

**ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ**

TRẦN THỊ HOÀN

**NGHIÊN CỨU MỘT SỐ PHƯƠNG PHÁP PHÁT
HIỆN CHUYỂN ĐỘNG TRONG VIDEO VÀ ỨNG
DỤNG**

Ngành: Công nghệ thông tin
Chuyên ngành: Công nghệ phần mềm
Mã số: 60 48 10

TÓM TẮT LUẬN VĂN THẠC SĨ

MỞ ĐẦU

Giám sát tự động là một hướng mới và có nhiều triển vọng trong sự phát triển tiếp theo của 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 PGS.TS. Đỗ Năng Toàn tôi đã tiến hành nghiên cứu đề tài ***“Nghiên cứu một số phương pháp phát hiện đối tượng chuyển động trong video và ứng dụng”***

Các vấn đề cần giải quyết trong phạm vi luận văn này 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, phân loại các đối tượng chuyển động và tạo kho cơ sở dữ liệu mẫu.
- Cài đặt một số phương pháp phát hiện, đánh dấu các đối tượng chuyển động dựa theo các phương pháp đã nêu ở trên.

Chương 1 - TỔNG QUAN VỀ XỬ LÝ VIDEO VÀ PHÁT HIỆN ĐỐI TƯỢNG CHUYỂN ĐỘNG

1.1. Tổng quan về xử lý Video

1.1.1. Sơ lược 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.

1.1.2. Các dạng Video

1.1.2.1 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 lẻ, mỗi trường bao gồm 312.5 dòng.

1.1.2.2 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 601525/60 NTSC	CCIR 601625/50 PAL/SECAM	CIF	QCIF
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	4:2:0	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

1.1.3. Các chuẩn video

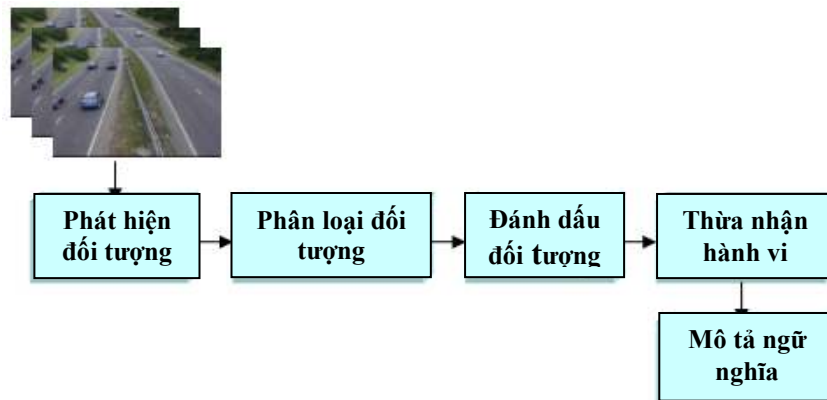
Chuẩn MJPEG	Chuẩn H.120
Chuẩn Video MPEG-1	Chuẩn Video H.261
Chuẩn Video MPEG-2	Chuẩn Video H.263
Chuẩn Video MPEG-4	Chuẩn Video H.263
Chuẩn Video MPEG-7	Chuẩn Video H.264

1.2. Một số vấn đề cơ bản trong xử lý Video

1.2.1. Phát hiện đối tượng chuyển động trong Video

Các ứng dụng dựa trên xử lý video đều có đặc điểm chung, đó là các đối tượng chuyển động.

Các công nghệ thường xuyên được sử dụng để phát hiện đối tượng chuyển động là phép trừ nền, các phương pháp tĩnh, sự khác biệt theo thời gian và optical flow,...



Sơ đồ 1.1: Sơ đồ chung cho các thuật toán xử lý video

1.2.1.1. Phép trừ nền

Phép trừ nền là một phần trong công nghệ phân đoạn chuyển động trong các cảnh tĩnh. Nó cố gắng phát hiện các vùng chuyển động bằng cách trừ điểm ảnh cho điểm ảnh từ ảnh hiện thời đến một ảnh nền cơ sở đã được tạo bởi trung bình các ảnh nền trong một khoảng thời gian của một chu kỳ khởi tạo.

Có một số hướng tiếp cận đã được trình bày như sau.

Heikkila và Silven sử dụng một cách đơn giản phép phối hợp này, với mỗi điểm ảnh được biểu diễn bởi một tọa độ (x,y) trong ảnh hiện thời I_t sẽ được đánh dấu là điểm nổi bật nếu

$$|I_t(x,y) - B_t(x,y)| > \tau \quad (1.1)$$

được thỏa mãn, trong đó, τ là ngưỡng xác định trước. Ảnh nền B_T được cập nhật bởi việc sử dụng phép lọc Infinite Impulse Response (IIR) như sau :

$$B_{t+1} = \alpha I_t + (1 - \alpha) B_t \quad (1.2)$$

1.2.1.2. Các phương pháp tính

Phương pháp W4 sử dụng một mô hình nền tĩnh, ở đó mỗi pixel được biểu diễn lại với giá trị nhỏ nhất (M) và giá trị lớn nhất (N) của nó về cường độ và sự sai khác cường độ lớn nhất (D) giữa các frame liên tiếp bất kỳ quan sát được trong suốt chu kỳ khởi tạo huấn luyện mà ở đó cảnh không chứa các đối tượng chuyển động. Một điểm trong ảnh hiện thời I_t được phân loại như là điểm nổi trội nếu nó thỏa mãn:

$$|M(x,y) - I_t(x,y)| > D(x,y) \text{ hoặc } |N(x,y) - I_t(x,y)| > D(x,y) \quad (1.3)$$

Stauffer và Grimson đã miêu tả một sự tương thích của mô hình pha trộn nền để đánh dấu thời gian thực. Mọi điểm ảnh thực sự được mô hình hóa bởi sự pha trộn của phương pháp Gaussians và được cập nhật trực tuyến bởi dữ liệu ảnh đầu vào. Các phân bố Gaussian sẽ đánh giá xem một điểm sẽ thuộc về việc xử lý điểm nổi trội hay là thuộc về xử lý nền.

