

**BỘ GIÁO DỤC VÀ ĐÀO TẠO  
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ TP.HCM**

---



**PHẠM BẢO QUỐC**

**THEO DÕI ĐỐI TƯỢNG CHUYỂN ĐỘNG  
BẰNG PHƯƠNG PHÁP LỘC TÍCH HỢP**

**LUẬN VĂN THẠC SĨ**

**Chuyên ngành: Công Nghệ Thông Tin**

**Mã ngành: 60480201**

**Hướng dẫn khoa học: TS. NGUYỄN THANH BÌNH**

**TP.HCM, THÁNG 11 NĂM 2014**

**CÔNG TRÌNH ĐƯỢC HOÀN THÀNH TẠI**  
**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ TP. HCM**

Cán bộ hướng dẫn khoa học: **TS. NGUYỄN THANH BÌNH**  
(Ghi rõ họ, tên, học hàm, học vị và chữ ký)

**TS. NGUYỄN THANH BÌNH**

Luận văn Thạc sĩ được bảo vệ tại Trường Đại học Công nghệ TP.HCM ngày 09 tháng 11 năm 2014.

Thành phần Hội đồng đánh giá Luận văn Thạc sĩ gồm:  
(Ghi rõ họ, tên, học hàm, học vị của Hội đồng chấm bảo vệ Luận văn Thạc sĩ)

<b>TT</b>	<b>Họ và tên</b>	<b>Chức danh Hội đồng</b>
1	PGS. TSKH. Nguyễn Xuân Huy	Chủ tịch
2	PGS. TS. Lê Hoàng Thái	Phản biện 1
3	TS. Lư Nhật Vinh	Phản biện 2
4	TS. Lê Mạnh Hải	Ủy viên
5	TS. Võ Đình Bảy	Ủy viên, Thư ký

Xác nhận của Chủ tịch Hội đồng đánh giá Luận văn sau khi Luận văn đã được sửa chữa (nếu có).

**Chủ tịch Hội đồng đánh giá LV**

**PGS. TSKH. Nguyễn Xuân Huy**

TP. HCM, ngày 27 tháng 03 năm 2014

## NHIỆM VỤ LUẬN VĂN THẠC SĨ

Họ tên học viên: PHAM BẢO QUỐC      Giới tính: Nam  
Ngày, tháng, năm sinh: 14/02/1979      Nơi sinh: TP.HCM  
Chuyên ngành: Công Nghệ Thông Tin      MSHV: 1241860016

### I. Tên đề tài:

Theo dõi đối tượng chuyển động bằng phương pháp lọc tích hợp

### II. Nhiệm vụ và nội dung:

- Nghiên cứu phương pháp tích hợp các bộ lọc phù hợp vào các phương pháp hiện tại để nâng cao hiệu quả của bài toán.
- Nghiên cứu các phương pháp xử lý ảnh và video. Nghiên cứu về bài toán phát hiện chuyển động và bài toán theo vết đối tượng. Nghiên cứu các phương pháp phát hiện chuyển động và theo vết đối tượng phổ biến. Nghiên cứu phương pháp tích hợp các bộ lọc phù hợp vào các phương pháp hiện tại và đề xuất giải thuật hiệu quả.

III. Ngày giao nhiệm vụ: 27/03/2014

IV. Ngày hoàn thành nhiệm vụ: 27/09/2014

V. Cán bộ hướng dẫn: TS. NGUYỄN THANH BÌNH

CÁN BỘ HƯỚNG DẪN

KHOA QUẢN LÝ CHUYÊN NGÀNH

TS. NGUYỄN THANH BÌNH

.....

## LỜI CAM ĐOAN

Tôi xin cam đoan đây là công trình nghiên cứu của riêng tôi. Các số liệu, kết quả nêu trong Luận văn là trung thực và chưa từng được ai công bố trong bất kỳ công trình nào khác.

Tôi xin cam đoan rằng mọi sự giúp đỡ cho việc thực hiện Luận văn này đã được cảm ơn và các thông tin trích dẫn trong Luận văn đã được chỉ rõ nguồn gốc.

**Học viên thực hiện Luận văn**

*(Ký và ghi rõ họ tên)*

**PHẠM BẢO QUỐC**

## LỜI CẢM ƠN

Để hoàn thành luận văn này tuy gặp nhiều khó khăn nhưng tác giả đã nhận được rất nhiều sự giúp đỡ và lời động viên từ những người thân trong gia đình, từ Thầy Cô, từ các giáo viên đồng nghiệp và từ cả các học sinh thân yêu.

Đầu tiên em xin gửi lời cảm ơn đến TS. Nguyễn Thanh Bình – Cán bộ hướng dẫn khoa học, trường ĐH Bách Khoa TP.HCM. Thầy đã nhiệt tình hướng dẫn em trong nghiên cứu, nhiệt tình giúp em chỉnh sửa luận văn để có một luận văn tốt nhất.

Em xin gửi lời cảm ơn đến TS. Nguyễn Chánh Thành – Trưởng khoa CNTT, trường ĐH Công Nghệ TP.HCM. Cảm ơn Thầy đã quan tâm hướng dẫn, giải đáp mọi thắc mắc cho em trong suốt quá trình thực hiện luận văn. Em xin cảm ơn những tình cảm mà Thầy đã dành cho em.

Em xin gửi lời cảm ơn đến các Thầy, Cô đang công tác tại phòng QLKH – ĐTSĐH đã quan tâm giúp đỡ, hướng dẫn các thủ tục để em có thể hoàn thành luận văn đúng yêu cầu và đúng tiến độ. Em xin chân thành cảm ơn các Thầy, Cô đang công tác tại khoa CNTT của các trường đại học đã giảng dạy em trong suốt quá trình học tập tại lớp cao học này.

Em xin chân thành cảm ơn BGH trường THPT Thanh Đa và các giáo viên đồng nghiệp đã quan tâm động viên giúp đỡ để em có thời gian nghiên cứu. Xin chân thành cảm ơn các thành viên trong gia đình đã động viên, tạo mọi điều kiện để tôi yên tâm nghiên cứu. Và cuối cùng xin cảm ơn các em học sinh đã quan tâm đến Thầy trong suốt thời gian học cao học và nghiên cứu luận văn.

Xin chân thành cảm ơn.

**PHẠM BẢO QUỐC**

## TÓM TẮT

Theo dõi đối tượng chuyển động là một ứng dụng phổ biến trong lĩnh vực thị giác máy tính. Theo dõi đối tượng chuyển động phân tích các khung hình trong chuỗi hình ảnh nhận được từ camera quan sát, tách đối tượng chuyển động ra khỏi ảnh nền, để từ đó theo dõi bám sát đối tượng. Theo dõi đối tượng chuyển động được ứng dụng rộng rãi trong các hệ thống giám sát an ninh, điều khiển giao thông, điều khiển tự động...

Giải thuật theo dõi đối tượng chuyển động gồm hai bước chính là phát hiện chuyển động và theo vết đối tượng. Độ chính xác của bước phát hiện chuyển động ảnh hưởng rất lớn đến bước theo vết đối tượng và kết quả của bài toán theo dõi đối tượng chuyển động. Có nhiều phương pháp để phát hiện chuyển động. Tuy nhiên mỗi phương pháp đều có ưu nhược điểm riêng. Hầu hết các phương pháp phát hiện chuyển động đều cho ra kết quả thiếu chính xác khi có nhiễu và thay đổi ánh sáng.

Trong luận văn, tác giả nghiên cứu đề xuất phương pháp tích hợp các bộ lọc phù hợp vào các phương pháp hiện tại để nâng cao hiệu quả của bài toán theo dõi đối tượng chuyển động. Đặc biệt ở bước phát hiện chuyển động, tác giả đã đề xuất được phương pháp tỷ lệ cường độ ảnh (Intensity Ratio Image) không bị ảnh hưởng bởi sự thay đổi ánh sáng. Ảnh tỷ lệ cường độ được tính toán dựa trên cường độ trung bình của khung hình hiện tại và cường độ của từng điểm ảnh trong khung hình đó. Tỷ lệ cường độ của một điểm ảnh gần như không thay đổi giữa hai khung hình liên tiếp. Do đó nó khắc phục được nhược điểm nhạy cảm với thay đổi ánh sáng của các phương pháp phát hiện chuyển động hiện tại.

Để theo dõi chuyển động, tác giả đã áp dụng phương pháp dựa trên bộ lọc Kalman để theo vết đối tượng chuyển động. Giải thuật mà đề tài sử dụng có khả năng chống nhiễu tốt và không bị tác động bởi sự thay đổi ánh sáng.

## **ABSTRACT**

Moving object tracking is a popular application of the computer vision. Moving object tracking analyzes frame by frame in image sequences obtained from observation cameras, separating moving objects from the background, and tracking them. Moving object tracking is widely used in security monitoring systems, traffic control systems, automatic control systems...

Two main steps of moving object tracking algorithm are motion detection and object tracking. The accuracy of motion detection step greatly affects to object tracking step and the results of moving object tracking. There are many methods for motion detection. However, each method has advantages and disadvantages. In almost methods, if the input video has noise and light change, moving objects will not be extracted accurately.

The thesis proposes an algorithm to track moving objects by integrating appropriate filtering methods. Especially in motion detection step, the thesis proposed a new combination method based on the intensity ratio image (IRI) concept that is not affected by light change. The intensity ratio image is computed by the average intensity of current frame and the intensity of each pixel in that frame. The intensity ratio of a pixel is almost unchanged between two consecutive frames. Therefore, the sensitivity with light change is overcome.

This thesis applied the Lucas-Kanade optical flow method based on the proposed intensity ratio image to detect moving objects. Then apply some common methods to track moving objects. The proposed algorithm has good noise tolerance and is not affected by light change.

# MỤC LỤC

<b>CHƯƠNG 1</b>	<b>GIỚI THIỆU TỔNG QUAN.....</b>	<b>1</b>
1.1	GIỚI THIỆU ĐỀ TÀI .....	1
1.2	MỤC TIÊU, NỘI DUNG CỦA ĐỀ TÀI .....	2
1.3	GIỚI HẠN ĐỀ TÀI .....	3
1.4	PHƯƠNG PHÁP NGHIÊN CỨU .....	3
1.5	Ý NGHĨA KHOA HỌC VÀ THỰC TIỄN CỦA ĐỀ TÀI .....	3
1.6	CẤU TRÚC LUẬN VĂN.....	4
<b>CHƯƠNG 2</b>	<b>CƠ SỞ LÝ THUYẾT VÀ CÁC NGHIÊN CỨU LIÊN QUAN.....</b>	<b>5</b>
2.1	CÁC KHÁI NIỆM CƠ BẢN VỀ VIDEO .....	5
2.1.1	<i>Video</i> .....	5
2.1.2	<i>Xử lý ảnh và video</i> .....	5
2.1.3	<i>Bộ lọc trung vị</i> .....	7
2.1.4	<i>Bộ lọc tương quan</i> .....	7
2.1.5	<i>Bộ lọc hình thái</i> .....	8
2.2	BÀI TOÁN THEO DÕI ĐỐI TƯỢNG CHUYỂN ĐỘNG .....	10
2.2.1	<i>Giới thiệu bài toán theo dõi đối tượng chuyển động</i> .....	10
2.2.2	<i>Phương pháp trừ nền</i> .....	11
2.2.3	<i>Phương pháp sai khác ảnh</i> .....	12
2.2.4	<i>Phương pháp Gaussian Mixture Model</i> .....	13
2.2.5	<i>Phương pháp Lucas-Kanade Optical Flow</i> .....	14
2.2.6	<i>Thuật toán Grass-Fire</i> .....	17
2.2.7	<i>Bộ lọc Kalman</i> .....	19
2.3	CÁC CÔNG TRÌNH NGHIÊN CỨU LIÊN QUAN .....	21
2.3.1	<i>Tình hình nghiên cứu trên thế giới:</i> .....	21
2.3.2	<i>Tình hình nghiên cứu trong nước:</i> .....	23
<b>CHƯƠNG 3</b>	<b>GIẢI THUẬT ĐỀ XUẤT .....</b>	<b>25</b>
3.1	TỶ LỆ CƯỜNG ĐỘ ẢNH .....	25
3.2	PHƯƠNG PHÁP LỌC TÍCH HỢP .....	28
3.2.1	<i>Phát hiện chuyển động với phương pháp IRI-LK</i> .....	28
3.2.2	<i>Theo vết đối tượng dựa trên bộ lọc Kalman</i> .....	31
<b>CHƯƠNG 4</b>	<b>THỰC NGHIỆM VÀ ĐÁNH GIÁ.....</b>	<b>36</b>
4.1	KẾT QUẢ ĐẠT ĐƯỢC QUA THỰC NGHIỆM.....	36
4.1.1	<i>Môi trường và dữ liệu thực nghiệm:</i> .....	36
4.1.2	<i>Kết quả thực nghiệm phát hiện chuyển động:</i> .....	38
4.1.3	<i>Kết quả thực nghiệm theo vết đối tượng:</i> .....	41
4.2	ĐÁNH GIÁ KẾT QUẢ.....	44
4.2.1	<i>Đánh giá kết quả phát hiện chuyển động:</i> .....	44
4.2.2	<i>Đánh giá kết quả theo vết đối tượng:</i> .....	45
<b>CHƯƠNG 5</b>	<b>KẾT LUẬN.....</b>	<b>46</b>
5.1	KẾT QUẢ ĐẠT ĐƯỢC CỦA ĐỀ TÀI.....	46
5.2	ƯU NHƯỢC ĐIỂM CỦA GIẢI THUẬT ĐỀ XUẤT .....	46
5.3	HƯỚNG PHÁT TRIỂN MỞ RỘNG .....	47



## DANH MỤC CÁC CHỮ VIẾT TẮT

Thuật ngữ hoặc chữ viết tắt	Diễn giải
2D	Two Dimensional – 2 chiều
3D	Three Dimensional – 3 chiều
BG	Background - Ảnh nền
BLOB	Binary Large Object – Đối tượng nhị phân lớn
CNTT	Công Nghệ Thông Tin
CV	Computer Vision – Thị giác máy tính
FG	Foreground – Cận ảnh
FPS	Frame Per Second – Khung hình trên giây
HSI	Hue Saturation Intensity – Màu sắc, độ bão hoà, cường độ
HSV	Hue Saturation Value – Màu sắc, độ bão hoà, giá trị
IRI	Intensity Ratio Image - Ảnh tỷ lệ cường độ
MO	Moving Object – Đối tượng chuyển động
PTZ	Pan Tilt Zoom – Quay, lật, thu phóng
RGB	Red Green Blue – Đỏ, xanh lá, xanh dương
ROI	Region of Interest - Vùng quan tâm
TO	Tracked Object – Đối tượng được theo dõi

## DANH MỤC CÁC HÌNH VẼ

Hình 2.1: Cấu trúc video.....	5
Hình 2.2: Sơ đồ cấu trúc của hệ thống xử lý ảnh và video tổng quát .....	5
Hình 2.3: Hình minh hoạ hoạt động của bộ lọc trung vị.....	7
Hình 2.4: Hình minh hoạ hoạt động của bộ lọc tương quan .....	8
Hình 2.5: Hình minh hoạ một số mặt nạ của bộ lọc tương quan .....	8
Hình 2.6: Hình minh hoạ hiệu ứng Dilation của thao tác Hit.....	9
Hình 2.7: Hình minh hoạ hiệu ứng Erosion của thao tác Fit.....	9
Hình 2.8: Mô hình hệ thống theo dõi đối tượng chuyển động .....	10
Hình 2.9: Hình minh hoạ 5 bước của phương pháp trừ nền.....	11
Hình 2.10: Hình minh hoạ 5 bước của phương pháp sai khác ảnh.....	12
Hình 2.11: Hình minh hoạ kết quả của Gaussian Mixture Model .....	14
Hình 2.12: Hình minh hoạ kết quả của Lucas-Kanade Optical Flow .....	17
Hình 2.13: Hình minh hoạ thuật toán Grass-Fire.....	19
Hình 2.14: Chu trình bộ lọc Kalman .....	19
Hình 3.1: Các khối cơ bản của phương pháp lọc tích hợp .....	28
Hình 3.2: Mô hình phương pháp IRI-LK .....	29
Hình 3.3: Hình minh hoạ mặt nạ đối tượng chuyển động.....	30
Hình 3.4: Hình minh hoạ giải thuật theo vết đối tượng .....	31
Hình 4.1: Kết quả phương pháp IRI-LK với ánh sáng thay đổi tự nhiên .....	39
Hình 4.2: Kết quả phương pháp IRI-LK với thay đổi ánh sáng được thêm vào .....	41
Hình 4.3: Minh hoạ kết quả theo vết đối tượng .....	42
Hình 4.4: Minh hoạ kết quả theo vết đối tượng có che khuất và giao cắt.....	43

## **Chương 1 GIỚI THIỆU TỔNG QUAN**

### **1.1 GIỚI THIỆU ĐỀ TÀI**

Trong thời đại hiện nay, việc ứng dụng công nghệ thông tin (CNTT) đã và đang đem lại rất nhiều lợi ích cho con người. Thị giác là một trong các giác quan quan trọng giúp con người tiếp nhận thông tin từ thế giới xung quanh. Thị giác máy tính (Computer Vision) là một trong các lĩnh vực của CNTT về trí tuệ nhân tạo. Thị giác máy tính cung cấp các phương pháp xử lý, phân tích hình ảnh từ thế giới thực tương tự như cách con người thực hiện, từ đó rút ra thông tin để có những quyết định phù hợp.

Theo dõi đối tượng chuyển động là một trong những bài toán cơ bản của lĩnh vực thị giác máy tính. Theo đó, sự di chuyển của đối tượng được phân tích qua các khung hình trong chuỗi khung hình nhận được từ camera quan sát để phát hiện chuyển động. Sau đó, dựa vào các vùng chuyển động đã phân tích được để xác định đối tượng chuyển động và theo dõi bám sát chúng. Bài toán theo dõi đối tượng chuyển động là bài toán cơ sở cho nhiều bài toán lớn hơn. Sau khi xác định và theo dõi được đối tượng chuyển động, một số thuật toán đặc biệt có thể được áp dụng để phân loại đối tượng, nhận diện đối tượng, phân tích hành vi của đối tượng, đo đếm đối tượng và rất nhiều ứng dụng thực tiễn khác.

Nghiên cứu bài toán theo dõi đối tượng chuyển động sẽ mang lại rất nhiều ứng dụng thiết thực. Các ứng dụng này giúp con người tiết kiệm được nhiều thời gian lẫn công sức nhờ vào khả năng điều khiển tự động và tính toán nhanh của máy tính. Trong một vài trường hợp nó còn có thể cho ra kết quả chính xác hơn những gì con người có thể làm và có thể xử lý tự động các vấn đề phức tạp mà không cần đến sự can thiệp của con người. Một số ứng dụng phổ biến của bài toán theo dõi đối tượng chuyển động là các ứng dụng về giám sát an ninh, điều khiển giao thông, thiết bị tự hành, điều khiển bằng cử chỉ...

Tuy nhiên bài toán theo dõi đối tượng chuyển động không phải là bài toán dễ giải quyết. Các đối tượng chuyển động có nhiều hình dạng khác nhau, chúng có

thể vừa chuyển động vừa thay đổi hình dạng, đa sắc màu và luôn ở trong một khung cảnh phức tạp, đầy nhiễu động. Để nhận diện được đối tượng chuyển động trong bối cảnh như thế là rất khó khăn. Kết quả nhận diện bị ảnh hưởng rất nhiều bởi ánh sáng môi trường thay đổi, các nhiễu động từ thiết bị giám sát, các chuyển động hỗn loạn trong ngữ cảnh bị giao cắt, bị che khuất... Những điều này khiến cho bài toán theo dõi đối tượng chuyển động trở nên phức tạp và kết quả có độ chính xác không cao. Vì vậy cần phải có nhiều nghiên cứu về bài toán theo dõi đối tượng chuyển động để đạt được kết quả có độ chính xác cao hơn. Đồng thời giúp tạo ra các ứng dụng tốt hơn, hiệu quả hơn.

Hiện nay trên thế giới có rất nhiều nghiên cứu về bài toán theo dõi đối tượng chuyển động. Các nghiên cứu này rất đa dạng và đã đạt được một số kết quả nhất định. Tuy nhiên, với nhu cầu tìm ra một giải pháp tốt và hiệu quả hơn nữa khiến cho bài toán theo dõi đối tượng chuyển động vẫn được tiếp tục nghiên cứu.

## **1.2 MỤC TIÊU, NỘI DUNG CỦA ĐỀ TÀI**

Mục tiêu của đề tài là nghiên cứu và đề xuất phương pháp để nâng cao hiệu quả của bài toán theo dõi đối tượng chuyển động. Cụ thể, nghiên cứu phương pháp tích hợp các bộ lọc phù hợp vào các phương pháp hiện tại để theo dõi, truy bắt đối tượng chính xác hơn. Đề tài sẽ là một ứng dụng thiết thực trong đời sống của lĩnh vực thị giác máy tính. Do đó, đề tài sẽ tập trung nghiên cứu các vấn đề sau:

- Tìm hiểu các phương pháp xử lý ảnh và video hiện có. Tìm hiểu các công trình nghiên cứu liên quan đến đề tài hiện có trong và ngoài nước.
- Từ những ưu nhược điểm đã phân tích được, nghiên cứu và đề xuất giải pháp phát hiện chuyển động của đối tượng trong các môi trường phức tạp (ánh sáng thay đổi, nhiễu...) thông qua việc tích hợp các bộ lọc phù hợp vào các phương pháp hiện tại.
- Để đánh giá kết quả đạt được, giải thuật đề xuất sẽ được thực nghiệm với các tập dữ liệu hiện có đã được các nhóm nghiên cứu trên thế giới công nhận. Kết quả thực nghiệm sẽ được so sánh với các giải thuật hiện có.

