**BỘ CÔNG THƯƠNG**

**TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP HÀ NỘI**



**BÁO CÁO**

**THỰC TẬP TỐT NGHIỆP**

**TÌM HIỂU MỘT SÔ PHƯƠNG PHÁP PHÁT HIỆN ĐỐI TƯỢNG CHUYỂN ĐỘNG TRONG VIDEO**

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

Sinh viên thực hiện: **Vũ Tiến Dũng**

*Hà Nội, tháng 2 – 2016*

**MỤC LỤC**

[LỜI CẢM ƠN 3](#_Toc444581798)

[TỔNG QUAN VỀ ĐỀ TÀI 4](#_Toc444581799)

[1. Lý do chọn đề tài 5](#_Toc444581800)

[2. Mục tiêu nghiên cứu 5](#_Toc444581801)

[3. Phương pháp nghiên cứu 5](#_Toc444581802)

[4. Đối tượng nghiên cứu 6](#_Toc444581803)

[Chương I 7](#_Toc444581804)

[TỔNG QUAN VỀ NỘI DUNG NGHIÊN CỨU 7](#_Toc444581805)

[1. Tổng quan về ảnh và video 8](#_Toc444581806)

[1.1. Tổng quan về ảnh 8](#_Toc444581807)

[1.2. Tổng quan về video 16](#_Toc444581808)

[1.3. Một số kỹ thuật phá hiện chuyển động 19](#_Toc444581809)

[Chương III 30](#_Toc444581810)

[THỰC NGHIỆM 30](#_Toc444581811)

[1. Môi trường cài đặt 31](#_Toc444581812)

[2. Xây dựng ứng dụng 31](#_Toc444581813)

[3. Kết quả thực nghiệm 31](#_Toc444581814)

[KẾT LUẬN 32](#_Toc444581815)

[TÀI LIỆU THAM KHẢO 33](#_Toc444581816)

**DANH MỤC CÁC HÌNH**

[Hình 1.1: Quan hệ trong vùng lân cận giữa các điểm ảnh. 12](#_Toc444583524)

[Hình 1.2: Cấu trúc phân đoạn video 19](#_Toc444583525)

[Hình 1.3: Biểu diễn video dựa trên đối tượng 21](#_Toc444583526)

[Hình 1.4: Ví dụ phương pháp khác biệt thời gian 24](#_Toc444583527)

[Hình 1.5: Mô hình hoạt động của phương pháp trừ nền 27](#_Toc444583528)

[Hình 1.6: Luồng quang học 28](#_Toc444583529)

**DANH SÁCH BẢNG BIỂU**

[Bảng 1.1: Các tiêu chuẩn của Video số 19](#_Toc444583421)

[Bảng 1.2: Bảng phân loại các thuật toán phát hiện đối tượng 21](#_Toc444583422)

# LỜI CẢM ƠN

Trước tiên tôi xin gửi lời cảm ơn sâu sắc đến Thầy Ngô Đức Vĩnh và các thầy cô giáo bộ môn nghành công nghệ thông tin đã tạo mọi điều kiện về cơ sở vật chất và tinh thần giúp đở hướng dẫn tôi trong trong thời gian làm thực tập tốt nghiệp.

Tôi xin cảm ơn các thầy giáo, cô giáo Khoa Công Nghệ Thông Tin Trường Đại học Công Nghiệp Hà Nội đã trang bị kiến thức cho tôi những kiến thức cần thiết và bổ ích để hoàn thành thực tập.

Do thời gian và kiến thức còn hạn chế nên báo cáo không tránh khỏi những sai sót. Tôi mong nhận được sự đóng góp bổ sung của thầy cô giáo và các bạn.

Cuối cùng xin chân thành cảm ơn tất cả các bạn đã đóng góp ý kiến và hỗ trợ tôi trong quá trình thực hiện báo cáo này.

*Hà Nội, tháng 2 năm 2016*

***Vũ Tiến Dũng***

# TỔNG QUAN VỀ ĐỀ TÀI

**Nội dung chính**

* Lý do chọn đề tài
* Mục tiêu nghiên cứu
* Phương pháp nghiên cứu
* Đối tượng nghiên cứu

## Lý do chọn đề tài

Giám sát tự động là một lĩnh vực nhận dạng và xử lý ảnh 2 chiều. Đồng thời, đó cũng là một hướng đi cho mảng phần mềm thiết kế chuyên dụng cho các thiết bị giám sát tự động. Việc phát hiện ra các đối tượng chuyển động trong video nhờ các kỹ thuật xử lý ảnh, trên cơ sở đó đoán nhận một số hành vi của đối tượng là một việc làm có ý nghĩa khoa học và thực tiễn. Nhất là trong hoàn cảnh Việt Nam chưa có nhiều những nghiên cứu và ứng dụng theo hướng này.

Xuất phát từ thực tế đó, việc nghiên cứu và đưa ra các phương pháp để xử lý video là vô cùng thiết thực. Được sự hướng dẫn của thầy Ngô Đức Vĩnh tôi đã tiến hành nghiên cứu đề tài “***Tìm hiểu một số phương pháp phát hiện đối tượng chuyển động trong video***” Các vấn đề cần giải quyết bao gồm:

- Nghiên cứu tổng quan về video và phát hiện đối tượng chuyển động.

- Nghiên cứu và đề xuất một số hướng khắc phục các nhược điểm trong việc phát hiện, đánh dấu các đối tượng chuyển động.

- Cài đặt một số phương pháp phát hiện, đánh dấu các đối tượng chuyển động dựa theo các phương pháp đã nêu ở trên.

## Mục tiêu nghiên cứu

* Tìm hiểu về phát hiện đối tượng
* Tìm hiểu về các kỹ thuật theo dõi đối tượng chuyển động .
* Phân tích đánh giá ưu nhược điểm của từng phương pháp.
* Tìm hiểu về các thuật toán liên quan đến lĩnh vực xử lý ảnh.
* Đánh giá hiệu quả của phương pháp cải tiến.

## Phương pháp nghiên cứu

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

## Đối tượng nghiên cứu

* Kỹ thuật phát hiện đối tượng chuyển động trong video.

