được dùng để truyền tải các tập tin đa phương tiện thông qua internet.

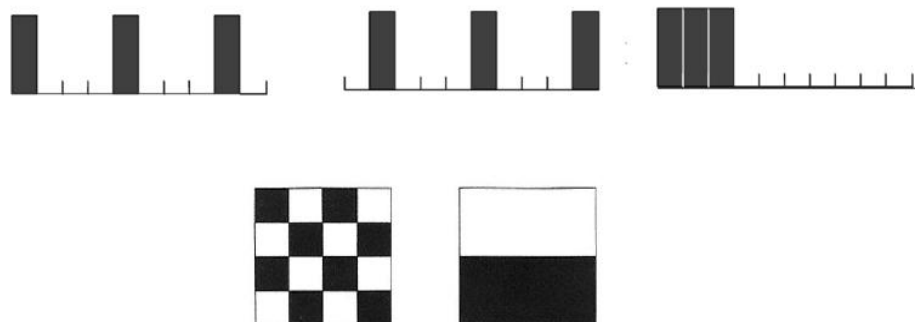
- **VOB – DVD VIDEO OBJECT FILE** - Định dạng VOB chủ yếu liên quan đến *DVD Video Movie File*. Một tập tin VOB thường chứa các luồng đa công (multiplex) gồm: video, âm thanh và phụ đề.

- **WMV – WINDOWS MEDIA VIDEO FILE** - Windows Media Video là một định dạng file chính mà bạn hay gặp hàng ngày. File Windows Media chứa video được mã hóa theo bộ codec Windows Media Video và âm thanh được mã hóa theo codec Windows Media Audio codec.

#### 1.2.4 Các đặc trưng của video.

##### 1.2.4.1 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 cùng biểu đồ màu nhưng nội dung có thể sẽ rất khác nhau.



Hình 1.15: Các hình ảnh khác nhau nhưng có cùng biểu đồ

##### 1.2.4.2 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.

Ngược lại người ta thấy rằng Entropi và Momen chênh lệch nghịch đảo lại có khả năng phân biệt tốt nhất. 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...

#### **1.2.4.3 Hình dạng (shape).**

Các đặc trưng hình dáng có thể được biểu 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.

#### **1.2.4.4 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 được 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.

## CHƯƠNG 2. HỆ THỐNG GIÁM SÁT BẰNG CAMERA

**Tóm tắt nội dung chính:** Trong chương này sẽ đề cập đến hệ thống giám sát sử dụng camera nói chung ở mức tổng quan. Chương cũng đưa ra được ứng dụng và các phương pháp, cách thức thực hiện của hệ thống giám sát. Từ đó nêu ra mô hình của hệ thống. Tạo tiền đề cho việc xây dựng và phát triển “Hệ thống kiểm soát vào ra” sẽ được đề cập đến trong chương sau.

### 2.1 Tổng quan hệ thống giám sát bằng camera.

#### 2.1.1 Giới thiệu chung.

Giám sát tự động là một hướng mới 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. 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.

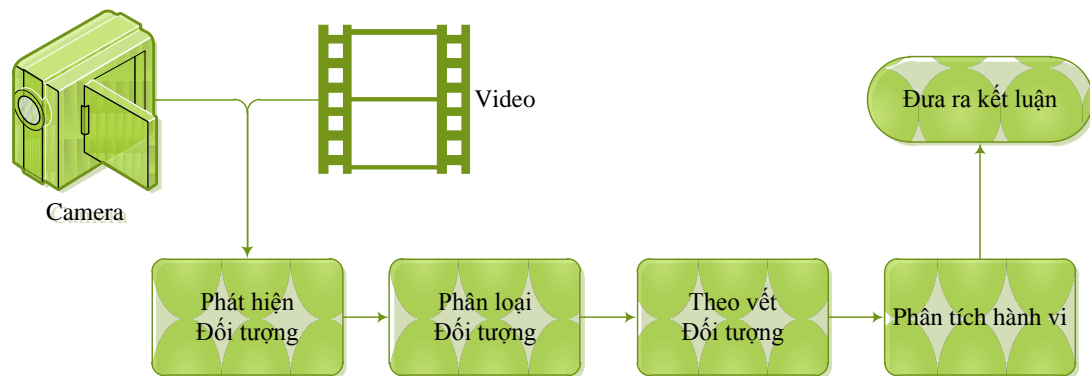
Hệ thống giám sát thế hệ thứ ba (3GSS, 2000) cung cấp cho hệ thống kỹ thuật số điểm tới điểm. Hình ảnh thu thập và xử lý ở các cảm biến, thông tin được truyền thông qua các mạng di động băng thông rộng và cố định không đồng nhất và hình ảnh được lưu trữ tại các máy chủ trung tâm.

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

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 công hiến cho sự phát triển của các thuật toán thông minh[3].

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 2.1: Mô hình của hệ thống giám sát*

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[3]. 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, PCA.
- **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 hướng chuyển động của chúng
  - + Dữ liệu ra: Danh sách các đối tượng kèm theo hành vi được định nghĩa từ trước: VD: Vào, ra...
  - + Vai trò: Module sẽ dựa vào hướng cũng như cách chuyển động của đối tượng, từ đó kết luận và phân tích xem thực ra đối tượng đó đang làm gì?
  - + Phương pháp sử dụng: Các phương pháp liên quan đến hình học.
- **Module đưa ra kết luận:**
  - + Phụ thuộc vào từng hệ thống cụ thể mà module sẽ có vai trò và đầu ra cụ thể.
  - + Ví dụ: Hệ thống giám sát vào ra có đầu ra là số người vào ra, và phương pháp sử dụng là đương ranh giới ảo, với hệ thống phát hiện đột nhập thì đầu ra là số lượng đối tượng đột nhập, vị trí và có thể là khuôn mặt

### 2.1.2 Ứng dụng của hệ thống giám sát.

- 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
  - + Kiểm soát truy cập
- 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
  - + Đếm loài có nguy cơ tuyệt chủng
- Thực thi pháp luật:

- + Đo tốc độ xe
- + Phát hiện vượt đèn đỏ và lấn làn đường
- An ninh quân sự:
  - + Tuần tra biên giới quốc gia
  - + Đo dòng người tị nạn
  - + Giám sát các điều ước hòa bình quốc tế
  - + Cung cấp các khu vực an toàn xung quanh các căn cứ quân sự
  - + Hỗ trợ kiểm soát chiến trường

### 2.1.3 Các bài toán cần phải 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, giao diện người và máy, 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.

## 2.2 Các phương pháp phát hiện đối tượng chuyển động.

### 2.2.1: Tổng quan về các phương pháp phát hiện đối tượng chuyển động.

Vấn đề phát hiện đối tượng đang được nghiên cứu và có nhiều ứng dụng trong cuộc sống. Các đối tượng được phát hiện nhờ những thông tin trong một frame ảnh. Có rất nhiều hướng tiếp cận để giải quyết vấn đề trên. Các tác giả Alper Yilmaz, Omar Javed và Mubarak Shah đã phân loại các hướng tiếp cận này được trình bày [4]:

*Bảng 2.1: Bảng phân loại các thuật toán phát hiện đối tượng*

Loại	Những nghiên cứu có liên quan
<b>Point detectors</b>	<ol style="list-style-type: none"> <li>1. Moravec's detector</li> <li>2. Harris detector</li> <li>3. Scale Invariant Feature Transform Affine</li> <li>4. Invariant Point Detector</li> </ol>
<b>Segmentation</b>	<ol style="list-style-type: none"> <li>1. Mean-shift</li> <li>2. Graph-cut</li> <li>3. Active contours</li> </ol>
<b>Background Modeling</b>	<ol style="list-style-type: none"> <li>1. Mixture of Gaussians</li> <li>2. Eigenbackground</li> <li>3. Wall flower</li> <li>4. Dynamic texture background</li> </ol>
<b>Supervised Classifier</b>	<ol style="list-style-type: none"> <li>1. Support Vector Machines</li> <li>2. Neural Networks</li> <li>3. Adaptive Boosting</li> </ol>

Việc lựa chọn phương pháp áp dụng phải dựa vào tình huống cụ thể, đối với trường hợp có ảnh nền không thay đổi việc phát hiện đối tượng chuyển động có thể bằng các phương pháp trừ nền. Các giải thuật này sẽ được trình bày sau đây. Hướng giải quyết là xây dựng mô hình nền, sau đó sử dụng mô hình này cùng với frame hiện tại để rút ra được các foreground chuyển động. Để có thể tiếp cận cần phải xây dựng được mô hình background. Có nhiều phương pháp xây dựng mô hình background bởi các tác giả: Anurag Mittal dùng adaptive kernel density estimation được tính bằng[5]. Kết quả tốt tuy nhiên khó khăn về không gian lưu trữ, tính toán

phức tạp, tốc độ không đáp ứng thời gian thực. Haritaoglu dùng giải thuật W4, Stauffer sử dụng Mixture of Gaussian [6] để xây dựng mô hình nền... Nhằm phát hiện được các đối tượng chuyển động, xác định xem những đối tượng này có đúng là những đối tượng ta cần phát hiện hay không. Đây là các khó khăn cần khắc phục.

Trong các lĩnh vực về phát hiện phần đầu của người thì Wei Qu, Nidhal Bouaynaya and Dan Schonfeld đề ra hướng tiếp cận bằng cách kết hợp mô hình màu da cùng với mô hình màu tóc (skin and hair color model). Những màu này được phát hiện dựa vào mô hình Gauss. Sau đó bằng cách áp dụng phương pháp so khớp mẫu (template matching) để đạt được mục đích phát hiện phần đầu người đáp ứng thời gian thực. Khó khăn trong hướng tiếp cận này thường gặp ở việc thu thập dữ liệu huấn luyện màu da và màu tóc, độ chính xác dễ bị ảnh hưởng bởi độ sáng của môi trường.

Việc phát hiện đối tượng có thể được thực hiện bằng các phương pháp máy học. Các phương pháp này có thể kể đến như: mạng neural, adaptive boosting, cây quyết định, support vector machines. Điểm chung của các phương pháp này đều phải trải qua giai đoạn huấn luyện trên một tập dữ liệu. Tập dữ liệu này phải đủ lớn, bao quát hết được các trạng thái của đối tượng. Sau đó các đặc trưng sẽ được rút trích ra trên bộ dữ liệu huấn luyện này. Việc lựa chọn đặc trưng sử dụng đóng vai trò quan trọng ảnh hưởng đến hiệu quả của các phương pháp máy học. Một số đặc trưng thường được sử dụng như: đặc trưng về màu sắc, đặc trưng về góc cạnh, đặc trưng histogram... Sau khi đã có được đặc trưng, ta sẽ đánh nhãn lớp cụ thể cho các đặc trưng đó để sử dụng trong việc huấn luyện. Trong quá trình huấn luyện, các phương pháp máy học sẽ sinh ra một hàm để ánh xạ những đặc trưng đầu vào tương ứng với nhãn lớp cụ thể. Sau khi đã huấn luyện xong thì các phương pháp máy học trên sẽ được dùng để phân lớp cho những đặc trưng mới. Đặc điểm của phương pháp này là độ chính xác cao. Tuy nhiên nó gặp phải khó khăn trong việc thu thập dữ liệu huấn luyện ban đầu, tốn thời gian và chi phí cho quá trình học máy.

### **2.2.2 Phương pháp trừ nền.**

Trừ nền là một kỹ thuật thường được sử dụng cho việc phân đoạn chuyển động trong những cảnh tĩnh. 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 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[7] 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  $I_t$  thuộc về foreground nếu:  $|I_t(x,y) - B_t(x,y)| > \tau$  được thỏa mãn trong đó  $\tau$  là một ngưỡng được xác định trước. Hình nền  $B_t$  được cập nhật bằng cách sử dụng bộ lọc Infinite Impulse Response (IIR) như sau:

$$B_{t+1} = \alpha I_t + (1 - \alpha) B_t \quad (2.1)$$

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



a-ảnh nền



b-ảnh chứa đối tượng



c-phát hiện sự sai khác

*Hình 2.2: Mô hình hoạt động của phương pháp 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. Mặc dù kỹ thuật trừ thực hiện rất tốt việc tách được hầu hết của các điểm ảnh có nằm trong vùng di chuyển ngay cả khi đối tượng dừng lại, nhưng lại thường nhạy cảm với những thay đổi đột ngột, ví dụ, đối tượng đứng yên ra khỏi nền (ví dụ như một chiếc xe đậu di chuyển ra khỏi bãi đậu xe) hoặc ánh sáng bị thay đổi đột ngột.

### **2.2.3 Phương pháp thống kê.**

Một phương pháp tiên tiến hơn là việc sử dụng các đặc tính thống kê của các điểm ảnh riêng lẻ đã được phát triển để khắc phục những thiếu sót của phương pháp trừ nền cơ bản. Những phương pháp thống kê chủ yếu được lấy tư tưởng từ các

phương pháp trừ nền trong việc lưu giữ và tự động cập nhật các số liệu thống kê của các điểm ảnh thuộc quá trình xử lý nền. Các điểm ảnh foreground được xác định bằng cách so sánh số liệu thống kê của mỗi điểm ảnh với mô hình nền. Cách tiếp cận này đang trở nên phổ biến hơn do độ tin cậy của nó trong những cảnh chứa các thay đổi tiếng ồn chiếu sáng, và bóng tối [8].

Các hệ thống W4 [9] sử dụng một mô hình nền thống kê mà mỗi điểm ảnh được biểu diễn bởi giá trị cường độ tối thiểu (M) và giá trị cường độ tối đa (N) của nó và sự khác biệt cường độ tối đa (D) giữa bất kỳ khung hình liên tiếp quan sát trong suốt quá trình huấn luyện ban đầu mà cảnh không có các đối tượng chuyển động. Một điểm ảnh trong  $I_t$  thuộc về foreground nếu thỏa mãn:

$$|M(x,y) - I_t(x,y)| > D(x,y) \text{ or } |N(x,y) - I_t(x,y)| > D(x,y) \quad (2.2)$$

Sau khi tách ngưỡng, thực hiện một số lần phép toán hình thái xói mòn cho ảnh foreground để loại bỏ các điểm ảnh nhiễu. Để phát triển các khu vực bị xói mòn về kích thước ban đầu, nối tiếp sau phép xói mòn là thực hiện phép giãn nở trên ma trận các điểm ảnh foreground. Ngoài ra, khu vực có kích thước nhỏ được loại bỏ. Sau đó áp dụng dán nhãn các thành phần được ghép nối để phát hiện các vùng. Số liệu thống kê của các điểm ảnh nền thuộc các khu vực không chuyển động của hình ảnh hiện tại được cập nhật với các dữ liệu hình ảnh mới.