1.2.1.3. Sự khác biệt theo thời gian

Sự khác biệt theo thời gian cố gắng phát hiện các vùng chuyển động bằng cách sử dụng sự khác nhau giữa một điểm ảnh ở các frame liên tiếp nhau (hai hoặc ba) trong một dãy video. Phương pháp này có khả năng thích ứng cao với các cảnh động, tuy nhiên, nó thường mắc một số lỗi trong việc phát hiện đối tượng chuyển động.

Vùng màu đơn của người bên ảnh trái làm cho thuật toán khác biệt về thời gian bị lỗi trong việc trích ra tất cả các điểm ảnh của vùng chuyển động của con người.

Lipton đã trình bày cách phối hợp giữa hai frame khác nhau, ở đó, các pixel thỏa mãn các biểu thức sau :

$$|I_t(x,y) - I_{t-1}(x,y)| > \tau \quad (1.4)$$

Để khắc phục các thiếu sót của hai frame khác nhau, trong một số trường hợp, ba frame khác nhau có thể được sử dụng.

1.2.1.4. Optical Flow

Phương pháp 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.

1.2.1.5. Phát hiện sự thay đổi ánh sáng và bóng

Horprasert đưa ra một cách giải quyết về phép trừ nền và phương pháp phát hiện bóng, với phương pháp đó, mỗi pixel được diễn tả bởi một mô hình màu phân biệt bởi độ sáng từ các tín hiệu chứa màu sắc thành phần. Mỗi điểm ảnh đã định rõ được phân chia vào bốn loại khác nhau (nền, nền bóng hoặc bóng, ánh sáng nền hoặc đối tượng chuyển động nổi trội) bằng cách tính toán sự chênh lệch của màu sắc và các tín hiệu chứa màu sắc giữa các điểm ảnh nền và ảnh hiện thời.

1.2.2. Phân loại đối tượng trong video

Vùng chuyển động được phát hiện trong video có thể tương ứng với các đối tượng khác nhau trong thế giới thực như người, nhóm người, xe cộ, sự hỗn loạn,... Việc thừa nhận kiểu của đối tượng là rất quan trọng để xác nhận kiểu của đối tượng được phát hiện để đánh dấu nó một cách chính xác và phân tích các hoạt động một cách đúng đắn.

1.2.2.1. Phân loại dựa theo hình dạng

Các đặc trưng chung chung sử dụng trong phân chia các đối tượng theo hình dạng là tạo các hình chữ nhật bao quanh, tạo các vùng, hình chiếu và độ nghiêng của các vùng chứa đối tượng được phát hiện.

Phương pháp này phụ thuộc vào các giả định, chẳng hạn như, nếu là con người thì sẽ nhỏ hơn xe cộ và có bóng phức tạp. Mức độ rải rác được dùng như là một ma trận phân loại và nó được định nghĩa trong các giới hạn của các vùng của đối tượng và độ dài đường viền (chu vi) như sau:

$$\text{Mức độ rải rác} = \frac{\text{chu vi}^2}{\text{vùng}} \quad (1.5)$$

1.2.2.2. Phân loại dựa theo chuyển động

Các phương pháp này dùng để phân biệt các đối tượng cứng (như xe cộ) và không cứng (như người). Phương pháp này dựa trên cơ sở tính chất đặc biệt theo thời gian của các đối tượng chuyển động. Với đối tượng được đưa ra ở chu kỳ chuyển động, đặc tính của nó đo được cũng như hiển thị một chu kỳ chuyển động. Phương pháp này khai thác đầu mối này để phân loại các đối tượng chuyển động sử dụng chu kỳ.

1.2.3. Phát hiện ánh sáng

Liu và Ahuja đưa ra một phương pháp định nghĩa về các mô hình quang phổ, không gian và thời gian của các điểm ánh sáng để phát hiện ra sự có mặt của nó trong video.

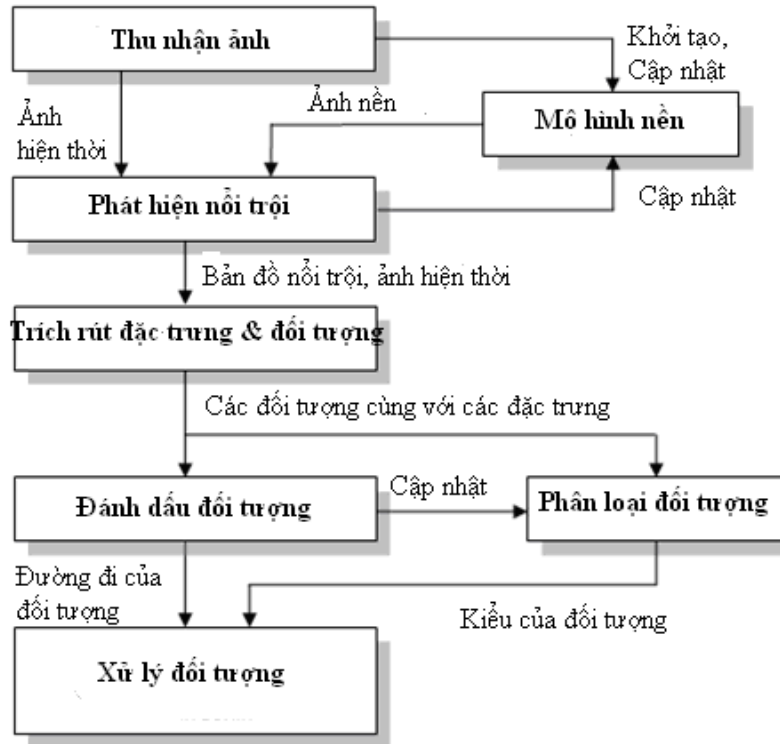
1.2.4. Đánh dấu đối tượng

Đánh dấu trong video có thể được phân biệt theo sự cần thiết của các ứng dụng sử dụng nó hoặc theo các phương pháp sử dụng giải pháp của nó.