### **1.3 GIỚI HẠN ĐỀ TÀI**

Đề tài tập trung nghiên cứu cải thiện kết quả của bước phát hiện chuyển động trong giải thuật theo dõi đối tượng chuyển động vì bước này có ảnh hưởng lớn đến kết quả của cả bài toán. Phương pháp dòng quang học được lựa chọn nghiên cứu kỹ hơn các phương pháp khác để tìm ra phương pháp cải tiến hiệu quả. Đề tài chỉ nghiên cứu giải thuật áp dụng trên các video đầu vào là các đoạn video giám sát thu được từ camera tĩnh có độ phân giải trung bình. Đề tài không nghiên cứu sâu việc xử lý chồng lấp trong theo dõi đa đối tượng.

### **1.4 PHƯƠNG PHÁP NGHIÊN CỨU**

Đề tài sử dụng phương pháp nghiên cứu định tính và phương pháp nghiên cứu định lượng.

Ở phương pháp nghiên cứu định tính, các công trình nghiên cứu liên quan trong và ngoài nước sẽ được nghiên cứu. Phương pháp nghiên cứu tài liệu được sử dụng để tìm hiểu về các vấn đề lý thuyết của đề tài. Sau khi nắm vững lý thuyết, sử dụng phương pháp phân tích và tổng kết kinh nghiệm để nghiên cứu kết quả của các công trình nghiên cứu đã tìm hiểu. Dựa trên các kết quả phân tích được, tìm kiếm và đưa ra một vài giả thuyết để cải thiện kết quả của các công trình nghiên cứu đã tìm hiểu. Sử dụng phương pháp toán học để chứng minh tính đúng đắn của giả thuyết. Từ đó loại bỏ các giả thuyết không phù hợp và tìm thêm các giả thuyết mới.

Ở phương pháp nghiên cứu định lượng, đề tài sử dụng phương pháp thống kê, đo đạt kết quả từ đó rút ra kết luận để đánh giá hiệu quả của thuật toán. Các thực nghiệm được thực hiện dựa trên ứng dụng minh họa bằng MATLAB. Kết quả thực nghiệm được so sánh với kết quả của các công trình nghiên cứu đã tìm hiểu, từ đó đánh giá tính hiệu quả của phương pháp đề xuất.

### **1.5 Ý NGHĨA KHOA HỌC VÀ THỰC TIỄN CỦA ĐỀ TÀI**

Về mặt ý nghĩa khoa học, đề tài đã nghiên cứu và đánh giá được một số phương pháp phát hiện chuyển động phổ biến. Đề tài đã đề xuất được phương pháp tỷ lệ cường độ ảnh. Tỷ lệ cường độ ảnh có ưu điểm là không bị ảnh hưởng bởi sự

thay đổi ánh sáng môi trường. Tác giả đã đề xuất phương pháp phát hiện chuyển động hiệu quả dựa trên tỷ lệ cường độ ảnh và phương pháp dòng quang học. Đây là nguồn tài liệu tham khảo tốt cho những ai quan tâm đến lĩnh vực theo dõi đối tượng.

Hơn thế nữa, kết quả khoa học của đề tài được minh chứng qua bài báo “Motion Detection Based on Intensity Ratio Image” được hội nghị quốc tế “International Conference on Nature of Computation and Communication” chấp nhận báo cáo tại hội nghị vào ngày 24 và 25/11/2014 tại TP.HCM. Bài báo này được nhà xuất bản Springer chấp nhận đăng tải vào cuối năm nay.

Về mặt ý nghĩa thực tiễn, đề tài góp phần tạo nền tảng cho các ứng dụng trong lĩnh vực thị giác máy tính. Đề tài có thể được ứng dụng ngay vào lĩnh vực giám sát an ninh để quan sát các đối tượng ra vào cơ quan, ứng dụng vào lĩnh vực giám sát giao thông để theo dõi và đếm các phương tiện, ứng dụng vào lĩnh vực sản xuất để đo đếm hàng hoá, sản phẩm...

## **1.6 CẤU TRÚC LUẬN VĂN**

Cấu trúc của luận văn được chia thành năm chương. Chương 1 là phần giới thiệu tổng quan về đề tài, mục tiêu và nội dung của đề tài, giới hạn đề tài, phương pháp nghiên cứu, ý nghĩa khoa học và thực tiễn của đề tài.

Các cơ sở lý thuyết và các công trình nghiên cứu có liên quan đến bài toán theo dõi đối tượng chuyển động sẽ được trình bày ở chương 2. Từ những ưu nhược điểm của các công trình nghiên cứu có liên quan này, giải thuật theo dõi đối tượng chuyển động bằng phương pháp lọc tích hợp sẽ được đề xuất để giải quyết vấn đề mà luận văn đặt ra. Giải thuật đề xuất này sẽ được trình bày trong chương 3.

Để đánh giá kết quả đạt được của giải thuật đề xuất, các thực nghiệm được thực hiện dựa trên giải thuật đề xuất và so sánh kết quả với các giải thuật hiện có. Tất cả vấn đề này sẽ được làm rõ trong chương 4.

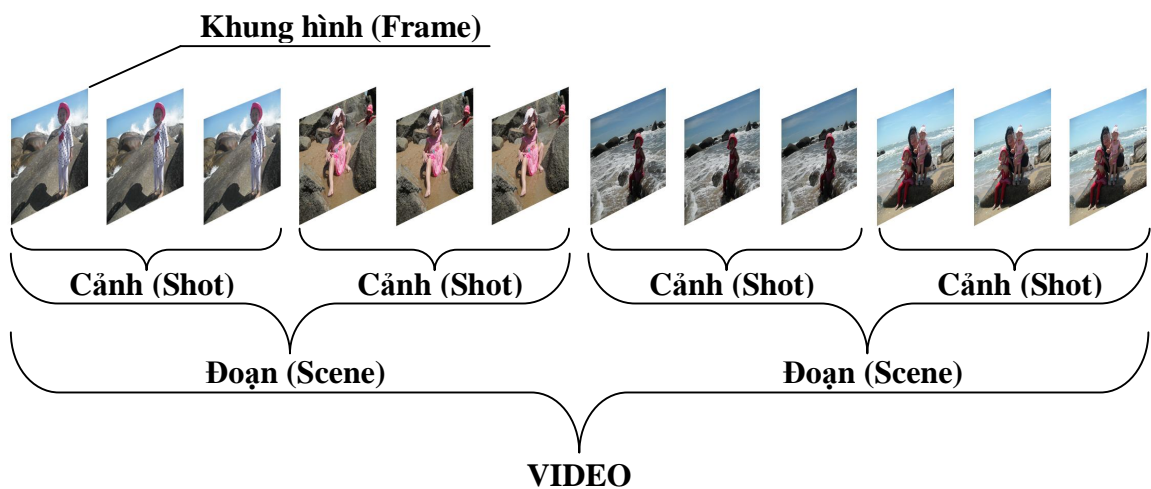
Chương cuối là phần tóm tắt lại các kết quả đã đạt được trong quá trình nghiên cứu, từ đó đưa ra các đề xuất để giải quyết những vấn đề chưa giải quyết được trong quá trình nghiên cứu.

## Chương 2 CƠ SỞ LÝ THUYẾT VÀ CÁC NGHIÊN CỨU LIÊN QUAN

### 2.1 CÁC KHÁI NIỆM CƠ BẢN VỀ VIDEO

#### 2.1.1 Video

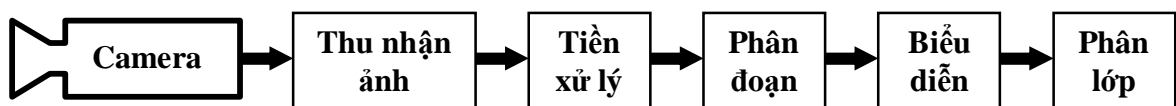
Video là một chuỗi hình ảnh cung cấp thông tin về không gian lẫn thời gian. Một video có thể được biểu diễn như là một cấu trúc đa lớp có nhiều đoạn (scene), mỗi đoạn có nhiều cảnh (shot) và mỗi cảnh có nhiều khung hình (frame) [2]. Số khung hình xuất hiện trong thời gian một giây gọi là tỉ lệ khung hình trên giây (frame rate). Tỉ lệ khung hình trên giây thông dụng là 24 fps (frame per second). Hình 2.1 minh họa cấu trúc của đoạn video.



Hình 2.1: Cấu trúc video

#### 2.1.2 Xử lý ảnh và video

Xử lý ảnh và video là quá trình tác động lên hình ảnh và video cung cấp ở đầu vào để cho ra các hình ảnh và video thoả mãn các yêu cầu đặt ra hoặc để rút trích ra các đặc trưng từ các hình ảnh và video đó.



Hình 2.2: Sơ đồ cấu trúc của hệ thống xử lý ảnh và video tổng quát

Sơ đồ cấu trúc của hệ thống xử lý ảnh và video tổng quát được giới thiệu trong hình 2.2 phù hợp với hầu hết các ứng dụng trong lĩnh vực này. Trong một số hệ thống đặc biệt có thể bỏ bớt một số khối trong sơ đồ trên [1].

**Thu nhận ảnh:** các thao tác liên quan đến camera và các thao tác liên quan đến thiết lập hệ thống để thu nhận được hình ảnh đối tượng từ thế giới thực thông qua camera.

**Tiền xử lý:** các thao tác xử lý tác động lên hình ảnh và video để chúng trở nên thích hợp hơn cho các thao tác xử lý chính. Ví dụ như các thao tác giảm nhiễu, làm sắc nét, chuyển từ ảnh màu sang ảnh đa cấp xám, cắt xén vùng ảnh cần tập trung xử lý...

**Phân đoạn:** các thông tin cần quan tâm trong ảnh và video sẽ được trích xuất ra. Ví dụ như các vùng ảnh chuyển động trong video sẽ được phân đoạn bằng cách thể hiện dưới dạng ảnh nhị phân. Vùng ảnh chuyển động sẽ có màu trắng và nền sẽ có màu đen.

**Biểu diễn:** các đối tượng được tách ra từ khối phân đoạn sẽ được biểu diễn một cách chính xác với các đặc trưng của từng đối tượng riêng biệt. Ví dụ như các đối tượng chuyển động sẽ được biểu diễn bằng đường bao quanh đối tượng và được gán nhãn phân biệt cho từng đối tượng.

**Phân lớp:** xác định các đối tượng được tách ra từ các khối trước đó có phải là đối tượng cần quan tâm hay không. Ví dụ như xác định xem các đối tượng chuyển động có phải là xe ô tô hoặc khách bộ hành hay không.

Tuỳ theo hệ thống mà một thao tác xử lý có thể thuộc về khối này hoặc khối kia. Trong một số hệ thống mà khối phân đoạn cho ra kết quả là ảnh nhị phân với màu trắng và đen thì khối biểu diễn và khối phân lớp được ghép thành khối phân tích BLOB (Binary Large Object). BLOB là đối tượng nhị phân lớn. BLOB mang ý nghĩa là chỉ quan tâm đến các nhóm điểm ảnh nhị phân lớn có kích thước nhất định. Còn những nhóm điểm ảnh nhị phân nhỏ xem như là nhiễu [1].



### 2.1.3 Bộ lọc trung vị

Các hình ảnh thô nhận được qua camera thường bị nhiễu và cần xử lý để loại bỏ nhiễu. Nhiễu phổ biến là nhiễu muối tiêu (salt and pepper noise). Nhiễu là các điểm ảnh có giá trị tách biệt với các điểm ảnh lân cận.

Bộ lọc trung vị (Median Filter) là một bộ lọc khá hiệu quả dùng để khử nhiễu mà không làm mờ hình ảnh. Giá trị trung vị của một nhóm giá trị là giá trị nằm ở vị trí giữa sau khi đã sắp xếp các giá trị theo thứ tự tăng dần. Bộ lọc trung vị quét qua từng điểm ảnh của toàn bộ hình ảnh và thay thế giá trị cường độ của điểm ảnh đang xử lý bằng giá trị trung vị của các điểm ảnh lân cận và điểm ảnh đang xử lý. Trong trường hợp sử dụng 8 điểm lân cận ta có kích thước lọc 3 x 3. Kích thước lọc là cặp các số lẻ 3 x 3, 5 x 5, 7 x 7, 9 x 9 ... Kích thước lọc càng lớn thì giá trị trung vị thu được dùng để loại nhiễu càng chính xác hơn. Tuy nhiên kích thước lọc càng lớn thì tốc độ xử lý càng chậm đi đáng kể [1]. Như hình minh họa bên dưới, điểm ảnh nhiễu có giá trị 0 được thay thế bằng giá trị trung vị có giá trị 204 của 8 điểm lân cận nên đã loại bỏ được nhiễu. Hình 2.3 minh họa hoạt động của bộ lọc trung vị.

205	204	204
206	<b>0</b>	208
201	119	205

205	204	204
206	<b>204</b>	208
201	119	205

**Hình 2.3: Hình minh họa hoạt động của bộ lọc trung vị**

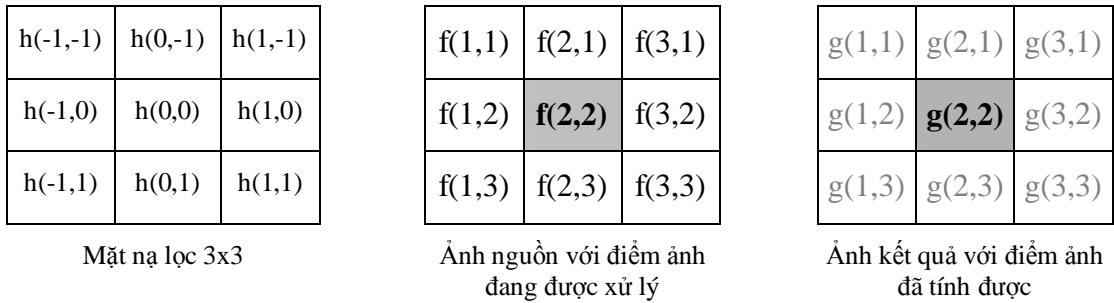
### 2.1.4 Bộ lọc tương quan

Bộ lọc tương quan (Correlation Filter) hoạt động tương tự bộ lọc trung vị. Bộ lọc tương quan quét qua từng điểm ảnh của toàn bộ hình ảnh và thay thế giá trị cường độ của điểm ảnh đang xử lý bằng giá trị mới. Giá trị mới là tổng các tích giữa các hệ số tương ứng trong mặt nạ lọc với giá trị của các điểm ảnh lân cận và điểm ảnh đang xử lý theo công thức sau [1]:

$$g(x, y) = \sum_{j=-R}^R \sum_{i=-R}^R h(i, j) f(x+i, y+j) \quad (2.1)$$

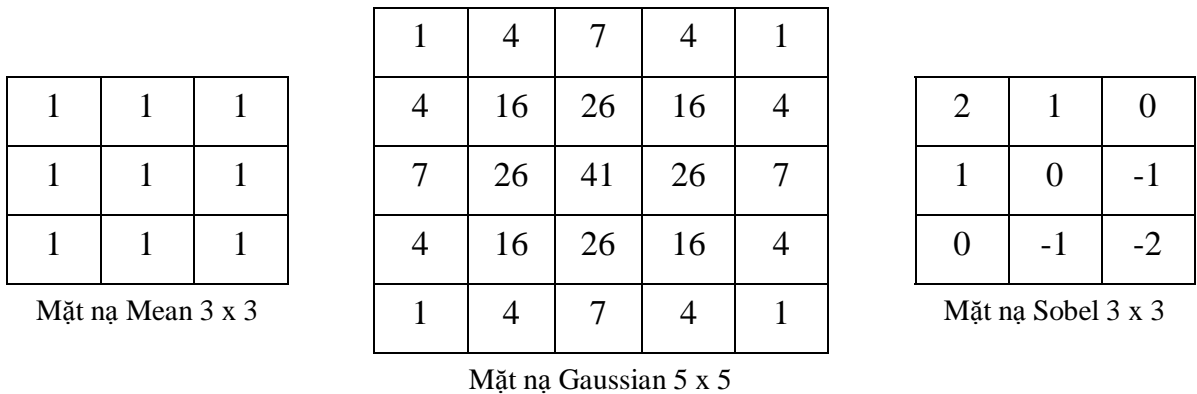
Trong công thức (2.1),  $f(x, y)$  là điểm ảnh đang được xử lý,  $f(x + i, y + j)$  là các điểm ảnh lân cận và điểm ảnh đang xử lý,  $g(x, y)$  là điểm ảnh mới,  $h(i, j)$  là các hệ số trong mặt nạ lọc,  $R$  là một nửa kích thước mặt nạ lọc. Hình 2.8 minh họa cách hoạt động của bộ lọc tương quan với kích thước mặt nạ lọc là  $3 \times 3$ , áp dụng cho điểm ảnh đang xử lý là  $f(2, 2)$ .

Tuỳ theo hệ số và kích thước của mặt nạ lọc mà bộ lọc tương quan sẽ cho ra kết quả khác nhau. Mặt nạ lọc Mean sẽ làm mờ (blur) hình ảnh. Mặt nạ lọc Gaussian sẽ làm nhoè hình ảnh. Mặt nạ lọc Sobel sẽ tách biên các đối tượng trong hình ảnh [1]. Hình 2.4 và hình 2.5 minh họa hoạt động của bộ lọc tương quan.



$$g(2, 2) = h(-1, -1)f(1, 1) + h(0, -1)f(2, 1) + h(1, -1)f(3, 1) + h(-1, 0)f(1, 2) + h(0, 0)f(2, 2) + h(1, 0)f(3, 2) + h(-1, 1)f(1, 3) + h(0, 1)f(2, 3) + h(1, 1)f(3, 3)$$

**Hình 2.4: Hình minh họa hoạt động của bộ lọc tương quan**



**Hình 2.5: Hình minh họa một số mặt nạ của bộ lọc tương quan**

### 2.1.5 Bộ lọc hình thái

Tương tự bộ lọc trung vị và bộ lọc tương quan, bộ lọc hình thái (Morphology Filter) quét qua từng điểm ảnh, áp mặt nạ lọc sao cho tâm trùng điểm

ảnh đang xử lý, dựa vào các điểm ảnh tương ứng với mặt nạ lọc để tính toán giá trị mới. Tuy nhiên bộ lọc hình thái chủ yếu làm việc trên ảnh nhị phân với giá trị của điểm ảnh là 0 hoặc 1. Mặt nạ lọc là một ma trận có kích thước 3 x 3, 5 x 5, 7 x 7... Mặt nạ lọc được điền đầy các giá trị 1 nếu cần bảo toàn độ sắc nét của các góc. Hoặc các giá trị 1 được điền thành dạng hình đĩa nếu cần làm tròn các góc.

0	0	0	0	0	0	0
0	1	0	0	0	1	0
0	1	0	0	0	0	0
0	1	1	1	0	0	0
0	1	0	1	0	0	0
0	1	1	1	1	1	0
0	0	0	0	0	0	0

0	0	0	0	0	0	0
0	1	1	0	1	1	0
0	1	1	1	1	1	0
0	1	1	1	1	0	0
0	1	1	1	1	1	0
0	1	1	1	1	1	0
0	0	0	0	0	0	0

**Hình 2.6: Hình minh họa hiệu ứng Dilation của thao tác Hit**

0	0	0	0	0	0	0
0	1	1	0	1	0	0
0	0	0	0	0	1	0
0	1	1	1	0	0	0
0	1	1	1	0	1	0
0	1	1	1	0	0	0
0	0	0	0	0	0	0

0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	1	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0

**Hình 2.7: Hình minh họa hiệu ứng Erosion của thao tác Fit**

Bộ lọc hình thái có hai thao tác cơ bản là Hit và Fit. Thao tác Hit sẽ trả về giá trị 1 cho điểm ảnh đang xử lý nếu có ít nhất một điểm ảnh tương ứng mang giá trị 1 trùng với giá trị 1 trên mặt nạ lọc. Thao tác Fit sẽ trả về giá trị 1 cho điểm ảnh đang xử lý nếu tất cả các điểm ảnh tương ứng đều mang giá trị 1 trùng với giá trị 1 trên mặt nạ lọc. Thao tác Hit khi áp dụng lần lượt cho tất cả các điểm ảnh trên ảnh cần xử lý sẽ tạo ra hiệu ứng giãn nở (Dilation). Thao tác Fit khi áp dụng lần lượt

cho tất cả các điểm ảnh trên ảnh cần xử lý sẽ tạo ra hiệu ứng xói mòn (Erosion) [1]. Hình 2.6 và 2.7 minh họa hiệu ứng Dilation của thao tác Hit và Fit.

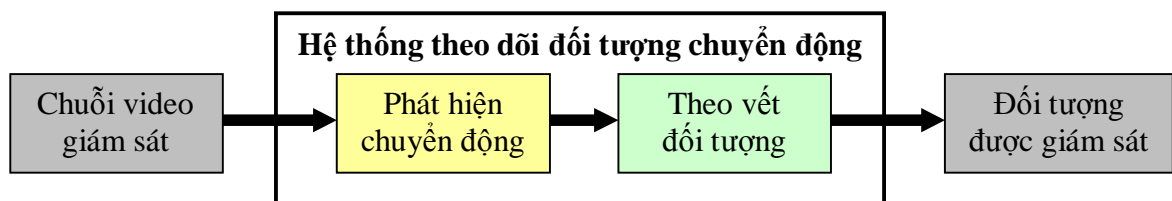
Bộ lọc Closing là một kết hợp giữa hiệu ứng Dilation và hiệu ứng Erosion. Đầu tiên hiệu ứng giãn nở Dilation sẽ lấp đầy các lỗ hổng nhỏ trong vùng ảnh vì ảnh bị phình to lên. Sau đó hiệu ứng Erosion sẽ co ảnh trở về kích thước ban đầu. Kết quả là bộ lọc Closing sẽ lấp đầy các lỗ hổng nhỏ bên trong đối tượng.