Một ví dụ về phương pháp thống kê, Stauffer và Grimson[10] mô tả một mô hình nền hỗn hợp thích ứng để theo dõi trong thời gian thực. Trong công việc của họ, mỗi điểm ảnh riêng biệt được mô hình hóa bằng một hỗn hợp của Gaussians được cập nhật liên tục từ dữ liệu hình ảnh hiện tại. Để phát hiện xem một pixel thuộc foreground hay background, sẽ dựa vào đánh giá phân phối Gaussian của mô hình hỗn hợp cho điểm ảnh đó.

#### **2.2.4 Phương pháp luồng quang học.**

Phương pháp luồng quang học dựa vào việc sử dụng của các vector luồng của vật thể chuyển động theo thời gian để phát hiện các khu vực di chuyển trong một hình ảnh. Phương pháp này có thể phát hiện chuyển động trong chuỗi video từ một máy ảnh chuyển động, tuy nhiên, hầu hết các phương pháp luồng quang học tính toán phức tạp và không thể được sử dụng thời gian thực mà không cần phần cứng chuyên biệt [12].



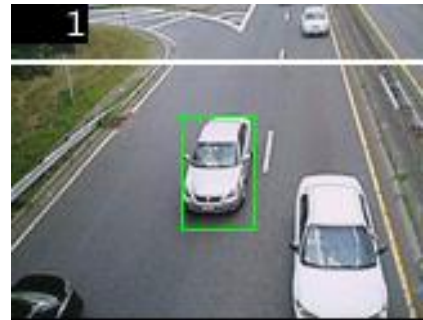
a-Frame video hiện tại



b-Sau khi áp dụng phương pháp luồng quang học



c-Phát hiện các khối di chuyển

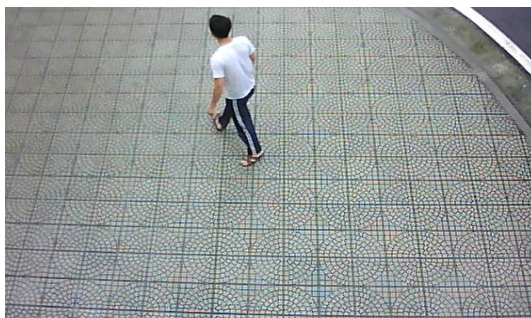


d-Khoanh vùng các khối chuyển động

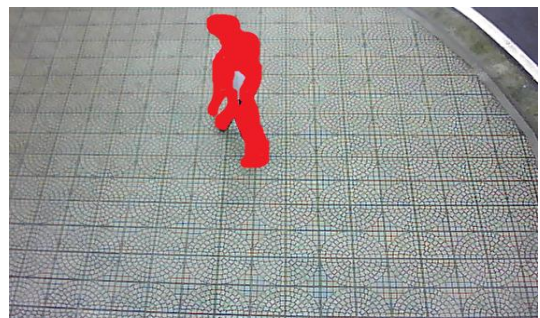
*Hình 2.3: Phương pháp luồng quang học phát hiện chuyển động của ô tô*

### 2.2.5 Phương pháp khác biệt thời gian.

Khác biệt thời gian thực hiện việc phát hiện các khu vực di chuyển bằng cách sử dụng các điểm ảnh khác biệt của các khung hình liên tiếp (hai hoặc ba) trong một chuỗi ảnh. Phương pháp này thích nghi rất tốt với những thay đổi trong những cảnh động, tuy nhiên, nó thường mắc lỗi trong việc phát hiện toàn bộ các điểm ảnh của một số loại đối tượng chuyển động.



a-Frame ảnh hiện tại



b-Frame ảnh khác biệt sau vài khung hình

*Hình 2.4: Ví dụ phương pháp khác biệt thời gian*

Khu vực giữa của hình 2.4b thể hiện nhược điểm của thuật toán khác biệt thời gian trong việc phát hiện tất cả các điểm ảnh của khu vực di chuyển. Ngoài ra,

phương pháp này không thể phát hiện các đối tượng dừng lại. Phương pháp cần được bổ sung để cải thiện khả năng phát hiện các đối tượng dừng lại.

Lipton et al. trình bày một phương thức khác biệt hai-frame mà các điểm ảnh mà thỏa mãn phương trình sau đây được đánh dấu như foreground.[11]

$$|I_t(x,y) - I_{t-1}(x,y)| > \tau \quad (2.3)$$

Để khắc phục thiếu sót của phương pháp khác biệt hai khung hình trong một số trường hợp, có thể sử dụng sự khác biệt của ba khung hình[18]. Ví dụ, Collins et al. phát triển một phương pháp lai kết hợp khác biệt ba khung hình với một mô hình thích ứng trừ nền cho dự án VSAM của họ [13]. Các thuật toán lai đã thành công trong việc phân đoạn vùng di chuyển trong video mà không có các lỗi của khác biệt thời gian và trừ nền.

## **2.3 Các thuật toán theo vết đối tượng chuyển động.**

### **2.3.1 Tổng quan về theo vết đối tượng.**

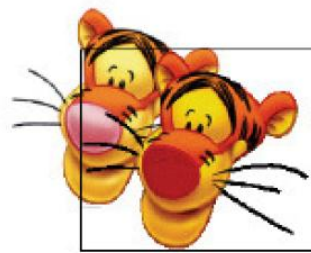
Theo vết đối tượng (Object Tracking) là bài toán quan trọng trong lĩnh vực thị giác máy tính. Do nhu cầu phân tích video ngày càng cao cũng với khả năng tính toán của máy tính cũng được nâng cao, các thuật toán theo vết đối tượng cũng thu hút được rất nhiều sự quan tâm.

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. Chính vì vậy, các lĩnh vực mà theo vết đối tượng có thể sử dụng tới là rất đa dạng và phong phú, trong số đó có thể liệt kê ra một số ứng dụng điển hình như sau:

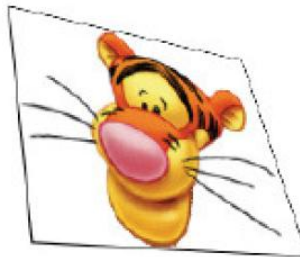
- Nhận dạng dựa trên chuyển động: Ví dụ nhận dạng con người dựa trên dáng đi, tự động phát hiện đối tượng.
- Giám sát tự động.
- Đánh chỉ mục video, tự động chú thích và truy vấn cơ sở dữ liệu video.
- Tương tác giữa người và máy: ví dụ nhận dạng ký hiệu người cầm.
- Giám sát giao thông, thu thập thông tin giao thông trong thời gian thực để phân tích luồng giao thông.
- Hệ thống lái xe, ví dụ lập lịch đường đi, tránh chướng ngại vật.

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 các đối tượng frame của 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. 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
- Nhiều
- 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....



(a) Chuyển động



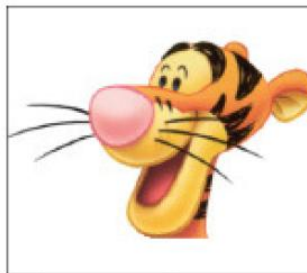
(b) Biến đổi hình học



(c) Nền hỗn loạn



(d) Che khuất đối tượng



(e) Thay đổi hình dạng



(f) Thay đổi độ sáng

Hình 2.5: Các khó khăn trong bài toán theo vết đối tượng chuyển động [14]

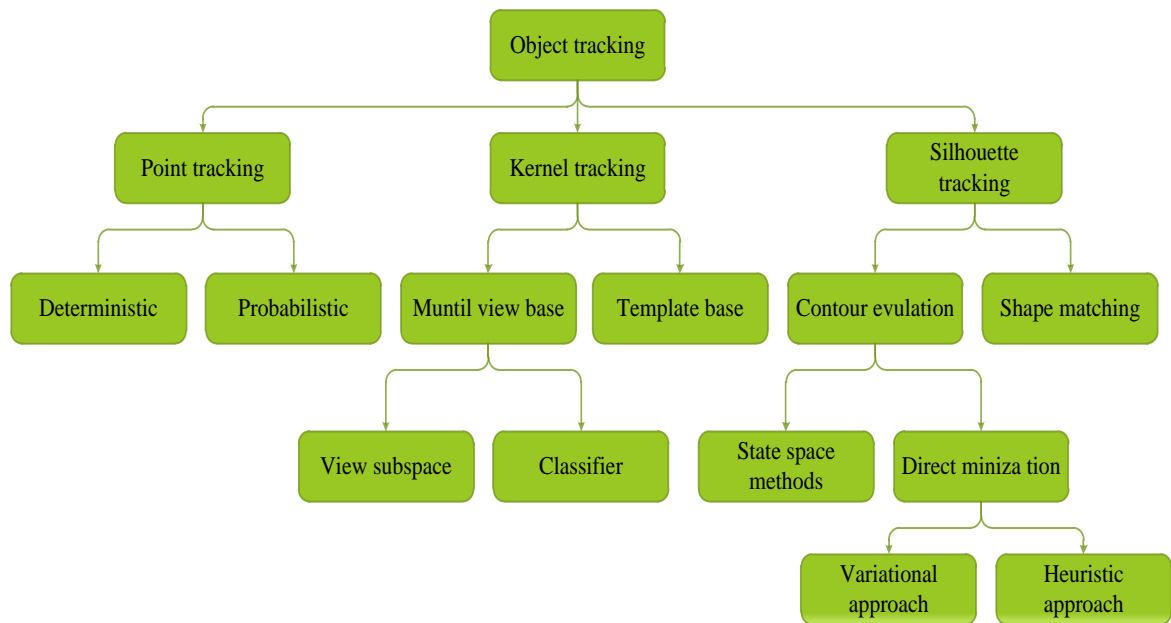
Có rất nhiều thuật toán theo vết đối tượng đã được đề xuất. Mỗi thuật toán dựa trên những giả thiết và ứng dụng khác nhau. Các thuật toán có thể được phân loại dựa trên cách mô hình hóa đối tượng được quan tâm, các loại đặc trưng dùng để phân biệt đối tượng đó. Trong [15] các phương pháp theo vết đối tượng được phân loại thành ba nhóm chính dựa trên mô hình đối tượng được sử dụng như hình bên dưới:

- Theo vết dựa trên điểm – point tracking : Mỗi đối tượng được quan tâm được biểu diễn bằng một tập các điểm. Các điểm này có thể được phát hiện trong mỗi Frame bằng một bộ phát hiện điểm (point detector), chẳng hạn như [16],[17] 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[18], Greedy Optimal Assignment [19].



- Theo vết dựa trên nhân – 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[20], KLT[21], SVT[22].

- Theo vết dựa trên bóng – 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)[23], hoặc tiến hóa đường viền (contour evolution) [24].



Hình 2.6: Phân loại các phương pháp theo vết đối tượng

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, nhóm tác giả 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 thuật toán cụ thể áp dụng cho module theo vết đối tượng của đề tài.

### 2.3.2 Thuật toán Meanshift.

Dorin Comaniciu đã giới thiệu phương pháp theo vết màu Meansshift [25]. Đây là một phương pháp theo vết tối ưu hóa tối thiểu cục bộ. Mỗi vị trí  $x_i$  trong miền ứng viên của theo vết sẽ tương ứng với một trọng số  $w_i$ .

$$w_i = \sum_{u=1}^m \sqrt{\frac{q_u}{p_u(y_0)}} \delta(b(x_i) - u) \quad (2.4)$$

Với  $b(x_i)$  là giá trị màu tại  $x_i$  và  $q_u$  là giá trị màu tại  $u$  của mô hình đích, và  $p_u(y_0)$  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:

$$\hat{y}_1 = \frac{\sum_i x_i w_i g\left(\left\|\frac{y_0 - x_i}{h}\right\|^2\right)}{\sum_i w_i g\left(\left\|\frac{y_0 - x_i}{h}\right\|^2\right)} \quad (2.5)$$

Với  $g(x) = -k'(x)$  và  $k(x)$  là một hàm nhân (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  $\{q_u\}_{u=1..m}$  và vị trí của nó trong khung hình trước  $y_0$

**Bước 1:** Khởi tạo vị trí của đích trong khung hiện tại với  $y_0$ , tính mô hình ứng viên  $\{p_u(y_0)\}_{u=1..m}$  và tính  $p[p(y_0), q] = \sum_{u=1}^m \sqrt{p_u(y_0) q_u}$

**Bước 2:** Tính được các trọng số  $w_i$  theo

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

**Bước 4:** Tính ra mô hình mới  $\{p_u(y_1)\}_{u=1..m}$  và tính được

$$p[p(y_1), q] = \sum_{u=1}^m \sqrt{p_u(y_1) q_u}$$

**Bước 5:** Trong khi  $p[p(y_1), q] < p[p(y_0), q]$ , thực hiện  $y_1 \leftarrow \frac{1}{2}(y_1 + y_0)$

**Bước 6:** Nếu  $|y_1 - y_0| < \varepsilon$  thì dừng, nếu không  $y_0 \leftarrow y_1$ , tiếp tục bước 2

Meanshift là một trong phương pháp đơn giản và hiệu quả cho theo vết thời gian thực. Nhưng nó chỉ tối ưu cục bộ chứ không toàn cục. Khi màu nền và màu đối tượng giống nhau, phương pháp này sẽ không còn tác dụng.

### 2.3.3 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 thích ứng các bước thay đổi kích cỡ của vùng. Các bước thực thi thuật toán:

**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 đến khi độ dịch chuyển vị trí tâm nhỏ hơn một ngưỡng đặt trước).

### **2.3.4 Thuật toán lọc Kalman.**

#### **2.3.4.1 Tổng quan**

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 50 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 có thể vẫn luôn còn được nhiều ứng dụng mới nữa ở tương lai (trừ khi có một bộ lọc hay giải pháp nào đó tốt hơn ra đời).

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).

#### **2.3.4.2 Ứng dụng của bộ lọc Kalman.**

Bởi vì bộ lọc Kalman giải quyết một vấn đề cơ bản là lọc nhiễu và tối ưu cho các ước lượng nên nó được áp dụng rất rộng rãi.

Một số ứng dụng được liệt kê sau đây:

- Lái tự động của máy bay (Autopilot)
- Ước lượng trạng thái của Pin
- Ước lượng trạng thái sạc của pin (Battery state of charge (SoC) estimation)
- Giao diện tương tác với máy tính bằng não (Brain–computer interface)
- Định vị chuyển động (Dynamic positioning)
- Hệ thống dẫn đường quán tính (Inertial guidance system)
- Theo dõi bằng radar (Radar tracker)
- Hệ thống định vị vệ tinh (Satellite navigation systems)
- Dự báo thời tiết (Weather forecasting)
- Hệ thống định vị (Navigation Systems)
- Mô hình hóa 3 chiều (3D-Modelling)
- Ứng dụng lọc Kalman trong phân tích biến dạng nhà cao tầng do bức xạ nhiệt mặt trời.
- Cải thiện chất lượng truyền động không đồng bộ bằng cấu trúc tách kênh trực tiếp sử dụng kalman filter để quan sát từ thông.
- Ứng dụng Kalman Filter cho dự báo nhiệt độ 2m từ sản phẩm mô hình HRM.
- Hệ thống dẫn đường quán tính INS/GPS.