Chương 2 - PHÁT HIỆN, ĐÁNH DẤU VÀ PHÂN LOẠI ĐỐI TƯỢNG CHUYỂN ĐỘNG TRONG VIDEO

2.1. Phát hiện đối tượng

Tổng quan về phát hiện, phân loại và đánh dấu đối tượng chuyển động trong video với thời gian thực được mô tả như sau



Sơ đồ 2.1: Sơ đồ khối hệ thống

Hệ thống này có thể phân biệt được các đối tượng thoáng qua hoặc đứng yên từ các đối tượng nền tĩnh trong các cảnh động; phát hiện và phân biệt các đối tượng di chuyển và biến mất; phân loại các đối tượng đã được phát hiện vào trong các nhóm khác nhau, như người, nhóm người, xe cộ,...; đánh dấu các đối tượng và tạo ra thông tin về đường đi ngay cả các trường hợp bị che khuất và phát hiện ánh sáng trong hình ảnh của video.

2.1.1. Phát hiện nổi trội

2.1.1.1 Mô hình trừ nền có khả năng thích ứng

Gọi $I_n(x)$ là biểu diễn của giá trị cường độ mức xám ở điểm ảnh có vị trí (x) và ở trường hợp thứ n của dãy video I thuộc trong đoạn $[0,255]$. Gọi $B_n(x)$ là giá trị cường độ nền tương ứng cho điểm ảnh ở vị trí (x) ước lượng theo thời gian từ ảnh video I_0 đến I_{n-1} . Một điểm ảnh ở vị trí (x) trong ảnh hiện thời thuộc thành phần nổi trội nếu nó thỏa mãn

$$|I_n(x) - B_n(x)| > T_n(x) \quad (2.1)$$

Trong đó $T_n(x)$ là giá trị ngưỡng có khả năng thích hợp được khởi tạo cùng với ảnh video đầu tiên I_0 , $B_0 = I_0$, và ảnh ngưỡng được khởi tạo bởi giá trị đã được xác định trước.

Nền cơ sở và các ảnh ngưỡng phải được cập nhật liên tục từ các ảnh đầu vào. Sự phối hợp cập nhật này là khác nhau đối với các vị trí điểm, chẳng hạn như một điểm $x \in FG$ thì sẽ khác với $x \in BG$:

$$B_{n+1}(x) = \begin{cases} \alpha B_n(x) + (1-\alpha)I_n(x), & x \in BG \\ \beta B_n(x) + (1-\beta)I_n(x), & x \in FG \end{cases} \quad (2.2)$$

$$T_{n+1}(x) = \begin{cases} \alpha T_n(x) + (1-\alpha)(\gamma \times |I_n(x) - B_n(x)|), & x \in BG \\ T_n(x), & x \in FG \end{cases} \quad (2.3)$$

Trong đó $\alpha, \beta (\in [0.0, 1.0])$, các giá trị α , β và γ được định rõ qua thực nghiệm bởi các video clip ở cả bên trong và bên ngoài nhà.

Ta cập nhật nền cho tất cả các kiểu của các điểm ảnh ($x \in FG$, $x \in BG$). Trong các phương pháp trừ nền, ảnh nền nguồn chỉ được cập nhật cho các điểm ảnh thuộc nền ($x \in BG$).

2.1.1.2. Mô hình hỗn hợp Gaussian tương thích

Trong mô hình này, các giá trị của một điểm ảnh riêng lẻ theo thời gian được coi như là một “xử lý điểm” và sự kiện gần đây của các điểm ảnh, $\{X_1, \dots, X_t\}$, được mô hình bởi hỗn hợp của phân phối Gaussian K. Khả năng của việc quan sát giá trị của điểm ảnh hiện thời trở thành:

$$P(X_t) = \sum_{i=1}^K \omega_{i,t} * \eta(X_t, \mu_{i,t}, \Sigma_{i,t}) \quad (2.4)$$

Trong đó, $\omega_{i,t}$ là một ước lượng trọng số thứ i của Gaussian ($G_{i,t}$) trong hỗn hợp ở thời điểm t , $\mu_{i,t}$ là giá trị trung bình của $G_{i,t}$ và $\Sigma_{i,t}$ là ma trận hiệp phương sai của $G_{i,t}$ và η là một hàm mật độ khả năng Gaussian:

$$\eta(X_t, \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(X_t - \mu)^T \Sigma^{-1} (X_t - \mu)} \quad (2.5)$$

Việc giải quyết trên K được quyết định bởi sự sẵn sàng của bộ nhớ và sức mạnh của tính toán. Ngoài ra, ma trận hiệp phương sai nhất được thừa nhận là mẫu sau

$$\Sigma_{k,t} = \alpha_k^2 I \quad (2.6)$$

Một điểm tương ứng được định nghĩa như là một giá trị điểm ảnh không vượt quá chuẩn $\gamma (=2.5)$ của một phân bố. tiếp theo, các trọng số chu kỳ của các phân bố K ở thời điểm t , $\omega_{k,t}$ được cập nhật như sau

$$\omega_{k,t} = (1 - \alpha) \omega_{k,t-1} + \alpha (M_{k,t}) \quad (2.7)$$

Trong đó α là tỷ lệ và $M_{k,t}$ bằng 1 tương ứng với phân bố Gaussian và bằng 0 tương ứng với các phân bố còn lại. Sau bước này, các trọng số ban đầu của các phân bố được bình thường hóa và các tham biến của Gaussian phù hợp được cập nhật với sự theo dõi mới như sau :

$$\mu_t = (1 - \rho) \mu_{t-1} + \rho (X_t) \quad (2.8)$$

$$\sigma_t^2 = (1 - \rho) \sigma_{t-1}^2 + \rho (X_t - \mu_t)^T (X_t - \mu_t) \quad (2.9)$$

$$\text{Trong đó : } \rho = \alpha \eta(X_t | \mu_k, \sigma_k) \quad (2.10)$$

Để phát hiện ra kiểu của điểm ảnh mới, các phân bố Gaussian K được sắp xếp theo giá trị của ω / σ . Danh sách đã sắp xếp của các phân bố này tương ứng với hầu hết các nền có khả năng từ đỉnh tới đáy trong công thức (2.7), các xử lý điểm ảnh nền thực hiện phân bố Gaussian tương ứng khi khởi tạo với giá trị trọng số ban đầu lớn và sự cách biệt nhỏ