# Chương I

# TỔNG QUAN VỀ NỘI DUNG NGHIÊN CỨU

**Nội dung chính**

* Tổng quan về ảnh
* Tổng quan về video

## Tổng quan về ảnh và video

### Tổng quan về ảnh

#### Ảnh và biểu diễn ảnh

Ảnh trong thực tế là một ảnh liên tục cả về không gian và giá trị độ sáng . Để có thể xử lý ảnh bằng máy tính thì cần thiết phải tiến hành số hóa ảnh. Quá trình số hóa biến đổi các tín hiệu liên tục sang tín hiệu rời rạc thông qua quá trình lấy mẫu ( rời rạc hóa về không gian) và lượng tử hóa các thành phần giá trị mà về nguyên tắc mắt thường không thể phân biệt được hai điểm gần kề nhau. Các điểm như vậy gọi là pixel hay các phần tử ảnh, điểm ảnh.

Như vậy mỗi ảnh là tập hợp các điểm ảnh. Khi được số hóa nó thường được biểu diễn bởi mảng 2 chiều I(n,p): n là dòng và p là cột.

Về mặt toán học có thể xem ảnh là một hàm hai biến f(x,y) với x, y là các biến tọa độ. Giá trị số ở điểm (x,y) tương ứng với giá trị xám hoặc độ sáng của ảnh (x là các cột còn y là các hàng). Giá trị của hàm ảnh f(x,y) được hạn chế trong phạm vi của các số nguyên dương.

0 ≤ f(x, y ≤ fmax)

Với ảnh đen trắng mức xám của ảnh có thể được biểu diễn bởi một số như sau:

Trong đó SBW( ) là đặc tính phổ của cảm biến được sử dụng và k là hệ số tỷ lệ xích. Vì sự cảm nhận độ sáng có tầm quan trọng hàng đầu đối với ảnh đen trắng nên SBW( ) được chọn giống như là hiệu suất sáng tương đối. Vì f biểu diễn công suất trên đơn vị diện tích, nên nó bao giờ cũng không âm và hữu hạn.

*0≤ f ≤ fmax*

Trong đó fmax là giá trị lớn nhất mà f đạt được. Trong xử lý ảnh, f được chia thang sao cho nó nằm trong một phạm vi thuận lợi nào đó.

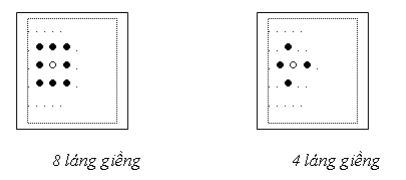
Thông thường đối với ảnh xám, giá trị fmax là 255 ( 28=256) bởi vì mỗi phần tử ảnh được mã hóa bởi một byte. Khi quan tâm đến ảnh màu ta có thể mô tả màu qua ba hàm số: thành phần màu đỏ qua R(x,y), thành phần màu lục qua G(x,y) và thành phần màu lam qua B(x,y). Bộ ba giá trị R, G, và B nhận được từ:

Ở đó SR ( ),SG( ) và SB( ) theo thứ tự là những đặc tính phổ của các cảm biến (bộ lọc) đỏ, lục và lam. R, G, B cũng không âm và hữu hạn.

Ảnh có thể được biểu diễn theo một trong hai mô hình: mô hình Vector hoặc mô hình Raster.

Mô hình Vector: ngoài mục đích tiết kiệm không gian lưu trữ, dễ dàng hiển thị và in ấn, các ảnh biểu diễn theo mô hình vector còn có ưu điểm cho phép dễ dàng lựa chọn, sao chép, di chuyển, tìm kiếm…Theo những yêu cầu này thì kỹ thuật biểu diễn vector tỏ ra ưu việt hơn. Trong mô hình này, người ta sử dụng hướng vector của các điểm ảnh lân cận để mã hóa và tái tạo lại hình ảnh ban đầu. Các ảnh vector được thu nhận trực tiếp từ các thiết bị số hóa như Digitalize hoặc được chuyển đổi từ các ảnh Raster thông qua các chương trình vector hóa.

Mô hình Raster: là mô hình biểu diễn ảnh thông dụng nhất hiện nay. Ảnh được biểu diễn dưới dạng ma trận các điểm ảnh. Tùy theo nhu cầu thực tế mà mỗi điểm ảnh có thể được biểu diễn bởi một hay nhiều bit. Mô hình Raster thuận lợi cho việc thu nhận, hiển thị và in ấn. Các ảnh được sử dụng trong phạm vi của đề tài này cũng là các ảnh được biểu diễn theo mô hình Raster.

Khi xử lý các ảnh Raster chúng ta có thể quan tâm đến mối quan hệ trong vùng lân cận của các điểm ảnh. Các điểm ảnh có thể xếp hàng trên một lưới (raster) hình vuông, lưới hình lục giác hoặc theo một cách hoàn toàn ngẫu nhiên với nhau.

Hình 1.1: Quan hệ trong vùng lân cận giữa các điểm ảnh.

Cách sắp xếp theo hình vuông là được quan tâm đến nhiều nhất và có hai loại: điểm 4 láng giềng (4 liền kề) hoặc 8 láng giềng (8 liền kề). Với điểm 4 láng giềng, một điểm ảnh I(i, j) sẽ có điểm kế cận theo 2 hướng i và j; trong khi đó với điểm 8 láng giềng, điểm ảnh I(i, j) sẽ có 4 điểm kế cận theo 2 hướng i, j và 4 điểm kế cận theo hướng chéo 45o.

#### Các định dạng ảnh thông dụng

##### PNG

Là một dạng hình ảnh sử dụng phương pháp nén dữ liệu mới – không làm mất đi dữ liệu gốc. PNG được tạo ra nhằm cải thiện và thay thế định dạng ảnh GIF với một định dạng hình ảnh không đòi hỏi phải có giấy phép sáng chế sử dụng. PNG được hỗ trợ bởi thư viện tham chiếu lib png, một thư viện nền độc lập bao gồm các hàm của C để quản lý các hình ảnh PNG.

Những tập tin PNG thường có phần mở rộng là PNG và đã được gán kiểu chuẩn MIME là image/png.

Một tập tin PNG bao gồm 8 – byte kí hiệu (89 50 4E 47 0D 0A 1A) được viết trong hệ thống có cơ số 16, chứa các chữ “PNG” và 2 dấu xuống dòng, ở giữa là xếp theo số lượng của các thành phần, mỗi thành phần đều chứa thông tin về hình ảnh. Cấu trúc dựa trên các thành phần được thiết kế cho phép định dạng PNG có thể tương thích với các phiên bản cũ khi sử dụng các “thành phần” trong tập tin.