**Ví dụ:** Chúng ta lấy thước đo chiều dài của một cây bút. Lần thứ nhất chúng ta đo được là 10cm, lần thứ hai chúng ta đo được là 10.05cm chẳng hạn. Vậy hỏi chiều dài cây bút là bao nhiêu?

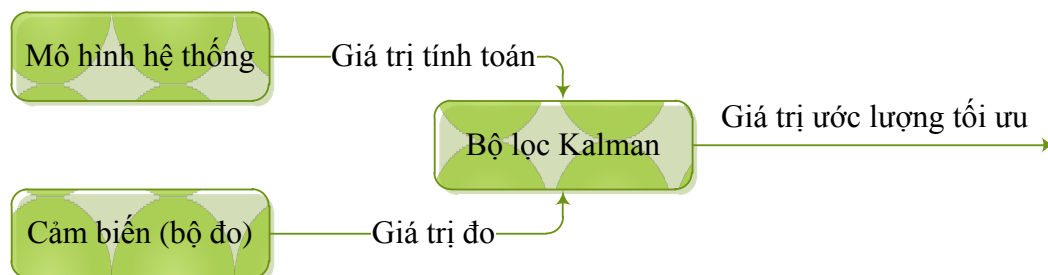
Có phải trong đầu chúng ta luôn cho một giá trị ước lượng tốt nhất là 10.025cm? Điều đó đúng, nhưng căn cứ vào đâu để các bạn ước lượng như vậy?

Nếu bây giờ, một anh F lấy cây thước đo chiều dài, và được kết quả là 10.02 cm, và một bạn T khác cũng lấy cây thước đo, và được chiều dài 10.07cm. Nếu chỉ có hai kết quả này thôi, chúng ta sẽ lại ước lượng rằng chiều dài của cây thước là 10.045 cm! Vậy thì, thực chất là không thể đánh đồng được chuyện đo đạc này, mà cần có một hệ số đánh giá. Chiều dài cây thước là:

- chiều dài =  $\alpha * 10.02$  (của F đo) +  $(1-\alpha)*10.07$  (của T đo)
- Vậy  $\alpha$  sẽ là bao nhiêu? Căn cứ vào đâu để lựa chọn  $\alpha$ ?
- Ví dụ trên để thấy được khái niệm về **ước lượng**.

Đến đây, bộ lọc Kalman xuất hiện. 2 vấn đề chúng ta vừa nêu chính là đối tượng giải quyết của bộ lọc Kalman.

### 2.3.4.3 Mô hình tổng quát của ước lượng dùng bộ lọc Kalman.



Hình 2.7: Mô hình tổng quát của một ước lượng dùng bộ lọc Kalman

Chúng ta có thể ánh xạ các phần trong ví dụ ở mục c vào sơ đồ trên như sau:

- Mô hình hệ thống: phương trình động học  $x(k) = x(k-1) + v$ .
- Cảm biến: bộ định vị gắn trên vật.
- Giá trị ước lượng tối ưu:  $X3$ .
- Bộ lọc Kalman: “hộp đen” tính các hệ số  $\alpha$  và  $\beta$ .

Một ước lượng thực sự sử dụng bộ lọc Kalman cũng bao gồm các phần như trên.

Như vậy đến đây chúng ta đã có đầy đủ các khái niệm cần thiết để bước vào xây dựng bộ lọc Kalman (thực chất là tìm và giải các phương trình để tính toán các tham số cho ước lượng tối ưu).

### 2.3.4.4 Mô hình toán học.

Gọi  $x_k$  là vector giá trị thực sự của một cái gì đó (ví dụ như vị trí của tên lửa) tại thời điểm thứ k. Ta sẽ giả sử  $x_k$  biến đổi theo qui luật sau:

$$x_k = F_k x_{k-1} + B_k u_{k-1} + w_k \quad (2.6)$$

trong đó  $F_k$  là ma trận thay đổi trạng thái (state transition matrix),  $u_k$  là vector điều khiển,  $B_k$  là ma trận điều khiển, còn  $w_k$  là nhiễu ngẫu nhiên, với giả sử là nó có phân bố Gaussian (phân bố normal nhiều chiều)  $N(0, Q_k)$ , trong đó  $Q_k$  là ký hiệu của ma trận hiệp phương sai tương ứng. Tại mỗi thời điểm thứ  $k$ , có một đo đạc (measurement) trạng thái  $x_k$  cho kết quả là:

$$z_k = H_k x_k + v_k \quad (2.7)$$

trong đó  $H_k$  là ma trận của mô hình quan sát, còn  $v_k$  là nhiễu trong lúc đo đạc, và ta giả sử nhiễu này cũng tuân theo một phân bố Gaussian  $N(0, R_k)$ . Các ma trận  $F_k B_k H_k$  được coi là đã biết. Ta giả sử thêm là các nhiễu  $\{v_1, \dots, v_k, w_k, \dots, w_k\}$  là một bộ biến ngẫu nhiên độc lập và cũng độc lập với trạng thái ban đầu  $x_0$ . Câu hỏi đặt ra là làm sao ước lượng được các trạng thái  $x_k$  từ các quan sát  $z_k$ ? Nếu không hề có nhiễu, thì ta chỉ cần đặt  $x_k = H_k^{-1}$ . Nhưng vì có nhiễu nên không tính được chính xác  $x_k$  mà chỉ có thể ước lượng nó. Kalman filter là một hàm ước lượng đệ quy (recursive estimator) cho phép làm việc này. Thuật toán ước lượng như sau: Có thể chia nó thành 2 bước, bước dự đoán ban đầu (predict) và bước điều chỉnh sau đó (update) Dự đoán:

$$\hat{x}_{k|k-1} = F_k \hat{x}_{k-1|k-1} + B_k u_k + w_k \quad (2.8)$$

$$P_{k|k-1} = F_k P_{k-1|k-1} F_k^T + Q_k \quad (2.9)$$

Ở đây  $\hat{x}_{k|k-1}$  là ký hiệu dự đoán giá trị của  $x_k$  dựa trên thông tin về giá trị tại thời điểm  $k-1$ , còn  $P_{k|k-1}$  là ước lượng giá trị của  $x_k$  sau khi đã sử dụng mọi thông tin tại thời điểm  $k$ . Ma trận  $P$  dùng để chỉ (ước lượng) ma trận hiệp phương sai của ước lượng của  $x$ . Cập nhật :

$$\text{- Độ lệch so với quan sát (measurement residual): } y_k = z_k - H_k \hat{x}_{k|k-1} \quad (2.10)$$

$$\text{- Thặng dư hiệp phương sai (residual covariance): } S_k = R_k + H_k \hat{P}_{k|k-1} H_k^T \quad (2.11)$$

$$\text{- Kalman tối ưu: } K_k = P_{k|k-1} H_k^T S_k^{-1} \quad (2.12)$$

$$\text{- Ước lượng được điều chỉnh (updated estimate): } \hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k y_k \quad (2.13)$$

- Hiệp phương sai cho ước lượng mới (updated estimate covariance):

$$P_{k|k} = (1 - K_k H_k) P_{k|k-1} \quad (2.14)$$

### 2.3.5 Theo vết đối tượng bằng so khớp mẫu.

Thủ tục trong so khớp mẫu như sau: một miền nhỏ xung quanh điểm cần theo vết sẽ được dùng làm mẫu. Mẫu này sau đó được dùng để tìm ra trong khung ảnh kế tiếp bằng cách sử dụng các kỹ thuật tương quan [26]. Vị trí với kết quả cao nhất sẽ được so khớp tốt nhất giữa mẫu và ảnh.

Bằng cách cập nhật các mẫu theo chuỗi ảnh, các biến dạng lớn cũng có thể được theo vết[ 27 ]. Sử dụng một trong 3 luật cập nhật cơ bản như sau:

- Nếu có một sự thay đổi lớn giữa vị trí mẫu ban đầu và vị trí mới thì mẫu mới được chọn. Trong trường hợp này, các mẫu được hoán đổi hoàn toàn bởi các hình dạng mới của chúng.

- Nếu có các thay đổi nhỏ giữa vị trí mẫu ban đầu và vị trí mới của mẫu, một phiên bản trung bình của mẫu mới và cũ sẽ được tính và cập nhật như mẫu mới. Bằng cách này, các đạo hàm nhỏ liên quan tới nhiễu sẽ là trung bình, do đó gia tăng được khả năng theo vết tránh nhiễu.

- Nếu chỉ có các thay đổi quá nhỏ giữa các vị trí ba đầu và vị trí mới, thì mẫu cũ sẽ được sử dụng. Điều này rất quan trọng cho các đối tượng tĩnh tiến bởi các lượng nhỏ hơn một pixel: nếu như ta cập nhật lại thì sẽ mất đi thông tin dịch chuyển điểm ảnh nhỏ.

Ưu điểm: Không chịu ảnh hưởng bởi nhiễu và hiệu ứng chiếu sáng, theo vết được các đối tượng biến dạng

Khuyết điểm: Độ phức tạp tính toán cao, chất lượng so khớp phụ thuộc vào chi tiết và độ chính xác của mẫu đối tượng.

## **2.4 Các kỹ thuật đường ranh giới ảo.**

Đường ranh giới ảo là một phần quan trọng của hệ thống giám sát vào/ ra. Chúng được sử dụng để nhận biết hướng chuyển động của các đối tượng, từ đó đưa ra thông tin về lượng người vào/ ra tương ứng. Có các kỹ thuật khác nhau, nhưng cơ bản trong hệ thống giám sát vào ra người ta thường sử dụng 3 kỹ thuật sau.

### 2.4.1 Kỹ thuật ranh giới ảo một đường



*Hình 2.8: Mô hình ranh giới ảo 1 đường*

Với kỹ thuật này, khu vực được giám sát sẽ được chia thành hai khu vực con bởi một đường ranh giới ảo. Vị trí của đường ranh giới này phụ thuộc vào vị trí của camera. Khi có đối tượng chạm qua vạch ranh giới này, dựa vào vị trí hiện tại và vị trí trước đó của đối tượng để xác định hướng di chuyển của đối tượng. Nếu là đi vào thì tăng số lượng người vào lên 1 đơn vị. Ngược lại thì tăng lượng người đi ra lên 1 đơn vị[3].

**Ưu điểm:** Cài đặt dễ dàng, đơn giản và tốc độ nhanh .

**Nhược điểm:** Nếu đối tượng chạm vạch rồi đứng im hoặc quay ngược trở lại, thì hệ thống sẽ bị nhầm.

#### 2.4.2 Kỹ thuật ranh giới ảo hai đường.



*Hình 2.9: Mô hình ranh giới ảo 2 đường*

Với kỹ thuật này, khu vực giám sát được chia thành 3 khu vực bởi 2 đường ranh giới ảo. Khi đối tượng chạm vào ranh giới vào/ra thì trạng thái sẽ được lưu lại cho đối tượng đó. Nếu trạng thái trước đó của đối tượng khác rỗng( tức là đã vào hoặc ra) thì xảy ra trường hợp

- Nếu trước là vào và hiện tại là ra thì tăng lượng người đi ra lên 1 đơn vị
- Nếu trước là ra và hiện tại là vào thì tăng lượng người đi vào lên 1 đơn vị[28]

**Ưu điểm:** Cải thiện việc quay ngược trở lại hoặc đứng im trên vạch. Hệ thống sẽ không còn nhầm lẫn

**Nhược điểm:** Hệ thống sẽ tốn chi phí thời gian hơn vì phải thực hiện so sánh vị trí từng đối tượng với 2 vạch ranh giới.



### 2.4.3 Kỹ thuật ranh giới ảo ba đường.



Hình 2.10: Mô hình ranh giới ảo 3 đường

Để khắc phục nhược điểm các kỹ thuật trên, kỹ thuật 3 đường ra đời. Tương tự như hệ thống 2 đường nhưng hệ thống này có thêm 1 đường ranh giới chính ở giữa hai ranh giới vào ra. Khi đối tượng đi qua hai đường ranh giới vào ra thì trạng thái tương ứng được lưu vào đối tượng. Khi đối tượng đi qua ranh giới chính thì hệ thống kiểm tra nếu:

- Trạng thái trước đó là vào thì tăng lượng người vào lên 1 đơn vị
- Trạng thái trước đó là ra thì tăng lượng người ra lên 1 đơn vị

**Ưu điểm:** Khắc phục việc người đi quay ngược lại và việc dừng giữa chừng.

**Nhược điểm:** Tốn chi phí thời gian hơn.

## 2.5 Các thuật toán nhận dạng đối tượng.

### 2.5.1 Tổng quan về bài toán nhận dạng.

#### 2.5.1.1 Khái niệm.

Nhận dạng là quá trình phân loại các đối tượng được biểu diễn theo một mô hình nào đó và gán chúng vào một lớp (gán cho đối tượng một tên gọi) dựa theo những quy luật và những mẫu chuẩn. Quá trình nhận dạng dựa vào những mẫu học biết trước gọi là nhận dạng có thầy hay **“học có giám sát”**, trong trường hợp ngược lại gọi là **“học không có giám sát”**[1]. Các khái niệm này sẽ được làm rõ hơn thông qua các thuật toán sẽ được nêu ra bên dưới.

### 2.5.1.2 Bản chất quá trình nhận dạng.

Quá trình nhận dạng gồm 3 giai đoạn chính:

- Lựa chọn mô hình biểu diễn đối tượng.
- Lựa chọn luật ra quyết định (phương pháp nhận dạng) và suy diễn quá trình học.
- Học nhận dạng.

Khi mô hình biểu diễn đối tượng đã được xác định, có thể là định lượng (mô hình tham số) hay định tính (mô hình cấu trúc), quá trình nhận dạng chuyển sang giai đoạn học. Học là giai đoạn rất quan trọng. Thao tác học nhằm cải thiện, điều chỉnh việc phân hoạch tập đối tượng thành các lớp.

Việc nhận dạng chính là tìm ra quy luật và các thuật toán để có thể gán đối tượng vào một lớp hay nói một cách khác gán cho đối tượng một tên.

#### *Học có giám sát*

Kỹ thuật phân loại nhờ kiến thức biết trước gọi là học có giám sát. Đặc điểm cơ bản của kỹ thuật này là người ta có một thư viện các mẫu chuẩn. Mẫu cần nhận dạng sẽ được đem sánh với mẫu chuẩn để xem nó thuộc loại nào. Thí dụ như trong một ảnh viễn thám, người ta muốn phân biệt một cánh đồng lúa, một cánh rừng hay một vùng đất hoang mà đã có các miêu tả về các đối tượng đó. Vấn đề chủ yếu là thiết kế một hệ thống để có thể đối sánh đối tượng trong ảnh với mẫu chuẩn và quyết định gán cho chúng vào một lớp. Việc đối sánh nhờ vào các thủ tục ra quyết định dựa trên một công cụ gọi là *hàm phân lớp* hay *hàm ra quyết định*. Hàm này sẽ được đề cập trong phần sau.