Khi đó các phân bố B đầu tiên được chọn như là mô hình nền, trong đó :

$$B = \underset{b}{\operatorname{argmin}} \left(\sum_{k=1}^b \omega_k > T \right) \quad (2.11)$$

Và T là phần nhỏ nhất của dữ liệu điểm ảnh được cho là của nền. Nếu một giá trị nhỏ được chọn cho T , nền thường được mô hình thống nhất.

2.1.1.3. Sự khác biệt theo thời gian

Đặt $I_n(x)$ là giá trị cường độ của mức xám của pixel ở vị trí (x) và ở thời điểm n của dãy ảnh video I thuộc vào đoạn $[0, 255]$. Sự phối hợp khác biệt theo thời gian của 2 frame liên tiếp được thừa nhận là một điểm ảnh đang chuyển động nếu nó thỏa mãn như sau

$$|I_n(x) - I_{n-1}(x)| > T_n(x) \quad (2.12)$$

Ngưỡng của mỗi điểm ảnh, T , khởi tạo thiết lập một giá trị quyết định trước và sau đó cập nhật như sau

$$T_{n+1}(x) = \begin{cases} \alpha T_n(x) + (1 - \alpha)(\gamma \times |I_n(x) - I_{n-1}(x)|), & x \in BG \\ T_n(x), & x \in FG \end{cases} \quad (2.13)$$

2.1.2. Bước hậu xử lý điểm ảnh

Các nhân tố khác nhau là nguyên nhân gây nhiễu trong việc phát hiện sự nổi trội như là

- Nhiễu Camera
- Nhiễu do phản xạ
- Nhiễu đối tượng có cùng màu với nền
- Sự thay đổi ánh sáng đột ngột và bóng

2.1.2.1. Phát hiện bóng và sự thay đổi ánh sáng đột ngột

Đặt I_x là màu RGB của một điểm ảnh của ảnh hiện thời ở vị trí x , và B_x là màu RGB của điểm ảnh nền tương ứng. Hơn nữa, đặt \hat{I}_x là vector mà được bắt đầu ở gốc $O(0,0,0)$ trong không gian màu RGB và kết thúc ở điểm I_x , đặt \hat{B}_x là vector của điểm nền tương ứng với B_x và

đặt d_x là phép nhân (.) giữa \hat{I}_x và \hat{B}_x . Sự phối hợp phát hiện bóng ở đây phân loại một điểm ảnh là một phần của vùng nổi bật được phát hiện như là bóng nếu thỏa mãn

$$\left(d_x = \frac{\hat{I}_x}{\|\hat{I}_x\|} \cdot \frac{\hat{B}_x}{\|\hat{B}_x\|} \right) < \tau \quad (2.14)$$

Và

$$\|\hat{I}_x\| < \|\hat{B}_x\| \quad (2.15)$$

Trong đó τ là ngưỡng định nghĩa trước gần với một. Phép nhân (.) được dùng để kiểm tra xem \hat{I}_x và \hat{B}_x có cùng hướng hay không. Nếu phép nhân này (d_x) của \hat{I}_x và \hat{B}_x gần đến một, nó chỉ ra rằng chúng hầu hết ở cùng một hướng với lượng khác biệt nhỏ. Lần kiểm tra thứ hai được thực hiện để bảo đảm rằng giá trị ánh sáng của I_x là nhỏ hơn B_x .

Để phát hiện thay đổi ánh sáng tổng thể từ chuyển động của một đối tượng, ta làm một kiểm tra khác bằng cách dựa vào thực tế rằng trong trường hợp thay đổi ánh sáng tổng thể, sự thay đổi về hình dáng và kích cỡ của các biên đối tượng trong một cảnh không thay đổi quá nhiều và đường bao xung quanh của các vùng nổi trội được phát hiện không phù hợp với các biên thực tế trong ảnh trong khi trong trường hợp chuyển động của một đối tượng lớn thì đường bao các vùng nổi trội được phát hiện phù hợp với các đường biên thực tế trong ảnh.

2.1.3. Phát hiện các vùng liên tục

Sau khi phát hiện các vùng nổi trội và ứng dụng các phép toán hậu xử lý để xóa các vùng nhiễu và bóng. Các điểm ảnh nổi trội đã được lọc và nhóm vào các vùng liên tục (các khối) và được đánh nhãn bằng cách sử dụng thuật toán đánh nhãn thành phần liên tục 2 bước. Sau khi tìm các khối màu độc lập tương ứng với các đối tượng, các hộp bao xung quanh của các vùng đó được tính toán.

2.1.4. Bước hậu xử lý vùng

Ngay sau bước loại bỏ nhiễu, một vài vùng tự tạo nhỏ còn lại làm cho việc phân đoạn đối tượng không đúng. Để khử các vùng loại này,

kích cỡ vùng trung bình (γ) trong giới hạn của các điểm ảnh được tính toán cho mỗi frame và các vùng có kích cỡ nhỏ hơn một số thập phân (α) của kích cỡ vùng trung bình ($Size(region) < \alpha * \gamma$) bị xóa khỏi bản đồ điểm nổi bật.