Bộ lọc Opening là một kết hợp giữa hiệu ứng Erosion và hiệu ứng Dilation. Đầu tiên hiệu ứng xói mòn Erosion sẽ làm biến mất các nhiễu kích thước nhỏ vì ảnh bị co lại. Sau đó hiệu ứng Dilation sẽ phình to ảnh về kích thước ban đầu. Kết quả là bộ lọc Opening sẽ loại bỏ được các nhiễu có kích thước nhỏ [1].

## 2.2 BÀI TOÁN THEO DÕI ĐỐI TƯỢNG CHUYỂN ĐỘNG

### 2.2.1 Giới thiệu bài toán theo dõi đối tượng chuyển động

Theo dõi đối tượng chuyển động là một tập các thao tác xử lý tác động lên các khung hình của chuỗi video, giúp xác định được các đối tượng chuyển động xuyên suốt qua các khung hình. Theo dõi đối tượng chuyển động trong chuỗi video nhằm tách ra chính xác vùng ảnh có chứa đối tượng giúp rút trích các thông tin về đối tượng dễ dàng và nhanh chóng. Hai bước chính để theo dõi đối tượng chuyển động là phát hiện chuyển động (Motion Detection) và theo vết đối tượng (Object Tracking). Hình 2.8 giới thiệu mô hình hệ thống theo dõi đối tượng chuyển động.



**Hình 2.8: Mô hình hệ thống theo dõi đối tượng chuyển động**

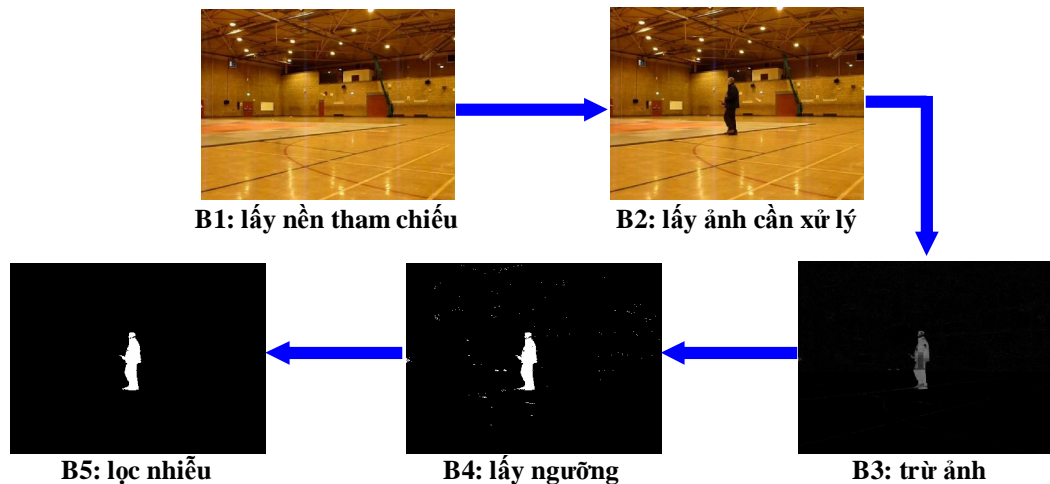
Phần quan trọng nhất trong theo dõi đối tượng chuyển động là phát hiện chuyển động. Nếu kết quả phát hiện chuyển động không tốt sẽ ảnh hưởng đến tốc độ và độ chính xác của cả hệ thống. Kết quả phát hiện chuyển động cũng phụ thuộc phương pháp phát hiện chuyển động được sử dụng. Các yếu tố thường ảnh hưởng đến kết quả phát hiện chuyển động là ánh sáng, nhiễu, độ ổn định của khung nhìn...

Phương pháp trừ nền (Background Subtraction), sai khác ảnh (Image Difference), mô hình Gaussian hỗn hợp (Gaussian Mixture Model) và dòng quang học Lucas-Kanade (Lucas-Kanade Optical Flow) là các phương pháp phát hiện chuyển động phổ biến.

### 2.2.2 Phương pháp trừ nền

Trừ nền (Background Subtraction) là phương pháp sử dụng kỹ thuật trừ ảnh (Image Subtraction). Phương pháp trừ nền đơn giản nhưng hiệu quả để tách đối tượng ra khỏi nền. Với trường hợp nền đồng nhất như sử dụng phòng xanh trong phim trường thì phương pháp trừ nền cực kỳ hiệu quả. Phương pháp trừ nền phù hợp với trường hợp sử dụng camera cố định đặt trong nhà có phòng nền cố định, kết cấu đơn giản, nguồn chiếu sáng ít thay đổi. Phương pháp trừ nền nhận diện vùng chuyển động thông qua việc giám sát sự thay đổi của ảnh nền. Để phát hiện các vùng ảnh thay đổi trên ảnh nền, phương pháp trừ nền lấy ảnh đang quan sát trừ cho ảnh dùng làm ảnh nền tham chiếu. Gọi  $f(x, y)$  là ảnh đang quan sát,  $r(x, y)$  là ảnh nền tham chiếu và  $g(x, y)$  là ảnh thể hiện kết quả xử lý, phương pháp trừ nền thể hiện qua công thức [1]:

$$g(x, y) = f(x, y) - r(x, y) \quad (2.2)$$



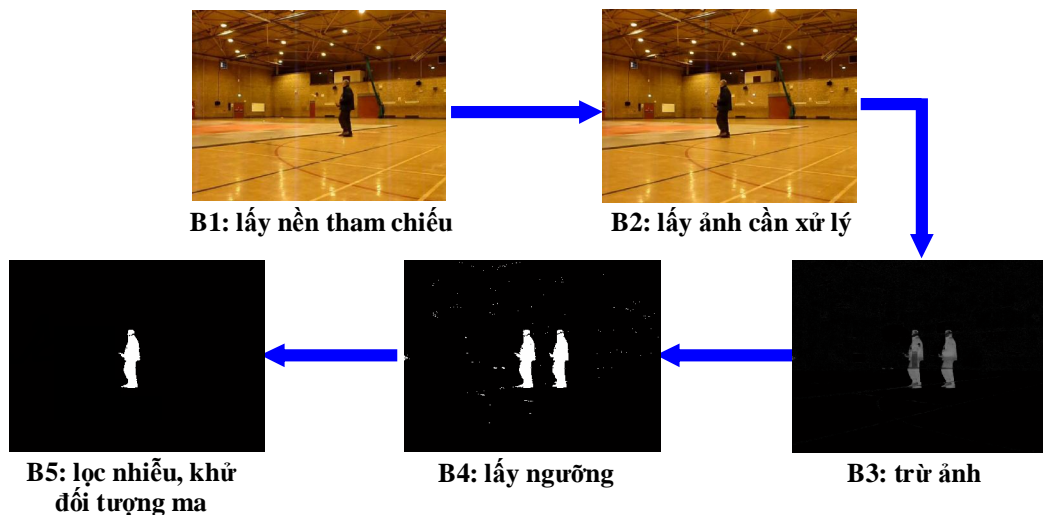
**Hình 2.9: Hình minh họa 5 bước của phương pháp trừ nền**

Để tách được các đối tượng chuyển động ra khỏi ảnh nền. Phương pháp trừ nền được thực hiện qua 5 bước như minh họa trong hình 2.9:

- **Bước 1:** lấy khung hình làm ảnh nền tham chiếu.
- **Bước 2:** lấy khung hình hiện tại làm ảnh cần xử lý.
- **Bước 3:** thực hiện trừ nền theo công thức (2.2) ở trên.
- **Bước 4:** lấy ngưỡng để làm nổi bật vùng ảnh chuyển động.
- **Bước 5:** lọc nhiễu để vùng ảnh chuyển động thu được chính xác hơn.

### 2.2.3 Phương pháp sai khác ảnh

Tương tự phương pháp trừ nền, phương pháp sai khác ảnh (Image Differencing) cũng sử dụng kỹ thuật trừ ảnh (Image Subtraction). Tuy nhiên phương pháp sai khác ảnh có điểm khác biệt là sử dụng khung hình liền trước khung hình đang xử lý làm ảnh tham chiếu. Phương pháp sai khác ảnh hiệu quả hơn so với phương pháp trừ nền trong trường hợp ảnh nền thường xuyên bị thay đổi do thay đổi sự chiếu sáng, thay đổi cấu trúc của ảnh nền... Tuy nhiên phương pháp sai khác ảnh không hiệu quả trong trường hợp đối tượng chuyển động rất chậm vì khi đó hai khung hình liên tiếp không có sự khác biệt đáng kể. Ngoài ra phương pháp sai khác ảnh còn tồn tại vấn đề “đối tượng ma” khiến không xác định được đối tượng và vấn đề chồng lấn khiến chỉ phát hiện được một phần đối tượng [1].



**Hình 2.10: Hình minh họa 5 bước của phương pháp sai khác ảnh**

Phương pháp sai khác ảnh được minh họa trong hình 2.10 cũng trải qua 5 bước tương tự phương pháp trừ nền. Tuy nhiên ở bước 1 thì ảnh tham khảo sẽ lấy khung hình trước khung hình đang xử lý và bước 5 thì sau khi loại bỏ nhiễu sẽ loại

bỏ đối tượng ma chỉ chừa lại đối tượng thật. Để loại bỏ đối tượng ma phải xác định được hướng di chuyển của đối tượng từ đó kết luận đâu là đối tượng thật đâu là bóng ma.

## 2.2.4 Phương pháp Gaussian Mixture Model

Phương pháp Gaussian Mixture Model là phương pháp mô hình hoá nền để phát hiện đối tượng chuyển động [4]. Yếu tố chính của phương pháp này là xử lý giá trị của một điểm ảnh với một mô hình Gaussian hỗn hợp. Nếu một điểm ảnh không khớp với phân phối nền thì nó thuộc đối tượng chuyển động. Stauffer và Grimson đã đề xuất cách tiếp cận xác suất sử dụng mô hình Gaussian hỗn hợp để xác định ảnh nền và đối tượng chuyển động [3]. Xác suất  $P$  của một điểm ảnh khả kiến được cho bởi công thức (2.3):

$$P(I_t) = \sum_{i=1}^K w_{i,t} h(I_t, m_{i,t}, s_{i,t}) \quad (2.3)$$

Trong đó  $I_t$  là điểm ảnh tại thời gian  $t$ .  $K$  là tổng số thành phần của hỗn hợp Gaussian được sử dụng.  $w_{i,t}$  là trọng số chỉ ra độ ảnh hưởng của hàm Gaussian  $h$  thứ  $i$  tại thời điểm  $t$ .  $m_{i,t}$  là cường độ trung bình của Gaussian thứ  $i$  tại thời điểm  $t$ .  $s_{i,t}$  là độ lệch chuẩn của Gaussian thứ  $i$  tại thời điểm  $t$ . Thuật toán sẽ duyệt qua từng điểm ảnh để cập nhật các tham số của toàn bộ mô hình Gaussian hỗn hợp cho khung hình đang xử lý. Khi một điểm ảnh được xử lý, nó sẽ được kiểm tra xem có khớp với bất kỳ thành phần nào trong  $K$  thành phần của hỗn hợp Gaussian không. Một điểm ảnh gọi là khớp với thành phần Gaussian thứ  $i$  tại thời điểm  $t$  khi  $|I_t - m_{i,t}| \leq 2.5s_{i,t}$ . Nếu khớp thì các tham số được cập nhật theo các công thức (2.4), (2.5), (2.6):

$$w_{i,t} = (1-a)w_{i,t-1} + a \quad (2.4)$$

$$m_{i,t} = (1-r)m_{i,t-1} + rI_t \quad (2.5)$$

$$s_{i,t}^2 = (1-r)s_{i,t-1}^2 + r(I_t - m_{i,t})^2 \quad (2.6)$$

Trong đó  $a$  là tỷ lệ học và  $r$  được tính xấp xỉ gần bằng  $a/w_{i,t}$ . Và nếu điểm ảnh không khớp với thành phần Gaussian thứ  $i$  tại thời điểm  $t$  thì các tham số được cập nhật theo các công thức (2.7), (2.8), (2.9):

$$w_{i,t} = (1-a)w_{i,t-1} \quad (2.7)$$

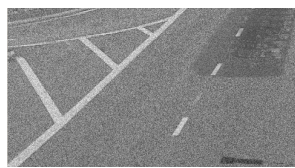
$$m_{i,t} = m_{i,t-1} \quad (2.8)$$

$$s^2_{i,t} = s^2_{i,t-1} \quad (2.9)$$

Sau quá trình cập nhật trên, tất cả các trọng số được chuẩn hoá để có tổng các trọng số là 1. Tiếp theo,  $K$  thành phần của hỗn hợp Gaussian sẽ được xếp hạng theo tỷ số  $w_{i,t}/s_{i,t}$ . Thành phần nào có thứ hạng cao hơn sẽ có độ biến thiên thấp và xác suất xuất hiện cao. Điều này thể hiện tính chất của nền. Vì vậy các thành phần Gaussian có thứ hạng cao mà trọng số của nó lớn hơn một ngưỡng  $T$  thì được xem như mô hình nền. Các thành phần Gaussian còn lại là mô hình cho đối tượng. Hình 2.11 minh hoạ kết quả của Gaussian Mixture Model.



Ảnh nguồn



Ảnh nền



Đối tượng chuyển động

**Hình 2.11: Hình minh hoạ kết quả của Gaussian Mixture Model**

### 2.2.5 Phương pháp Lucas-Kanade Optical Flow

Dòng quang học (Optical Flow) là mô hình chuyển động biểu kiến của các đối tượng, các bề mặt, các cạnh... trong một quang cảnh bởi chuyển động tương đối giữa vật quan sát và quang cảnh đó. Khái niệm dòng quang học được đề xuất bởi tiến sĩ James (Jerome) Gibson nhà tâm lý học người Mỹ vào những năm 1940 để mô tả kích thích thị giác ở các động vật đang di chuyển.

Trong lĩnh vực thị giác máy tính, phương pháp dòng quang học thử tính chuyển động của các điểm ảnh giữa hai khung hình ở thời điểm  $t$  và  $t + Dt$ . Giả sử rằng cường độ sáng không thay đổi, ràng buộc về cường độ sáng của điểm ảnh trong không gian 2D sau khoảng thời gian  $Dt$  được cho bởi công thức (2.10):



$$I(x, y, t) = I(x + \Delta x, y + \Delta y, t + \Delta t) \quad (2.10)$$

Với  $I(x, y, t)$  là cường độ sáng của điểm ảnh tại vị trí  $x, y$  ở thời điểm  $t$ . Giả sử sự khác biệt vị trí của điểm ảnh sau thời gian  $\Delta t$  rất nhỏ, áp dụng triển khai Taylor để có công thức (2.11) tính xấp xỉ như sau:

$$I(x + \Delta x, y + \Delta y, t + \Delta t) \approx I(x, y, t) + \frac{dI}{dx} \Delta x + \frac{dI}{dy} \Delta y + \frac{dI}{dt} \Delta t \quad (2.11)$$

Từ công thức (2.10) và (2.11) ta có công thức (2.12):

$$\frac{dI}{dx} \Delta x + \frac{dI}{dy} \Delta y + \frac{dI}{dt} \Delta t = 0 \quad (2.12)$$

Hay:

$$\frac{dI}{dx} \frac{\Delta x}{\Delta t} + \frac{dI}{dy} \frac{\Delta y}{\Delta t} + \frac{dI}{dt} \frac{\Delta t}{\Delta t} = 0 \quad (2.13)$$

Suy ra:

$$\frac{dI}{dx} V_x + \frac{dI}{dy} V_y + \frac{dI}{dt} = 0 \quad (2.14)$$

Trong đó  $V_x$  và  $V_y$  là hai thành phần vận tốc hay còn gọi là dòng quang học của  $I(x, y, t)$ .  $\frac{dI}{dx}, \frac{dI}{dy}, \frac{dI}{dt}$  lần lượt là các đạo hàm của ảnh tại  $x, y, t$ . Sử dụng  $I_x, I_y, I_t$  để

ký hiệu cho các đạo hàm trên, từ công thức (2.14) ta có công thức (2.15):

$$I_x V_x + I_y V_y = -I_t \quad (2.15)$$

Công thức (2.15) là phương trình ràng buộc dòng quang học. Đây là phương trình bậc nhất hai ẩn  $V_x$  và  $V_y$ . Giả sử các điểm ảnh lân cận cũng chuyển động cùng vận tốc với điểm ảnh đang xét, giá trị các điểm ảnh lân cận này có thể được dùng để giải phương trình trên tìm ra  $V_x$  và  $V_y$ . Gọi  $m \geq 2$  là số điểm ảnh cần sử dụng để giải công thức dòng quang học, khi đó ta có hệ  $m$  phương trình như sau:

$$\begin{cases} I_{x_1} V_x + I_{y_1} V_y = -I_{t_1} \\ \mathbf{M} \\ I_{x_m} V_x + I_{y_m} V_y = -I_{t_m} \end{cases} \quad (2.16)$$

Hệ  $m$  phương trình (2.16) có thể viết lại dưới dạng ma trận như sau:

$$\begin{bmatrix} I_{x_1} & I_{y_1} \\ \mathbf{M} & \mathbf{M} \\ I_{x_m} & I_{y_m} \end{bmatrix} \begin{bmatrix} V_x \\ V_y \end{bmatrix} = \begin{bmatrix} -I_{t_1} \\ \mathbf{M} \\ -I_{t_m} \end{bmatrix} \quad (2.17)$$

Đặt:  $A = \begin{bmatrix} I_{x_1} & I_{y_1} \\ \mathbf{M} & \mathbf{M} \\ I_{x_m} & I_{y_m} \end{bmatrix}$ ,  $v = \begin{bmatrix} V_x \\ V_y \end{bmatrix}$ ,  $b = \begin{bmatrix} -I_{t_1} \\ \mathbf{M} \\ -I_{t_m} \end{bmatrix}$  ta có công thức  $Av = b$ . Nhân hai

vế với ma trận chuyển vị  $A^T$  ta có công thức  $A^T Av = A^T b$ . Từ đó ta có công thức:

$$v = (A^T A)^{-1} A^T b \quad (2.18)$$

Để giải công thức (2.18), Lucas-Kanade [6, 8] sử dụng phương pháp Least Quares với kết quả như sau:

$$\begin{bmatrix} V_x \\ V_y \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^m I_{x_i}^2 & \sum_{i=1}^m I_{x_i} I_{y_i} \\ \sum_{i=1}^m I_{y_i} I_{x_i} & \sum_{i=1}^m I_{y_i}^2 \end{bmatrix}^{-1} \begin{bmatrix} -\sum_{i=1}^m I_{x_i} I_{t_i} \\ -\sum_{i=1}^m I_{y_i} I_{t_i} \end{bmatrix} \quad (2.19)$$

Đầu tiên sử dụng toán tử Sobel để tính trước các đạo hàm không gian  $I_x$ ,  $I_y$  và trừ hai ảnh liên tiếp để tính đạo hàm thời gian  $I_t$  [7, 8]. Với  $I$  là ảnh đang xét và  $I'$  là ảnh trước đó ta có các công thức tính các đạo hàm của ảnh như sau:

$$I_x = I * \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} \quad (2.20)$$

$$I_y = I * \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \quad (2.21)$$

$$I_t = I - I' \quad (2.22)$$

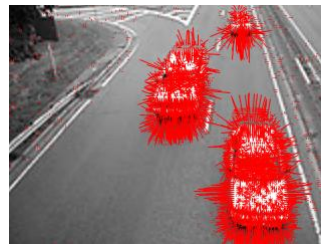
Sau đó áp dụng công thức tính ma trận nghịch đảo để tính công thức (2.19) như sau:

$$\begin{bmatrix} V_x \\ V_y \end{bmatrix} = \frac{\begin{bmatrix} \sum_{i=1}^m I_{y_i}^2 & -\sum_{i=1}^m I_{x_i} I_{y_i} \\ -\sum_{i=1}^m I_{y_i} I_{x_i} & \sum_{i=1}^m I_{x_i}^2 \end{bmatrix} \begin{bmatrix} -\sum_{i=1}^m I_{x_i} I_{t_i} \\ -\sum_{i=1}^m I_{y_i} I_{t_i} \end{bmatrix}}{\sum_{i=1}^m I_{x_i}^2 \sum_{i=1}^m I_{y_i}^2 - \sum_{i=1}^m I_{x_i} I_{y_i} \sum_{i=1}^m I_{y_i} I_{x_i}} \quad (2.23)$$

Để phương pháp dòng quang học đạt kết quả tốt, cần áp dụng thêm bộ lọc Gaussian trong bước tính đạo hàm và thêm vào công thức (2.23) một hằng số Tikhonov [9]. Lưu ý là phương pháp dòng quang học chỉ cho ra kết quả tốt khi đối tượng chuyển động với khoảng cách rất nhỏ giữa các khung hình đủ để có thể lấy được đạo hàm. Hình 2.12 minh họa kết quả của Lucas-Kanade Optical Flow.



Khung hình hiện tại



Hình ảnh dòng quang học

**Hình 2.12: Hình minh họa kết quả của Lucas-Kanade Optical Flow**

### 2.2.6 Thuật toán Grass-Fire

Kết quả trả về của các phương pháp phát hiện chuyển động thường là một ảnh nhị phân gọi là mặt nạ đối tượng chuyển động. Trong ảnh nhị phân này, các điểm ảnh thuộc về đối tượng chuyển động sẽ mang giá trị 1 và các điểm ảnh thuộc về nền sẽ mang giá trị 0. Để có thể xử lý được từng đối tượng đã phát hiện được cần phải trích xuất đối tượng nhị phân (BLOB Extraction) từ mặt nạ đối tượng chuyển động đã nhận được.