PNG là cấu trúc như một chuỗi các thành phần, mỗi thành phần chứa kích thước, kiểu, dữ liệu, và mã sửa lỗi CRC ngay trong nó.

Chuỗi được gán tên bằng 4 chữ cái phân biệt chữ hoa chữ thường. Sự phân biệt này giúp bộ giải mã phát hiện bản chất của chuỗi khi nó không nhận dạng được.

Với chữ cái đầu, viết hoa thể hiện chuỗi này là thiết yếu, nếu không thì ít cần thiết hơn ancillary. Chuỗi thiết yếu chứa thông tin cần thiết để đọc được tệp và nếu bộ giải mã không nhận dạng được chuỗi thiết yếu,việc đọc tệp phải được hủy.

Về cơ bản, định dạng PNG đtôi lại cho ta những ưu thế vƣợt trội hơn so với các định dạng phổ thông khác hiện nay như JPG, GIF, BMP…Những ưu thế tỏ rõ sức mạnh hơn khi được sử dụng trong môi trường đồ họa web.

* Giảm thiểu dung lượng: Trong tất cả các định dạng ảnh phổ thông hiện nay thì hình ảnh PNG có thể coi là dung lượng nhỏ nhất. Điều này rất quan trọng khi sử dụng PNG trong môi trường web.
* Độ sâu của màu: Ảnh PNG hỗ trợ đến true color 48 bit màu. Trong khi đóảnh gif chỉ ở mức 256 màu.

##### GIF

Cách lưu trữ kiểu PCX có lợi về không gian lưu trữ: với ảnh đen trắng kích thước tệp có thể nhỏ hơn bản gốc từ 5 đến7 lần. Với ảnh 16 màu, kích thước ảnh nhỏ hơn ảnh gốc 2-3 lần, có trường hợp có thể xấp xỉ ảnh gốc. Tuy nhiên, với ảnh 256 màu thì nó bộc lộ rõ khả năng nén rất kém. Điều này có thể lý giải như sau: khi số màu tăng lên, các loạt dài xuất hiện ít hơn và vì thế, lưu trữ theo kiểu PCX không còn lợi nữa. Hơn nữa, nếu ta muốn lưu trữ nhiều đối tượng trên một tệp ảnh như kiểu định dạng TIFF, đòi hỏi có một định dạng khác thích hợp.

Định dạng ảnh GIF do hãng ComputServer Incorporated (Mỹ) đề xuất lần đầu tiên vào năm 1990. Với định dạng GIF, những vướng mắc mà các định dạng khác gặp phải khi số màu trong ảnh tăng lên không còn nữa. Khi số màu càng tăng thì ưu thế của định dạng GIF càng nổi trội. Những ưu thế này có được là do GIF tiếp cận các thuật toán nén LZW(Ltôipel-Ziv-Welch). Bản chất của kỹ thuật nén LZW là dựa vào sự lặp lại của một nhóm điểm chứ không phải loạt dài giống nhau. Do vậy, dữ liệu càng lớn thì sự lặp lại càng nhiều. Dạng ảnh GIF cho chất lượng cao, độ phân giải đồ họa cũng đạt cao, cho phép hiển thị trên hầu hết các phần cứng đồ hoạ.

Định dạng tổng quát của ảnh GIF như sau:

* Chữ ký của ảnh
* Bộ mô tả hiển thị
* Bản đồ màu tổng thể
* Mô tả một đối tượng của ảnh
* Dấu phân cách
* Bộ mô tả ảnh
* Bản đồ màu cục bộ
* Dữ liệu ảnh

Phần mô tả này lặp n lần nếu ảnh chứa n đối tượng.

* Phần đầu cuối ảnh GIF(terminator)

|  |
| --- |
| GIF Note |
| GIF header(7 bytes) |
| Global palerre |
| Header image(10 bytes) |
| Palerre of image 1 (nếu có) |
| Data of image 1 |
| ‘,’ ký tự liên kết 1 |
| .............................................. |
| ‘:’ GIF ttôiinator |

* Chứ ký của ảnh GIF có giá trị là GIF87a. Nó gồm 6 ký tự, 3 kí tự đầu chỉ ra kiểu định dạng, 3 ký tự sau chỉ ra version của ảnh.
* Bộ hình hiển thị: chứa mô tả các thông số cho toàn bộ ảnh GIF:
* Độ rộng hình raster theo pixel: 2 byte;
* Độ cao hình raster theo pixel: 2 byte;
* Các thông tin về bản đồ màu, hình hiển thị,...
* Thông tin màu nền: 1 byte;
* Phần chưa dùng: 1 byte.
* Bản đồ màu tổng thể: mô tả bộ màu tối ưu đòi hỏi khi bit M = 1. Khi bộ màu tổng thể được thể hiện, nó sẽ xác lập ngay bộ mô tả hình hiển thị. Số lượng thực thể bản đồ màu lấy theo bộ mô tả hình hiển thị ở trên và bằng 2m, với m là lượng bit trên một pixel khi mỗi thực thể chứa đựng 3 byte (biểu diễn cường độ màu của ba màu cơ bản Red-Green-Blue). Cấu trúc của khối này như sau:

|  |  |  |
| --- | --- | --- |
| **Bit** | **Thứ tự bytes** | **Mô tả** |
| Màu Red | 1 | Giá trị đỏ theo index 0 |
| Màu Blue | 2 | Giá trị xanh lục theo index 0 |
| Màu Green | 3 | Giá trị xanh lơ theo index 0 |
| Màu Red | 4 | Giá trị đỏ theo index 1 |
| Màu Blue | 5 | Giá trị xanh lục theo index 1 |
| Màu Green | 6 | Giá trị xanh lơ theo index 0 |
| .................................. | .................................... |  |