2.1.5. Trích rút các đặc trưng của đối tượng

Ngay khi ta phân đoạn các vùng, chúng ta trích rút các điểm đặc trưng của các đối tượng tương ứng trong ảnh hiện thời. Các điểm đặc trưng đó là kích thước (S), điểm trung tâm của khối (C_m), lược đồ màu (H_c) và đường viền các đốm màu của đối tượng.

Để tính toán điểm chính giữa của khối điểm, $C_m = (xC_m, yC_m)$, của một đối tượng O , ta dùng công thức sau [42]

$$xC_m = \frac{\sum_i^n x_i}{n}, \quad yC_m = \frac{\sum_i^n y_i}{n} \quad (2.16)$$

Trong đó n là số điểm ảnh trong O

Lược đồ màu được tính toán bởi sự lặp đi lặp lại trên các điểm ảnh của O và tốc độ tăng lên của giá trị lưu trữ của khoảng cách màu tương ứng trong lược đồ H_c . Với mỗi đối tượng O , lược đồ màu được cập nhật như sau

$$H_c \left[\frac{c_i}{N} \right] = H_c \left[\frac{c_i}{N} \right] + 1, \quad \forall c_i \in O \quad (2.17)$$

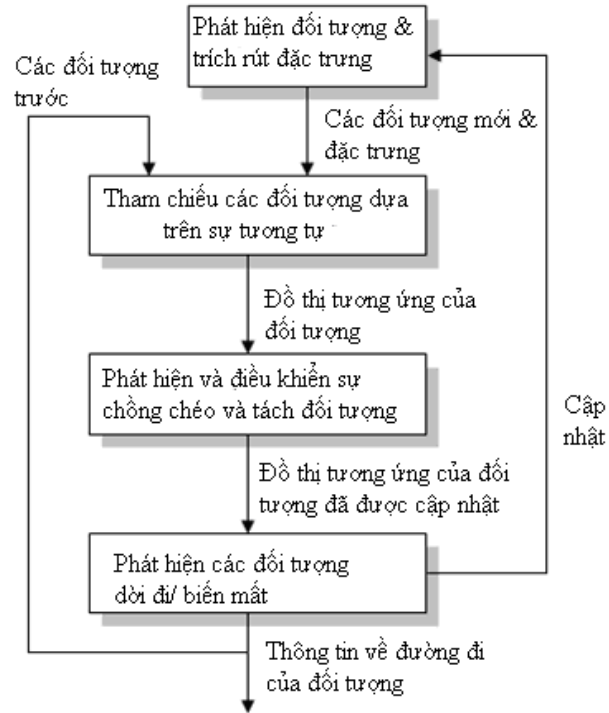
Trong đó c_i mô tả giá trị màu sắc của điểm ảnh thứ i . Trong bước tiếp theo, lược đồ màu sắc được chuẩn hóa cho phép thích hợp để so sánh với các lược đồ khác trong các bước sau. Lược đồ chuẩn hóa \hat{H}_c được tính như sau

$$\hat{H}_c[i] = \frac{H_c[i]}{\sum_i H_c[i]} \quad (2.18)$$

2.2. Đánh dấu đối tượng

Mục đích của đánh dấu đối tượng là thiết lập sự tương ứng giữa các đối tượng và các phần của đối tượng trong các frame liên tục và để

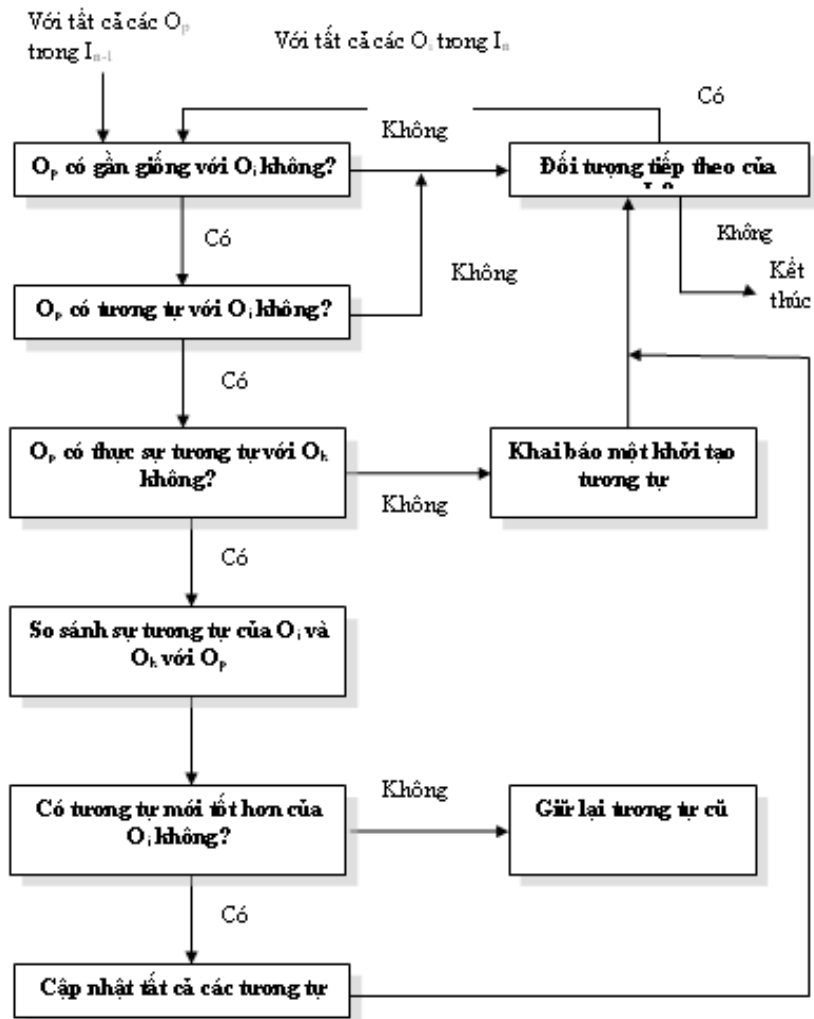
trích rút thông tin theo thời gian về các đối tượng như đường đi, tư thế, tốc độ và phương hướng