Có nhiều thuật toán để tìm và tách đối tượng nhị phân. Một trong số đó là thuật toán Grass-Fire [1]. Thuật toán bắt đầu từ điểm ảnh đầu tiên ở góc trên trái ảnh nhị phân. Sau đó thuật toán sẽ lần lượt duyệt qua từ điểm ảnh từ trái sang phải và từ trên xuống dưới. Khi phát hiện điểm ảnh thuộc về đối tượng (điểm ảnh có giá trị 1), thuật toán sẽ tìm kiếm các điểm ảnh lân cận theo 4 hướng hoặc 8 hướng lan rộng ra đến khi gặp điểm ảnh thuộc nền thì quay lui lại. Thuật toán thực hiện giống

như đốt cỏ khô. Ở đây cỏ khô là các điểm ảnh thuộc cùng một đối tượng. Các điểm ảnh thuộc nền sẽ được xem như là nước.

Trong khi tìm kiếm các điểm ảnh lân cận thuộc về đối tượng (đốt cỏ khô), thuật toán sẽ thực hiện hai thao tác. Thứ nhất, thuật toán sẽ đánh số các điểm ảnh thuộc một đối tượng với cùng một giá trị gọi là gán nhãn đối tượng. Việc gán nhãn đối tượng được thực hiện trên ảnh đầu ra của thuật toán. Thứ hai, tại vị trí tương ứng đã được gán nhãn trên ảnh đầu vào, điểm ảnh thuộc đối tượng sẽ được gỡ bỏ thay bằng điểm ảnh thuộc nền. Thao tác này giúp tránh việc lặp lại thao tác tìm kiếm trên các điểm ảnh đã xử lý xong.

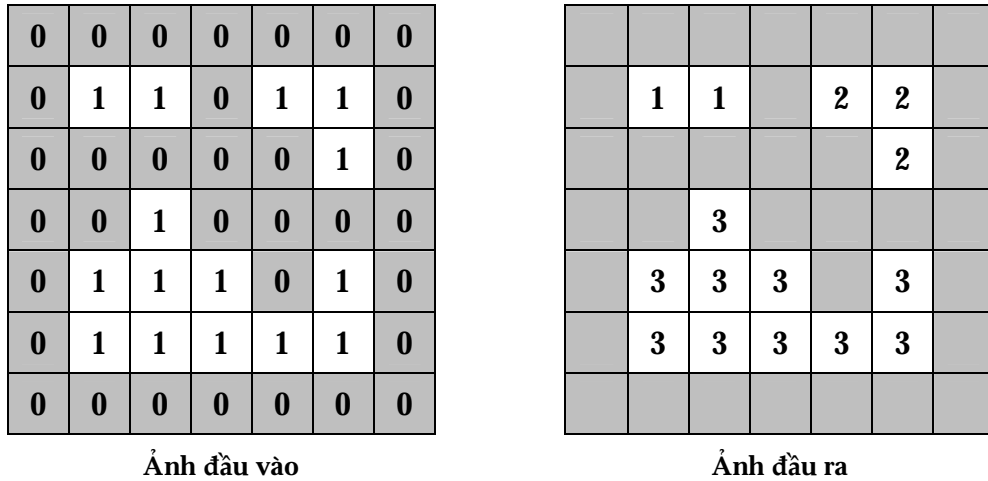
Sau khi gán nhãn xong một đối tượng, thuật toán lại tiếp tục duyệt qua các điểm ảnh theo trình tự từ trái sang phải, từ trên xuống dưới để tìm điểm ảnh thuộc đối tượng tiếp theo. Vì các điểm ảnh thuộc đối tượng ban đầu đã bị gỡ bỏ khỏi ảnh đầu vào nên nếu tìm thấy điểm ảnh thuộc đối tượng thì nó sẽ thuộc về đối tượng tiếp theo. Khi đó việc gán nhãn sẽ được thực hiện với giá trị mới.

Trong ảnh đầu ra của thuật toán Grass-Fire các đối tượng đã được gán nhãn nên dễ dàng tách được đối tượng cần xử lý. Thông thường các đối tượng đã gán nhãn sẽ được vẽ đường bao chữ nhật, tính tâm điểm của đường bao, tính tâm điểm của đối tượng cũng như thực hiện các thao tác rút trích đặc trưng khác. Hình 2.13 minh họa kết quả gán nhãn của thuật toán Grass-Fire.

Đường bao chữ nhật (Bounding Box) của đối tượng nhị phân là một hình chữ nhật nhỏ nhất có thể bao trọn đối tượng nhị phân. Đường bao nhị phân được xác định bằng tọa độ góc trên trái và kích thước dài rộng của hình chữ nhật. Đường bao chữ nhật quanh đối tượng nhị phân giúp xác định được kích thước của đối tượng cũng như vùng quan tâm (ROI) của đối tượng.

Tâm điểm của đường bao là giao điểm hai đường chéo của đường bao. Tâm điểm của đường bao chữ nhật được tính dựa vào tọa độ đỉnh trên trái và kích thước dài rộng của hình chữ nhật. Tâm điểm của đường bao có thể dùng để xác định vị trí của đường bao.

Tâm điểm của đối tượng có thể không trùng với tâm điểm của đường bao. Tâm điểm của đối tượng được tính bằng cách lấy trung bình cộng tọa độ tất cả các điểm ảnh thuộc đối tượng. Tâm điểm của đối tượng sẽ lệch về phía có nhiều điểm ảnh nhất của đối tượng và có thể xem như là trọng tâm của đối tượng.



**Hình 2.13: Hình minh họa thuật toán Grass-Fire**

### 2.2.7 Bộ lọc Kalman

Bộ lọc Kalman [11] được R. E. Kalman đề xuất năm 1960. Bộ lọc Kalman dự đoán (predict) trạng thái của tiến trình tại một thời điểm sau đó ghi nhận phản hồi (bao gồm cả nhiễu) từ các đo lường thực tế để hiệu chỉnh (correct). Chu trình của bộ lọc Kalman được mô tả trong hình 2.14. Các phương trình của bộ lọc Kalman được chia thành hai nhóm là dự đoán (cập nhật theo thời gian) và hiệu chỉnh (cập nhật theo đo lường). Cập nhật theo đo lường dùng để cung cấp một giá trị phản hồi kết hợp với ước lượng trạng thái tiên nghiệm cho ra một ước lượng trạng thái hậu nghiệm [12].



**Hình 2.14: Chu trình bộ lọc Kalman**

Nhóm công thức dùng để cập nhật theo thời gian sẽ tìm ước lượng trạng thái tiên nghiệm (2.24) và tính hiệp phương sai của ước lượng tiên nghiệm (2.25):

$$\hat{x}_k^- = A\hat{x}_{k-1} + Bu_{k-1} \quad (2.24)$$

$$P_k^- = AP_{k-1}A^T + Q \quad (2.25)$$

Nhóm công thức cập nhật theo đo lường sẽ tính độ lợi của bộ lọc Kalman (2.26), sau đó tìm ước lượng trạng thái hậu nghiệm dựa trên đo đạc quá trình thật sự (2.27), và cuối cùng là tính hiệp phương sai của ước lượng hậu nghiệm (2.28).

$$K_k = P_k^- H^T (HP_k^- H^T + R)^{-1} \quad (2.26)$$

$$\hat{x}_k = \hat{x}_k^- + K_k(z_k - H\hat{x}_k^-) \quad (2.27)$$

$$P_k = (I - K_k H)P_k^- \quad (2.28)$$

Bước thứ nhất là tìm ước lượng trạng thái tiên nghiệm. Đầu tiên phải định nghĩa vector trạng thái có kích thước  $n \times 1$ . Sau đó áp dụng công thức (2.24) để tìm vector ước lượng trạng thái tiên nghiệm  $\hat{x}_k^-$  ở bước thứ  $k$  dựa vào vector ước lượng trạng thái hậu nghiệm  $\hat{x}_{k-1}$  ở bước  $k-1$ . Ở bước  $k=1$  phải khởi tạo giá trị ban đầu cho  $\hat{x}_{k-1}$  tức là  $\hat{x}_0$ . Trong công thức (2.24), ma trận  $A$  kích thước  $n \times n$  là ma trận biến đổi trạng thái. Ma trận  $A$  thể hiện mối quan hệ giữa vector trạng thái ở bước  $k-1$  và vector trạng thái ở bước  $k$ . Ma trận  $B$  kích thước  $n \times l$  là ma trận điều khiển. Ma trận  $B$  liên kết đầu vào điều khiển tùy biến với vector trạng thái. Vector đầu vào điều khiển  $u_{k-1}$  có kích thước  $l \times 1$ .

Bước thứ hai là tính hiệp phương sai lỗi tiên nghiệm bằng cách áp dụng công thức (2.25). Trong công thức (2.25), hiệp phương sai lỗi tiên nghiệm  $P_k^-$  là một ma trận có kích thước  $n \times n$ . Ma trận  $Q$  kích thước  $n \times n$  là ma trận hiệp phương sai nhiễu tiến trình. Trong trường hợp nhiễu tiến trình là nhiễu trắng Gaussian, ma trận  $Q$  được tính bằng công thức (2.29) với  $w_k$  là nhiễu tiến trình có kích thước  $n \times 1$ . Tại bước  $k=1$  cần phải khởi tạo giá trị ban đầu cho  $P_{k-1}$  tức là  $P_0$ .

$$Q = \text{cov}(w_k) = E[w_k w_k^T] \quad (2.29)$$

Bước thứ ba là tính độ lợi bộ lọc Kalman bằng cách áp dụng công thức (2.26). Trong công thức (2.26), độ lợi bộ lọc Kalman  $K_k$  là một ma trận có kích thước  $n \times m$ . Ma trận  $H$  kích thước  $m \times n$  là ma trận chuyển đổi. Ma trận  $H$  liên kết vectơ đo lường với vectơ trạng thái. Ma trận  $R$  kích thước  $n \times m$  là ma trận hiệp phương sai nhiễu đo lường. Tương tự, ma trận  $R$  được tính bằng công thức (2.30) với  $v_k$  là nhiễu đo lường có kích thước  $m \times 1$ . [13]

$$R = \text{cov}(v_k) = E[v_k v_k^T] \quad (2.30)$$

Bước thứ tư là tìm ước lượng trạng thái hậu nghiệm bằng cách áp dụng công thức (2.27). Trong công thức (2.27),  $\hat{x}_k$  là vectơ ước lượng trạng thái hậu nghiệm được tính dựa vào vectơ đo lường  $z_k$  và ước lượng trạng thái tiên nghiệm  $\hat{x}_k^-$ . Vectơ đo lường  $z_k$  có kích thước  $m \times 1$ .

Bước cuối là tính ma trận hiệp phương sai lỗi hậu nghiệm bằng cách áp dụng công thức (2.28). Trong công thức (2.28), Ma trận hiệp phương sai lỗi hậu nghiệm  $P_k$  có kích thước  $n \times n$ . Ma trận  $I$  kích thước  $n \times n$  là ma trận đơn vị.

## 2.3 CÁC CÔNG TRÌNH NGHIÊN CỨU LIÊN QUAN

### 2.3.1 Tình hình nghiên cứu trên thế giới:

Trên thế giới có rất nhiều nghiên cứu về theo dõi đối tượng chuyển động. Nhiều bài báo đã đề ra các phương pháp khác nhau với hiệu quả khác nhau. Swantje Johnsen [15] đã đưa ra mô hình theo dõi đối tượng qua camera tĩnh với phần phát hiện chuyển động sử dụng kỹ thuật trừ nền. Ảnh nền tham khảo được sinh ra bằng cách sử dụng bộ lọc Approximated Median Filter. Sau khi đã phát hiện được vùng chuyển động thì sử dụng phương pháp Two-Pass Connected Component Labeling để xác định và đóng khung vùng ảnh chuyển động làm dữ liệu cho bước theo vết đối tượng. Tiếp theo mô hình biên đối tượng được tách ra sử dụng biểu đồ màu RGB chuẩn rồi dùng bộ lọc Kalman để dự đoán vị trí tiếp theo của đối tượng. Cuối cùng sử dụng Apriori Assignment kết hợp với các độ đo khoảng cách Euclidean, Bhattacharya để theo vết đối tượng. Nhược điểm của phương pháp này là dùng kỹ

thuật trừ nền để phát hiện chuyển động nên đòi hỏi phải có được ảnh nền tham khảo phù hợp. Nếu có sự thay đổi ánh sáng so với ảnh nền tham khảo thì toàn bộ khung ảnh đang xét sẽ bị xem là vùng chuyển động nên không thể phát hiện được đối tượng chuyển động. Nếu có đối tượng xuất hiện sau đó đứng yên trong các khung hình đang xét hoặc đối tượng đứng yên trong ảnh nền di chuyển ra khỏi ảnh nền thì tại vị trí đó sẽ luôn bị nhận diện sai là đối tượng chuyển động.

Để khắc phục nhược điểm của phương pháp trừ nền nêu trên, Ajit Ranjan [16] đã đề xuất một tiếp cận mới để giải quyết bài toán phát hiện và theo vết đối tượng. Đó là sử dụng phương pháp trừ ảnh tương tự như trừ nền nhưng sử dụng kỹ thuật sai khác ảnh. Thay vì thực hiện phép trừ ảnh giữa khung hình đang xử lý và ảnh nền tham khảo thì thực hiện phép trừ ảnh giữa khung hình đang xử lý và khung hình trước đó. Sau đó kết hợp với phương pháp dòng quang học để theo vết đối tượng. Phương pháp sai khác ảnh đã khắc phục được nhiều nhược điểm của phương pháp trừ nền nhưng vẫn có nhược điểm là không phát hiện được trường đối tượng chuyển động chậm khiến cho giữa hai khung hình liên tiếp không có sự sai khác đáng kể để có thể nhận diện được đối tượng chuyển động.

Một phương pháp mới để xây dựng mô hình nền để phát hiện đối tượng chuyển động dựa vào xác suất đã được Stauffer và Grimson [3] đề xuất. Phương pháp Gaussian Mixture Model là phương pháp mô hình hoá nền để phát hiện đối tượng chuyển động. Yếu tố chính của phương pháp này là xử lý giá trị của một điểm ảnh với một mô hình Gaussian hỗn hợp. Nếu một điểm ảnh không khớp với phân phối nền thì nó thuộc đối tượng chuyển động. Kỹ thuật này rất tốt trong trường hợp xuất hiện một đối tượng cố định trong ảnh nền hoặc một đối tượng cố định trong ảnh nền biến mất. Vì sau một thời gian nhất định ảnh nền sẽ được cập nhật lại. Tuy nhiên kỹ thuật này có nhược điểm là không phù hợp với môi trường có ánh sáng thay đổi liên tục và đột ngột.

Một phương pháp khác để theo dõi đối tượng chuyển động được Shamshad Shirageri [7] sử dụng là phương pháp dòng quang học. Phương pháp dòng quang học được đề xuất bởi Lucas-Kanade [6]. Phương pháp dòng quang học thử tính



vector chuyển động của các điểm ảnh giữa các khung hình. Dựa vào các vector chuyển động này có thể phát hiện được đối tượng chuyển động. Phương pháp này khá tốt vì không nhạy cảm với nhiễu nhưng vẫn có nhược điểm là không phù hợp với trường hợp ánh sáng thay đổi đột ngột. Khi đó toàn bộ khung hình sẽ bị nhận diện sai là vùng chuyển động nên không thể phát hiện được đối tượng chuyển động.

Nhìn chung trên thế giới đã có rất nhiều nghiên cứu về theo dõi đối tượng chuyển động. Trong đó có nhiều nghiên cứu đã đạt được hiệu quả khá cao dù rằng còn một số vấn đề chưa khắc phục được. Tuy nhiên đây là các nghiên cứu của người nước ngoài. Cần phải có nghiên cứu của người Việt về theo dõi đối tượng chuyển động để có các ứng dụng phù hợp hơn.

### **2.3.2 Tình hình nghiên cứu trong nước:**

Các nghiên cứu ở Việt Nam về theo dõi đối tượng chuyển động được tìm thấy chủ yếu trong các luận văn tốt nghiệp và một số bài báo. Trong số đó có nghiên cứu của nhóm tác giả trường đại học Bách Khoa TP.HCM [17]. Các tác giả đề xuất phương pháp theo vết đối tượng trong miền Curvelet. Đầu tiên thực hiện phép biến đổi Curvelet rời rạc trên ảnh gốc để lấy các hệ số Curvelet. Sau đó thực hiện lấy ngưỡng và tái dựng hệ số Curvelet ở tỉ lệ thô nhất để lấy bản đồ cạnh của nó. Thực hiện lặp lại với các bản đồ cạnh ở tỉ lệ mịn hơn đến khi đạt được sự hội tụ. Phương pháp này đòi hỏi tính toán khá phức tạp.

Bài báo về kỹ thuật phát hiện, bám sát đối tượng và ứng dụng của nhóm tác giả trường đại học Lạc Hồng [18] đã đề xuất giải pháp kết hợp giữa phương pháp dòng quang học với phương pháp trích chọn mẫu. Đầu tiên, thuật toán tạo ra tập đối tượng ban đầu, sau đó sử dụng phương pháp dòng quang học để lấy mẫu đối tượng. Tiếp theo xác định các đối tượng lân cận có hệ số khoảng cách giữa hai đối tượng giống nhau. Tính và chuẩn hoá các trọng số sau đó quay lại bước lấy lại mẫu và ước lượng kích thước mẫu. Thực hiện lặp lại N lần lấy mẫu và thay thế để đạt được các mẫu mới. Giải pháp này vẫn có nhược điểm là chưa phù hợp với môi trường ánh sáng thay đổi.

Ngoài ra còn có các nghiên cứu trong luận văn tốt nghiệp của Trần Thái Nhân (2006), Nguyễn Thị Lan Hương (2006), Nguyễn Quỳnh Nga (2007), Bùi Cao Phát (2010)... đa số đều sử dụng phương pháp trừ nền kết hợp với một số bộ lọc phổ biến để phát hiện đối tượng chuyển động và theo vết đối tượng. Các phương pháp này còn tồn tại nhiều nhược điểm cần phải khắc phục.

Các nghiên cứu trong nước cơ bản đã giải quyết được bài toán theo dõi đối tượng chuyển động và đã có những ứng dụng thiết thực. Tuy nhiên các phương pháp giải quyết bài toán còn nhiều nhược điểm và cần được cải tiến để nâng cao hiệu quả. Đề tài theo dõi đối tượng chuyển động bằng phương pháp lọc tích hợp sẽ là một đóng góp làm phong phú thêm cho lĩnh vực nghiên cứu này tại Việt Nam.

## Chương 3 GIẢI THUẬT ĐỀ XUẤT

### 3.1 TỶ LỆ CƯỜNG ĐỘ ẢNH

Trong hệ thống theo dõi đối tượng chuyển động, bước phát hiện chuyển động là bước quan trọng và phức tạp. Ở bước phát hiện chuyển động, đối tượng chuyển động sẽ được tách ra khỏi nền. Các đối tượng chuyển động thu được từ bước phát hiện chuyển động sẽ được sử dụng cho bước theo vết đối tượng. Nếu kết quả phát hiện đối tượng chuyển động không chính xác nó sẽ ảnh hưởng đến kết quả của cả hệ thống. Tuy nhiên đây không phải là việc dễ dàng vì trong video giám sát luôn có nhiễu và thay đổi ánh sáng do tác động bởi môi trường bên ngoài. Các tác động này sẽ làm cho việc phát hiện đối tượng chuyển động trở nên thiếu chính xác. Một số phần thuộc nền cố định sẽ bị nhận diện nhầm là đối tượng chuyển động.

Có nhiều phương pháp để phát hiện chuyển động. Như đã đề cập trong phần 2.2 và 2.3, một số phương pháp phát hiện chuyển động phổ biến mà đề tài đã nghiên cứu gồm trừ nền, sai khác ảnh, Gaussian Mixture Model, Lucas-Kanade Optical Flow. Các phương pháp này đều đạt hiệu quả cao, đặc biệt là phương pháp Gaussian Mixture Model và Lucas-Kanade Optical Flow. Tuy nhiên các phương pháp này đều có nhược điểm là phát hiện sai trong trường hợp ánh sáng môi trường thay đổi đột ngột. Để khắc phục nhược điểm cho ra kết quả sai khi ánh sáng môi trường thay đổi, đề tài đề xuất giải pháp tỷ lệ cường độ ảnh (Intensity Ratio Image) không bị tác động bởi thay đổi ánh sáng đột ngột của môi trường.

Khi phân tích các khung hình có sự thay đổi độ sáng so với khung hình trước đó do bị tác động bởi môi trường, kết quả phân tích cho thấy hầu hết cường độ sáng của các điểm ảnh cố định đều bị thay đổi gần như theo cùng một hệ số  $I$ . Giả sử trong điều kiện lý tưởng, khi ánh sáng môi trường thay đổi, tất cả các điểm ảnh trong khung hình đều thay đổi cùng với một hệ số  $I$ . Khi đó, vấn đề đặt ra là cần tìm một đại lượng đặc trưng cho độ sáng của cả khung hình và đại lượng này cũng bị thay đổi giá trị với một hệ số  $I$  so với khung hình trước đó. Để giải quyết

vấn đề này, tổng cường độ sáng của các điểm ảnh đã được nghĩ đến. Tuy nhiên tổng cường độ sáng của các điểm ảnh trong cả khung hình sẽ cho ra giá trị khá lớn, gây khó khăn trong việc biểu diễn và tính toán. Để khắc phục nhược điểm này, tổng cường độ sáng của các điểm ảnh được chia cho số điểm ảnh cố định của khung hình để cho ra giá trị cường độ trung bình. Khi đó, nếu cường độ sáng của tất cả các điểm ảnh đều thay đổi với cùng một hệ số  $I$  thì giá trị cường độ sáng trung bình cũng thay đổi cùng một hệ số  $I$ .