#### *Học không có giám sát*

Kỹ thuật học này phải tự định ra các lớp khác nhau và xác định các tham số đặc trưng cho từng lớp. Học không có giám sát đương nhiên là khó khăn hơn. Một mặt, do số lớp không được biết trước, mặt khác những đặc trưng của các lớp cũng không biết trước. Kỹ thuật này nhằm tiến hành mọi cách gộp nhóm có thể và chọn lựa cách tốt nhất. Bắt đầu từ tập dữ liệu, nhiều thủ tục xử lý khác nhau nhằm phân lớp và nâng cấp dần để đạt được một phương án phân loại. Bảng bên dưới liệt kê ra một số thuật toán nhận dạng, phân lớp phổ biến mà ta thường hay áp dụng.

*Bảng 2.2 : Một số thuật toán nhận dạng, phân lớp đối tượng*

STT	Tên thuật toán	Tóm tắt cách thực hiện
1	Mạng nơ ron	Áp dụng nguyên lý hoạt động của nơ ron thần kinh cho bài toán nhận dạng.
2	K-means	Sử dụng khoảng cách ngắn nhất đến các mẫu đã

		được huấn luyện.
3	Knn	Sử dụng k khoảng cách gần nhất đến các mẫu đã được huấn luyện
4	PCA	Thuật toán phân tích thành phần chính
5	Bayes	Thuật toán sử dụng lý thuyết xác suất
6	Đối sánh ảnh	Thuật toán sử dụng sai số giữa hai đối tượng
7	.....	.....

Trong đề tài này nhóm tác giả xin phép không đưa hết lý thuyết của các thuật toán nhận dạng, phân lớp đối tượng vào trong báo cáo. Một số thuật toán theo chương trình giảng dạy trên lớp sẽ được đề cập cụ thể hơn ở các mục bên dưới.

## 2.5.2 Nhận dạng đối tượng bằng thuật toán K-Means.

### 2.5.2.1 Lý thuyết thuật toán.

K-Means là thuật toán rất quan trọng và được sử dụng phổ biến trong kỹ thuật phân cụm. Tư tưởng chính của thuật toán K-Means là tìm cách phân nhóm các đối tượng (objects) đã cho vào K cụm (K là số các cụm được xác định trước, K nguyên dương) sao cho tổng bình phương khoảng cách giữa các đối tượng đến tâm nhóm (centroid) là nhỏ nhất.

Thuật toán K-Means thực hiện qua các bước chính sau:

- Chọn ngẫu nhiên K tâm (centroid) cho K cụm (cluster). Mỗi cụm được đại diện bằng các tâm của cụm.
- Tính khoảng cách giữa các đối tượng (objects) đến K tâm (thường dùng khoảng cách Euclidean)
- Nhóm các đối tượng vào nhóm gần nhất
- Xác định lại tâm mới cho các nhóm
- Thực hiện lại bước 2 cho đến khi không có sự thay đổi nhóm nào của các đối tượng

Thuật toán K-Means có ưu điểm là đơn giản, dễ hiểu và cài đặt. Tuy nhiên, một số hạn chế của K-Means là hiệu quả của thuật toán phụ thuộc vào việc chọn số nhóm K (phải xác định trước) và chi phí cho thực hiện vòng lặp tính toán khoảng cách lớn khi số cụm K và dữ liệu phân cụm lớn.

### 2.5.2.2 Ví dụ dẫn nhập.

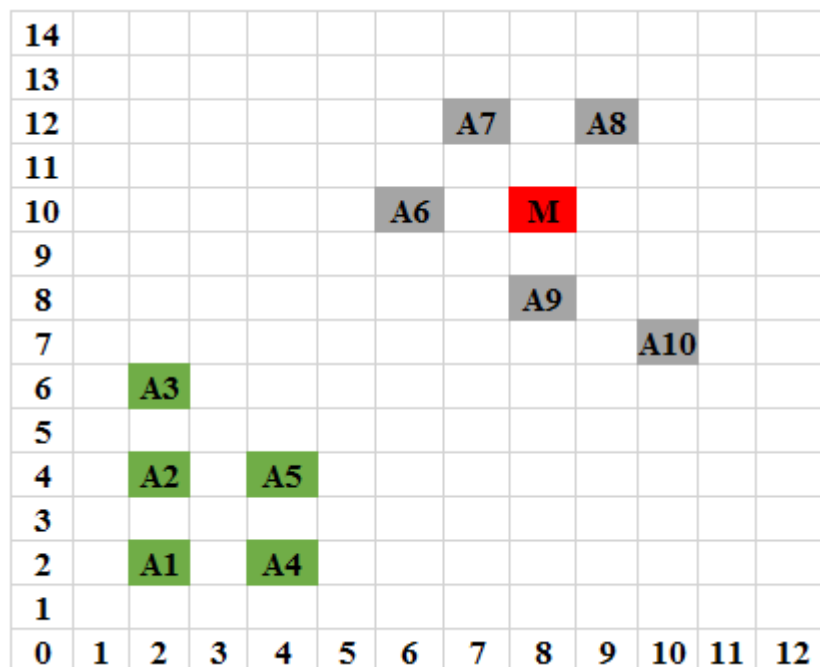
Ta giả sử trên mặt phẳng tọa độ OXY có 10 điểm và được chia là 2 cụm. Cụm thứ nhất bao gồm các điểm A1...A5, cụm thứ hai bao gồm các điểm A6...A10. Với tọa độ của các điểm được cho bởi bảng sau:

*Bảng 2.3: Danh sách các tọa độ điểm*



Điểm	X	Y	Điểm	X	Y
A1	2	2	A6	6	10
A2	2	4	A7	7	12
A3	2	6	A8	9	12
A4	4	2	A9	8	8
A5	4	4	A10	10	7

Với điểm M có tọa độ M(8,10) thì bảo toán đặt ra là M sẽ thuộc cụm nào? Nếu nhìn trên đồ thị thì ta hoàn toàn có thể nhận ra được M đang ở đâu và gần cụm nào hơn. Với tư tưởng đó, ta sẽ sử dụng K-mean để tìm ra câu trả lời. (Hình bên dưới mô tả vị trí của các điểm A1..A10 trên mặt phẳng tọa độ)



Hình 2.11: Mô tả các điểm trên mặt phẳng tọa độ

Cách thực hiện như sau:

**B1:** Tính tọa độ trung bình của cụm 1:  $K1=(A1+A2+A3+A4+A5)/5=(2.8, 3.6)$

**B2:** Tính tọa độ trung bình của cụm 2:  $K2=(A6+A7+A8+A9+A10)/5=(8,9.8)$

**B3:** Tính khoảng cách X1 từ điểm M đến K1:  $X1=8.246$

**B4:** Tính khoảng cách X2 từ điểm M đến K2:  $X2=0.2$

**B5:** So sánh  $X1>X2$ , suy ra M thuộc cụm 2.

### 2.5.3 Nhận dạng đối tượng bằng thuật toán KNN.

#### 2.5.3.1 Lý thuyết thuật toán.

K-Nearest Neighbors algorithm (KNN) được sử dụng rất phổ biến trong lĩnh vực Data Mining. KNN là phương pháp để phân lớp các đối tượng dựa vào khoảng cách gần nhất giữa đối tượng cần xếp lớp với tất cả các đối tượng trong Training Data.

Một đối tượng được phân lớp dựa vào k láng giềng của nó. K là số nguyên dương được xác định trước khi thực hiện thuật toán. Người ta thường dùng khoảng cách Euclidean để tính khoảng cách giữa các đối tượng.

Các bước thực hiện của thuật toán như sau:

- Xác định giá trị tham số K (số láng giềng gần nhất)
- Tính khoảng cách giữa đối tượng cần phân lớp với tất cả các đối tượng trong training data (thường sử dụng khoảng cách Euclidean, Cosine...)
- Sắp xếp khoảng cách theo thứ tự tăng dần và xác định k láng giềng gần nhất với đối tượng cần phân lớp
- Lấy tất cả các lớp của k láng giềng gần nhất đã xác định
- Dựa vào phần lớn lớp của láng giềng gần nhất để xác định lớp cho đối tượng.

#### 2.5.3.2 Ví dụ dẫn nhập.

Ta sẽ áp dụng lý thuyết KNN cho bài toán phân loại văn bản.

Ý tưởng: Khi cần phân loại một văn bản mới, thuật toán sẽ tính khoảng cách (khoảng cách Euclidean, Cosine...) của tất cả các văn bản trong tập huấn luyện đến văn bản này để tìm ra k văn bản gần nhất (gọi là k “láng giềng”), sau đó dùng các khoảng cách này đánh trọng số cho tất cả chủ đề. Trọng số của một chủ đề chính là tổng tất cả các văn bản trong k láng giềng có cùng chủ đề, chủ đề nào không xuất hiện trong k láng giềng sẽ có trọng số bằng 0. Sau đó các chủ đề sẽ được sắp xếp theo mức độ giảm dần và các chủ đề có trọng số cao sẽ được chọn là chủ đề của văn bản cần phân loại.

Khoảng cách giữa 2 văn bản chính là độ tương tự giữa 2 văn bản đó, 2 văn bản có giá trị độ tương tự càng lớn thì khoảng cách càng gần nhau.

Ví dụ: Dùng công thức Cosine để tính độ tương tự giữa 2 văn bản:

$$\text{sim}(\vec{x}, \vec{y}) = \cos(\vec{x}, \vec{y}) = \frac{\vec{x} \cdot \vec{y}}{\|\vec{x}\| \cdot \|\vec{y}\|}$$

Văn bản A: Tôi là học sinh.

Văn Bản B: Tôi là sinh viên.

Văn bản C: Tôi là giáo viên.

Bảng 2.4: Biểu diễn văn bản theo vector

	Tôi	Là	Học	Sinh	Viên	Giáo
<b>Văn bản A</b>	1	1	1	1	0	0
<b>Văn bản B</b>	1	1	0	1	1	0
<b>Văn bản C</b>	1	1	0	0	1	1

Vector A = (1,1,1,1,0,0)

Vector B = (1,1,0,1,1,0)

Vector C = (1,1,0,0,1,1)

$$\sin(\vec{A}, \vec{B}) = \cos(\vec{A}, \vec{B}) = \frac{3}{\sqrt{4 * 4}} = 0.75$$

$$\sin(\vec{A}, \vec{C}) = \cos(\vec{A}, \vec{C}) = \frac{2}{\sqrt{4 * 4}} = 0.5$$

Điều đó cho thấy văn bản A tương tự văn bản B hơn so với C.

## CHƯƠNG 3. BÀI TOÁN KIỂM SOÁT VÀO RA BẰNG CAMERA

**Tóm tắt nội dung chính:** Trong chương này sẽ đề cập đến nội dung chính của đề tài. Đó là việc nghiên cứu hệ thống giám sát tổng quát để xây dựng nên hệ thống kiểm soát vào ra. Các bài toán cũng như phương pháp cụ thể sẽ được áp dụng trong đề tài. Cùng với đó là những đề xuất, cải tiến, đánh giá thực nghiệm để cải thiện hệ thống.

### 3.1 Phát biểu bài toán.

#### 3.1.1 Giới thiệu chung.

Theo dõi lưu lượng người là một thông tin rất hữu ích cho các ứng dụng như bảo mật như quản lý như giao thông cho người đi bộ, quản lý khách du lịch. Việc sử dụng camera để theo dõi và đếm người đang tăng đáng kể trong những năm gần đây do sự tiến bộ của các thuật toán xử lý hình ảnh và khả năng của máy tính. Có nhiều cách để thực hiện việc này nhưng chủ yếu được phân thành 3 loại:

- Phương pháp sử dụng tính năng theo dõi vùng. Để cải thiện phương pháp này cần thêm một chương trình phân loại của các điểm ảnh dựa trên màu sắc hoặc kết cấu.
- Phương pháp sử dụng xuất hiện 2D của con người (bằng cách sử dụng các mô hình khác nhau của con người).
- Phương pháp sử dụng nhiều camera để thực hiện một mô hình 3D đầy đủ.

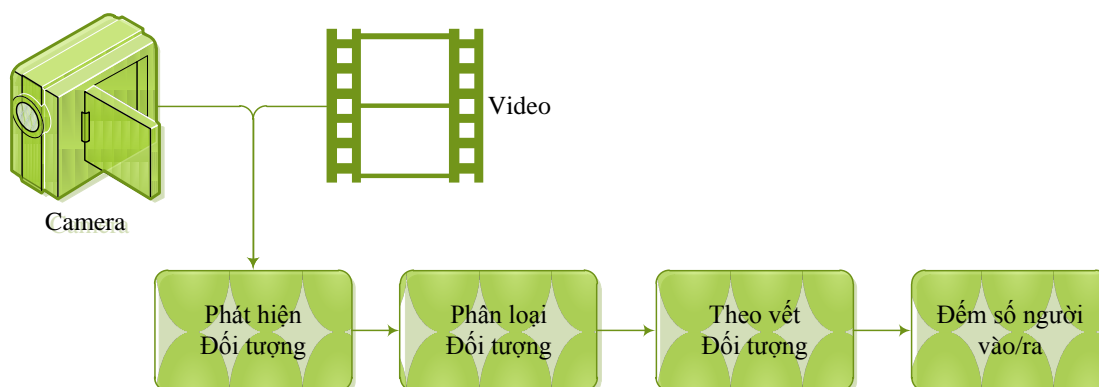
Loại thứ ba là chính xác hơn hơn so với hai phương pháp còn lại vì nó xây dựng lại chính xác ngữ cảnh, nhưng mặc khác nó cũng gặp khó khăn với các thuật toán phức tạp. Điều này đòi hỏi một hệ thống thiết lập máy ảnh phức tạp (hiệu chuẩn) và không thể hoạt động trong thời gian thực bởi vì các mô hình 3D quá chậm. Đây là lý do tại sao hầu hết các hệ thống sử dụng hai loại còn lại.