Sơ đồ 2.2: Sơ đồ hệ thống đánh dấu đối tượng

2.2.1. Tham chiếu tương tự với các đối tượng cơ bản

Biểu đồ hoạt động của thuật toán tham chiếu đối tượng thích hợp cơ bản được thể hiện trong sơ đồ 2.3. Bước đầu tiên trong thuật toán đánh dấu đối tượng này là tham chiếu đến các đối tượng (O_p 's) trong ảnh trước (I_{n-1}) đến các đối tượng mới (O_i 's) phát hiện được trong ảnh hiện thời (I_n).

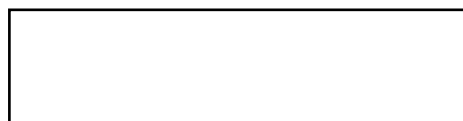


Sơ đồ 2.3: Phương pháp tham chiếu đối tượng phù hợp cơ bản

Hai đối tượng với các điểm trọng tâm c_p và c_i được coi là gần nhau nếu thỏa mãn

$$Dist(c_p, c_i) < \tau \quad (2.19)$$

Trong đó hàm $Dist()$ được định nghĩa như là khoảng cách Euclidean giữa hai điểm



$$Dist(c_p, c_i) = \sqrt{(x_{c_p} - x_{c_i})^2 + (y_{c_p} - y_{c_i})^2} \quad (2.20)$$

Hai đối tượng khác biệt được coi là tương tự nếu nó thỏa mãn

$$\frac{s_p}{s_i} < \mu \quad \text{hoặc} \quad \frac{s_i}{s_p} < \mu \quad (2.21)$$

Trong đó s_i là kích cỡ của đối tượng O_i và μ là một ngưỡng được định nghĩa trước.

Trong khi xác lập tham chiếu giữa các đối tượng trước và các đối tượng mới có 5 trường hợp tham chiếu khác nhau, ta có thể nêu ra như sau:

- One – to – one: Mỗi đối tượng trước O_p được tham chiếu với một đối tượng đơn O_i . Các đặc điểm của O_p được cập nhật thêm thông tin từ O_i
- One – to – many: Mỗi đối tượng trước O_p được tham chiếu với nhiều hơn một đối tượng mới.
- One – to – none: Mỗi đối tượng trước O_p không tham chiếu đến bất kỳ đối tượng mới nào.
- None – to – one: Trường hợp một đối tượng mới O_i không tham chiếu đến bất kỳ một đối tượng nào trong các đối tượng đã tồn tại.
- Many – to – one: Trường hợp một đối tượng mới O_i được tham chiếu với nhiều hơn một đối tượng trước đó.

2.2.2. Điều khiển sự chồng chéo của các đối tượng

2.2.2.1 Phát hiện các đối tượng bị che khuất

Khi một đối tượng O_p được phát hiện biến mất bằng cách dùng thuật toán tham chiếu đối tượng ban đầu, ta kiểm tra xem có một đối tượng mới O_i nào có đường bao xung quanh che lên O_p và nó được tham chiếu bởi đối tượng trước là O_i . Trong trường hợp thế này, có khả năng cao là O_p và O_i bị che khuất bởi một đối tượng khác có hình dạng mới là O_i .

2.2.2.2. Phát hiện các đối tượng tách ra

Khi một đối tượng O_i được tìm thấy trong một cảnh bởi thuật toán tham chiếu đối tượng, ta kiểm tra xem có đối tượng trước O_l có đường bao che khuất O_i và có giá trị trong nhóm ID và được tham chiếu bởi một đối tượng mới là O_k . Trong trường hợp này có thể coi như là một đối tượng vừa tách ra. Ta kiểm tra danh sách đối tượng trước về các đối tượng có cùng nhóm bị che khuất ID với O_p xem liệu có phải O_p và O_l cùng bị che khuất bởi một đối tượng trước đó không. Khi đó ta có 2 đối tượng đánh dấu $TO=\{O_p, O_l\}$ và hai đối tượng mới $NO=\{O_i, O_k\}$. Bây giờ, ta cần nhận dạng đối tượng trong TO tương ứng với đối tượng trong NO .

2.2.2.3. Lược đồ cơ sở tham chiếu tương ứng

Khoảng cách d_{ab} giữa hai lược đồ màu thông thường H_a và H_b với N bin được tính toán bằng cách sử dụng ma trận L_l như sau

$$d_{ab} = \sum_i^N |H_a[i] - H_b[i]| \quad (2.22)$$

Từ giờ ta giữ hai lược đồ của mỗi đối tượng, một cao hơn phần thân của đối tượng và một thấp hơn, ta tính toán toàn bộ khoảng cách bằng cách tổng kết các khoảng cách giữa các lược đồ màu tương ứng của các đối tượng. Đó là

$$d_{total} = d_{upper\ histogram} + d_{lower\ histogram} \quad (2.23)$$

2.2.3. Phát hiện các đối tượng dời đi và các đối tượng biến mất

Khả năng phát hiện sự dời đi và biến mất của các đối tượng là vấn đề sống còn của một số ứng dụng giám sát.

Hệ thống ta trình bày ở đây có thể phát hiện và phân biệt các đối tượng dời đi và biến mất trong video. Có ba bước để phát hiện sự dời đi và sự biến mất của các đối tượng như sau:

1. Phát hiện một sự thay đổi giữa ảnh hiện thời với ảnh nền cơ sở bằng cách sử dụng phối hợp trừ nền tương ứng.
2. Quyết định rằng vùng chuyển động được phát hiện có phù hợp với một sự dời đi và biến mất của đối tượng bằng cách sử dụng phương pháp đánh dấu đối tượng.

