

ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

TRẦN VĂN THÀNH

ph, t hiÖn chuyÓn ®éng B»NG

PH|ÝNG PH, P trõ nÒn

Chuyên ngành: KHOA HỌC MÁY TÍNH

Mã số: 60 48 0101

TÓM TẮT LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

Thái Nguyên - 2015

MỞ ĐẦU

Trong thời đại ngày nay công nghệ thông tin hầu như đã thâm nhập vào toàn bộ các lĩnh vực đời sống xã hội. Xã hội càng phát triển thì nhu cầu về công nghệ thông tin ngày càng cao, do vậy dữ liệu số hầu như không còn xa lạ đối với mỗi người chúng ta. Trong mọi lĩnh vực các ứng dụng công nghệ thông tin đã trợ giúp con người rất nhiều.

Hiện nay, thông tin hình ảnh đóng vai trò rất quan trọng trong trao đổi thông tin, bởi phần lớn các thông tin mà con người thu nhận được đều thông qua thị giác. Trong các lĩnh vực công nghệ thông tin thì lĩnh vực giám sát tự động đã và đang thu hút được nhiều sự quan tâm của các nhóm nghiên cứu trong và ngoài nước. Cùng với sự phát triển của sức mạnh máy tính, các hệ thống giám sát tự động ngày càng tinh vi và hiện đại đã trợ giúp con người rất nhiều trong lĩnh vực an ninh, giám sát giao thông v.v.

Thời gian qua, hệ thống giám sát bằng camera đã trở thành lĩnh vực nghiên cứu rất phát triển. Trước các nhu cầu giám sát, đảm bảo an ninh mục tiêu, trụ sở, nhà riêng... của các tổ chức, cá nhân ngày càng cao đã thúc đẩy việc nghiên cứu, ứng dụng các sản phẩm giám sát thông qua camera theo dõi. Bài toán đặt ra là xây dựng và đưa vào sử dụng các hệ thống giám sát đủ mạnh, có độ chính xác cao, phù hợp với từng mục đích cũng như đối tượng sử dụng. Một số ứng dụng quan trọng của hệ thống giám sát camera trong các lĩnh vực như: Các thiết bị giám sát an ninh chuyên dụng sử dụng trong lực lượng vũ trang; các thiết bị giám sát an ninh dân dụng; các thiết bị giám sát, phân luồng giao thông; công nghệ nghiên cứu, chế tạo Robot thông qua việc phân tích các cử động của con người...

Ở nước ta hiện nay, lĩnh vực giám sát Video cũng có những bước phát triển đáng kể. Tuy nhiên nó chỉ mới dựa trên nền tảng phần cứng và cũng chưa áp dụng nhiều trong thực tế. Việc giải quyết bài toán này theo hướng tiếp cận sử dụng phần mềm chưa được quan tâm phát triển.

Một hệ thống giám sát camera bao gồm nhiều chức năng: Phát hiện chuyển động, phân loại, theo dấu, cảnh báo, phân tích hành động. Phát hiện chuyển động được xác định là một trong những chức năng đầu tiên và quan trọng nhất của một hệ thống giám sát camera. Nó không chỉ có chức năng trích xuất đối tượng chuyển động mà còn có ý

nghĩa quan trọng đối với ứng dụng thị giác máy tính như mã hóa video dựa trên đối tượng, phân tích chuyển động của con người và tương tác người – máy.

Xuất phát từ các nhu cầu thực tế, học viên đã báo cáo xin ý kiến thầy giáo hướng dẫn nghiên cứu luận văn của mình là: **“Phát hiện chuyển động bằng phương pháp trừ nền”**. Kết quả của nghiên cứu là xây dựng giải thuật phát hiện chuyển động và phần mềm phát hiện và cảnh báo khi có đối tượng đột nhập vào khu vực giám sát một cách nhanh chóng và hiệu quả. Trên cơ sở đó nhằm xây dựng hệ thống giám sát bằng camera trụ sở của đơn vị sử dụng các nghiên cứu về thuật toán phát hiện chuyển động.