Do đặc thù của bài toán đếm người vào ra, yêu cầu độ chính xác và cả tốc độ tính toán nhanh nên trong đề tài này nhóm tác giả xin đề cập tới mô hình sử dụng tính năng theo dõi vùng

#### 3.1.2 Mô hình của hệ thống.

Là một trong những loại thuộc hệ thống giám sát thông minh, hệ thống kiểm soát vào ra cũng có 4 thành phần chính đó là: Phát hiện đối tượng, phân loại đối tượng, theo vết đối tượng và đếm lượng người vào/ra. 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. Bản chất module phân tích hành vi ở đây đã được chỉ rõ là dùng để đếm lượng người vào ra. Chính vì vậy thay vì có 5 module thì hệ thống kiểm soát vào ra sẽ chỉ còn 4 module. 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 trong điều kiện vị trí thực nghiệm là trong trường học, 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 3.1: Mô hình hệ thống giám sát vào ra bằng camera*

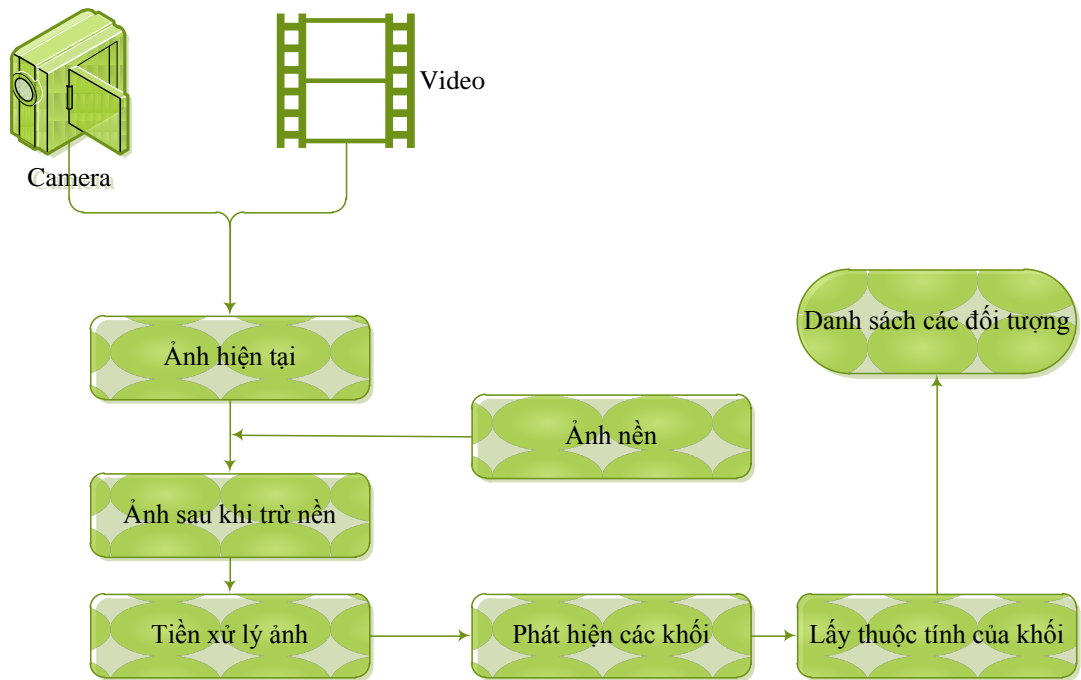
Trong bốn thành phần trên thì phần phát hiện đối tượng chuyển động và phần theo vết đối tượng là hai thành phần cực kỳ quan trọng của hệ thống, chúng sẽ ảnh hưởng tới hiệu quả cũng như tốc độ của hệ thống. Đối với đặc thù của hệ thống là xử lý trong thời gian thực nên sẽ rất khó khăn nếu ta đi sâu về 1 module nào đó. Sau đây sẽ là các thuật toán cơ bản thường được sử dụng trong nhận dạng đối tượng chuyển động và theo vết đối tượng chuyển động

### **3.2: Các kỹ thuật triển khai.**

#### **3.2.1 Thuật toán trừ nền.**

Phát hiện các đối tượng chuyển động từ ảnh nền đứng yên là cả một vấn đề nghiên cứu khó. Hầu hết trong tất cả các hệ thống giám sát, bước đầu tiên là phát hiện các đối tượng chuyển động. Bước này sẽ tác động rất lớn tới hiệu quả của các bước sau như bước theo dõi, phân lớp, phân tích hành vi. Và nó cũng giảm chi phí thời gian tính toán của hệ thống đi đáng kể vì chỉ các điểm ảnh thuộc đối tượng mới cần xử lý. Tuy nhiên việc phát hiện đối tượng chuyển động gặp rất nhiều khó khăn ví dụ như trong các cảnh thường xuất hiện các chuyển động lặp đi lặp lại (ví dụ như lá cây rơi), phản xạ ánh sáng, bóng, nhiễu và các thay đổi ánh sáng đột ngột làm cho việc phát hiện đối tượng trở nên khó khăn. Do đó, việc phát hiện đối tượng từ nền là bước rất quan trọng cần được chú ý để cho hệ thống giám sát trở nên đáng tin cậy, mạnh mẽ và nhanh chóng.

Sơ đồ hệ thống phương pháp phát hiện đối tượng của nhóm tác giả được thể hiện trong hình 3.2



*Hình 3.2: Sơ đồ hệ thống phát hiện đối tượng chuyển động*

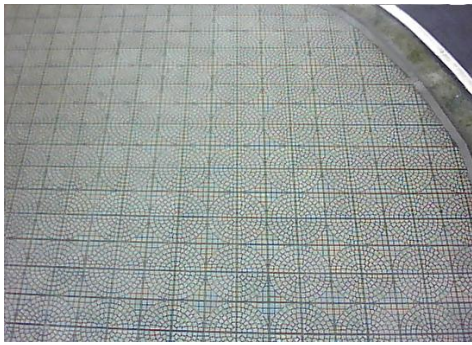
Phương pháp này gồm sáu giai đoạn xử lý để trích xuất các đối tượng với các đặc trưng của chúng từ hình ảnh video. Bước đầu tiên là khởi tạo cảnh nền. Có những kỹ thuật khác nhau được sử dụng để mô hình nền trong các tài liệu.

Bước tiếp theo trong phương pháp này là phát hiện các điểm ảnh “chuyển động” bằng cách sử dụng mô hình nền và hình ảnh hiện tại từ video. Mức độ phát hiện điểm ảnh phụ thuộc vào mô hình nền được sử dụng và nó được dùng để cập nhật mô hình nền để thích ứng với các đoạn cảnh chuyển động. Ngoài ra, do nhiễu máy ảnh hoặc các tác động môi trường nên ma trận điểm ảnh foreground phát hiện được sẽ chứa nhiễu. Quá trình xử lý được thực hiện để loại bỏ nhiễu trong ma trận điểm ảnh foreground.

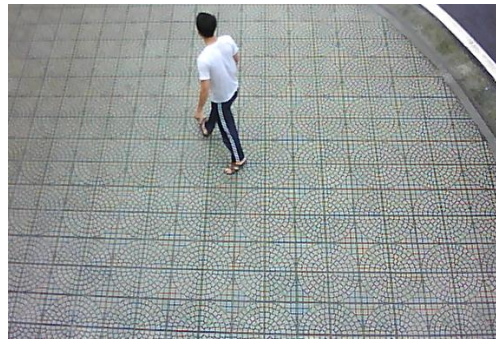
Một khi có được các điểm ảnh foreground đã được lọc, trong bước tiếp theo, các khu vực liên quan được gắn kết bằng cách sử dụng một thuật toán kết nối thành phần và tìm hình chữ nhật bao quanh. Một số vùng tương đối nhỏ do nhiễu môi trường gây ra được loại bỏ trong bước sau.

Trong bước cuối cùng của quá trình phát hiện, một số các đặc trưng của đối tượng được phát hiện từ hình ảnh hiện tại bằng cách sử dụng ma trận điểm ảnh foreground. những thuộc tính này là vùng trung tâm, lược đồ màu sắc của các vùng tương ứng với các đối tượng.

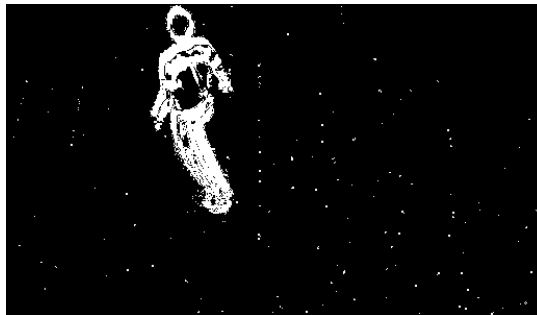
### Phát hiện ảnh foreground



a - Ảnh nền



b - Ảnh hiện tại



c - Ảnh sau khi trừ nền

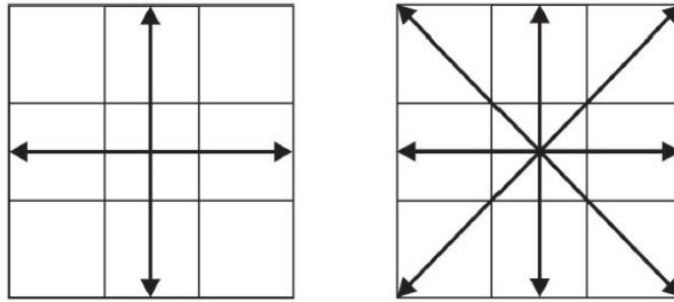
*Hình 3.3: Phát hiện ảnh Foreground*

#### 3.2.2 Thuật toán phát hiện các khối.

Sau khi phát hiện foreground và loại bỏ nhiễu, các điểm ảnh foreground được nhóm lại thành các khối (blobs) và dán nhãn. Sau khi tìm thấy các khối riêng rẽ tương ứng với các đối tượng, thuật toán sẽ sinh ra hình chữ nhật bao quanh đối tượng.

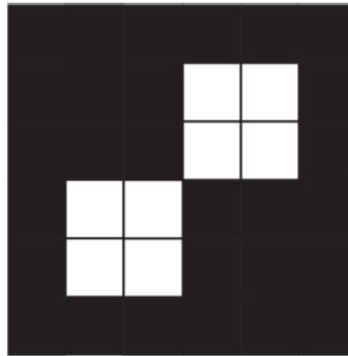
- Đầu vào: Ảnh nhị phân cấp xám foreground của quá trình trước
- Đầu ra: Danh sách các blob với các đặc trưng như vị trí, diện tích.
- Phương pháp: Sử dụng thuật toán loang với các liên thông 4 hoặc liên thông 8.

Việc này bao gồm phân tích hình ảnh nhị phân, tìm thấy tất cả các blobs và tính toán số liệu thống kê cho mỗi một blob. Thông thường, các đặc trưng tính năng thường được tính là diện tích (số lượng điểm ảnh có trong các blob), chu vi, vị trí và hình dạng blob. Trong quá trình, nó có thể để loại các đặc trưng khác nhau bởi tính năng của chúng. Ví dụ, nếu blobs tìm kiếm có để khu vực quan sát, một số đặc trưng có thể được loại bỏ nếu chúng không thỏa mãn các điều kiện (nó cho phép hạn chế số lượng của các blob, do đó làm giảm chi phí tính toán). Có hai cách để tìm điểm liên thông. Một là bao gồm có các điểm ảnh lân cận dọc theo chiều thẳng đứng, và các điểm ảnh ngang dính liền. Và một là bằng cách bao gồm các điểm ảnh lân cận theo đường chéo (hình minh họa).



Hình 3.4: Các phép liên thông

Thiết lập các quy tắc cho việc liên kết các điểm ảnh là rất quan trọng bởi vì kết quả của việc tìm các blob có thể khác nhau. Ví dụ (hình minh họa), nhóm các điểm ảnh sẽ là xem xét như là một blob nếu thuật toán sử dụng mạng với tám kết nối và hai đốm màu khác nhau nếu nó được sử dụng mạng tinh thể khác.



Hình 3.5: Ví dụ sự khác biệt các phép liên thông

Kết quả của việc thực hiện các thuật toán phân tích blob phụ thuộc hoàn toàn vào chất lượng của các phân khúc. Với phân khúc xấu, phân tích blob có thể phát hiện một số blob không liên quan hoặc tồi tệ hơn có thể hợp nhất một số blob khác nhau do điều kiện ánh sáng hoặc tiếng ồn trong hình ảnh.



a - Ảnh trước khi phát hiện các khối



b - Ảnh sau khi phát hiện khối

Hình 3.6: Phép bóc tách các block



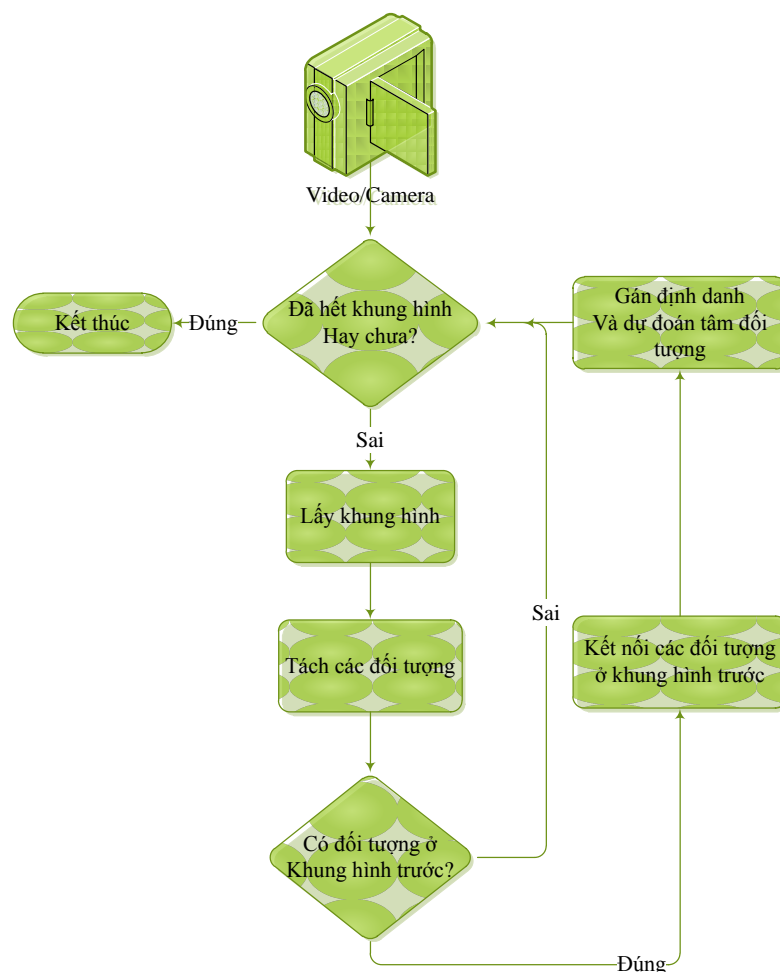
### 3.2.3 Thuật toán xử lý các vùng đối tượng nhỏ.

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 các khu vực này, mỗi khung hình, cần tính toán kích thước trung bình của các khối ( $y$ ). Khối nào có kích thước nhỏ hơn ( $\alpha$ ) lần của kích thước trung bình (Kích thước khối  $< \alpha * \gamma$ ) được xóa bỏ khỏi ma trận điểm ảnh foreground. Ngoài ra, do lỗi phân khúc, một số bộ phận 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.

### 3.2.4 Thuật toán theo vết đối tượng.

Trong hệ thống này, chúng ta sử dụng thuật toán theo vết đối tượng ở mức độ toàn bộ đối tượng. Nghĩa là chúng ta theo dõi toàn bộ đối tượng, các thông tin thu được từ việc theo dõi này cũng đủ cung cấp cho hệ thống.