Vấn đề tìm đại lượng đặc trưng cho độ sáng của khung hình đã được giải quyết bằng giá trị cường độ sáng trung bình. Vấn đề tiếp theo là làm thế nào biểu diễn được sự tương quan cường độ sáng của các điểm ảnh theo giá trị cường độ sáng trung bình để khi cường độ sáng của các điểm ảnh và cường độ sáng trung bình thay đổi với cùng một hệ số  $I$  thì sự tương quan cường độ sáng giữa các điểm ảnh vẫn không thay đổi. Điều này nghĩa là không bị ảnh hưởng bởi sự thay đổi ánh sáng của môi trường. Nếu biểu diễn sự tương quan cường độ sáng của các điểm ảnh bằng cách sử dụng độ lệch cường độ sáng  $d$  giữa cường độ sáng điểm ảnh và cường độ sáng trung bình thì nó vẫn phụ thuộc vào hệ số  $I$ . Để loại bỏ sự phụ thuộc vào hệ số  $I$  này, phương pháp lấy tỷ lệ giữa cường độ sáng điểm ảnh và cường độ sáng trung bình được sử dụng. Để làm sáng tỏ lập luận trên, các công thức tính cường độ sáng trung bình và tỷ lệ cường độ sáng sẽ được phân tích trong phần bên dưới.

Trong video giám sát, giả sử tổng số điểm ảnh của tất cả các đối tượng chuyển động giữa hai khung hình liên tiếp có sự khác biệt không quá lớn. Nếu không có sự thay đổi ánh sáng giữa hai khung hình liên tiếp, thì trung bình giá trị cường độ của toàn bộ ảnh giữa hai khung hình liên tiếp cũng sẽ có khác biệt không đáng kể. Ngược lại nếu có sự thay đổi ánh sáng giữa hai khung hình liên tiếp thì giá trị cường độ của hầu hết các điểm ảnh trong khung hình cũng có sự thay đổi. Điều này dẫn đến trung bình giá trị cường độ của khung hình cũng thay đổi tương ứng với sự thay đổi của các điểm ảnh trong khung hình.

Giả sử khi có sự thay đổi ánh sáng, giá trị cường độ của tất cả các điểm ảnh đều bị thay đổi với cùng một hệ số  $I$ . Gọi  $I(x, y, t)$  và  $I(x + \Delta x, y + \Delta y, t + \Delta t)$  là cường độ của các điểm ảnh tại thời điểm  $t$  và  $t + \Delta t$ , ta có công thức:

$$I(x, y, t) = I I(x + \Delta x, y + \Delta y, t + \Delta t) \quad (3.1)$$

Gọi  $\bar{I}(t)$  và  $\bar{I}(t + \Delta t)$  là trung bình giá trị cường độ của khung hình có kích thước  $m \times n$  tại thời điểm  $t$  và  $t + \Delta t$ , ta có công thức sau:

$$\bar{I}(t) = \frac{\sum_{y=1}^n \sum_{x=1}^m I(x, y, t)}{m.n} \quad (3.2)$$

Tương tự ta có công thức:

$$\bar{I}(t + \Delta t) = \frac{\sum_{y=1}^n \sum_{x=1}^m I(x + \Delta x, y + \Delta y, t + \Delta t)}{m.n} \quad (3.3)$$

Từ công thức (3.1), (3.2), (3.3) ta có:

$$\bar{I}(t) = I \bar{I}(t + \Delta t) \quad (3.4)$$

Gọi  $R(x, y, t)$  và  $R(x + \Delta x, y + \Delta y, t + \Delta t)$  là tỷ lệ cường độ điểm ảnh tại thời điểm  $t$  và  $t + \Delta t$ . Ta có công thức:

$$R(x, y, t) = \frac{I(x, y, t)}{\bar{I}(t)} \quad (3.5)$$

Tương tự ta cũng có công thức:

$$R(x + \Delta x, y + \Delta y, t + \Delta t) = \frac{I(x + \Delta x, y + \Delta y, t + \Delta t)}{\bar{I}(t + \Delta t)} \quad (3.6)$$

Từ công thức (3.1), (3.4), (3.5), (3.6) ta có:

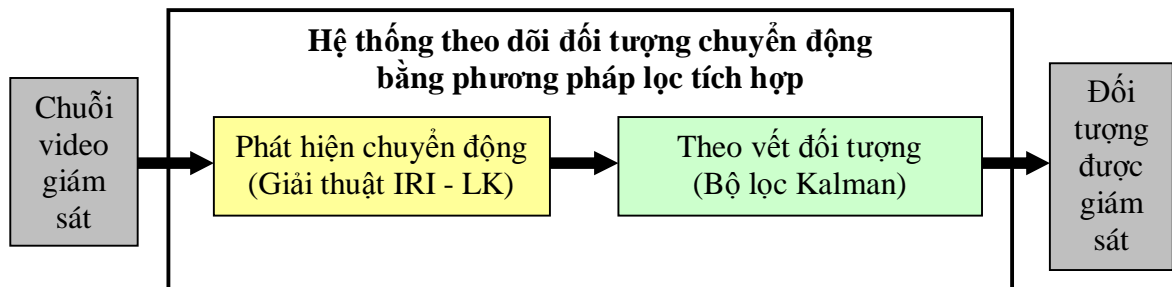
$$R(x, y, t) = R(x + \Delta x, y + \Delta y, t + \Delta t) \quad (3.7)$$

Công thức (3.7) chứng tỏ là tỷ lệ cường độ của điểm ảnh không phụ thuộc vào hệ số thay đổi ánh sáng  $I$ . Do đó tỷ lệ cường độ không bị thay đổi khi ánh sáng thay đổi. Vì vậy ảnh tỷ lệ cường độ không bị tác động bởi sự thay đổi ánh sáng.

### 3.2 PHƯƠNG PHÁP LỌC TÍCH HỢP

Theo dõi đối tượng chuyển động là một ứng dụng phổ biến trong lĩnh vực thị giác máy tính. Theo dõi đối tượng chuyển động gồm hai bước chính. Bước phát hiện chuyển động giúp tách các đối tượng chuyển động ra khỏi nền. Bước theo vết đối tượng sẽ quản lý và theo dõi các đối tượng chuyển động qua các khung hình.

Phương pháp lọc tích hợp sử dụng các bộ lọc phù hợp dựa trên đánh giá thực nghiệm và kết hợp chúng với các phương pháp theo dõi đối tượng chọn lọc theo một trình tự tối ưu để đạt được hiệu quả cao nhất. Trong phần phát hiện chuyển động, đề tài đã đề xuất cách kết hợp tỷ lệ cường độ ảnh với phương pháp Lucas-Kanade Optical Flow - giải thuật kết hợp giữa tỷ lệ cường độ ảnh và phương pháp Lucas-Kanade Optical Flow được gọi tắt là phương pháp IRI-LK (Intensity Ratio Image – Lucas Kanade)- dùng để tách các đối tượng chuyển động ra khỏi nền, phần theo vết đối tượng sẽ sử dụng các phương pháp phổ biến dựa trên bộ lọc Kalman để dự đoán và theo vết đối tượng. Trình tự thực hiện giải thuật được biểu diễn như trong hình 3.1.



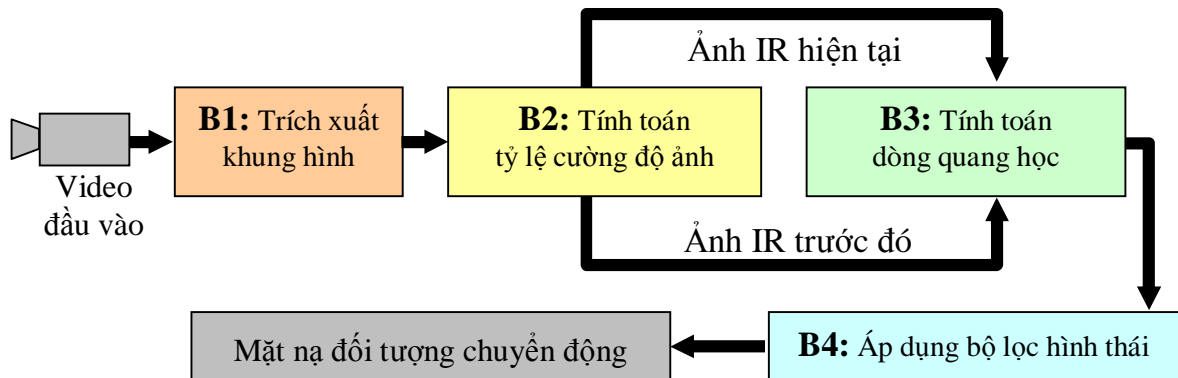
**Hình 3.1: Các khối cơ bản của phương pháp lọc tích hợp**

#### 3.2.1 Phát hiện chuyển động với phương pháp IRI-LK

Để khắc phục nhược điểm phát hiện chuyển động sai khi có sự thay đổi ánh sáng, đề tài đã đề xuất phương pháp tỷ lệ cường độ ảnh không bị tác động bởi sự thay đổi ánh sáng như đã trình bày trong phần 3.1. Tỷ lệ cường độ ảnh có thể tích hợp vào các phương pháp phát hiện chuyển động để cải thiện kết quả phát hiện chuyển động. Phương pháp IRI-LK kế thừa được ưu điểm của phương pháp Lucas-Kanade Optical Flow như khả năng chịu nhiễu tốt, phù hợp để theo dõi đối tượng

chuyển động. Đồng thời phương pháp IRI-LK khắc phục được nhược điểm nhạy cảm với ánh sáng thay đổi đột ngột của phương pháp Lucas-Kanade Optical Flow.

Phương pháp IRI-LK được mô tả trong hình 3.2. Mô hình gồm có bốn bước. Trong đó hai bước chính là tạo tỷ lệ cường độ ảnh từ các khung hình của video cần xử lý và tính toán các vectơ dòng quang học bằng phương pháp Lucas-Kanade Optical Flow.



**Hình 3.2: Mô hình phương pháp IRI-LK**

### **Bước 1 - Trích xuất khung hình:**

Các khung hình tuần tự được trích ra từ video giám sát. Từng khung hình sẽ lần lượt được chuyển sang ảnh đa cấp xám. Tiếp theo các ảnh đa cấp xám sẽ được chuẩn hoá. Cường độ của từng điểm ảnh sẽ được biểu diễn dưới dạng một con số thực có giá trị từ 0 đến 1.

### **Bước 2 - Tính toán ảnh tỷ lệ cường độ:**

Tại bước này, từng ảnh đa cấp xám lần lượt được tính toán để chuyển sang ảnh tỷ lệ cường độ. Đầu tiên, công thức (3.2) sẽ được dùng để tính cường độ trung bình của ảnh cường độ hiện tại. Sau đó, công thức (3.5) được dùng để tính ảnh tỷ lệ cường độ. Cuối cùng ảnh tỷ lệ cường độ có thể được nhân với một hệ số  $\alpha$  để giá trị tỷ lệ cường độ không vượt quá ngưỡng.

### **Bước 3 - Tính toán vectơ dòng quang học:**

Phương pháp Lucas-Kanade Optical Flow được áp dụng để tính vectơ vận tốc chuyển động của các điểm ảnh. Các đạo hàm không gian và thời gian được tính

dựa vào ảnh tỷ lệ cường độ hiện tại và ảnh tỷ lệ cường độ trước đó bằng hàm Sobel hoặc Gaussian như đã mô tả trong công thức (2.20), (2.21), (2.22). Công thức (2.23) được dùng để tính vector vận tốc  $\begin{bmatrix} V_x \\ V_y \end{bmatrix}$  của các điểm ảnh. Độ lớn vận tốc  $V(x, y)$  được tính bằng công thức sau:

$$V(x, y) = \sqrt{V_x^2 + V_y^2} \quad (3.8)$$

Để cho ra mặt nạ nhị phân của các đối tượng chuyển động  $B(x, y)$ , độ lớn vận tốc  $V(x, y)$  được lấy ngưỡng như sau:

$$B(x, y) = \begin{cases} 1 & (V(x, y) \geq Threshold) \\ 0 & (V(x, y) < Threshold) \end{cases} \quad (3.9)$$

#### Bước 4 – Áp dụng bộ lọc hình thái:

Bước cuối cùng là áp dụng các bộ lọc hình thái (như đã trình bày trong phần 2.1.7). Các bộ lọc hình thái như Closing, Opening sẽ được áp dụng để loại bỏ nhiễu, lấp lỗ trống và loại bỏ các khối chuyển động có kích thước khá nhỏ so với các đối tượng cần xử lý.

Sau các bước trên, kết quả nhận được là mặt nạ đối tượng chuyển động. Mặt nạ đối tượng chuyển động là ảnh nhị phân với các điểm ảnh thuộc về đối tượng chuyển động sẽ mang giá trị 1 và các điểm ảnh thuộc về nền sẽ mang giá trị 0. Mặt nạ này sẽ được dùng để xác định, quản lý đối tượng chuyển động và theo vết chúng trong các bước tiếp theo. Hình 3.3 minh họa mặt nạ đối tượng chuyển động.



Khung hình cần xử lý



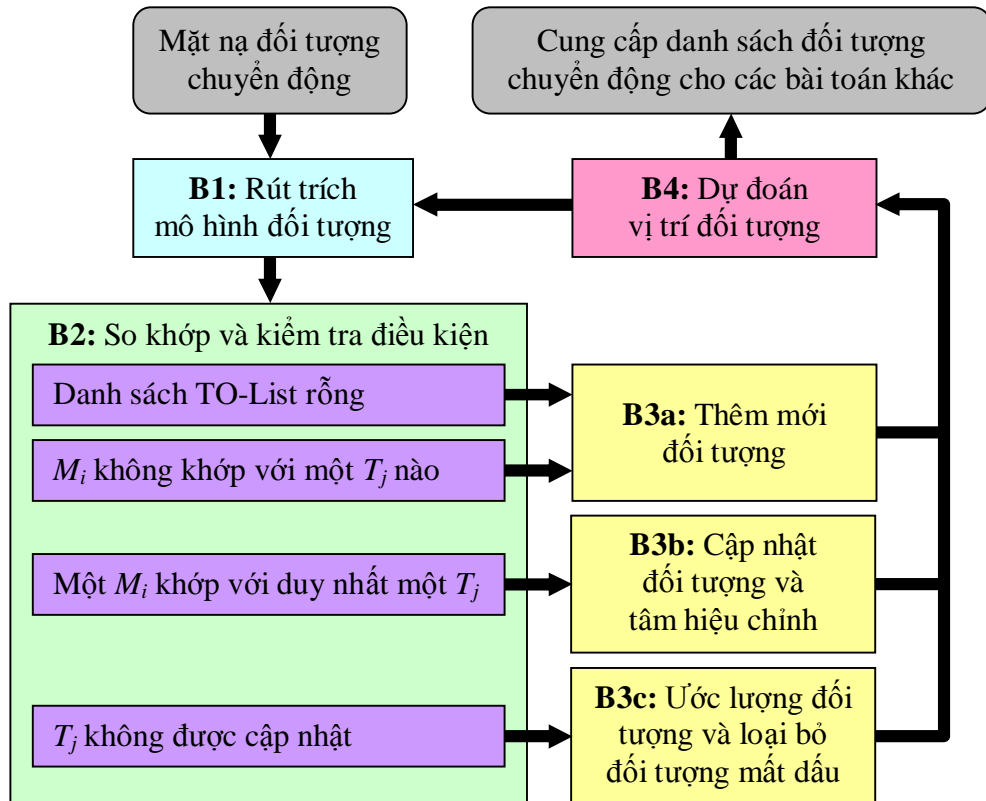
Mặt nạ đối tượng chuyển động

**Hình 3.3: Hình minh họa mặt nạ đối tượng chuyển động**



### 3.2.2 Theo vết đối tượng dựa trên bộ lọc Kalman

Sau khi nhận được mặt nạ đối tượng chuyển động từ bước phát hiện chuyển động, bước theo vết đối tượng được thực hiện để cho ra danh sách các đối tượng chuyển động đang được theo dõi. Danh sách đối tượng này sẽ phục vụ cho các bài toán khác như phân lớp, nhận diện... Giải thuật theo vết đối tượng được mô tả trong hình 3.4 với phần trọng tâm là dự đoán vị trí đối tượng bằng bộ lọc Kalman.



**Hình 3.4: Hình minh họa giải thuật theo vết đối tượng**

#### Bước 1 – Rút trích mô hình đối tượng:

Sau khi nhận được mặt nạ đối tượng chuyển động từ phương pháp IRI-LK, mô hình đối tượng sẽ được rút trích từ mặt nạ nhị phân này để phục vụ theo vết đối tượng. Mô hình đối tượng mang thông tin về các đặc trưng của đối tượng như nhãn đối tượng, tâm đối tượng, đường bao chữ nhật...

Để rút trích được mô hình đối tượng, đầu tiên thuật toán Grass-Fire (như đã trình bày ở phần 2.2.6) sẽ được áp dụng để gán nhãn cho các điểm ảnh trong mặt nạ nhị phân. Các điểm ảnh liên nhau thuộc cùng một đối tượng sẽ được gán cùng một

nhãn. Nhờ đó vùng ảnh thuộc từng đối tượng sẽ được tách ra và tính toán các đặc trưng của đối tượng. Tâm đối tượng được tính bằng cách lấy trung bình cộng tọa độ tất cả các điểm thuộc đối tượng theo từng thành phần của tọa độ. Đường bao chữ nhật được tính dựa vào tọa độ nhỏ nhất và lớn nhất của các điểm ảnh thuộc đối tượng theo từng thành phần của tọa độ.

Mô hình đối tượng rút trích từ mặt nạ đối tượng chuyển động sẽ được quản lý bởi danh sách mô hình đối tượng chuyển động ký hiệu là MO-List. Các đối tượng thuộc MO-List ký hiệu là  $M_i$  sẽ được so khớp với các đối tượng đang theo dõi để xác định xem mỗi một đối tượng đang theo dõi sẽ là đối tượng nào trong khung hình tiếp theo.

### **Bước 2 – So khớp và kiểm tra điều kiện:**

Các đối tượng chuyển động cần theo vết được quản lý bởi danh sách đối tượng đang theo dõi ký hiệu là TO-List. Các đối tượng thuộc TO-List được phân biệt thông qua chỉ số đối tượng. Các đối tượng thuộc TO-List là mô hình đối tượng chuyển động tương tự như các đối tượng trong MO-List. Ngoài các thuộc tính thể hiện đặc trưng của đối tượng như chỉ số đối tượng, nhãn đối tượng, đường bao chữ nhật, tâm đối tượng... thì các đối tượng thuộc TO-List còn có các thuộc tính hỗ trợ theo vết đối tượng như bộ lọc Kalman, tâm dự đoán, tâm hiệu chỉnh, đường bao ước lượng, trạng thái theo vết, chỉ số khung hình hiện tại... Các đối tượng thuộc TO-List ký hiệu là  $T_j$  sẽ dùng để phục vụ cho các bài toán khác như phân lớp đối tượng, nhận diện đối tượng...

Sau khi có được MO-List từ mặt nạ đối tượng chuyển động của khung hình hiện tại, nếu TO-List rỗng thì  $M_i$  sẽ được thêm ngay vào TO-List để quản lý và theo vết. Ngược lại, nếu TO-List đã có đối tượng thì thao tác so khớp đối tượng và kiểm tra điều kiện sẽ được thực hiện. Để so khớp các  $M_i$  với các  $T_j$ , khoảng cách Euclide được sử dụng để đánh giá khoảng cách giữa tâm của từng  $M_i$  với tâm dự đoán của tất cả các  $T_j$ . Nếu  $M_i$  và  $T_j$  có khoảng cách tâm nhỏ nhất và phải nhỏ hơn một ngưỡng  $d$  cho trước thì  $M_i \equiv T_j$ . Thông thường ngưỡng  $d$  được xác định là khoảng cách lớn nhất của tâm đối tượng giữa hai khung hình.

Ngoài ra để tăng thêm độ chính xác, đường bao của  $M_i$  được so sánh với đường bao ước lượng của  $T_j$ . Nếu độ chênh lệch kích thước giữa hai đường bao nhỏ hơn một ngưỡng  $s$  cho trước thì mới khẳng định  $M_i \equiv T_j$ .

Sau khi so khớp, nếu  $M_i$  không trùng với bất kỳ  $T_j$  nào thì  $M_i$  sẽ được xem là đối tượng mới và sẽ được thêm vào TO-List. Nếu tồn tại một  $M_i$  trùng với duy nhất một  $T_j$  thì  $M_i$  sẽ được xem như là đối tượng đang theo dõi và  $T_j$  sẽ được cập nhật lại thông tin tại vị trí mới. Nếu tồn tại các  $T_j$  không được cập nhật lại thông tin thì  $T_j$  sẽ được ước lượng nếu bị che khuất hoặc xoá bỏ nếu đã ra khỏi vùng quan sát.

### **Bước 3a – Thêm mới đối tượng:**

Trong bước 2, nếu xác định được là TO-List rỗng hoặc  $M_i$  không trùng với bất kỳ  $T_j$  nào thì bước 3a sẽ được thực hiện để thêm  $M_i$  vào TO-List. Khi thêm  $M_i$  vào TO-List, chỉ số đối tượng của  $T_j$  sẽ là chỉ số lớn nhất của các  $T_j$  đang quản lý tăng thêm 1. Nếu  $T_j$  là đối tượng đầu tiên thì sẽ có chỉ số là 1. Nhãn đối tượng của  $T_j$  sẽ được gán dựa vào chỉ số đối tượng của  $T_j$ . Tâm đối tượng và đường bao chữ nhật của  $T_j$  sẽ được lấy từ  $M_i$ .