* Bộ mô tả ảnh: định nghĩa vị trí thực tế và phần mở rộng của ảnh trong phạm vi không gian ảnh đã có trong phần mô tả hình hiển thị. Nếu ảnh biểu diễn theo ánh xạ bản đồ màu cục bộ thì cờ định nghĩa phải được thiết lập. Mỗi bộ mô tả ảnh được chỉ ra bởi ký tự kết nối ảnh. Ký tự này chỉ được dùng khi định dạng GIF có từ 2 ảnh trở lên. Ký tự này có giá trị 0x2c (ký tự dấu phảy). Khi ký tự này được đọc qua, bộ mô tả ảnh sẽ được kích hoạt. Bộ mô tả ảnh gồm 10 byte và có cấu trúc như sau:

|  |  |  |
| --- | --- | --- |
| Các bit | Thứ tự bytes | Mô tả |
| 0010110 | 1 | Ký tự liên kết ảnh |
| Căn trái ảnh | 2.3 | Pixel bắt đầu ảnh tính từ trái hình hiển thị |
| Căn đỉnh trên | 4.5 | Pixel bắt đầu ảnh tính từ đỉnh trên hình hiển thị |
| Độ rộng ảnh | 6.7 | Chiều rộng ảnh tính theo pixel |
| Độ cao ảnh | 8.9 | Chiều rộng cao tính theo pixel |
| MI000pixel | 10 | Khi bit M = 0: sử dụng bản đồ màu tổng thể  Khi bit M = 0:sử dụng bản đồ màu cục bộ  Khi bit I = 0: định dạng ảnh theo thứ tự liên tục  Khi bit I = : định dạng ảnh theo thứ tự xen kẽ  Khi bit Pixel + 1: số bit /pixel của ảnh này. |

* Bản đồ màu cục bộ: bản đồ màu cục bộ chỉ được chọn khi bit M của byte thứ 10 là 1. Khi bản đồ màu được chọn, bản đồ màu sẽ chiếu theo bộ mô tả ảnh mà lấy vào cho đúng. Tại phần cuối ảnh, bản đồ màu sẽ lấy lại phần xác lập sau bộ mô tả hình hiển thị. Lưu ý là trường “pixel “ của byte thứ 10 chỉ được dùng khi bản đồ màu được chỉ định. Các tham số này không những chỉ cho biết kích thước ảnh theo pixel mà còn chỉ ra số thực thể bản đồ màu của nó.
* Dữ liệu ảnh: chuỗi các giá trị có thứ tự của các pixel màu tạo nên ảnh. Các pixel được xếp liên tục trên một dòng ảnh, từ trái qua phải. Các dòng ảnh được viết từ trên xuống dưới.
* Phần kết thúc ảnh: cung cấp tính đồng bộ cho đầu cuối của ảnh GIF. Cuối của ảnh sẽ xác định bởi kí tự “;” (0x3b).
* Định dạng GIF có rất nhiều ưu điểm và đã được công nhận là chuẩn để lưu trữ ảnh màu thực tế (chuẩn ISO 10918-1). Nó được mọi trình duyệt Web (Web Browser) hỗ trợ với nhiều ứng dụng hiện đại. Cùng với nó có chuẩn JPEG (Joint Photograph Expert Group). GIF dùng cho các ảnh đồ hoạ (Graphic), còn JPEG dùng cho ảnh chụp(Photographic).

##### IMG

Ảnh IMG là ảnh đen trắng, phần đầu file IMG có 16 bytes chứa các thông tin cần thiết:

* 6 bytes đầu dùng để đánh dấu nhận dạng file IMG. Giá trị của 6 bytes đầu này viết dưới dạng hexa: 0x0001 0x0008 0x0001.
* 2 bytes chứa độ dài các mẫu tin. Đó là độ dài của một dãy các bytes lặp lại một số lần nào đó, số lần lặp này sẽ được lưu trong một file đếm. Nhiều dãy giống nhau đượclưu trong một bytes. Đó chính là cách lưu trữ nén
* 4 bytes tiếp theo mô tả kích cỡ của pixel
* 2 bytes tiếp mô tả số pixel trên một dòng
* 2 bytes cuối cho biết số dòng trong ảnh

Các dòng giống nhau được nén thành một pack. Có 4 loại pack sau:

* Loại 1: Gói các dòng giống nhau.Quy cách gói tin này 0x00 0x00 0xFF Count. 3 bytes đầu cho biết số các dãy giống nhau ,bytes cuối cho biết số các dòng giống nhau.
* Loại 2: Gói các dãy giống nhau. Quy cách gói này 0x00 Count. Bytes thứ hai cho số các dãy giống nhau được nén trong gói. Độ dài cử dãy được ghi đầu file.
* Loại 3:Dãy các pixel không giống nhau, không lặp lại và không nén được. Quy cách như sau: 0x80 Count. Bytes thứ hai cho biết độ dài dãy các pixel không giống nhau, không nén được.
* Loại 4:Dãy các pixel giống nhau. Tuỳ theo các bit cao của bytes đầu được bật hay tắt, nếu bit cao được bật (giá trị 1) thì đây là gói nén các bytes chỉ gồm bit 0, số các bytes được nén được tính bởi 7 bit thấp còn lại. Nếu bit cao tắt (giá tri 0) thì đay là gói nén các bytes toàn bit 1. Số các bytes được nén được tính bởi 7 bit thấp còn lại. Các cảu file IMG phong phú như vậy là do ảnh IMG là ảnh đen trắng nên chỉ cần 1 bit cho một pixel thay vì 4 hoặc 8 bit như đã nói ở trên toàn bộ ảnh chỉ có điểm sáng hoặc tối tương ứng với 1 hoặc 0. Tỷ lệ nén của file này là khá cao.

### Tổng quan về video

Video là sự tái tạo ảnh tự nhiên theo không gian và thời gian hoặc cả hai, thực chất là một dãy ảnh liên tục theo thời gian nhằm mô phỏng sự chuyển động.

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

#### Video tương tự

**NTSC Video**: Đây là dạng Video tương tự với 525 dòng trên một khung hình, 30 khung hình trong một giây, quét cách dòng, chia làm hai trường (mỗi trường 262.5 dòng), có 20 dòng dự trữ cho thông tin điều khiển tại thời điểm bắt đầu mỗi trường.

**PAL Video**: Dạng Video này có 625 dòng trên một khung hình, 25 khung hình trong một giây, quét cách dòng. Khung gồm hai trường chẵn lẽ, mỗi trường bao gồm 312.5 dòng.