Nội dung chính của luận văn gồm 3 chương:

Chương 1: Khái quát về Video và phát hiện chuyển động

Chương 2: Nội dung giải thuật phát hiện chuyển động

Chương 3: Chương trình thử nghiệm phát hiện chuyển động

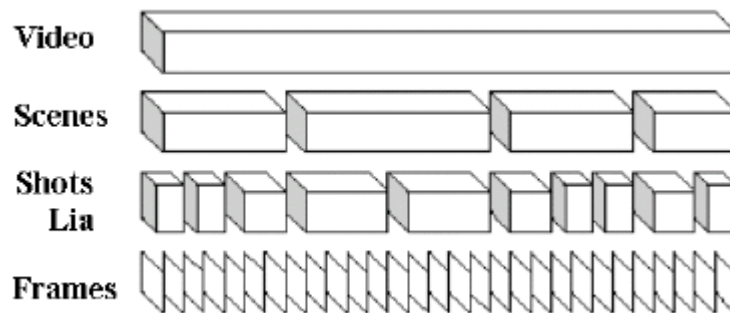
Chương 1

KHÁI QUÁT VỀ VIDEO VÀ PHÁT HIỆN CHUYỂN ĐỘNG

1.1. Các khái niệm cơ bản về video [3]

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

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



Hình 1.1: Cấu trúc phân đoạn video

Khi phim được chiếu, các khung hình lần lượt được hiển thị ở tốc độ nhất định. Tốc độ thường thấy ở các định dạng video khác nhau là 30 và 25 hình/s. Như vậy một giờ video sẽ có số khung hình tương ứng là 108000 hoặc là 90000. Dù là video ở định dạng nào thì nó cũng có dung lượng rất lớn và nếu xử lý với tất cả các khung hình thì thật không hiệu quả.

Phân đoạn là quá trình phân tích và chia nội dung hình ảnh video thành các đơn vị cơ sở gọi là các lia (shot). Việc lấy mẫu chính là chọn gần đúng một khung video đại diện cho mỗi lia (hoặc nhiều hơn tùy theo mức độ phức tạp của nội dung hình ảnh của lia) và được gọi là các khung - khóa.

Khung – khóa là khung hình đại diện mô tả nội dung chính của shot.

Quá trình phân đoạn dữ liệu video tiến hành phân tích, phát hiện sự chuyển đổi từ lia này sang lia khác hay chính là sự phát hiện ranh giới giữa các lia (đó chính là

đo sự khác nhau giữa các khung hình liên kế). Trong hình dưới đây là ví dụ về sự chuyển đổi giữa các lia:

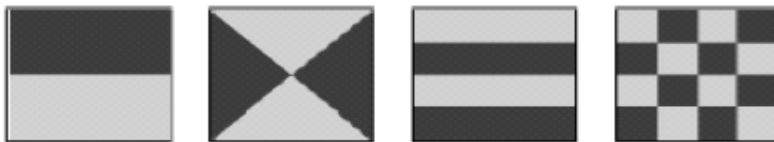


Hình 1.2: Chuyển đổi Lia giữa khung hình thứ 3 và thứ 4

Một số thuộc tính đặc trưng của video: Video có 4 đặc trưng chính đó là: Color (màu), Texture (kết cấu), Shape (hình dạng), Motion (chuyển động).

1.1.1. Màu (Color)

Màu là một thuộc tính quan trọng của ảnh. Biểu đồ màu, biểu diễn sự phân bố màu, là một đặc trưng màu phổ biến nhất hiện nay. Biểu đồ màu không phụ thuộc vào sự quay, dịch chuyển ảnh cũng như chiều nhìn ảnh. Tính hiệu quả của nó lại phụ thuộc vào hệ màu và phương pháp định lượng được dùng. Có một vấn đề với biểu đồ màu là nó không biểu diễn thông tin về không gian phân bố các điểm ảnh, do đó hai ảnh có cùng biểu đồ màu có thể có nội dung rất khác nhau.