Sau khi cập nhật các thông tin cơ bản, các thuộc tính hỗ trợ theo vết của  $T_j$  sẽ được tính toán. Bộ lọc Kalman (như đã trình bày trong phần 2.2.7) sẽ được khởi tạo các giá trị ban đầu. Vectơ trạng thái có dạng  $x_k = \begin{bmatrix} x \\ y \end{bmatrix}$  với  $x, y$  là tâm đối tượng của  $T_j$ . Ma trận biến đổi trạng thái được thiết lập giá trị  $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ . Ma trận điều khiển được thiết lập giá trị  $B = [ ]$ . Ma trận hiệp phương sai trạng thái được khởi tạo giá trị  $P_0 = \begin{bmatrix} EE & 0 \\ 0 & EE \end{bmatrix}$  với  $EE$  là giá trị ước lượng lỗi. Thay đổi giá trị  $EE$  sẽ ảnh hưởng đến độ chính xác của tâm dự đoán trong vài khung hình đầu tiên. Ma trận hiệp phương sai nhiễu tiến trình được khởi tạo giá trị  $Q = \begin{bmatrix} MO & 0 \\ 0 & MO \end{bmatrix}$  với  $MO$  là giá trị nhiễu tiến trình. Giá trị  $MO$  càng nhỏ thì các tâm dự đoán càng gần với quỹ đạo chuyển động lý tưởng (không bị ảnh hưởng bởi nhiễu). Ma trận chuyển đổi được

thiết lập giá trị  $H = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ . Ma trận hiệp phương sai nhiễu đo lường được khởi tạo

giá trị  $R = \begin{bmatrix} ME & 0 \\ 0 & ME \end{bmatrix}$  với  $ME$  là giá trị nhiễu đo lường. Giá trị  $ME$  càng nhỏ thì các

tâm hiệu chỉnh càng gần với tâm đối tượng hơn.

Sau khi khởi tạo giá trị ban đầu cho bộ lọc Kalman, giá trị ban đầu của tâm dự đoán và tâm hiệu chỉnh được khởi tạo với giá trị là tâm đối tượng. Đường bao ước lượng được khởi tạo với giá trị là đường bao chữ nhật. Chỉ số khung hình hiện tại cũng được cập nhật để giúp xác định lần cập nhật cuối của  $T_j$ . Trạng thái theo vết được thiết lập giá trị là 0 mang ý nghĩa là trạng thái theo vết bình thường.

### **Bước 3b – Cập nhật đối tượng và tâm hiệu chỉnh:**

Trong bước 2, nếu xác định được là có tồn tại một  $M_i$  trùng với duy nhất một  $T_j$  thì  $M_i$  sẽ được xem như là đối tượng đang theo dõi và bước 3b sẽ được thực hiện để cập nhật lại thông tin tại vị trí mới cho  $T_j$ . Khi cập nhật lại thông tin cho  $T_j$ , chỉ số đối tượng và nhãn đối tượng của  $T_j$  sẽ giữ nguyên không thay đổi. Tâm đối tượng và đường bao chữ nhật của  $T_j$  sẽ được lấy từ  $M_i$ .

Tâm hiệu chỉnh của  $T_j$  được cập nhật bằng các công thức (2.26), (2.27), (2.28). Đây là quá trình hiệu chỉnh (correct) của bộ lọc Kalman. Đường bao ước lượng của  $T_j$  được tính dựa vào kích thước đường bao trung bình của  $T_j$  ở 3 khung hình liên tiếp với các trọng số khác nhau. Chỉ số khung hình hiện tại được cập nhật để giúp xác định lần cập nhật cuối của  $T_j$ . Trạng thái theo vết được thiết lập giá trị là 0 mang ý nghĩa là trạng thái theo vết bình thường.

### **Bước 3c – Ước lượng đối tượng và loại bỏ đối tượng mất dấu:**

Trong bước 2, nếu xác định được là có tồn tại các  $T_j$  không được cập nhật lại thông tin thì bước 3c sẽ được thực hiện để ước lượng các  $T_j$  bị che khuất và loại bỏ các  $T_j$  bị mất dấu. Khi  $T_j$  không được cập nhật nghĩa là nó bị mất dấu. Nếu tâm dự đoán của  $T_j$  nằm ngoài vùng quan sát thì  $T_j$  sẽ bị loại bỏ.

Nếu tâm dự đoán của  $T_j$  vẫn nằm trong vùng quan sát thì  $T_j$  được xem như bị che khuất và sẽ không bị loại bỏ mà vẫn được cập nhật thông tin ước lượng. Khi cập nhật lại thông tin cho  $T_j$ , chỉ số đối tượng và nhãn đối tượng của  $T_j$  sẽ giữ nguyên không thay đổi. Tâm đối tượng của  $T_j$  sẽ được lấy giá trị của tâm dự đoán. Đường bao chữ nhật của  $T_j$  sẽ được lấy giá trị của đường bao ước lượng.

Tâm hiệu chỉnh của  $T_j$  cũng sẽ được lấy giá trị của tâm dự đoán. Đường bao ước lượng của  $T_j$  không thay đổi giá trị. Chỉ số khung hình hiện tại được cập nhật để giúp xác định lần cập nhật cuối của  $T_j$ . Trạng thái theo vết được thiết lập giá trị là 1 mang ý nghĩa là trạng thái theo vết ước lượng.

Trong trường hợp các đối tượng chuyển động giao cắt nhau, tại thời điểm các đối tượng bị gộp lại thì xem như là đối tượng bị che khuất. Khi đó từng đối tượng riêng biệt trong nhóm bị gộp lại vẫn được xác định bằng cách dự đoán. Khi tách ra, các đối tượng sẽ được so khớp để xác định chính xác từng đối tượng.

#### **Bước 4 – Dự đoán vị trí đối tượng:**

Sau khi thêm mới  $T_j$  ở bước 3a, cập nhật  $T_j$  ở bước 3b, ước lượng  $T_j$  ở bước 3c, thì bước 4 được thực hiện để tính tâm dự đoán của các  $T_j$  bằng các công thức (2.24) và công thức (2.25). Đây là quá trình dự đoán (predict) của bộ lọc Kalman. Tâm dự đoán của các  $T_j$  sẽ được dùng để so khớp với tâm đối tượng của các  $M_i$ .

Sau khi tính xong tâm dự đoán, các  $T_j$  trong TO-List sẽ được dùng để phục vụ cho các bài toán khác như phân lớp đối tượng, nhận diện đối tượng... Trong phạm vi của đề tài, đặc trưng của các  $T_j$  sẽ được vẽ trở lại vào khung hình hiện tại để thấy được kết quả theo dõi đối tượng chuyển động. Sau cùng, bước 1 sẽ được thực hiện lại để xử lý khung hình tiếp theo.

## **Chương 4 THỰC NGHIỆM VÀ ĐÁNH GIÁ**

### **4.1 KẾT QUẢ ĐẠT ĐƯỢC QUA THỰC NGHIỆM**

Để nghiên cứu và đề xuất được phương pháp hiệu quả, đề tài đã thực hiện rất nhiều thực nghiệm. Các thực nghiệm được thực hiện nhằm mục đích đánh giá các phương pháp làm cơ sở lựa chọn, cải thiện và minh họa cho các phương pháp mà đề tài nghiên cứu. Các thực nghiệm được chia thành 4 nhóm:

- Nhóm các thực nghiệm về phương pháp xử lý ảnh, video và các bộ lọc.
- Nhóm thực nghiệm về các phương pháp phát hiện chuyển động phổ biến.
- Nhóm thực nghiệm phương pháp phát hiện chuyển động của đề tài.
- Nhóm thực nghiệm phương pháp theo vết đối tượng của đề tài.

#### **4.1.1 Môi trường và dữ liệu thực nghiệm:**

Các ứng dụng minh họa sử dụng trong thực nghiệm được viết trên nền tảng MATLAB R2013b. Thực nghiệm được thực hiện trên máy tính trang bị CPU lõi kép tốc độ 2.8GHz, RAM 2GB, hệ điều hành Windows XP SP3 hoặc Windows 7. Ngoài ra thực nghiệm còn sử dụng Webcam có độ phân giải 320x240 điểm ảnh để thực hiện các thực nghiệm theo thời gian thực.

Dữ liệu thực nghiệm bao gồm các tracking video được lấy từ các đoạn video chuẩn trong Computer Vision Toolbox của MATLAB, các tracking video lấy từ Tracking Dataset trên Internet và các tracking video do bạn bè trong nhóm nghiên cứu trường đại học Bách Khoa TP.HCM tự quay. Các tracking video này đều được quay từ các camera cố định.

#### **Các tracking video trong Computer Vision Toolbox của MATLAB:**

- Tên video: atrium.avi; kích thước khung hình: 640x360 điểm; tỷ lệ khung hình trên giây: 30fps; đặc điểm video: có chuyển động phức tạp của nhiều đối tượng, bị nhiễu bởi bóng đối tượng, bị che khuất bởi tán cây.
- Tên video: singleball.avi; kích thước khung hình: 480x360 điểm; tỷ lệ khung hình trên giây: 30fps; đặc điểm video: đối tượng chuyển động bị che khuất.

- Tên video: viptraffic.avi; kích thước khung hình: 160x120 điểm; tỷ lệ khung hình trên giây: 15fps; đặc điểm video: nhiều đối tượng chuyển động có kích thước thay đổi, có nhiều khung hình bị thay đổi ánh sáng liên tục.
- Tên video: viptraffic.avi; kích thước khung hình: 640x360 điểm; tỷ lệ khung hình trên giây: 29fps; đặc điểm video: nhiều đối tượng chuyển động có kích thước thay đổi.

**Các tracking video trong Tracking Dataset từ Internet:**

- Tên video: caroverhead.avi; kích thước khung hình: 320x240 điểm; tỷ lệ khung hình trên giây: 25fps; đặc điểm video: đối tượng chuyển động bị che khuất, ảnh nền có kết cấu phức tạp.
- Tên video: crossing.avi; kích thước khung hình: 352x288 điểm; tỷ lệ khung hình trên giây: 8fps; đặc điểm video: các đối tượng chuyển động cắt nhau, có nhiễu chuyển động của lá cây do gió, chất lượng video thấp.
- Tên video: parkinglot.mov; kích thước khung hình: 768x576 điểm; tỷ lệ khung hình trên giây: 25fps; đặc điểm video: có nhiều loại đối tượng chuyển động giao cắt phức tạp.
- Tên video: sequence.mpg; kích thước khung hình: 512x512 điểm; tỷ lệ khung hình trên giây: 30fps; đặc điểm video: chất lượng ảnh kém, nhiều đối tượng chuyển động phức tạp.

**Các tracking video của nhóm nghiên cứu trường ĐH Bách Khoa TP.HCM:**

- Tên video: bachkhoa2.avi; kích thước khung hình: 512x512 điểm; tỷ lệ khung hình trên giây: 30fps; đặc điểm video: có đối tượng chuyển động giao cắt nhau, một phần đối tượng có màu gần với nền.
- Tên video: bachkhoa5.mp4; kích thước khung hình: 720x576 điểm; tỷ lệ khung hình trên giây: 25fps; đặc điểm video: có nhiều đối tượng chuyển động giao cắt phức tạp.
- Tên video: bachkhoa11.mp4; kích thước khung hình: 720x576 điểm; tỷ lệ khung hình trên giây: 25fps; đặc điểm video: có nhiều đối tượng chuyển động giao cắt phức tạp, một phần đối tượng có màu gần với nền.

- Tên video: bachkhoa1835.mov; kích thước khung hình: 320x240 điểm; tỷ lệ khung hình trên giây: 29fps; đặc điểm video: đối tượng chuyển động phức tạp, các đối tượng có đặc điểm tương tự nhau.

#### 4.1.2 Kết quả thực nghiệm phát hiện chuyển động:

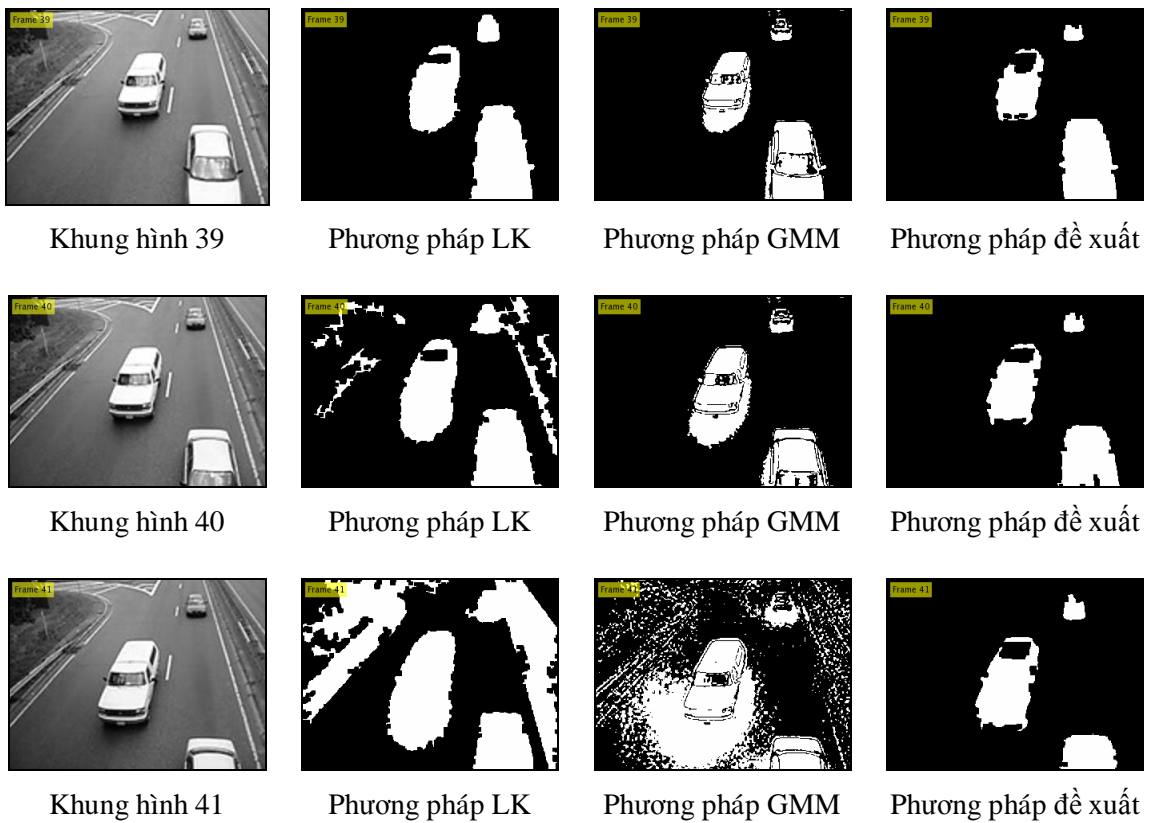
Trong ứng dụng MATLAB thực nghiệm phát hiện chuyển động, phần chính của phương pháp IRI-LK được viết hoàn toàn bằng các lệnh cơ bản của MATLAB không sử dụng các công cụ xây dựng sẵn trong MATLAB. Dù vậy tốc độ thực thi của ứng dụng thực nghiệm cũng tương đương với khi sử dụng các công cụ MATLAB xây dựng sẵn.

Trong phần thực nghiệm này, phương pháp IRI-LK do đề tài đề xuất sẽ được so sánh với phương pháp Lucas-Kanade (LK) và phương pháp Gaussian Mixture Model (GMM) là những phương pháp đã đạt được hiệu quả cao và được sử dụng nhiều trong lĩnh vực theo dõi đối tượng chuyển động. Thực nghiệm sử dụng các đoạn video có vài vị trí trong đó ánh sáng thay đổi đột ngột từ khung hình này sang khung hình khác.

Trong thực nghiệm thứ nhất, đoạn video viptraffic.avi được sử dụng. Trong thực nghiệm này hệ số  $a$  trong phương pháp IRI-LK được chọn là 6 và hằng số Tikhonov của phương pháp Lucas-Kanade được chọn là 0.005. Hình 4.1 minh họa kết quả trong một vài khung hình của video thực nghiệm. Trong các khung hình từ 39 đến 41 ánh sáng trở nên tối hơn. Với phương pháp Lucas-Kanade, kết quả nhận được là chớp sáng tại khung hình 40 và khung hình 41. Với phương pháp Gaussian Mixture Model, kết quả nhận được chỉ có một chớp sáng tại khung hình 41. Còn với phương pháp đề xuất IRI-LK, kết quả nhận được rất là tốt.

Ngoài việc đánh giá qua vài khung hình có lựa chọn, bảng 1 thống kê số khung hình có kết quả nhận diện sai trong toàn bộ đoạn video thực nghiệm tương với phương pháp Lucas-Kanade, Gaussian Mixture Model và phương pháp đề xuất. Sau đó kết quả thống kê của từng phương pháp sẽ được so sánh với nhau. Bảng 1 cho thấy kết quả phát hiện sai của các phương pháp lần lượt như sau: LK: 12.40%, GMM: 8.26%, IRI-LK: 0.00%.





**Hình 4.1: Kết quả phương pháp IRI-LK với ánh sáng thay đổi tự nhiên**

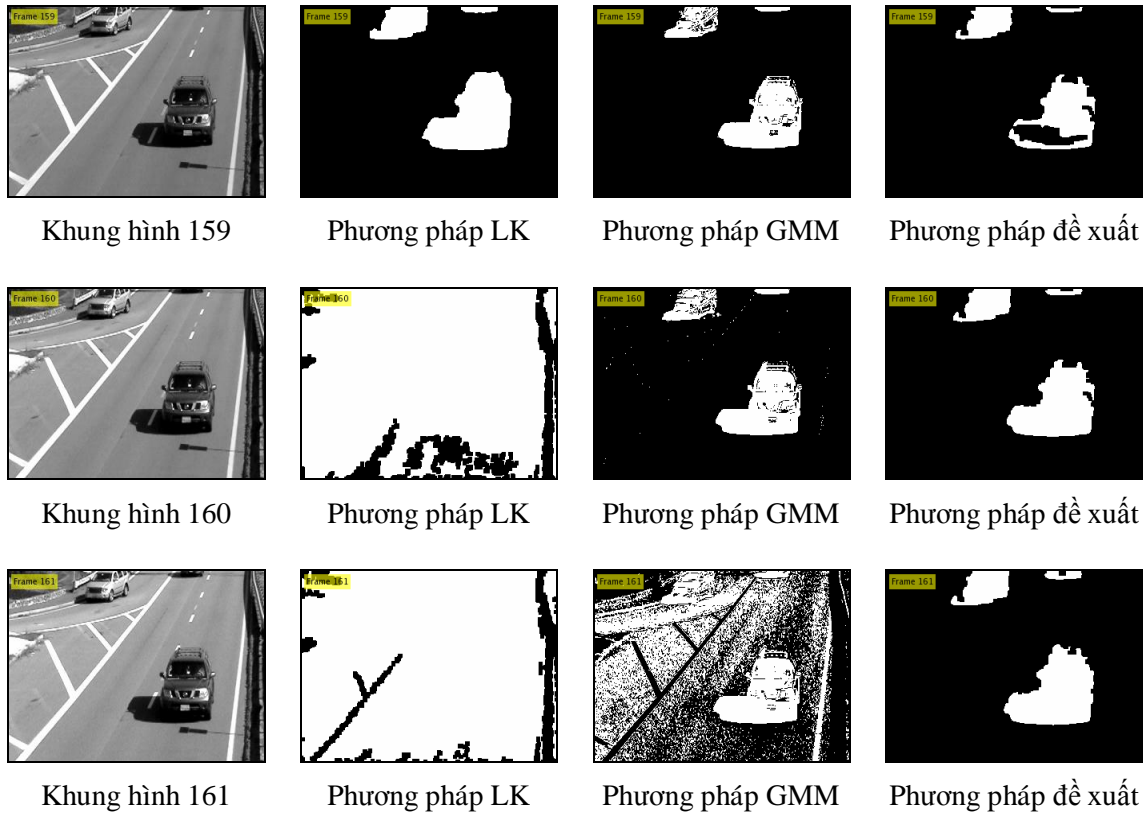
**Bảng 1: So sánh kết quả phát hiện chuyển động sai của LK, GMM và IRI-LK**

Khung hình	Cường độ trung bình	Phương pháp LK	Phương pháp GMM	Phương pháp đề xuất
1	0.5220			
...				
21	0.5030			
22	0.5056			
23	0.4628	Phát hiện sai		
24	0.4717			
25	0.4854			
...				
38	0.5165			
39	0.5035			
40	0.4643	Phát hiện sai		
41	0.4166	Phát hiện sai	Phát hiện sai	
42	0.3714	Phát hiện sai	Phát hiện sai	
43	0.3922		Phát hiện sai	
44	0.3374	Phát hiện sai	Phát hiện sai	
45	0.3814	Phát hiện sai	Phát hiện sai	

Khung hình	Cường độ trung bình	Phương pháp LK	Phương pháp GMM	Phương pháp đề xuất
46	0.4026	Phát hiện sai	Phát hiện sai	
47	0.4345	Phát hiện sai		
48	0.4829	Phát hiện sai		
49	0.4818			
50	0.4809			
...				
76	0.4764			
77	0.4975			
78	0.4472	Phát hiện sai	Phát hiện sai	
79	0.4056	Phát hiện sai	Phát hiện sai	
80	0.4237		Phát hiện sai	
81	0.4068		Phát hiện sai	
82	0.4551	Phát hiện sai		
83	0.4925	Phát hiện sai		
84	0.4890			
85	0.4892			
...				
98	0.4790			
99	0.4677			
100	0.4200	Phát hiện sai		
101	0.4282			
102	0.4330			
103	0.4267			
104	0.4704	Phát hiện sai		
105	0.4875			
106	0.4872			
...				
121	0.4891			
<b>Tổng cộng:</b>		<b>15</b>	<b>10</b>	<b>0</b>

Trong thực nghiệm thứ hai, đoạn video visiontraffic.avi được sử dụng. Trong thực nghiệm này hệ số  $a$  trong phương pháp IRI-LK được chọn là 6 và hằng số Tikhonov của phương pháp Lucas-Kanade được chọn là 0.001. Ngoài ra có vài sự thay đổi độ sáng được cố tình thêm vào khung hình để đánh giá hiệu quả thuật toán. Hình 4.2 minh họa kết quả trong một vài khung hình của video thực nghiệm. Trong đó, khung hình 160 được tăng 10% độ sáng, khung hình 161 được tăng 20% độ sáng. Với phương pháp Lucas-Kanade, kết quả nhận được là chớp sáng tại

khung hình 160 và khung hình 161. Với phương pháp Gaussian Mixture Model, kết quả nhận được chỉ có một chớp sáng tại khung hình 161. Còn với phương pháp đề xuất IRI-LK, kết quả nhận được rất là tốt.