#### Video số

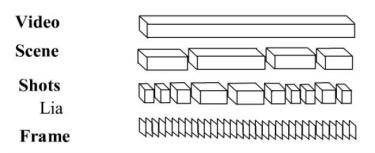
Một số tiêu chuẩn của Video số lấy theo tiêu chuẩn CCIR (Consultative Committee for International Radio)

Bảng 1.1: Các tiêu chuẩn của Video số

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | CCIR  601525/6  NTSC | CCIR  601625/50  PAL/SECAM | CFI | QCFI |
| Độ phân giải  độ chói | 720x480 | 360x576 | 352x288 | 176x144 |
| Độ phân giải  màu sắc | 360x480 | 360x576 | 176x144 | 88x72 |
| Lấy mẫu màu | 4:2:2 | 4:2:2 | 4:2:o | 4:2:0 |
| Số trường /s | 60 | 50 | 30,15,10,7.5 | 30,15,10,7.5 |
| Cách quét | Cách dòng | Cách dòng | Liên tục | Liên tục |

#### Cấu trúc của video

Video là tập hợp các khung hình, mỗi khung hình là một ảnh. Shot là đơn vị cơ sở của video. 1 lia là một đơn vị vật lý của dòng video, gồm các chuỗi các khung hình liên tiếp, không thể chia nhỏ hơn, ứng với một thao tác camera đơn. Sciene là đơn vị logic của dòng video, một cảnh gồm các lia liên quan về không gian và thời gian, cùng miêu tả một nội dung ngữ nghĩa hoặc một tình tiết.



Hình 1.2: Cấu trúc phân đoạn video

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

**Frame:** Video được tạo nên bởi chuỗi các ảnh tĩnh. Một chuỗi các khung hình tĩnh đặt cạnh nhau tạo nên các cảnh phim chuyển động. Một khung đơn là một ảnh tĩnh.

Để đoạn video có thể tạo cảm giác chuyển động, các khung hình phải được quay với tốc độ phù hợp. Mắt người chỉ có thể nhận được 24 hình/giây, nếu như lần lượt 24 hình hoặc nhiều hơn 24 hình được phát trong một giây thì mắt sẽ không nhận ra được sự rời rạc giữa những khung hình, mà chỉ thấy những cảnh liên tục. Có nhiều hệ video và mỗi hệ có tốc độ quay khác nhau như : NTSC 30 hình/giây, PAL 24 hình/giây, SECAM 29.99 hình/giây.

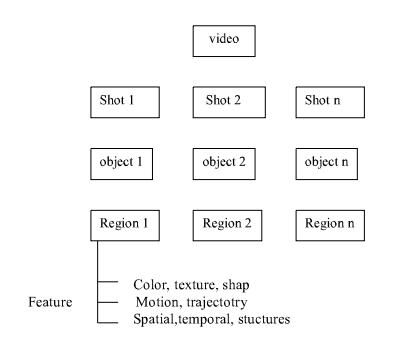
Khung hình là đơn vị cơ bản nhất của dữ liệu video. Theo chuẩn của hệ NTSC thì một giây có 30 khung hình, vậy một phút có 1800 khung hình, một giờ có 60x1800 = 108000 khung hình. Có thể thấy rằng số lượng khung hình cho một đoạn video thường là rất lớn, cần phải có một đơn vị cấp cao hơn cho video số.

Shot : là đơn vị cơ sở của video. Một lia là một đơn vị vật lý của dòng video, gồm các chuỗi một hay nhiều khung hình liên tiếp, không thể chia nhỏ hơn, ứng với một thao tác camera đơn.

Những khung hình này liên tiếp theo thời gian mô tả một hành động liên tục, và được giới hạn bởi 2 chuyển cảnh.

Một đoạn video có thể có nhiều đoạn cơ sở, mà cũng có thể chỉ là một đoạn cơ sở. Những đoạn cơ sở đại diện cho toàn bộ đoạn video, và truy xuất đến chúng cũng coi như là truy xuất đến chính đoạn video. Tổ chức những đoạn video ở mức độ đoạn cơ sở là thích hợp nhất cho việc duyệt và truy tìm thông tin dựa vào nội dung.

**Scene** : là các đơn vị logic của dòng video, một cảnh gồm một hay nhiều shot liên quan đến nhau về không gian và liền kề về thời gian, cùng mô tả một nội dung ngữ nghĩa hoặc một tình tiết.

Hình 1.3: Biểu diễn video dựa trên đối tượng

### Một số kỹ thuật phá hiện chuyển động

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

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

Bảng 1.2: Bảng phân loại các thuật toán phát hiện đối tượng

|  |  |
| --- | --- |
| Loại | Những nghiên cứu có liên quan |
| Point detectors | 1. Moravec’s detector  2. Harris detector  3. Scale Invariant Feature Transform Affine  4. Invariant Point Detector |
| Segmentation | 1. Mean-shift  2. Graph-cut  3. Active contours |
| Background Modeling | 1. Mixture of Gaussians  2. Eigenbackground  3. Wall flower  4. Dynamic texture background |
| Supervised Classifier | 1. Support Vector Machines  2. Neural Networks  3. Adaptive Boosting |

Việc lựa chọn phương pháp áp dụng phải dựa vào tình huống cụ thể, đối với trường hợp có ảnh nền không thay đổi việc phát hiện đối tượng chuyển động có thể bằng các phương pháp trừ nền. Các giải thuật này sẽ được trình bày sau đây. Hướng giải quyết là xây dựng mô hình nền, sau đó sử dụng mô hình này cùng với frame hiện tại để rút ra được các foreground chuyển động. Để có thể tiếp cận cần phải xây dựng được mô hình background. Có nhiều phương pháp xây dựng mô hình background bởi các tác giả: Anurag Mittal dùng adaptive kernel density estimation . Kết quả tốt tuy nhiên khó khăn về không gian lưu trữ, tính toán phức tạp, tốc độ không đáp ứng thời gian thực. Haritaoglu dùng giải thuật W4, Stauffer sử dụng Mixture of Gaussian để xây dựng mô hình nền… Nhằm phát hiện được các đối tượng chuyển động, xác định xem những đối tượng này có đúng là những đối tượng ta cần phát hiện hay không. Đây là các khó khăn cần khắc phục.