Hình 1.3: Bốn khung hình khác nhau song có cùng một biểu đồ màu

1.1.2. Kết cấu (Texture)

Đây là một đặc trưng quan trọng của bề mặt, nơi xảy ra việc lặp lại mẫu cơ bản. Có hai dạng biểu diễn texture phổ biến: biểu diễn dạng ma trận đồng thời và biểu diễn Tamura. Ma trận đồng thời mô tả hướng và khoảng cách giữa các điểm ảnh, nhờ đó các thống kê có nghĩa có thể được trích chọn. Ngược lại, người ta thấy rằng entropy và mô men chênh lệch nghịch đảo lại có khả năng phân biệt tốt nhất. Biểu diễn Tamura được thúc đẩy nhờ các nghiên cứu về tâm lý trong việc thu nhận trực giác của con người và nó bao gồm các đại lượng đo tính thô, độ tương phản, hướng, tính trơn,

tính cân đối và độ ráp. Các đặc trưng Tamura rất hấp dẫn trong việc hiểu nội dung ảnh vì nó biểu đạt trực quan. Ngoài ra còn có một số các dạng biểu diễn khác như trường ngẫu nhiên Markov, biến đổi Gabor, biến đổi gợn sóng,...

1.1.3. Hình dáng (Shape)

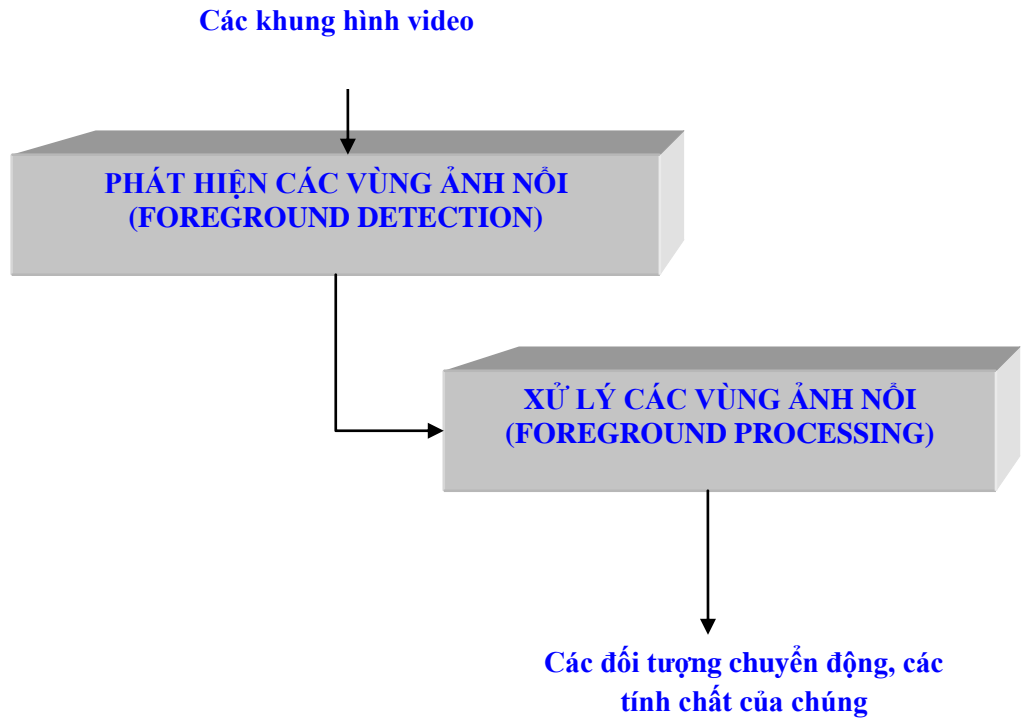
Các đặc trưng hình dáng có thể được biểu diễn sử dụng phân tích hình dáng truyền thống như bất biến mô men, mô tả Fourier, mô hình học tự động quay lui và các thuộc tính hình học. Các đặc trưng này có thể được phân chia thành đặc trưng toàn cục và đặc trưng cục bộ. Đặc trưng toàn cục là đặc trưng thuộc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

1.1.4. Chuyển động (Motion)

Motion là thuộc tính quan trọng của video. Thông tin về chuyển động có thể được sinh ra bằng các kỹ thuật ghép khối hoặc luồng ánh sáng. Các đặc trưng chuyển động như mô men của trường chuyển động, biểu đồ chuyển động hoặc là các tham số chuyển động toàn cục có thể được trích chọn từ vectơ chuyển động. Các đặc trưng mức cao phản ánh di chuyển camera như quét camera (pan), nghiêng (tilt), phóng to (zoom out), thu nhỏ (zoom in) cũng có thể được trích chọn.