3. Phân biệt các đối tượng dời đi và các đối tượng biến mất bằng cách sử dụng thuộc tính màu thống kê của các vùng được phát hiện và đường bao xung quanh của nó.

Để phân biệt kiểu của đối tượng (dời đi hay biến mất), ta sử dụng các đặc tính tĩnh của các giá trị màu ở bên trong và ở viền của vùng được phát hiện. Đặt R là vùng tương ứng với một sự thay đổi với thời hạn dài trong nền; S là đường bao xung quanh của R và đặt A_X là giá trị cường độ màu trung bình trong vùng X . Kinh nghiệm cho thấy từ các cuộc thử nghiệm về đối tượng dời đi/ biến mất là nếu các giá trị của A_R và A_S là gần nhau, khi đó cho biết rằng vùng các đối tượng được phát hiện và vùng bao của nó hầu hết có cùng màu cho nên vùng này tương ứng với đối tượng chuyển chỗ. Mặt khác, nếu A_R và A_S không gần nhau cho thấy rằng vùng này tương ứng với đối tượng biến mất. Việc xác định A_R và A_S có gần nhau hay không như sau:

$$\tau \leq \frac{A_R}{A_S} \leq 1, \text{ nếu } A_R \leq A_S \quad (2.24)$$

$$\tau \leq \frac{A_S}{A_R} \leq 1, \text{ nếu } A_S \leq A_R$$

Trong đó τ là một hằng được định nghĩa trước (≈ 0.85).

2.3. Phân loại đối tượng

2.3.1. Phân loại dựa trên mẫu hình chiếu

Toàn bộ quá trình của phương pháp phân loại đối tượng bao gồm hai bước

- Bước ngoại tuyến: Tạo ra một cơ sở dữ liệu mẫu chứa hình chiếu của đối tượng bằng đánh nhãn thủ công các kiểu của đối tượng.
- Bước trực tuyến: Trích rút hình chiếu của mỗi đối tượng được đánh dấu trong mỗi frame và thừa nhận kiểu của nó bằng cách so sánh các đặc trưng cơ bản của hình chiếu của nó với các hình chiếu mẫu trong cơ sở dữ liệu với thời gian thực theo hệ thống.

2.3.2. Cơ sở dữ liệu mẫu hình chiếu

Cơ sở dữ liệu mẫu hình chiếu được tạo ngoại tuyến bởi trích rút các các đường viền tổng thể đối tượng từ các cảnh khác nhau.

Trong bước phân loại, phương pháp này không sử dụng các hình chiếu có định dạng thô mà so sánh các tín hiệu khoảng cách hình chiếu đã được biến đổi. Do đó, trong cơ sở dữ liệu mẫu ta chỉ lưu trữ tín hiệu khoảng cách của hình chiếu và thông tin kiểu tương ứng cho khả năng tính toán và lưu trữ.

Đặt $S = \{p_1, p_2, \dots, p_n\}$ là hình chiếu của một đối tượng O bao gồm n điểm sắp xếp từ điểm trung tâm của vùng được phát hiện theo chiều kim đồng hồ và c_m là điểm trung tâm của khối O . Tín hiệu khoảng cách $DS = \{d_1, d_2, \dots, d_n\}$ được sinh ra bởi việc tính toán khoảng cách giữa c_m và mỗi p_i bắt đầu từ 1 đến n như sau

$$d_i = \text{Dist}(c_m, p_i), \quad \forall i \in [1..n] \quad (2.25)$$

Trong đó Dist là một hàm khoảng cách Euclidian giữa hai điểm a và b

$$\text{Dist}(a, b) = \sqrt{(x_a - x_b)^2 + (y_a - y_b)^2} \quad (2.26)$$

Các đối tượng khác nhau có các hình chiếu khác nhau trong video và như thế các hình chiếu có kích cỡ không ổn định. Đặt N là kích cỡ của một tín hiệu khoảng cách DS và đặt C là một hằng số về độ dài tín hiệu được quy định. Kích cỡ quy định của tín hiệu khoảng cách \widehat{DS} khi đó được tính toán bởi mẫu nhỏ hoặc mẫu lớn của tín hiệu ban đầu DS như sau

$$\widehat{DS}[i] = DS[i * \frac{N}{C}], \quad \forall i \in [1..C] \quad (2.27)$$

Trong bước tiếp theo, tín hiệu khoảng cách tỷ lệ \widehat{DS} được chuẩn hóa thành miền đơn vị đầy đủ. Tín hiệu khoảng cách chuẩn \overline{DS} được tính toán với công thức sau

$$\overline{DS}[i] = \frac{\widehat{DS}[i]}{\sum_1^n \widehat{DS}[i]} \quad (2.28)$$

2.3.3. Ma trận phân loại

Các yêu cầu quan trọng của ma trận so sánh hình ảnh là cố định tỉ lệ, sự dịch chuyển và quay. Phương pháp này thỏa mãn cả ba đặc điểm đó.

1. Cố định tỷ lệ: Ta sử dụng một độ dài quy định phạm vi cho các tín hiệu khoảng cách của các hình dáng đối tượng.
2. Cố định sự dịch chuyển: Tín hiệu khoảng cách là độc lập với vị trí hình học của hình dạng đối tượng bởi tín hiệu khoảng cách được tính với trọng tâm khối của hình dạng đối tượng. do đó ma trận so sánh sẽ không bị ảnh hưởng bởi sự dịch chuyển.
3. Cố định quay: Chúng ta không sử dụng đặc tính cố định quay của ma trận phân loại do chúng ta muốn phân biệt các tư thế khác biệt của một đối tượng đơn cho bước sau trong hệ thống giám sát.