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

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

#### Phương pháp sử dụng ngưỡng hình ảnh

Phương pháp sử dụng ngưỡng hình ảnh là phương pháp phát hiện đối tượng đơn giản. Trong kỹ thuật này, người ta cho rằng các đối tượng chuyển động và nền có màu sắc khác nhau. Ritter và Wilson đưa ra phương pháp ngưỡng hình ảnh để phân loại các điểm ảnh là đối tượng hoặc nền. Mỗi điểm ảnh trong khung hình được phân loại là đối tượng hoặc nền. Nếu giá trị màu của một điểm ảnh nằm trong giá trị ngưỡng màu đưa ra, thì sẽ được gán giá trị nhị phân 1 trái lại nó được coi như một điểm ảnh nền và được gán giá trị 0.

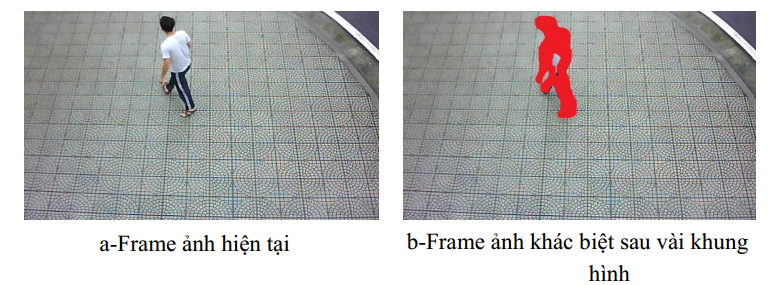
Thay vì sử dụng một giá trị ngưỡng toàn cục, thì phương pháp tích lũy, sử dụng ngưỡng cục bộ hoặc ngưỡng động thường được sử dụng.

Trong hầu hết các tình huống thực tế, các đối tượng và hình nền được chia sẻ nhiều màu sắc phổ biến, điều này gây ra khó khăn để chọn giá trị ngưỡng tối ưu để tách đối tượng từ nền. Su và Amer tập trung vào hai loại ngưỡng (ước lượng sự phân bố các vùng ảnh thay đổi do vị trí không gian màu) và đề xuất một thuật toán phi tham số để tính toán ngưỡng toàn cục. Phương pháp này cải thiện được việc phát hiện đối tượng. Petrou và Bosdogianni sử dụng biểu đồ màu nhằm tìm ra các giá trị ngưỡng để tách các đối tượng từ nền.

Ngưỡng hình ảnh là một phương pháp hữu ích cho phát hiện đối tượng chuyển động, khi các đối tượng và màu nền là rất khác nhau.

#### Phương pháp chênh lệch tạm thời

Phương pháp chênh lệch tạm thời hoặc khác biệt khung hình thực hiện việc phát hiện vùng chuyển động bằng cách tính toán sự khác nhau ở mức điểm ảnh giữa hai hoặc ba khung hình liên tiếp trong một chuỗi các khung hình video, chỉ những điểm ảnh có biến động đáng kể (lớn hơn một ngưỡng quy định) được đánh dấu là điểm ảnh của đối tượng chuyển động và điểm ảnh còn lại được phân loại là ảnh nền. Phương pháp này khá tốt đối với khung cảnh động, có nhiều thay đổi song nó lại có nhược điểm trong việc phát hiện ra các điểm ảnh có liên quan đến một số kiểu chuyển động đặc biệt của đối tượng trong cảnh quay.

Hình 1.4: Ví dụ phương pháp khác biệt thời gian

Phương pháp chênh lệch tạm thời được thực hiện qua công thức :

|I\_t (x,y)-I\_(t-1) (x,y)| > τ

Trong đó I\_t (x,y) là điểm ảnh có tọa độ (x,y) của khung hình tại thời điểm t, I\_(t-1) (x,y) là điểm ảnh có tọa độ (x,y) của khung hình tại thời điểm t-1. τ là ngưỡng được định nghĩa trước đó.

Verbeke và Vincent thay vì tính toán sự khác biệt ngay ở khung hình trước đó, họ tính toán sự khác biệt với mười khung hình đã tích lũy trước đó, kết hợp với phương pháp phân tích thành phần chính để giảm kích thước dữ liệu, kỹ thuật họ đề xuất được đánh giá tốt hơn so với kỹ thuật khác biệt khung hình đơn giản. Tuy nhiên với các điều kiện ánh sáng thay đổi, nhiễu và đặc biệt là hiện tượng dừng tạm thời của đối tượng thì kỹ thuật vẫn chưa giải quyết được. Như các kỹ thuật khác biệt khung hình khác, kỹ thuật này cũng không tách được đối tượng trên nền chuyển động.

#### Phương pháp trừ nền

Nhận dạng đối tượng chuyển động từ chuỗi video là nên tảng và tác vụ quan trọng trong giám sát video, kiểm tra và phân tích giao thông, dò tìm và theo dõi mỗi con người, nhận dạng cử chỉ trong giao diện người máy . Phương pháp chung để nhận dạng đôi tượng chuyển động là trừ nền, ở đây mỗi khung hình video được so sánh với mỗi mo hình tham chiếu hoặc mô hình nền. Các pixel trong frame hiện thời mà lệch đang kể so với nền sẽ được xem xét như là đối tượng chuyển động. Những pixel này được xử lý tiếp cho định vị và theo dõi đối tượng. Từ nền là bước đầu tiên và quan trọng trong nhiều ứng dụng thị giác máy tính, nó được sử dụng để trích chọn ra những điểm tương ứng đúng với những đối tượng chuyển động quan tâm. Mặc dù nhiều giải thuật trừ nền được đề xuất, những vấn đề của nhận dạng đối tượng chuyển động trong những môi trường phức tạp vẫn còn chưa được giải quyết hoàn toàn.