Cách tiếp cận ở đây là sử dụng các đặc trưng của đối tượng như kích thước, điểm trung tâm của khối, khung giới hạn và biểu đồ màu sắc được trích xuất ở bước trước để tạo sự liên kết giữa các đối tượng trong các khung hình liên tiếp. Sơ đồ của phương pháp được thể hiện ở hình dưới



Hình 3.7: Sơ đồ tổng quát thuật toán theo vết đối tượng

### Tách đối tượng chuyển động

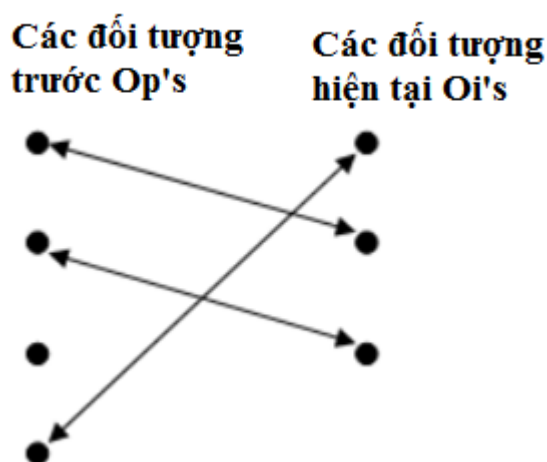
Ở bước này kỹ thuật trừ nền được đề cập phần 2.2.2 sẽ được áp dụng để tách lấy các đối tượng chuyển động.

### Gán định danh và dự đoán tâm đối tượng

Sử dụng các thuật toán đã nêu và trình bày ở trên để xác định vị trí chuyển động của đối tượng. Sau đó gán nhãn để phục vụ xử lý cho các bước phía sau.

### Kết nối các đối tượng ở hai khung hình liên tiếp

Bằng cách dựa trên sự tương đồng giữa hai đối tượng nhau chúng ta có thể kết nối các đối tượng ở hai khung hình khác nhau.



Hình 3.8: Ví dụ về kết nối các đối tượng

Bước đầu tiên của thuật toán theo dõi là nối các đối tượng ( $O_p$ 's) ở ảnh trước ( $I_{n-1}$ ) với các đối tượng mới ( $O_i$ 's) được phát hiện ở ảnh hiện tại ( $I_n$ ).

Chúng ta lưu trữ các kết hợp của các đối tượng trong một đồ thị hai phía  $G(m, n)$ . Trong biểu đồ này, đỉnh đại diện cho các đối tượng (một bên đại diện cho đỉnh các đối tượng trước  $O_p$  và bên còn lại đại diện cho các đối tượng mới,  $O_i$ ) và các cạnh đại diện cho một sự kết nối giữa hai đối tượng.  $G(m, n)$ ,  $m$  là số các đối tượng trước đó, và  $n$  là số các đối tượng mới. Một đồ thị kết hợp đơn giản được thể hiện trong hình 3.8. Để thực hiện kết nối các đối tượng, chúng ta lập qua danh sách các đối tượng trước và các đối tượng mới để đánh giá độ tương đồng. Với mỗi đối tượng trước đó ( $O_p$ ) chúng ta lập các đối tượng mới và đầu tiên kiểm tra xem một đối tượng mới  $O_i$  trong danh sách các đối tượng mới có gần  $O_p$  hay không. Các tiêu chí cho việc đánh giá là gần hay không được định nghĩa là khoảng cách giữa trung tâm của khối của hai đối tượng này ( $O_p$  và  $O_i$ ) nhỏ hơn một hằng số được xác định trước. Việc kiểm tra này được xuất phát từ thực tế là sự di chuyển của một đối

tượng giữa các hình ảnh liên tiếp là nhỏ. Nói cách khác, hai đối tượng với trung tâm  $c_p$  và  $c_i$  là gần nhau nếu thỏa mãn:

$$\text{Dist}(c_p, c_i) < T \quad (3.1)$$

Với hàm  $\text{Dist}()$  được định nghĩa là hàm khoảng cách Euclidean giữa hai điểm:

$$\text{Dist}(C_p, C_i) = \sqrt{(x_{cp} - x_{ci})^2 + (y_{cp} - y_{ci})^2} \quad (3.2)$$

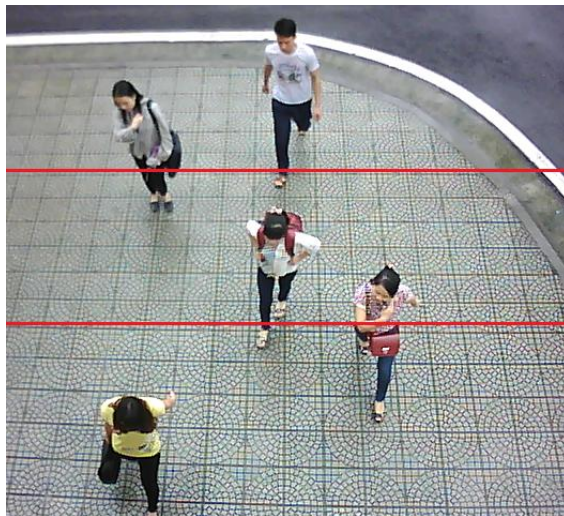
Bởi vì hai đối tượng gần nhau trong một ngưỡng không nhất thiết sẽ được kết nối, trong bước tiếp theo chúng ta kiểm tra khả năng tương đồng giữa hai đối tượng để làm việc kết hợp cho chính xác hơn. Các điều kiện để đánh giá sự tương đồng là tỷ lệ kích thước của ác đối tượng. Việc kiểm tra này cũng xuất phát từ thực tế là đối tượng không lớn hơn hoặc co lại quá nhiều giữa các khung hình liên tiếp. Do đó hai đối tượng được gọi là tương tự nhau nếu chúng thỏa mãn:

$$\frac{s_p}{s_i} < T \text{ hoặc } \frac{s_i}{s_p} > T \quad (3.3)$$

Với  $s_i$  là kích thước của đối tượng  $O_i$  và  $T$  là một ngưỡng được xác định trước. Kiểm tra kích thước các đối tượng đặc biệt hữu ích nếu một đối tượng trong khung trước chia tách thành một vùng rộng lớn và rất nhỏ do phân khúc không chính xác. Điều này kiểm tra loại bỏ khả năng kết hợp một khối lớn với một khối nhỏ.

### 3.2.5 Thuật toán ranh giới ảo hai đường.

Thuật toán đã được trình bày ở các mục trước. Đối với bài toán đếm số người vào ra sử dụng đường ranh giới ảo thì thường hay sử dụng ranh giới ảo một đường. Tuy nhiên nhận thấy nhược điểm của đường ranh giới ảo một đường cũng như đề tăng tính mới cho đề tài, nhóm tác giả đã lựa chọn đường ranh giới ảo hai đường. Với mong muốn tìm hiểu thêm về kỹ thuật này đồng thời xây dựng thuật toán sao cho đáp ứng được bài toán. Tạo tiền đề cho những nghiên cứu và phát triển sau này.



Hình 3.9: Áp dụng ranh giới ảo hai đường cho bài toán

### 3.2.6 Thuật toán KNN.

Thuật toán đã được trình bày ở chương trước. Trong chương này, nhóm tác giả đã áp dụng thuật toán cho bài toán nhận dạng đầu người – giải quyết vấn đề khó khăn trong việc có một nhóm người đi cùng với nhau. Với đặc điểm dễ cài đặt và tính hiệu quả ổn định, thích hợp với việc xử lý trong thời gian thực nên rất phù hợp với yêu cầu bài toán.

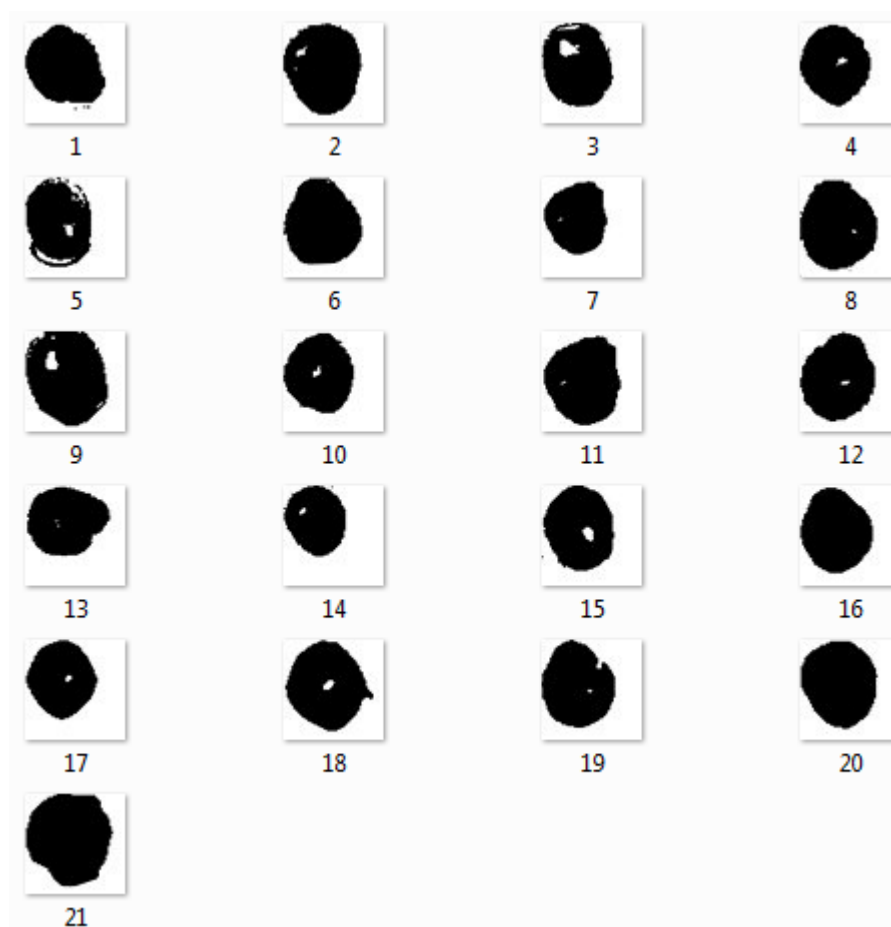
- **Dữ liệu huấn luyện:** 40 ảnh đen trắng có kích thước 50x50 (tương đương với vector gồm 2500 phần tử) được chia thành 2 bộ. Bộ I bao gồm 21 ảnh chứa đầu người, bộ II bao gồm 19 ảnh không chứa đầu người.

- **Dữ liệu vào:** Hình vuông X kích thước 50x50 lấy ra từ ảnh frame hiện tại.

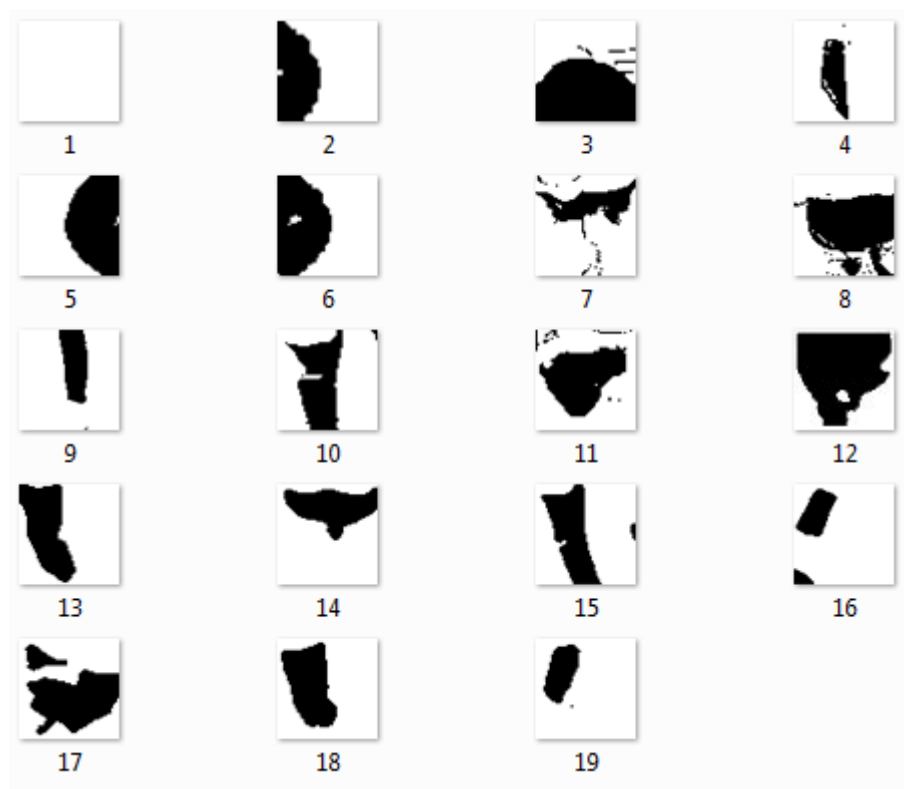
- **Dữ liệu ra:** X có phải là đầu người hay không.

Để thuật toán có thể áp dụng được ta cần hai giai đoạn chính: Giai đoạn huấn luyện và giai đoạn nhận dạng. Qua quá trình kiểm nghiệm nhóm tác giả đã xây dựng được các ảnh chứa và không chứa đầu người (tập huấn luyện) được mô tả như hình bên dưới. Bộ ảnh huấn luyện này bao gồm 21 ảnh chứa đầu người và 19 ảnh không chứa đầu người ( tất cả các ảnh đều có kích thước 50x50) và là ảnh đen trắng (0-255).

#### Giai đoạn huấn luyện (lưu trữ các mẫu học chuẩn)



Hình 3.10: Tập ảnh huấn luyện chứa đầu người



Hình 3.11: Tập ảnh huấn luyện không chứa đầu người

#### Giai đoạn nhận dạng:

Ban đầu từ ảnh màu, ta sẽ chuyển ảnh về dạng ảnh xám. Sau đó sử dụng một hình vuông có kích thước 50x50 và qua tất cả các điểm trên ảnh và sử dụng thuật toán KNN như đã trình bày trong chương trước để xác định xem khu vực ảnh hình vuông đó có là đầu người hay không. Bằng cách sử dụng 5 “láng giềng” gần nhất với khu vực ảnh đang xét, ta sẽ tính sai số đến từng tập ảnh huấn luyện và đi đến kết luận cuối cùng.



Hình 3.12: Phát hiện vùng đầu người bằng thuật toán KNN



### 3.3 Đề xuất, cải tiến.

#### 3.3.1 Cải tiến vùng quan sát của camera.

Qua thực nghiệm nhóm tác giả nhận thấy tốc độ của hệ thống còn chậm do vùng quan sát rộng có những vùng không cần thiết phải theo dõi đối tượng. Ví dụ:



*Hình 3.13: Sơ đồ cải tiến vùng quan sát*

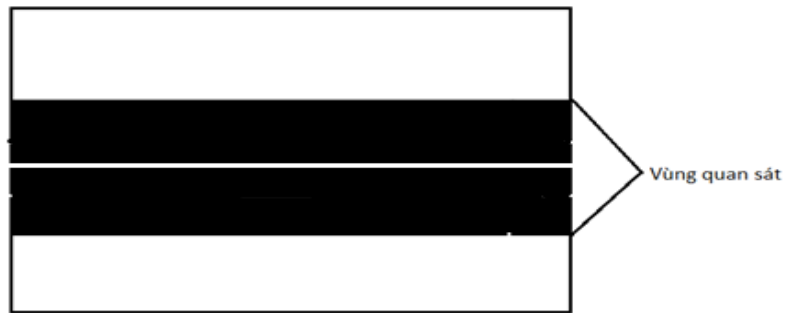
Trên hình ta thấy rằng, các khu vực được khoanh số 1 và 2 là không cần thiết phải theo dõi vì nó nằm ngoài phạm vi hai đường ranh giới ảo.