Ma trận phân loại so sánh sự tương tự giữa các hình dạng của 2 đối tượng, A và B, bằng cách tìm kiếm khoảng cách giữa các tín hiệu khoảng cách tương ứng của chúng, \overline{DS}_A và \overline{DS}_B . Khoảng cách giữa hai tỷ lệ và các tín hiệu khoảng cách chuẩn, \overline{DS}_A và \overline{DS}_B được tính như sau:

$$Dist_{AB} = \sum_{i=1}^n |\overline{DS}_A[i] - \overline{DS}_B[i]| \quad (2.29)$$

Để tìm kiểu T_O của một đối tượng O , ta so sánh tín hiệu khoảng cách của nó \overline{DS}_O với tất cả các tín hiệu khoảng cách của các đối tượng trong cơ sở dữ liệu mẫu. Kiểu T_P của mẫu đối tượng P được chỉ định như là kiểu của đối tượng yêu cầu O . $T_O = T_P$ trong đó P thỏa mãn:

$$Dist_{OP} \leq Dist_{Ob}, \forall \text{ đối tượng } I \text{ trong cơ sở dữ liệu mẫu} \quad (2.30)$$

Chương 3 - ỨNG DỤNG

3.1. Một số ứng dụng trong công nghệ phần mềm

3.1.1. Ứng dụng trong phần mềm nhúng (Embedded Software)

3.1.1.1. Khái niệm phần mềm nhúng

Phần mềm nhúng là một chương trình được viết, biên dịch trên máy tính và nạp vào một hệ thống khác (gọi tắt là KIT) bao gồm một hoặc nhiều bộ vi xử lý đã được cài sẵn một hệ điều hành, bộ nhớ ghi chép được, các cổng giao tiếp với các phần cứng khác

3.1.1.2. Mục đích của phần mềm nhúng

Nhằm hỗ trợ cho các sản phẩm phần cứng các chức năng hoàn hảo nhất, phục vụ tốt nhất các nhu cầu của người dùng với sự bảo mật về sản phẩm tốt nhất.

3.1.1.3. Tính chất của phần mềm nhúng

- Phụ thuộc vào hệ điều hành cài sẵn trên KIT
- Phụ thuộc vào các tính năng đặc trưng của từng sản phẩm phần cứng có trong KIT
- Phụ thuộc vào đặc tính của hệ thống

3.1.1.4. Lợi ích của phần mềm nhúng

Phần mềm nhúng là sự tích hợp của ngành tin học (phần mềm) với ngành điện tử (phần cứng). Với các các thiết bị điện tử, phần mềm nhúng mang lại nhiều sự hữu ích cần thiết cho người sử dụng và đồng thời giảm chi phí giá thành về phần cứng cho thiết bị.

3.1.1.5. Ứng dụng trong xử lý video

Hiện nay, các hệ thống giám sát hoạt động tốt dựa vào công nghệ phần cứng, tuy nhiên, như đã nói ở trên, phần mềm nhúng ra đời để làm giảm chi phí cho việc sản xuất các thiết bị phần cứng chuyên dụng vốn rất đắt đỏ. Việc cài đặt các phần mềm giám sát vào các thiết bị giám sát là rất cần thiết.

Với các phương pháp về phát hiện, đánh dấu, phân loại đối tượng đã nêu ở trên, ta có thể xây dựng một phần mềm giám sát để làm giảm thiểu chi phí cho các thiết bị phần cứng đồng thời có tính bảo mật cao.

3.1.2. Xây dựng các phần mềm điều khiển

Nhờ camera và các phương pháp phát hiện đối tượng chuyển động trong camera, ta có thể xây dựng các phần mềm điều khiển, các phần mềm hỗ trợ. Chẳng hạn như phần sử dụng tia laze để điều khiển các chương trình ứng dụng chạy trên máy tính. Phần mềm hỗ trợ giao tiếp giữa người khiếm thính với người thường,...

3.2. Cài đặt một số phương pháp phát hiện và đánh dấu đối tượng

Dựa vào một số phương pháp phát hiện và đánh dấu chuyển động như đã trình bày ở trên, trong phạm vi luận văn này tôi đã cài đặt được một phần mềm nhỏ có khả năng phát hiện đối tượng chuyển động và đánh dấu nó bằng cách đánh dấu toàn bộ đối tượng chuyển động, đánh dấu theo đường bao xung quanh đối tượng, đánh dấu bằng cách đóng khung đối tượng đã được phát hiện.

KẾT LUẬN

Trong phạm vi luận văn này, tôi đã nghiên cứu và trình bày lại một số phương pháp phát hiện đối tượng và một số bước hậu xử lý để làm cho đối tượng vừa được phát hiện rõ ràng hơn, tránh các trường hợp phát hiện nhầm do nhiễu. Sau khi phát hiện đối tượng, sẽ chuyển sang bước đánh dấu và phân loại đối tượng chuyển động.

Sau khi nghiên cứu các phương pháp đó, sẽ tìm ra phần ưu, phần nhược của từng phương pháp, từ đó có bước kết hợp giữa các phương pháp hoặc đưa ra các giải pháp để khắc phục các nhược điểm nhằm xây dựng một hệ thống giám sát tối ưu nhất có thể.

Trong luận văn này tôi cũng đã cài đặt thành công một số phương pháp phát hiện và đánh dấu đối tượng chuyển động. Hệ thống đầu vào trong bản demo này là lấy hình ảnh trực tiếp từ camera gắn vào máy tính, lấy một file video có phần đặc trưng là AVI từ nguồn có sẵn hoặc các video lấy từ nguồn trực tuyến.

Hướng nghiên cứu sâu hơn của đề tài này mà tôi quan tâm đó là phát hiện và phân loại từng phần chuyển động của đối tượng, từ đó xây dựng các ứng dụng hỗ trợ cho con người. Một hướng khác nữa là tìm hiểu rõ về đặc điểm của lửa, từ đó xây dựng các hệ thống cảnh báo hỏa hoạn.