Có một vài vấn đề mà một giải thuật trừ nền tốt phải giải quyết phù hợp. Xét mọi chuỗi video từ camera theo dõi cố định ở một nút giao thông. Đây là một môi trường ngoài trời, do đó một giải thuật trừ nền phải thích nghi với nhiều mức khác nhau với độ sáng tại những thời điểm khác nhau trong ngày và xử lý điều kiện thời tiết bất lợi như sương mù hoặc tuyết làm thay đổi nền. Việc thay đổi bóng, đổ bóng bởi sự di chuyển của đối tượng cần phải được loại bỏ để những đặc trưng thích hợp có thể được trích chọn từ những đối tượng trong quá trình xử lý tiếp theo. Luồng giao thông phức tạp tại chỗ giao nhau luôn đặt ra thách thức mới cho giải thuật trừ nền. Xe di chuyển tốc độ bình thường khi đèn xanh, nhưng dừng lại khi đèn bật đỏ. Xe cộ đứng yên không chuyển động cho đến khi đèn xanh bật lại. Một giải thuật trừ nền tốt nhưng phải xử lý những đối tượng chuyển động mà ban đầu nó hòa trộn với nền và sau đó nó không phải nền. Ngoài ra, nhiều ứng dụng cần phải xử lý thời gian thực, nên giải thuật trừ nền phải tính toán nhanh và có yêu cầu bộ nhớ ít, trong khi vẫn có khả năng xác định những đối tượng chính xác trong video.

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

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

Bt+1= αIt + (1- α)Bt

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

- Ảnh nền

- Ảnh chứa đối tượng

- Khoanh vùng đối tượng sau khi trừ nền

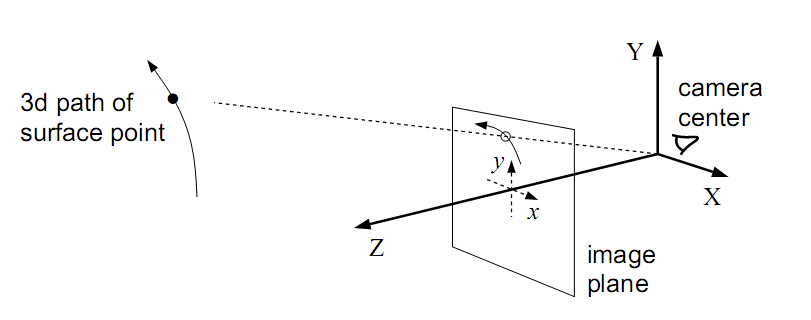
Hình 1.5: Mô hình hoạt động của phương pháp trừ nền

Vì kỹ thuật này khá đơn giản, dễ cài đặt và có thể thực hiện trong thời gian thực nên được sử dụng rất phổ biến. Mặc dù kỹ thuật trừ thực hiện rất tốt việc tác được hầu hết của các điểm ảnh có nằm trong vùng di chuyển ngay cả khi đối tượng dừng lại, nhưng lại thường nhạy cảm với những thay đổi đột ngột, ví dụ, đối tượng đứng yên ra khỏi nền (ví dụ như một chiếc xe đậu di chuyển ra khỏi bãi đậu xe hoặc ánh sáng bị thay đổi đột ngột.

#### Phương pháp luồng quang học

Phương pháp dựa vào luồng quang học (optical flow) thực hiện bằng cách sử dụng các vector có hướng của các đối tượng chuyển động theo thời gian để phát hiện các vùng chuyển động trong một ảnh. Chúng có thể phát hiện chuyển động trong các dãy video ngay cả các video thu được từ camera di chuyển.

Luồng quang học là khái niệm chỉ sự chuyển động tương đối của các điểm trên bề mặt một đối tượng, vật thể nào đó gây ra, dưới góc quan sát của một điểm mốc (mắt, camera…). Sự chuyển động của các vật thể (mà thực tế có thể coi là sự chuyển động của các điểm trên bề mặt của vật thể ấy) trong không gian 3 chiều, khi được chiếu lên một mặt phẳng quan sát 2D được gọi là trường chuyển động (motion field). Nói chung, mục đích của các phương pháp luồng quang học là để xác định (xấp xỉ) trường chuyển động từ một tập các khung hình thay đổi theo thời gian, chúng được sử dụng rộng rãi trong các bài toán phân đoạn đối tượng, phát hiện đối tượng chuyển động, theo vết đối tượng…



Hình 1.6: Luồng quang học

Có nhiều kĩ thuật xác định luồng quang học khác nhau, báo cáo sẽ tập trung vào phương pháp dựa trên gradient, chi tiết các phương pháp khác có thể tham khảo khảo sát của Raveshiya .

Ý tưởng quan trọng của phương pháp tính toán luồng quang học dựa trên gradient là giả định sau:

Bề ngoài của đối tượng không có nhiều thay đổi (về cường độ sáng) khi xét từ frame thứ n sang frame n+1. Nghĩa là:

|  |  |  |
| --- | --- | --- |
|  | (1.1) |  |

Trong đó I(x ⃗,t) là hàm trả về cường độ sáng của điểm ảnh tại thời điểm t (khung hình thứ t). là tọa độ của điểm ảnh trên bề mặt (2D), là vector vận tốc, thể hiện sự thay đổi vị trí của điểm ảnh từ frame thứ t sang frame t+1. Mặc dù có vẻ không thực tế, nhưng nhiều công trình cho thấy giả sử (1.2) làm việc khá tốt trong thực tế.

Sử dụng khai triển Taylor, ta xấp xỉ:

|  |  |  |
| --- | --- | --- |
|  | (1.2) |  |

với ∇I(x ⃗,t), và I\_t (x ⃗,t) là đạo hàm tức thời của I tại thời điểm t.

Bỏ qua các phần tử có bậc đạo hàm cao hơn trong (1.3), so sánh (1.2) và (1.3), đồng thời xấp xỉ bằng , ta nhận được phương trình sau, gọi là phương trình ràng buộc gradient (gọi là phương trình luồng quang học):

|  |  |  |
| --- | --- | --- |
|  | (1.3) |  |

Phương trình (1.4) còn có thể đạt được bằng cách xuất phát ngay từ giả thiết ban đầu là giá trị cường độ sáng của điểm ảnh không thay đổi từ frame t sang t+1, như vậy:

|  |  |  |
| --- | --- | --- |
|  | (1.4) |  |

Phương trình (1.4) là phương trình cơ bản của luồng quang học, trong đó chính là luồng cần tìm. Tuy nhiên phương trình này có dạng với 2 ẩn là u\_1 và u\_2 nên không giải được. Ta có nhận xét rằng phương trình (4), với ẩn u ⃗=(u\_1,u\_2 ), là phương trình đường thẳng trong 2D với vector pháp tuyến là ∇I.

Mọi phương pháp ước lượng luồng quang học đều xoay quanh việc cố gắng tìm thêm một phương trình nữa để tính u ⃗.