**Hình 4.2: Kết quả phương pháp IRI-LK với thay đổi ánh sáng được thêm vào**

#### 4.1.3 Kết quả thực nghiệm theo vết đối tượng:

Trong phần thực nghiệm theo vết đối tượng, phương pháp quan sát được sử dụng để đánh giá độ chính xác của giải thuật. Các đối tượng theo dõi được vẽ đường bao chữ nhật xung quanh để xác định phạm vi của đối tượng. Đồng thời mỗi đối tượng đều được gắn nhãn riêng biệt để tiện theo dõi. Qua quan sát hầu hết các đối tượng đều được gắn đúng nhãn và số nhãn được gắn luôn bằng tổng số đối tượng theo dõi.

Trong thực nghiệm thứ nhất đoạn video viptraffic.avi và visiontraffic.avi được sử dụng để đánh giá trường hợp theo vết nhiều đối tượng chuyển động không che khuất, không giao cắt. Hình 4.3 minh họa kết quả theo vết ở một vài khung hình của thực nghiệm theo vết đối tượng.

Trong video viptraffic.avi khung hình 23 là đối tượng số 1; khung hình 38 là đối tượng số 2 và 3; khung hình 41 là đối tượng số 3 và 4; khung hình 73 là đối tượng số 5, 6 và 7; khung hình 78 là đối tượng số 7; khung hình 94 là đối tượng số 8 và 9; khung hình 100 là đối tượng số 9; khung hình 116 là đối tượng số 10. Kết quả theo vết chính xác và đủ 10 đối tượng trong video.

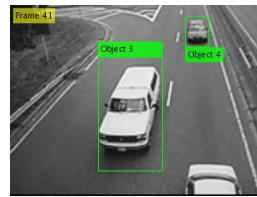
**Video viptraffic.avi**



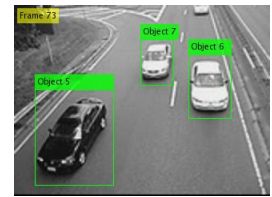
Khung hình 23



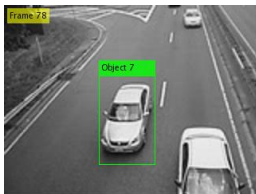
Khung hình 38



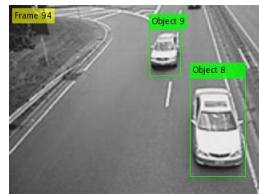
Khung hình 41



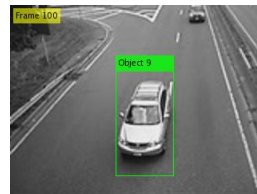
Khung hình 73



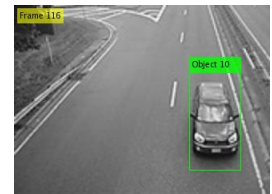
Khung hình 78



Khung hình 94



Khung hình 100

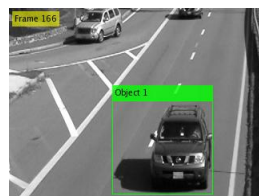


Khung hình 116

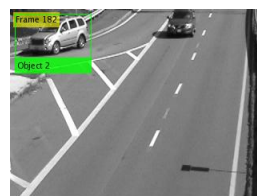
**Video visiontraffic.avi**



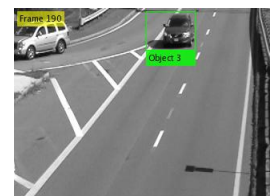
Khung hình 145



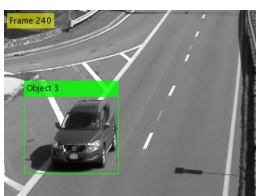
Khung hình 166



Khung hình 182



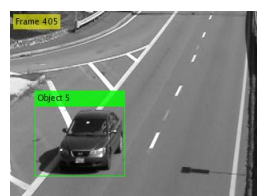
Khung hình 190



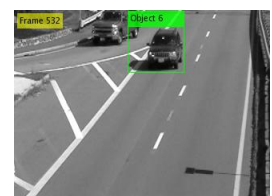
Khung hình 240



Khung hình 335



Khung hình 405



Khung hình 532

**Hình 4.3: Minh họa kết quả theo vết đối tượng**

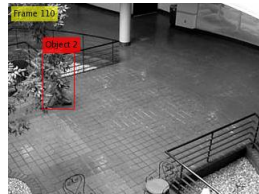
Trong video visiontraffic.avi khung hình 145 là đối tượng số 1; khung hình 166 là đối tượng số 1; khung hình 182 là đối tượng số 2; khung hình 190 là đối tượng số 3; khung hình 240 là đối tượng số 3; khung hình 335 là đối tượng số 4; khung hình 405 là đối tượng số 5; khung hình 532 là đối tượng số 6 và 3.

tượng số 3; khung hình 240 là đối tượng số 3; khung hình 335 là đối tượng số 4; khung hình 405 là đối tượng số 5; khung hình 532 là đối tượng số 6. Đối tượng số 7 chưa xuất hiện trọn vẹn nên không được theo vết. Kết quả theo vết chính xác và đủ 6 đối tượng trong video.

**Video atrium.avi**



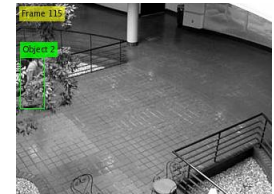
Khung hình 106



Khung hình 110



Khung hình 114



Khung hình 115



Khung hình 175



Khung hình 177



Khung hình 179

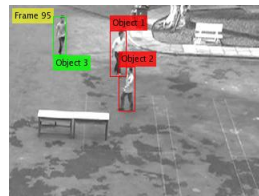


Khung hình 181

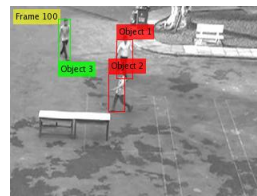
**Video bachkhoa11.mp4**



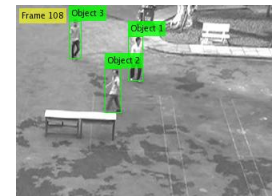
Khung hình 90



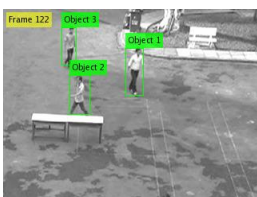
Khung hình 95



Khung hình 100



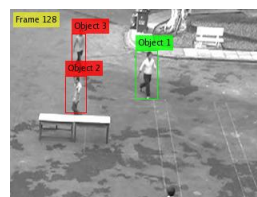
Khung hình 108



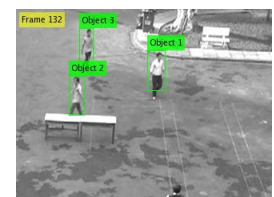
Khung hình 122



Khung hình 125



Khung hình 128



Khung hình 132

**Hình 4.4: Minh họa kết quả theo vết đối tượng có che khuất và giao cắt**

Trong thực nghiệm thứ hai đoạn video atrium.avi và bachkhoa11.mp4 được sử dụng để đánh giá trường hợp theo vết nhiều đối tượng chuyển động có che khuất,

và giao cắt. Hình 4.4 minh họa kết quả theo vết ở một vài khung hình của thực nghiệm này.

Trong video atrium.avi khung hình 106 đối tượng số 2 đang chuẩn bị di chuyển vào vùng bị che khuất bởi tán lá cây; khung hình 110 đối tượng số 2 đã bị che khuất bởi tán lá cây nhưng vẫn được ước lượng khá chính xác; khung hình 114 đối tượng số 2 đang chuẩn bị di chuyển ra khỏi vùng bị che khuất bởi tán lá cây, lúc này đối tượng đã bị tách thành 2 đối tượng nhỏ hơn vì bị lá cây chia cắt nên đối tượng vẫn được ước lượng; khung hình 115 đối tượng số 2 đã di chuyển ra khỏi vùng bị che khuất bởi tán lá cây và được theo vết chính xác trở lại; khung hình 175 đối tượng số 3 và 4 chuẩn bị giao cắt nhau; khung hình 177 đối tượng số 3 và 4 giao cắt nhau nhưng từng đối tượng vẫn được ước lượng chính xác; khung hình 179 đối tượng số 3 và 4 vẫn còn giao cắt nhau nên vẫn được ước lượng; khung hình 181 đối tượng số 3 và 4 đã tách ra và vẫn được theo vết chính xác.

Trong video bachkhoa11.mp4 khung hình 90 đối tượng số 1 và 2 chuẩn bị giao cắt nhau; khung hình 95 đối tượng số 1 và 2 giao cắt nhau nhưng từng đối tượng vẫn được ước lượng chính xác; khung hình 100 đối tượng số 1 và 2 vẫn còn giao cắt nhau nên vẫn được ước lượng; khung hình 108 đối tượng số 1 và 2 đã tách ra và vẫn được theo vết chính xác; khung hình 122 đối tượng số 2 và 3 chuẩn bị giao cắt nhau; khung hình 125 đối tượng số 2 và 3 giao cắt nhau nhưng từng đối tượng vẫn được ước lượng chính xác; khung hình 128 đối tượng số 2 và 3 vẫn còn giao cắt nhau nên vẫn được ước lượng; khung hình 132 đối tượng số 2 và 3 đã tách ra và vẫn được theo vết chính xác.

## **4.2 ĐÁNH GIÁ KẾT QUẢ**

### **4.2.1 Đánh giá kết quả phát hiện chuyển động:**

Dựa vào kết quả thực nghiệm, phương pháp IRI-LK mà đề tài đề xuất đã cho ra kết quả khá tốt. Trong thực nghiệm, khi 3 khung hình liên tiếp của video đầu vào có sự thay đổi độ sáng thì phương pháp Lucas-Kanade cho ra kết quả sai trong hai khung hình, phương pháp Gaussian Mixture Model cho ra kết quả sai trong một khung hình còn phương pháp IRI-LK thì hoàn toàn chính xác.



Với các thực nghiệm tương tự có sự thay đổi độ sáng khác nhau thì phương pháp IRI-LK vẫn cho ra kết quả chính xác. Khi thống kê số khung hình cho ra kết quả sai trong cả đoạn video thì hầu hết ở các đoạn video thực nghiệm phương pháp IRI-LK đều cho ra kết quả chính xác gần như không có khung hình nào cho ra kết quả sai.

Tuy nhiên với các video mà có vài khung hình được tăng độ sáng lên trên 20% để thực nghiệm thì kết quả không chính xác. Lúc này khung hình quá sáng, các vùng ảnh có độ sáng chạm ngưỡng thì không thể sáng hơn nữa. Trong khi một số vùng ảnh khác cũng sáng lên đến ngưỡng. Các vùng ảnh này hoàn toàn hoà lẫn vào nhau không còn sự khác biệt. Điều này dẫn đến mất thông tin nên ảnh ở khung hình sau không còn đầy đủ các đối tượng như khung hình trước đó. Do vậy không thể phát hiện đúng được đối tượng chuyển động trong trường hợp này. Tuy nhiên trường hợp này trong thực tế không phổ biến. Vì vậy phương pháp IRI-LK vẫn đạt được hiệu quả cao.

#### **4.2.2 Đánh giá kết quả theo vết đối tượng:**

Kết quả thực nghiệm theo vết đối tượng được đánh giá bằng phương pháp quan sát thực tế. Qua quan sát, việc theo vết đối tượng hoàn toàn chính xác trong các video có các đối tượng chuyển động không giao cắt lẫn nhau. Việc gán nhãn đối tượng cũng rất chính xác. Mỗi đối tượng được gán đúng một nhãn riêng biệt và số nhãn đã gán đúng bằng số đối tượng được theo dõi.

Trong các video có các đối tượng chuyển động bị che khuất hay giao cắt nhau, các đối tượng cũng được theo vết chính xác. Khi bị che khuất hay giao cắt, các đối tượng sẽ được ước lượng tương đối. Sau khi đối tượng xuất hiện trở lại thì các đối tượng vẫn được theo vết khá chính xác.

Tuy nhiên đối với trường hợp trong khi bị che khuất hoặc giao cắt mà đối tượng thay đổi hướng đi khiến cho khi xuất hiện trở lại đối tượng ở quá xa với đối tượng ước lượng nên không thể theo vết được nữa. Lúc này đối tượng ước lượng sẽ bị ước lượng mãi mãi và đối tượng xuất hiện trở lại sẽ là đối tượng mới.

## **Chương 5 KẾT LUẬN**

### **5.1 KẾT QUẢ ĐẠT ĐƯỢC CỦA ĐỀ TÀI**

Trong luận văn, tác giả đã nghiên cứu được các bộ lọc thông dụng trong xử lý ảnh và video. Đặc biệt là các bộ lọc hình thái rất hữu ích trong việc xử lý kết quả phát hiện chuyển động để cho ra mặt nạ đối tượng chuyển động chính xác hơn. Tác giả cũng đã nghiên cứu và cài đặt thực nghiệm nhiều phương pháp phát hiện chuyển động phổ biến như phương pháp trừ nền, phương pháp sai khác ảnh, phương pháp Gaussian Mixture Model, phương pháp Lucas-Kanade để đánh giá và lựa chọn phương pháp phù hợp.

Trong đề tài, tác giả nghiên cứu đề xuất phương pháp kết hợp tỷ lệ cường độ ảnh tích hợp vào phương pháp Lucas-Kanade (IRI-LK) đồng thời tích hợp thêm nhiều bộ lọc để cải thiện việc theo dõi đối tượng chính xác hơn trong trường hợp ánh sáng thay đổi liên tục trong video. Hơn thế nữa, đề tài cũng đã xử lý được một số trường hợp đối tượng bị che khuất và giao cắt đơn giản.

Các thực nghiệm của đề tài được áp dụng trên nhiều video đầu vào và đã cho ra kết quả chính xác đáng tin cậy hơn các phương pháp khác.

### **5.2 ƯU NHƯỢC ĐIỂM CỦA GIẢI THUẬT ĐỀ XUẤT**

Giải thuật do đề tài đề xuất có ưu điểm vượt trội so với các phương pháp khác trong trường hợp có ánh sáng thay đổi giữa các khung hình. Với ưu điểm này giải thuật đã khắc phục được việc nhận diện sai do ánh sáng môi trường thay đổi. Điều này giúp cho việc theo vết đối tượng chính xác hơn.

Đề tài xác định rằng phát hiện chuyển động là bước quan trọng nhất trong theo dõi đối tượng chuyển động nên tập trung vào việc cải thiện kết quả phát hiện chuyển động. Vì vậy, phần phát hiện chuyển động đã đạt được kết quả khá tốt. Còn phần theo vết đối tượng thì đề tài cũng đã giải quyết được các yêu cầu cơ bản.

Phương pháp IRI-LK hoạt động tốt trong trường hợp có sự thay đổi độ sáng trong khoảng 20% giữa các khung hình tùy theo độ sáng của khung hình đang xử lý.



Nếu sự thay đổi độ sáng vượt qua ngưỡng này thì ảnh sẽ bị mất chi tiết. Điều này xảy ra do các vùng ảnh có độ sáng chạm ngưỡng không thể sáng hơn nữa trong khi các vùng ảnh khác sáng lên đến ngưỡng nên không thể phân biệt. Tuy nhiên với sự thay đổi độ sáng khung hình thường gặp thì thuật toán IRI-LK đều xử lý tốt.

### **5.3 HƯỚNG PHÁT TRIỂN MỞ RỘNG**

Giải thuật đề xuất của đề tài đã đạt được một số kết quả nhất định. Tuy nhiên bài toán theo dõi đối tượng chuyển động là bài toán khá phức tạp, có nhiều vấn đề cần phải xử lý để nâng cao hiệu quả của bài toán. Vì vậy sau khi hoàn thành luận văn, đề tài vẫn được tiếp tục nghiên cứu phát triển mở rộng.

Hướng thứ nhất là nghiên cứu cải thiện thuật toán IRI-LK để kết quả phát hiện chuyển động sẽ nhận được mặt nạ đối tượng chuyển động chính xác. Các vùng chuyển động trong mặt nạ nhị phân nhận được sẽ bao chính xác lên đối tượng một cách liền lạc và không bao gồm bóng của đối tượng.

Hướng thứ hai là nghiên cứu cải thiện độ chính xác của các vectơ dòng quang học. Phối hợp các vectơ dòng quang học này và một số đặc trưng khác để theo vết đối tượng chính xác hơn. Cải thiện độ chính xác trong việc xử lý che khuất, gộp nhiều đối tượng và tách nhiều đối tượng.

Hướng thứ ba là sử dụng kết quả theo vết đối tượng chuyển động, phát triển thuật toán để phân lớp đối tượng và tiến đến là nhận diện đối tượng, nhận diện hành vi của đối tượng. Đồng thời cải thiện tốc độ thuật toán để có thể xử lý trong thời gian thực trên các máy tính thông dụng.

## TÀI LIỆU THAM KHẢO

- [1] Moeslund, T. B. (2012). 'Introduction to video and image processing: Building real systems and applications'. *Springer*.
- [2] Lin, Y. T., & Chang, C. H. (2011). 'User-aware Video Coding Based on Semantic Video Understanding and Enhancing'. *InTech*.
- [3] Stauffer, C., & Grimson, W. E. L. (1999). 'Adaptive background mixture models for real-time tracking'. In: *Computer Vision and Pattern Recognition*, 1999. IEEE Computer Society Conference on. (Vol. 2). IEEE.
- [4] Rout, R. K. (2013). 'A survey on object detection and tracking algorithms' (Doctoral dissertation).
- [5] Frantc, V. A., Voronin, V. V., Marchuck, V. I., & Egiazarian, K. O. (2013, February). 'Video inpainting using scene model and object tracking'. In: *IS&T/SPIE Electronic Imaging* (pp. 86550V-86550V). International Society for Optics and Photonics.
- [6] Lucas, B. D., & Kanade, T. (1981, August). 'An iterative image registration technique with an application to stereo vision'. In: *IJCAI* (Vol. 81, pp. 674-679).
- [7] Shirageri, M. S., Udupi, G. R., & Bidkar, G. A. (2013). 'Design and development of Optical flow based Moving Object Detection and Tracking (OMODT) System'. *Vectors*.
- [8] Lu, N., Wang, J., Yang, L., & Wu, Q. H. (2007). 'Motion Detection Based On Accumulative Optical Flow and Double Background Filtering'. In: *World Congress on Engineering* (pp. 602-607).
- [9] Karlsson, S. M., & Bigun, J. (2012, June). 'Lip-motion events analysis and lip segmentation using optical flow'. In: *Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2012 IEEE Computer Society Conference on (pp. 138-145). IEEE.

- [10] Otsu, N. (1975). 'A threshold selection method from gray-level histograms'. *Automatica*, 11(285-296), 23-27.
- [11] Kalman, R. E. (1960). 'A new approach to linear filtering and prediction problems'. *Journal of Fluids Engineering*, 82(1), 35-45.
- [12] Welch, G., & Bishop, G. (1995). 'An introduction to the Kalman filter'.
- [13] Elies Henar, F. 'Application of a Kalman Filter with Augmented Measurement Model in Non-Invasive Cardiac Imaging'.
- [14] Mehta, M., Goyal, C., Srivastava, M. C., & Jain, R. C. (2010, February). 'Real time object detection and tracking: Histogram matching and kalman filter approach'. In: *Computer and Automation Engineering (ICCAE)*, 2010 The 2nd International Conference on (Vol. 5, pp. 796-801). IEEE.
- [15] Johnsen, S., & Tews, A. (2009, May). 'Real-time object tracking and classification using a static camera'. In: *Proceedings of IEEE International Conference on Robotics and Automation, workshop on People Detection and Tracking*.
- [16] Ranjan, A., & Chaple, M. (2013, April). 'A Novel Approach For Object Detection And Tracking'. In: *International Journal of Engineering Research and Technology* (Vol. 2, No. 4 (April-2013)). ESRSA Publications.
- [17] Nguyen Thanh Binh, Ashish Khare (2013). 'Object tracking in curvelet domain'. In: *Image Denoising, Deblurring and Object Tracking, A new generation wavelet based approach* (pp. 87-117). LAP LAMBERT Academic Publishing.
- [18] Trần Thanh Việt, Trần Công Chiến, Huỳnh Cao Tuấn, Nguyễn Hữu Nam, Đỗ Năng Toàn, Trần Hành (2011, Oct). 'Một kỹ thuật phát hiện và bám sát đối tượng'. In: *Kỹ yếu hội thảo quốc gia lần thứ 14*, Cần Thơ.

## PHỤ LỤC

Trong quá trình làm luận văn, tác giả đã gửi kết quả nghiên cứu của mình (bài báo “Motion Detection Based on Intensity Ratio Image”) đến hội nghị quốc tế “International Conference on Nature of Computation and Communication” (hội nghị ICTCC 2014) và đã được chấp nhận báo cáo tại hội nghị vào ngày 24 và 25/11/2014 tại TP.HCM. Bài báo này được nhà xuất bản Springer chấp nhận đăng tải vào cuối năm nay.

### **Email chấp nhận báo cáo tại hội nghị:**

-----Original Message-----

From: [noreply@icst.org](mailto:noreply@icst.org) [mailto:[noreply@icst.org](mailto:noreply@icst.org)]

Sent: Wednesday, September 17, 2014 5:37 PM

To: Pham Bao Quoc

Subject: Your Paper has been accepted for ICTCC 2014

---

subject: Your Paper has been accepted for ICTCC 2014

---

Dear Pham Bao Quoc,

Congratulations, we are pleased to inform you that your paper

MOTION DETECTION BASED ON INTENSITY RATIO IMAGE

has been accepted for presentation as Full Paper in ICTCC 2014.