Chính vì vậy, ta chỉ cần thiết lập vùng quan sát về theo dõi các đối tượng trong phạm vi hai đường ranh giới ảo mà thôi. Điều này có ý nghĩa cực kỳ lớn, vì:

- Giảm chi phí thời gian cho việc trừ ảnh (kích thước ảnh nhỏ đi)
- Giảm chi phí thời gian cho việc theo dõi các đối tượng

Đề xuất này sẽ thể hiện được hiệu quả rất lớn nếu trong trường hợp trong khu vực không cần quan tâm xuất hiện một đám đông hay khu vực đi lại ngang qua.

Phương pháp đề xuất: Dựa vào vị trí của hai đường ranh giới ảo “vào/ra” ta mở rộng về hai phía một khoảng rộng k đơn vị. như vậy vùng quan sát sẽ chỉ giới hạn ở một khung hình nhỏ hơn khung hình của camera.



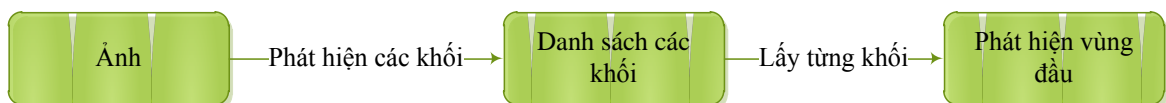
Hình 3.14: Vùng quan sát sau cải tiến

### 3.3.2 Cải tiến phát hiện các đối tượng dính liền nhau.

Như đã trình bày ở các chương trên, bài toán kiểm soát vào ra luôn gặp rất nhiều khó khăn. Trong đó phải kể đến việc các đối tượng bị dính liền vào nhau làm giảm khả năng kiểm soát của hệ thống. Bằng việc sử dụng thuật toán KNN để phát hiện vùng đầu người (trong điều kiện vị trí của camera giám sát đặt vuông góc với lối đi lại) đã giải quyết được phần nào khó khăn của bài toán. Tuy hiệu quả chưa được cao nhưng đó là tiền đề cho những nghiên cứu phát triển sau này để cải thiện tính chính xác hơn nữa.

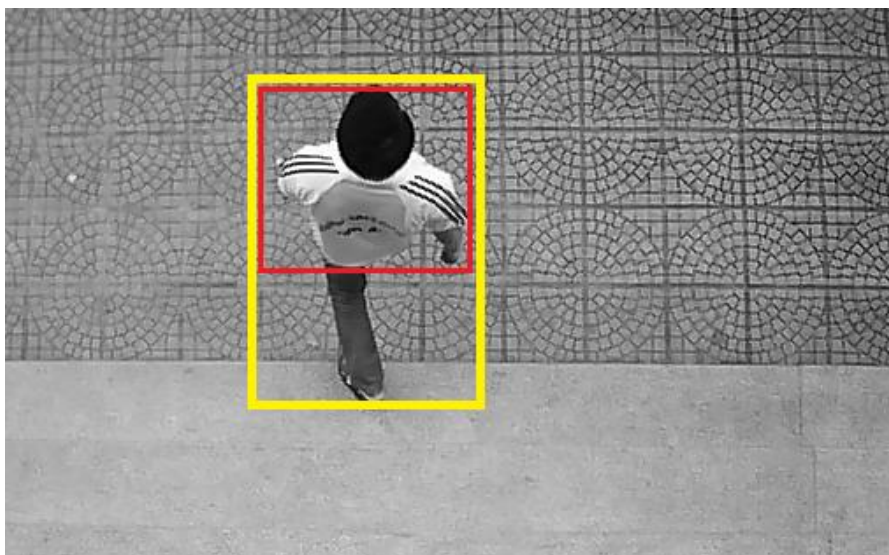
Tuy nhiên vấn đề ở đây đó là để nhận dạng được chính xác thì ta phải có bộ dữ liệu huấn luyện tương đối nhiều. Điều này đặt ra bài toán làm sao để có thể tính sai số đến các bộ huấn luyện trong thời gian thực? Đặc biệt đây là bài toán xử lý video nên yếu tố nhanh là rất cần thiết.

Để làm được điều này, qua quá trình thực nghiệm và quan sát nhóm tác giả đã rút ra được quy trình phát hiện đầu người nhanh đáp ứng trong thời gian thực được biểu diễn bằng sơ đồ bên dưới.



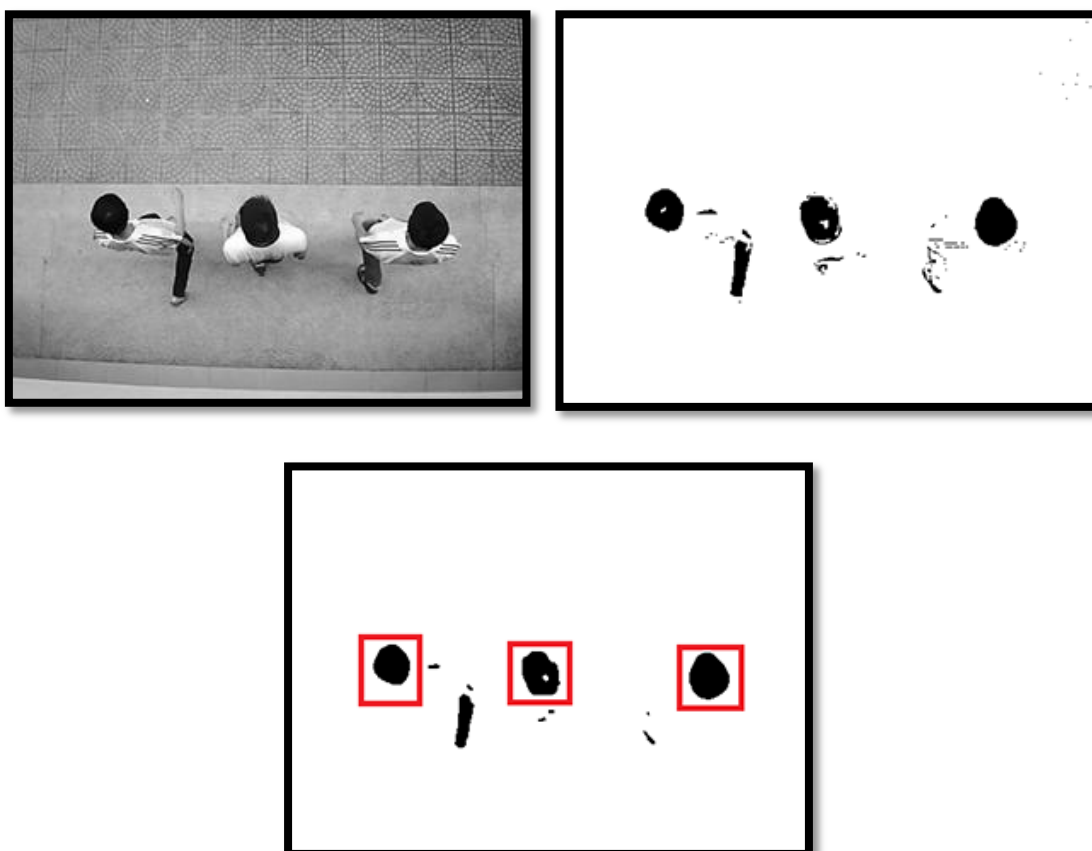
Hình 3.15: Sơ đồ tóm tắt quy trình phát hiện đầu người

Thay vì phải duyệt qua toàn bộ ảnh, ta chỉ cần phát hiện các khối do module 1 cung cấp sau đó lấy từng khối đưa vào bộ nhận dạng. Điều này làm giảm đáng kể quá trình nhận dạng. Bên cạnh đó, nhận thấy rằng vị trí của đầu người nằm phía trên cùng của cơ thể, nên ta chỉ cần xét phần nửa trên của khối ảnh chuyển động. Từ đó tăng tốc độ nhận dạng. Giảm 1 nửa thời gian tính toán. Hình bên dưới thể hiện việc phát hiện đầu người dựa vào khối chuyển động. Hình vuông màu vàng bên ngoài cùng biểu diễn khu vực phát hiện được chuyển động. Hình vuông màu đỏ bên trong là nửa trên khu vực chúng ta cần quan tâm để nhận diện được phần đầu người.



*Hình 3.16: Cải tiến khu vực chứa đầu người*

Bên cạnh đó, việc chỉ để ảnh xám để nhận diện vùng đầu người còn chưa được chính xác. Nhóm tác giả xin được đề xuất giải pháp chuyển ảnh xám về ảnh đen trắng sử dụng ngưỡng  $T=40$  (trong điều kiện thực nghiệm), sau đó tiếp tục loại bỏ nhiễu và các chi tiết không cần thiết để được ảnh cuối cùng chứa vùng đầu người được rõ nét nhất. Qua đó nâng cao độ chính xác của thuật toán.



*Hình 3.17: Tiền xử lý ảnh trước khi phát hiện đầu người*



Sau khi cải tiến bằng các biện pháp hàng loạt như đã nêu trên, chương trình đã phát hiện không những nhanh mà còn chuẩn xác vùng chứa đầu người. Giải quyết được vấn đề khó khăn trong việc các đối tượng dính liền nhau với bài toán kiểm soát vào ra.

### **3.4 Điều kiện thực nghiệm.**

#### **3.4.1 Môi trường thực nghiệm.**

- Các môi trường thông thường như: Nhà ở, phòng học, thang máy với các yếu tố
  - Ánh sáng vừa đủ
  - Không có khói bụi nhiều
  - Người đi lại không mang đồ đặc kích thước lớn
  - Không có cây cối trong khu vực giám sát
  - Ảnh nền không có nhiều vật chuyển động mạnh như quạt trần.
  - Màu nền không có quá nhiều màu sắc và vùng màu.

#### **3.4.2 Thiết bị thực nghiệm.**

##### **Thiết bị ghi hình hoặc file video:**

- Thiết bị ghi hình: độ phân giải  $\geq 3.0$  Mps
- File video : Kích thước lớn hơn 480x600 pixel

##### **Cấu hình máy**

- CPU tối thiểu: 2.0 GHz
- RAM tối thiểu 1Gb

#### **3.4.3 Khoảng cách thực nghiệm.**

- Thiết bị ghi hình cách mặt đất không quá 4m
- Không dưới 2m
- Góc quay từ 45-90 độ. (tùy thuộc vào tiêu cự của thiết bị ghi hình – Khoảng cách đề nghị trên thông dụng cho các camera an ninh tại trường học)

### **3.5: Chương trình demo.**

#### **3.5.1 Các bước thực hiện của chương trình.**

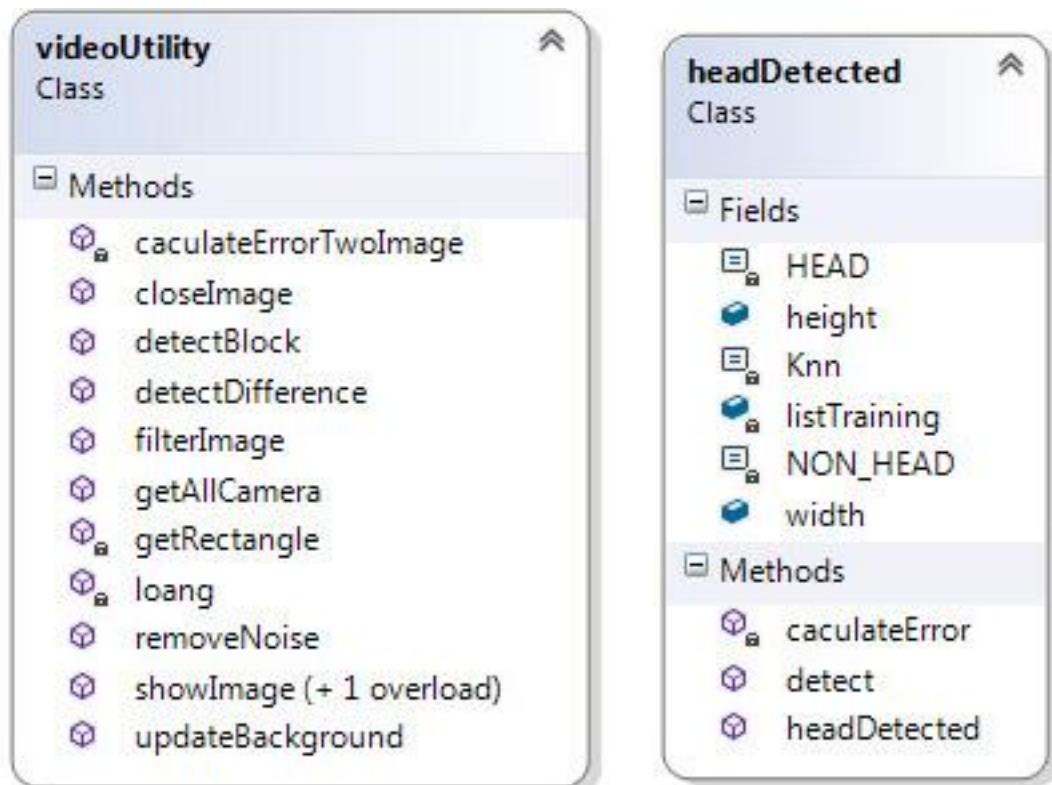
Để có thể xây dựng được chương trình “kiểm soát lượng người vào ra trong tòa nhà”, nhóm tác giả đã sử dụng thư viện hỗ trợ xử lý và thao tác với ảnh trên ngôn ngữ C Sharps, đó là thư viện EmguCV – Một thư viện mở rộng từ OpenCV (dành cho ngôn ngữ C/C++). Thư viện cung cấp các hàm thao tác với ảnh có sẵn và hiệu quả như tính Histogram của ảnh, biến đổi nhị phân, ảnh xám, mở ảnh, lưu ảnh, tương tác với các thiết bị ngoại vi như camera, sensor...

Chương trình được mô tả các thao tác thực hiện như sau:

- Mở video (.avi) hoặc kết nối với camera bên ngoài

- Lấy từng khung hình và cập nhật mô hình nền
- Trừ nền để phát hiện sự dịch chuyển
- Xử lý nhiễu và dẫn ảnh và khoanh vùng các khối di chuyển
- Phát hiện vùng đầu có trong khối dịch chuyển
- Theo dõi và kiểm soát lượng vào ra dựa vào đầu người.