Để tìm thêm một ràng buộc nữa cho u ⃗, ta có thể dùng đến các điểm ảnh lân cận với điểm ảnh đang xét, với giả sử là các điểm ảnh lân cận này cũng nằm trên vật thể và chuyển động với cùng vận tốc của điểm ảnh đang xét. Trong thực tế, các điểm ảnh lân cận có thể không chuyển động cùng vận tốc với điểm ảnh đang xét, do đó ta tìm vector vận tốc mà làm cực tiểu bình phương lỗi:

|  |  |  |
| --- | --- | --- |
|  | (1.5) |  |

Phương pháp này gọi là ước lượng bình phương tối thiểu (Least-squares estimation), trong đó g(u ⃗ ) xác định “độ rộng” mà hàm ước lượng này bao phủ. Thông thường người ta chọn hàm Gaussian. Vận tốc cần tìm là giá trị u ⃗ làm cực tiểu E(u ⃗ ).

Người ta chứng minh được E(u ⃗ ) đạt cực tiểu tại điểm đạo hàm riêng của nó bằng zero, nghĩa là:

|  |  |
| --- | --- |
|  | (1.6) |
|  | (1.7) |

Hai phương trình này có thể viết ở dạng ma trận:

|  |  |
| --- | --- |
|  | (1.8) |

Với :

|  |  |
| --- | --- |
|  | (1.9) |

Nghiệm của (1.9) là 

Các phương pháp luồng quang học rất phù hợp với tình huống camera di chuyển, tuy nhiên các phương pháp này luôn đòi hỏi sự tính toán khá phức tạp và khá nhạy với nhiễu.

#### Phương pháp thống kê

Một phương pháp tiên tiến hơn là việc sử dụng các đặc tính thống kê của các điểm ảnh riêng lẻ đã được phát triển để khắc phục những thiếu sót của phương phá trừ nền cơ bản. Những phương pháp thống kê chủ yếu được lấy tư tưởng từ các phương pháp trừ nền trong việc lưu giữ và tự động cập nhật các số liệu thống kê củ các điểm ảnh thuộc quá trình xử lý nền. Các điểm ảnh foreground được xác định bằng cách so sánh số liệu thống kê của mỗi điểm ảnh với mô hình nền. Cách tiếp cận này đang trở nên phổ biến hơn do độ tin cậy của nó trong những cảnh chứa các thay đổi tiếng ồn chiếu sáng, và bóng tối.

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

|M(x,y) – It(x,y)| > D(x,y) or |N(x,y) – It(x,y)| > D(x,y)

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

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

# Chương II

# THỰC NGHIỆM

**Nội dung chính**

* Môi trường thực nghiệm
* Xây dựng ứng dụng
* Kết quả thực nghiệm

## Môi trường cài đặt

* **Môi trường thực nghiệm**
* Các môi trường thông thường như: nhà ở, phòng học, thang máy với các yếu tố ánh sáng vừa đủ.
* Không có khói bụi nhiều.
* Người đi lại không mang đồ đạc kích thước lớn.
* Không có cây cối trong khu vực giám sát.
* Ảnh nền không có nhiều vật chuyển động mạnh như quạt trần.
* Màu nền không có quá nhiều màu sắc và vùng màu.
* **Thiết bị thực nghiệm**
* **Thiết bị ghi hình hoặc file video:**
* Thiết bị ghi hình: độ phân giải > = 3.0 Mps.
* File video : Kích thước lớn hơn 480x600 pixel.
* **Cấu hình máy**
* CPU tối thiểu: 1.8 GHz.
* RAM tối thiểu 4Gb.
* **Khoảng cách thực nghiệm**
* Thiết bị ghi hình cách mặt đất không quá 4m.
* Không dưới 2m.
* Góc quay từ 45-90 độ. (tùy thuộc vào tiêu cự của thiết bị ghi hình – Khoảng cách đề nghị trên thông dụng cho các camera an ninh tại trường học).

## Xây dựng ứng dụng

## Kết quả thực nghiệm

# KẾT LUẬN

Phát hiện chuyển động của đối tượng trong đoạn video là một đề tài với nhiều ứng dụng trong các lĩnh vực đời sống đặc biệt là ngành đảm bảo trật tự an ninh. Trong báo cáo này tôi đã trình bày một số phương pháp phát hiện đối tượng chuyển động cùng các ưu nhược điểm của từng phương pháp khi áp dụng chúng. Tùy từng yêu cầu cụ thể về thời gian, về địa điểm và chất lượng của video cần xử lý mà chúng ta chon sử dụng phương pháp phát hiện thích hợp.

Bên cạnh việc đưa ra một số cải tiến cho phương pháp phát hiện đối tượng chuyển động, báo cáo cũng đã cài đặt thử nghiệm thuật toán trừ nền và sự khác biệt thời gian cho bài toán phát hiện đối tượng chuyển động.

Những mục tiêu đạt được:

* Tìm hiểu các phương pháp phát hiện đối tượng chuyển động trong video.
* Cài đặt thử nghiệm một số phương pháp.
* Phân tích ưu nhược điểm của từng phương pháp.

Những mục tiêu chưa đạt được:

Hướng phát triển trong tương lai:

Tuy nhiên do hạn chế về điều kiện và thời gian , báo cáo sẽ không tránh khỏi những thiếu sót. Kính mong được sự đóng góp ý kiến của thầy cô và các bạn để đề tài nghiên cứu của tôi sẽ được hoàn thiện hơn.

# TÀI LIỆU THAM KHẢO

###### Ths. Lương Mạnh Bá, Pts. Nguyễn Thanh Thủy, Nhập Môn Xử Lý Ảnh Số, Hà Nội – 1999.

###### <https://vi.wikipedia.org/wiki/Avi>

###### Yi githan Dedeo glu, Moving Object Detection, Tracking And Classification For Smart Video Surveillance , August- 2004.

###### Alper Yilmaz, Omar Javed, and Mubarak Shah, "Object Tracking: A Survey" pp. 7-15.

###### Anurag Mittal and Mikos Paragios, "Motion Based Background Subtraction using Adaptive Kernel Density Estimation" pp. 302-309, 2004.

###### C.Stauffer and W.Grimson, "Adaptive Background mixture models for Realtime tracking" pp. 750-755, 2009.

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