Please visit <http://ictcc.org/2014/> to register for the conference (a registration webpage will be open shortly). Only after you have concluded the registration process and we have received your full member or non-member registration fee, will you be eligible to submit the camera ready version of your paper at <http://cameraready.eai.eu/>. Please note that if you chose to pay via bank transfer it may take up to 5 business days before your transfer arrives and your registration is concluded.

We wish you a very successful ICTCC 2014

Best regards,

Organizing Team  
ICTCC 2014

**Nội dung bài báo:**

# MOTION DETECTION BASED ON INTENSITY RATIO IMAGE

Pham Bao Quoc\* and Nguyen Thanh Binh

\* Faculty of Information of Technology,  
Ho Chi Minh City University of Technology, HUTECH, Vietnam  
Faculty of Computer Science and Engineering,  
Ho Chi Minh City University of Technology, VNUHCM, Vietnam

[nationpham@hotmail.com](mailto:nationpham@hotmail.com), [ntbinh@cse.hcmut.edu.vn](mailto:ntbinh@cse.hcmut.edu.vn)

**Abstract.** Motion detection is the first important step in large applications of computer vision. Motion detection extracts moving objects from the background. There are many methods to do that. However, in almost methods, if the input video has noise and light change, moving objects will not be extracted accurately. In this paper, we propose the method for motion detection which extracts moving objects from the background based on the intensity ratio image concept that is not affected by light change; therefore, the sensitivity with light change is overcome. The intensity ratio image is computed by the average intensity of current frame and the intensity of every pixel in that frame. The intensity ratio of a pixel is nearly unchanged between two frames. We apply the Lucas-Kanade optical flow method based on that intensity ratio image. Our proposed algorithm has good noise tolerance and is not affected by light change. For demonstrating the superiority of the proposed method, we have compared the results with the other recent methods available in literature.

**Keywords:** motion detection, intensity ratio image, moving object.

## 1 INTRODUCTION

Real-time object tracking is a popular application of computer vision. It faces up to complex problems. Although the algorithms have to do a lot of manipulation, they must be fast enough to finish processing a video frame in the extremely short time between two frames. The motion detection is a very important and complex step in the real-time object tracking system. In this step, moving objects will be extracted from the background. It is not easy to extract moving objects. In the input video, there is a lot of noise and light change because of effecting outdoor environment. They make the moving object extraction inaccurate. They make some parts of the background become moving objects.

In the past, there are many methods to extract moving objects from the background. In image subtraction methods [1], the current frame will be subtracted with a reference frame. The reference frame may be the background frame (background subtraction) or the previous frame (frame difference). This method is very sensitive with noise and light changing. Noise and light changing make the current frame different from the reference frame. These different pixels become the foreground and make the moving object extraction inaccurate. In Gaussian mixture model [2], Stauffer has proposed a probabilistic approach using a mixture of Gaussian for identifying the background and foreground. This method is not affected by noise and sudden light change. However, if the light changes

continuously, some parts of the background will become foreground objects. In Lucas-Kanade optical flow method [4, 5, 6, 7, 8], two continuous frames are used to compute the velocity of moving objects by the spatial and temporal derivatives. This method has good noise tolerance. However, if the light changes suddenly and continuously, some parts of the background will become moving objects.

In this paper, we propose the method for motion detection extracts moving objects from the background based on the intensity ratio image concept that is not affected by light change; therefore, the sensitivity with light change is overcome. The intensity ratio image is computed by the average intensity of current frame and the intensity of every pixel in that frame. The intensity ratio of a pixel is nearly unchanged between two frames. Then, we apply the Lucas-Kanade optical flow method based on that intensity ratio image. Our proposed algorithm has good noise tolerance and is not affected by light change. It is suitable for the real-time object tracking system. For demonstrating the superiority of the proposed method, we have compared the results with the other recent methods available in literature.

The rest of the paper is organized as follows: in section 2, we described the basic of Lucas-kanade optical flow; details of the proposed method are given in section 3; the results of the proposed method are presented in section 4 and our conclusions in section 5.

## 2 LUCAS-KANADE OPTICAL FLOW

The Lucas-Kanade optical flow method is proposed by Lucas and Takeo [4]. This optical flow method is used to compute velocity of moving objects between two continuous frames by the spatial and temporal derivatives. It is fast and has a low computational cost, and good noise tolerance [5]. It tries to compute the motion between two continuous frames at time  $t$  and  $t + Dt$ . Assuming the intensity of a pixel does not change between two frames, we have equation:

$$I(x, y, t) = I(x + \Delta x, y + \Delta y, t + \Delta t) \quad (2.1)$$

Assuming the movement between two frames is small, the equation (2.1) with Taylor series can be derived to give:

$$I(x + \Delta x, y + \Delta y, t + \Delta t) \approx I(x, y, t) + \frac{dI}{dx} \Delta x + \frac{dI}{dy} \Delta y + \frac{dI}{dt} \Delta t \quad (2.2)$$

From (2.1) and (2.2), we obtain the following:

$$\frac{dI}{dx} \Delta x + \frac{dI}{dy} \Delta y + \frac{dI}{dt} \Delta t = 0 \quad (2.3)$$

or

$$\frac{dI}{dx} \frac{\Delta x}{\Delta t} + \frac{dI}{dy} \frac{\Delta y}{\Delta t} + \frac{dI}{dt} \frac{\Delta t}{\Delta t} = 0 \quad (2.4)$$

and the result is:

$$\frac{dI}{dx}V_x + \frac{dI}{dy}V_y + \frac{dI}{dt} = 0 \quad (2.5)$$

where  $V_x$  and  $V_y$  are the  $x$  and  $y$  components of the velocity and  $\frac{dI}{dx}$ ,  $\frac{dI}{dy}$ ,  $\frac{dI}{dt}$  are the spatial and temporal derivatives at  $x$ ,  $y$ ,  $t$ . Set  $I_x, I_y, I_t$  as those derivatives, we have equation:

$$I_x V_x + I_y V_y = -I_t \quad (2.6)$$

Assuming the neighboring pixels move at the same velocity with the pixel under consideration, we have the following:

$$\begin{cases} I_{x_1} V_x + I_{y_1} V_y = -I_{t_1} \\ \mathbf{M} \\ I_{x_m} V_x + I_{y_m} V_y = -I_{t_m} \end{cases} \quad (2.7)$$

The (2.7) can be written as:

$$\begin{bmatrix} I_{x_1} & I_{y_1} \\ \mathbf{M} & \mathbf{M} \\ I_{x_m} & I_{y_m} \end{bmatrix} \begin{bmatrix} V_x \\ V_y \end{bmatrix} = \begin{bmatrix} -I_{t_1} \\ \mathbf{M} \\ -I_{t_m} \end{bmatrix} \quad (2.8)$$

Set  $A = \begin{bmatrix} I_{x_1} & I_{y_1} \\ \mathbf{M} & \mathbf{M} \\ I_{x_m} & I_{y_m} \end{bmatrix}$ ,  $v = \begin{bmatrix} V_x \\ V_y \end{bmatrix}$ ,  $b = \begin{bmatrix} -I_{t_1} \\ \mathbf{M} \\ -I_{t_m} \end{bmatrix}$ , we have equation:

$$Av = b \quad (2.9)$$

Use the least squares method to solve equation (2.9), we have the following:

$$v = (A^T A)^{-1} A^T b \quad (2.10)$$

And the result is:

$$\begin{bmatrix} V_x \\ V_y \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^m I_{x_i}^2 & \sum_{i=1}^m I_{x_i} I_{y_i} \\ \sum_{i=1}^m I_{y_i} I_{x_i} & \sum_{i=1}^m I_{y_i}^2 \end{bmatrix}^{-1} \begin{bmatrix} -\sum_{i=1}^m I_{x_i} I_{t_i} \\ -\sum_{i=1}^m I_{y_i} I_{t_i} \end{bmatrix} \quad (2.11)$$

We can use Gaussian function in computational derivatives step and add a Tikhonov constant to (2.11) for the better result [8].

### 3 THE PROPOSED METHOD

Real-time object tracking is a complex problem and a popular application of computer vision. In this section, we propose an approach for motion detection which extracts moving objects using an intensity ratio image based on Lucas-Kanade method

(IRI-LK). We apply the Lucas-Kanade optical flow method based on the intensity ratio image. We propose the intensity ratio image concept because it is not affected by light change.

In input video, assuming the different pixels of actual interested moving objects between two frames is small. If the light does not change between two continuous frames, the average intensity of two continuous frames is different slightly. Else, the intensity of all pixels is changed and the average intensity is significant different between two continuous frames, which means the average intensity depends on the light change. Assumingly, the intensity of all pixels will change with a same coefficient  $I$  when the light changes.

Set  $I(x, y, t)$  and  $I(x+Dx, y+Dy, t+Dt)$  as intensity of pixel at time  $t$  and  $t + Dt$ , we have equation:

$$I(x, y, t) = II(x + \Delta x, y + \Delta y, t + \Delta t) \quad (3.1)$$

Set  $\bar{I}(t)$  and  $\bar{I}(t + \Delta t)$  as the average intensity of the frame with the size  $m \times n$  at the time  $t$  and  $t+Dt$ , we have the following:

$$\bar{I}(t) = \frac{\sum_{y=1}^n \sum_{x=1}^m I(x, y, t)}{m.n} \quad (3.2)$$

and

$$\bar{I}(t + \Delta t) = \frac{\sum_{y=1}^n \sum_{x=1}^m I(x + \Delta x, y + \Delta y, t + \Delta t)}{m.n} \quad (3.3)$$

From (3.1), (3.2) and (3.3), we have equation:

$$\bar{I}(t) = I\bar{I}(t + \Delta t) \quad (3.4)$$

Set  $R(x, y, t)$  and  $R(x+Dx, y+Dy, t+Dt)$  as the intensity ratio of pixel at time  $t$  and  $t+Dt$ , we have the following:

$$R(x, y, t) = \frac{I(x, y, t)}{\bar{I}(t)} \quad (3.5)$$

and

$$R(x + \Delta x, y + \Delta y, t + \Delta t) = \frac{I(x + \Delta x, y + \Delta y, t + \Delta t)}{\bar{I}(t + \Delta t)} \quad (3.6)$$

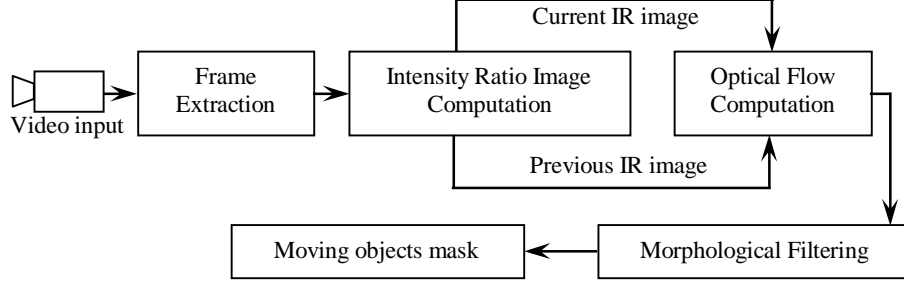
From (3.1), (3.4), (3.5) and (3.6), we have equation:

$$R(x, y, t) = R(x + \Delta x, y + \Delta y, t + \Delta t) \quad (3.7)$$

The equation (3.7) means the intensity ratio of pixels do not change when the light changes. Therefore, the intensity ratio image is not affected by light change. If the Lucas-Kanade optical flow method is based on the intensity ratio image, the sensitivity with light change of it is overcome.



The proposed method IRI-LK algorithm which we propose is depicted in figure 1. This model includes four steps. The two main steps are intensity ratio image computation and optical flow computation.



**Fig. 1.** Intensity ratio image based Lucas-Kanade model

Firstly, the frame extraction. Frame sequences are extracted from input video. Frame by frame are converted to intensity images. Then, the intensity image is standardized. The intensity value will be a real number between 0 and 1. In this step, we can use an image processing function to resize the frame in an appropriate size.

Secondly, the intensity ratio image computation. Each intensity image in turn is used to compute the intensity ratio image. We use the equation (3.2) to compute the average intensity of the current intensity image. Then, we use the equation (3.5) to compute the intensity ratio image. We can multiply the intensity ratio image with a coefficient  $a$  which depends on the highest intensity ratio. This makes the intensity ratio value always in a correct range.

Thirdly, optical flow computation. The Lucas-Kanade method is applied to compute the velocity of moving objects. The spatial and temporal derivatives are computed based on the current intensity ratio image and previous intensity ratio image using Gaussian function. We use the equation (2.11) to calculate the velocity vector  $\begin{bmatrix} V_x \\ V_y \end{bmatrix}$  of moving objects. The velocity magnitude  $V(x, y)$  is calculated by the following:

$$V(x, y) = \sqrt{V_x^2 + V_y^2} \quad (3.8)$$

For returning binary foreground mask  $B(x, y)$ , the velocity magnitude  $V(x, y)$  is taken threshold by the following:

$$B(x, y) = \begin{cases} 1 & (V(x, y) \geq Threshold) \\ 0 & (V(x, y) < Threshold) \end{cases} \quad (3.9)$$

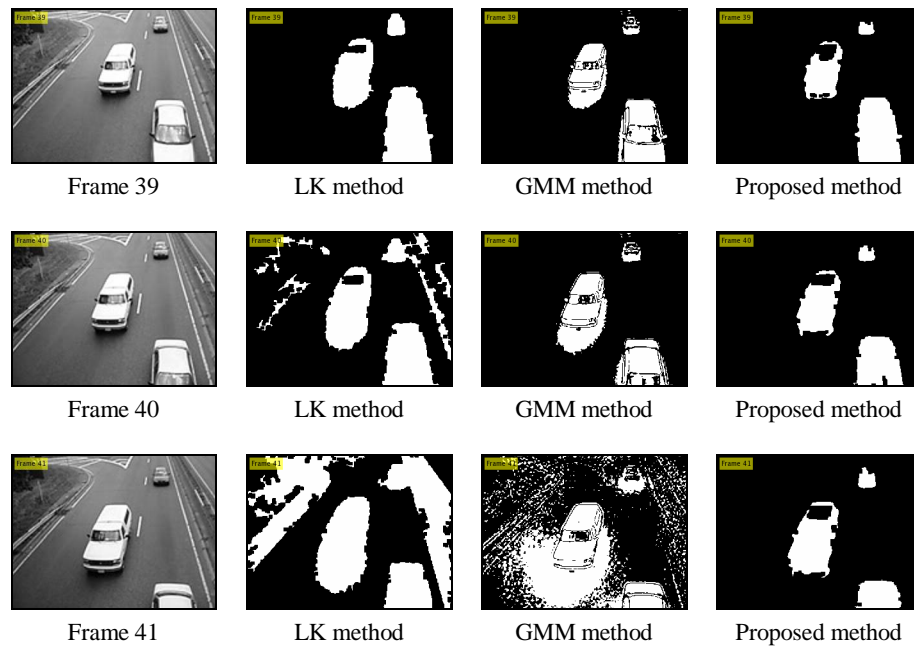
Finally, we apply some morphological filtering to remove noise and small blobs such as morphological closing, image fill, and binary area open.

## 4 EXPERIMENTAL RESULTS

In this section, we illustrate the results of IRI-LK algorithm. The implementation of developed algorithm has been tested under MATLAB platform. The input video is captured by a static camera with the resolution of 160 x120 pixels, at the frame rate of 15fps. We choose coefficient  $a = 6$  and Tikhonov constant = 0.005. We tested many video clips. In here, we present some frames. In the input video, there are many frames having light change. Video clips for testing are taken from standard datasets and some clips from cameras on the streets. For demonstrating the superiority of the proposed method, we have compared the results with the Lucas-Kanade (LK) method, Gaussian Mixture Model (GMM) method.

In figure 2, we show three input frames that have light change. The light is darker from frame 39 to frame 41. For the Lucas-Kanade optical flow method, we receive a flash at frame 40 and frame 41. For the Gaussian mixture model method, we receive a flash at frame 41. For the proposed method, the segment result is very good.

Table 1 compares the motion segment error between LK method, GMM, and the proposed method (IRI-LK). Look at table 1, we have the percentage of the motion segment error is LK: 12.40%, GMM: 8.26%, IRI-LK: 0.00%. The LK gets the motion segment error when the light changes suddenly and continuously. The GMM gets the motion segment error when the light changes continuously. The IRI-LK is very good for this input video.



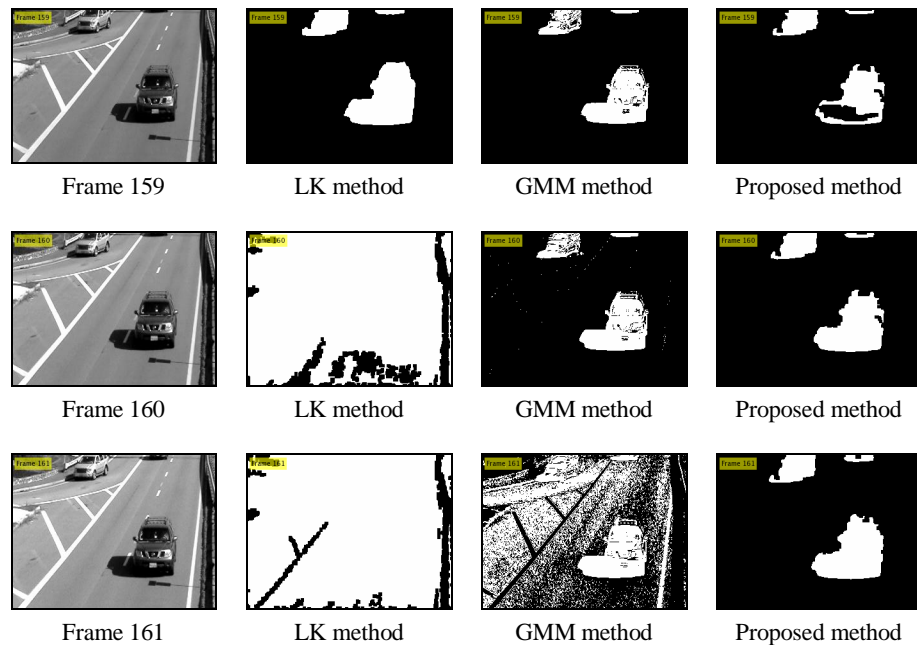
**Fig. 2.** The segment of Car video compared the proposed method with the other methods

**Table 1.** Comparing the motion segment error of LK, GMM, and IRI-LK

Frame	Average Intensity	LK method	GMM method	Proposed method
1	0.5220			
...				
21	0.5030			
22	0.5056			
23	0.4628	Error		
24	0.4717			
25	0.4854			
...				
38	0.5165			
39	0.5035			
40	0.4643	Error		
41	0.4166	Error	Error	
42	0.3714	Error	Error	
43	0.3922		Error	
44	0.3374	Error	Error	
45	0.3814	Error	Error	
46	0.4026	Error	Error	
47	0.4345	Error		
48	0.4829	Error		
49	0.4818			
50	0.4809			
...				
76	0.4764			
77	0.4975			
78	0.4472	Error	Error	
79	0.4056	Error	Error	
80	0.4237		Error	
81	0.4068		Error	
82	0.4551	Error		
83	0.4925	Error		
84	0.4890			
85	0.4892			
...				
98	0.4790			
99	0.4677			
100	0.4200	Error		
101	0.4282			
102	0.4330			

Frame	Average Intensity	LK method	GMM method	Proposed method
103	0.4267			
104	0.4704	Error		
105	0.4875			
106	0.4872			
...				
121	0.4891			

In another test, the input video is captured by a static camera with the resolution of 640x360 pixels, at the frame rate of 29fps. We choose coefficient  $a = 6$  and Tikhonov constant = 0.001. We add 10% brightness at frame 160 and 20% brightness at frame 161 for testing. Three input frames are shown in figure 3.



**Fig. 3.** The segment of clip, which was captured by a static camera, compared the proposed method with the other methods.

In figure 3, we also show three input frames that have light change. For the Lucas-Kanade optical flow method, we receive a flash at frame 160 and frame 161. For the Gaussian mixture model method, we receive a flash at frame 161. For the proposed method, the segment result is also very good.

## 5 CONCLUSIONS

In motion detection, the noise and light change makes moving object identification inaccurate because the Lucas-Kanade optical flow method works well with the assumption

that the intensity of pixels does not change. We propose the intensity ratio image concept that is not affected by light change in accordance with the Lucas-Kanade optical flow method. The proposed method has good noise tolerance and is not affected by light change. It is good for the real-time object tracking system. In fact, the light does not change with the same coefficient  $I$  at all pixels. The intensity ratio of a pixel has a small difference between two continuous frames. It makes the segment result not good in all cases. Despite that, our proposed method greatly improves the result of motion detection in the light change.

## REFERENCES

1. Moeslund, Thomas B.: Introduction to video and image processing: Building real systems and applications. Springer (2012)
2. Rout, Rupesh Kumar: A survey on object detection and tracking algorithms. Diss (2013)
3. Stauffer, Chris, and W. Eric L. Grimson: Adaptive background mixture models for real-time tracking. Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on. Vol. 2. IEEE (1999)
4. Lucas, Bruce D., and Takeo Kanade: An iterative image registration technique with an application to stereo vision. IJCAI. Vol. 81. (1981)
5. Lu, Nan, et al.: Motion Detection Based On Accumulative Optical Flow and Double Background Filtering. World Congress on Engineering (2007)
6. Frantc, V. A., et al.: Video inpainting using scene model and object tracking. IS&T/SPIE Electronic Imaging. International Society for Optics and Photonics (2013)
7. Shirageri, Ms Shamshad, G. R. Udupi, and G. A. Bidkar.: Design and development of Optical flow based Moving Object Detection and Tracking (OMODT) System. vectors 2.4 (2013)
8. Karlsson, Stefan M., and Josef Bigun.: Lip-motion events analysis and lip segmentation using optical flow. Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on. IEEE (2012)