### 3.5.2 Mô hình các lớp quan trọng được cài đặt.



Hình 3.18: Các lớp chính được cài đặt trong hệ thống

#### Lớp “videoUtility”

Đây là lớp chứa các hàm tương tác trực tiếp với frame ảnh. Lớp cung cấp các hàm như lấy tất cả các camera được kết nối với máy tính, cập nhật nền, tính toán sai số giữa hai ảnh, phát hiện các khối, khử nhiễu, lọc màu ảnh.... Đặc biệt class được thiết kế sao cho có thể sử dụng lại vào các hệ thống khác. Nghĩa là sau này khi có nhu cầu phát triển chức năng cũng như mô hình của hệ thống thì ta sẽ giảm chi phí và thời gian viết lại code cho chương trình. Class cũng được viết theo hướng “Lập trình hướng đối tượng”.

#### Lớp “headDetected”

Đây là lớp chứa các hàm thao tác với ảnh để tìm ra được vùng đầu có trong ảnh sử dụng thuật toán KNN. Bằng cách sử dụng và lưu trữ lại các danh sách file huấn luyện đã được lưu trước đó, hàm “detect” sẽ có nhiệm vụ tính sai số đến từng file và sử dụng K láng giềng gần nhất để kết luận xem vùng ảnh đang xét có là đầu người hay không là đầu người. Hàm cho độ chính xác khá cao.

### 3.5.3 Kết quả thực nghiệm và hình ảnh chương trình.

Qua thực nghiệm nhóm tác giả thấy rằng hệ thống thực hiện cho kết quả khá tốt trong điều kiện cho phép. Trong trường hợp đối tượng có nhiều và dính vào nhau thì vẫn chưa giải quyết được triệt để. Đây cũng là một hướng phát triển và nghiên cứu cho các hệ thống sau này.

*Bảng 3.1: Kết quả thực nghiệm của hệ thống*

Tên video	Độ dài Frame	Địa điểm thực th nghiệm	Vào/Ra thực tế	Vào/Ra đếm được	Độ chính xác
<b>Capture_20140401.avi</b>	1653	Thư viện khu A	15/10	13/10	85%
<b>Capture_20140402.avi</b>	2000	Thư viện khu A	17/8	15/5	70%
<b>Capture_20140403.avi</b>	1200	Thư viện khu A	10/14	10/14	100%
<b>Capture_20140404.avi</b>	3500	Khu A nhà A8	20/10	15/8	60%
<b>Capture_20140405.avi</b>	3000	Khu A nhà A8	18/15	16/14	70%
<b>Capture_20140406.avi</b>	1500	Khu A nhà A9	8/8	8/8	100%

Nhìn chung, qua thực nghiệm nhóm tác giả thấy rằng hệ thống chưa thực sự chính xác ở module đếm người vào ra sử dụng khoảng cách nhỏ nhất để liên kết các đối tượng ở frame khác nhau. Phương pháp này không hiệu quả khi đối tượng di chuyển quá nhanh hoặc đứt quãng. Một lý do khác dẫn đến hệ thống bị sai đó là sự nhập nhằng của các đối tượng. Nghĩa là trong cùng một thời điểm có nhiều đối tượng khác nhau di chuyển hỗn loạn quanh khu vực giám sát.

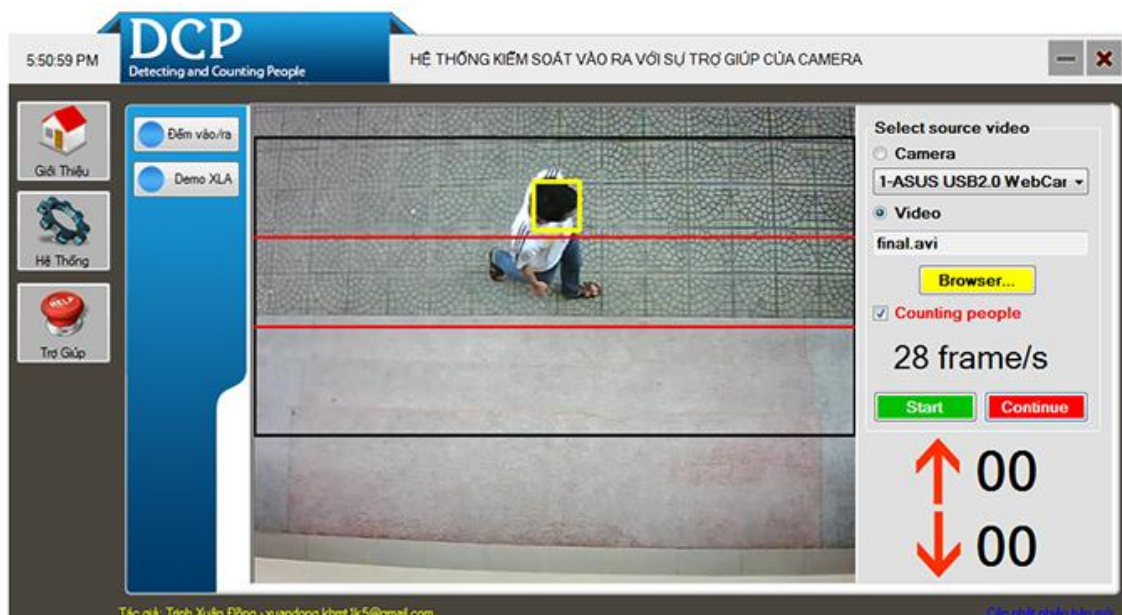
Tuy nhiên, hệ thống tỏ ra chính xác cao trong trường hợp có đơn lẻ đối tượng đi vào khu vực giám sát hoặc một khối người di chuyển. Nhưng nên hạn chế tối đa các đối tượng di chuyển hỗn loạn. Trong trường hợp đối tượng di chuyển quá chậm ta có thể tăng số khung hình/giây để nhanh chóng phát hiện và kiểm soát đối tượng.

Hình bên dưới là giao diện chính của hệ thống. Bao gồm form giới thiệu về đề tài và form đếm số người vào ra lấy dữ liệu từ video hoặc camera giám sát.



Hình 3.19: Form giới thiệu về hệ thống

Chương trình được thiết kế và lập trình trên ngôn ngữ C# (4.5). Giao diện tuân theo quy tắc thiết kế giao diện trong môn học “Giao diện người máy”, đảm bảo tính dễ sử dụng cũng như đơn giản đối với người sử dụng. Chương trình cũng sử dụng các hình ảnh trực quan, kết nối đến các camera được gắn vào máy tính. Nhờ đó có thể sử dụng với video lưu trên máy tính hoặc camera bên ngoài. Đáp ứng yêu cầu thực tế của bài toán đặt ra.



Hình 3.20: Form đếm người vào/ra

### **3.6 Đánh giá các kỹ thuật áp dụng.**

#### **3.6.1 Kết quả đạt được.**

- Hoàn thiện được hệ thống giám sát vào ra sử dụng camera (video).
- Tỷ lệ chính xác khá cao trong trường hợp đối tượng đơn lẻ và cả đối tượng dính liền nhau.
- Chương trình chạy tốt nhất cho môi trường trong nhà.
- Các thuật toán áp dụng cho hệ thống chạy rất nhanh, phù hợp với thời gian thực.

#### **3.6.2 Tồn tại.**

- Tốc độ xử lý còn chậm khi có nhiều đối tượng cùng đi vào khu vực giám sát
- Nhạy cảm với việc thay đổi ánh sáng đột ngột
- Chưa khoanh chính xác đối tượng khi đối tượng di chuyển rất chậm hoặc quá nhanh.

# KẾT LUẬN

## 1. Những mục tiêu đạt được.

- Tìm hiểu các kỹ thuật nhận dạng và theo dõi đối tượng trong video.
- Vận dụng xây dựng ứng dụng đếm số người ra vào trong một tòa nhà dưới sự hỗ trợ của camera trong ngôn ngữ C# trên nền. Net Framework 4.5 sử dụng thư viện hỗ trợ xử lý ảnh EmguCV.
- Phát hiện được đối tượng chuyển động và đếm lượng người vào, ra tương đối chính xác trong hai trường hợp đối tượng riêng lẻ và đối tượng dính liền nhau.
- Áp dụng thành công kỹ thuật đường ranh giới ảo hai đường.
- Áp dụng và cải tiến thành công kỹ thuật nhận dạng đầu người cho bài toán tách các đối tượng dính liền nhau trong cùng một khối.
- Đề xuất phương pháp cải tiến vùng quan sát giám sát và cải tiến tốc độ xử lý, độ chính xác của kỹ thuật nhận dạng vùng đầu người.
- Đánh giá hiệu quả của phương pháp cải tiến.
- Hệ thống chạy ổn định và có độ chính xác >80% trong điều kiện thực nghiệm.

## 2. Những mục tiêu chưa đạt.

- Chưa tìm hiểu và áp dụng các thuật toán khác (như Optical Flow) trong việc theo dõi đối tượng chuyển động.
- Chưa áp dụng và so sánh được các thuật toán nhận dạng và phân lớp đối tượng như mạng nơ ron, PCA, K-Means.

## 3. Hướng nghiên cứu trong tương lai.

- Tìm hiểu các thuật toán phân loại đối tượng để phân biệt đối tượng là người hay không, áp dụng cho bài toán giám sát ở môi trường bên ngoài khi có nhiều đối tượng.
- Tìm hiểu các thuật toán theo dõi chuyển động như Optical flow.
- Áp dụng thêm các thuật toán khác để cải thiện, nâng cao độ chính xác của hệ thống.
- Nghiên cứu, áp dụng thuật toán theo vết đối tượng camshift kết hợp với bộ lọc kalman dự đoán tâm đối tượng để nâng cao độ chính xác.
- Sử dụng thuật toán mạng nơ ron cho bài toán phân loại, nhận dạng đối tượng.

# TÀI LIỆU THAM KHẢO

- [1] Ths. Lương Mạnh Bá, Pts. Nguyễn Thanh Thủy, Nhập Môn Xử Lý Ảnh Số, Hà Nội - 1999
- [2] <http://vi.wikipedia.org/wiki/Avi>.
- [3]. Yi githan Dedeo glu, Moving Object Detection, Tracking And Classification For Smart Video Surveillance , August- 2004.
- [4] Alper Yilmaz, Omar Javed, and Mubarak Shah, "Object Tracking: A Survey" pp. 7-15.
- [5] Anurag Mittal and Mikos Paragios, "Motion Based Background Subtraction using Adaptive Kernel Density Estimation" pp. 302-309, 2004
- [6] C.Stauffer and W.Grimson, "Adaptive Background mixture models for Real-time tracking" pp. 750-755, 2009.
- [7] J. Heikkila and O. Silven, Fort Collins, Colorado. A real-time system for monitoring of cyclists and pedestrians. In Proc. of Second IEEE Workshop on Visual Surveillance, pages 74–81, June 1999.
- [8] L. Wang, W. Hu, and T. Tan. Recent developments in human motion analysis. Pattern Recognition, 36(3):585–601, March 2003.
- [9] I. Haritaoglu, D. Harwood, and L.S. Davis. W4: A real time system for detecting and tracking people. In Computer Vision and Pattern Recognition, pages 962–967, 1998.
- [10] C. Stauffer and W. Grimson. Adaptive background mixture models for real-time tracking. In Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, page 246252, 1999.
- [11] A. J. Lipton, H. Fujiyoshi, and R.S. Patil. Moving target classification and tracking from real-time video. In Proc. of Workshop Applications of Computer Vision, pages 129–136, 1998.
- [12] L. Wang, W. Hu, and T. Tan. Recent developments in human motion analysis. Pattern Recognition, 36(3):585–601, March 2003.
- [13] R. T. Collins et al. A system for video surveillance and monitoring: VSAM final report. Technical report CMU-RI-TR-00-12, Robotics Institute, Carnegie Mellon University, May 2000.
- [14] M. Godec, Robust Object Tracking using Semi-Supervised Online Boosting, 2008.
- [15] A. Yilmaz, Object Tracking: A Survey, ACM Computing Surveys, 2006.
- [16] C. Harris and M. Stephens, A combined corner and edge detector, Alvey Vision Conference, 1988.

- [17] D. G. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision, 2004.
- [18] T. J. Broida and R. Chellappa, Estimation of object motion parameters from noisy images, IEEE Trans. on Pattern Analysis and Machine Intelligence, 1986.
- [19] C. Veenman, M. Reinders, and E. Backer, Resolving motion correspondence for densely moving points, IEEE Transaction. on Pattern. and Analysis, 2001.
- [20] J. Shi and C. Tomasi, Good features to track, IEEE Conference on Computer Vision and Pattern Recognition., 1994.
- [21] D. Comaniciu, V. Ramesh, and P. Meer, Kernel-based object tracking, IEEE Transactions on Pattern Analysis an Machine Intelligence, 2003.
- [22] S. Avidan, Support vector tracking, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004.
- [23] J. Kang, I. Cohen, and G. Medioni, Object reacquisition using invariant appearance model, International Conference on Pattern Recongition, 2004.
- [24] M. Isard and A. Blake, Condensation - Conditional Density Propagation for Visual Tracking, International Journal of Computer Vision (IJCV), vol. 29, 1998, pp. 5-28.
- [25] Dorin Comaniciu, Visvanathan Ramesh, Peter Meer, *Kernel Base object tracking*, Real-time Vision and modeling Department, Siemens Corporate Research, Princeton, NJ, USA, 2003.
- [26] W. Krattenthaler, K.J Mayer, and M. Zeiller, *Point Correlation. A Reduced Cost Template matching technique*. Proc, ICIP, pp 208-212, 1994
- [27] A.K.Jain, Y. Zhong, and S. Lakshamanan, *Object matching using deformable templates*. IEEE Trans, Pattern Analysis and machine intelligence, vol 18 no 3, pp 267-278, Mar 1996.
- [28] Jae-Won Kim, Kang-Sun Choi, Byeong-Doo Choi, and Sung-Jea Ko, *Real-time Vision-based People Counting System for the Security Door*, Department of Electronics Engineering, Korea University, Anam-dong, Sungbuk-ku, Seoul, 136-701, Korea 2003.



## **BẢNG ĐỐI CHIẾU THUẬT NGỮ VIỆT – ANH**

<b>Phân tích thành phần chính</b>	Principal component analysis
<b>Hình nền</b>	Background
<b>So khớp mẫu</b>	Template patching
<b>Đường ranh giới ảo</b>	Virtual line
<b>Phân lớp đối tượng</b>	Object classssification
<b>Theo vết đối tượng</b>	Object tracking
<b>Phát hiện đối tượng</b>	Object detecting
<b>Nhận diện hành vi</b>	Action recognition
<b>Mạng noron nhân tạo</b>	Artificial neural network
<b>Tương tự</b>	Analog
<b>Điểm ảnh</b>	Pixel