1.2. Phát hiện đối tượng chuyển động là gì?

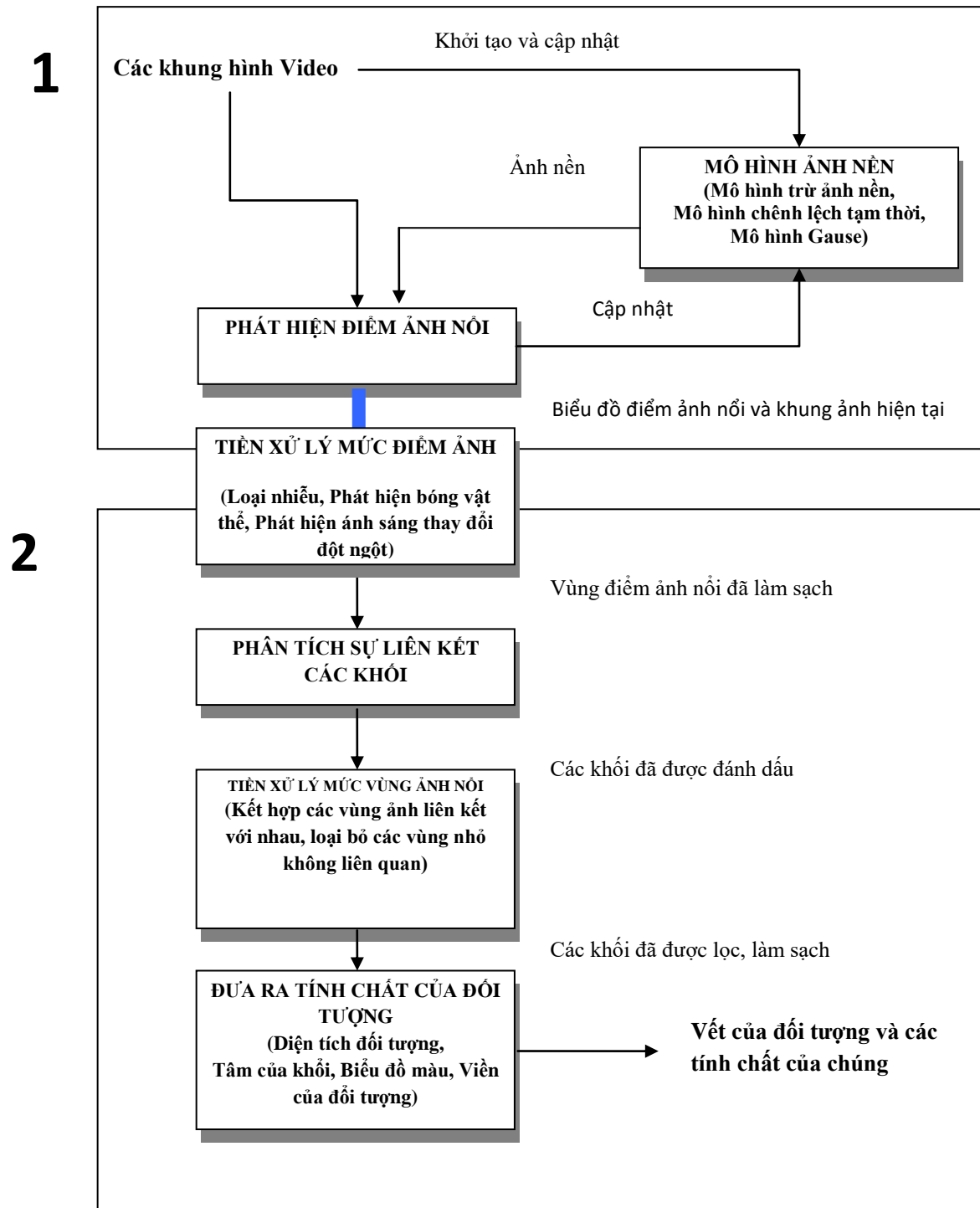
Đó là quá trình đưa ra vết các đối tượng chuyển động từ các khung hình video. Quá trình này thực chất là quá trình xử lý chuỗi ảnh liên tiếp trong một đoạn video để phát hiện ra các đối tượng chuyển động trong một đoạn hình ảnh theo mô hình dưới đây:



Hình 1.4: Tổng quan các khối xử lý trong bài toán phát hiện đối tượng

1.2.1. Thiết kế các khối xử lý

Mô hình tổng quan khối xử lý cho cả bài toán phát hiện đối tượng chuyển động

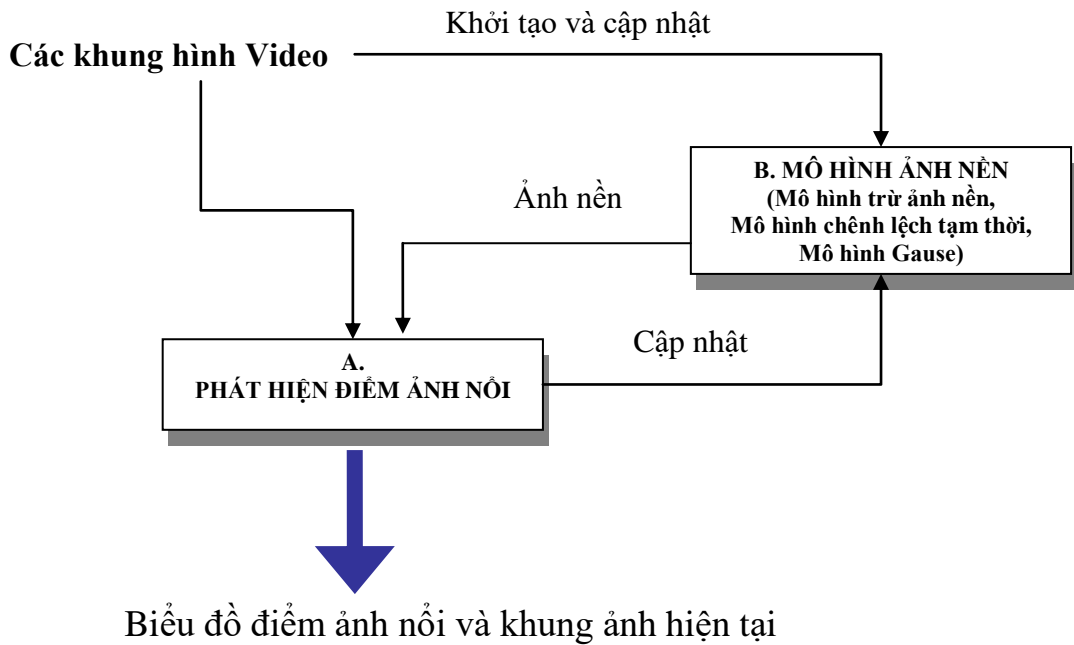


Hình 1.5: Mô hình chi tiết các khối xử lý

1.2.1.1. Khối phát hiện các vùng ảnh nổi

Trong phần thiết kế xử lý cho khối phát hiện vùng ảnh nổi ta sử dụng một mô hình ảnh nền (Background Model) nhằm tích hợp các phương pháp trừ ảnh nền, mô

hình chên lệch tạm thời và mô hình thống kê Gausse nhằm giải quyết vấn đề phát hiện các vùng ảnh nổi.



Hình 1.6: Khởi phát hiện vùng ảnh nổi

Sơ đồ trên chỉ ra các khối xử lý nhằm phát hiện các vùng ảnh nổi. Trong đó mô hình trừ ảnh nền với các phương thức khởi tạo và cập nhật các khung hình ảnh nền nhằm đưa đầu vào: các ảnh nền cho quá trình xử lý phát hiện điểm ảnh nổi.

Sau khi đã có các ảnh nền, thao tác xử lý bên trong khối phát hiện điểm ảnh nổi đơn giản chỉ là thực hiện việc trừ ảnh, lấy khung hình hiện tại so sánh với ảnh nền để đưa ra một biểu đồ các điểm ảnh nổi lên từ ảnh hiện tại. Biểu đồ điểm ảnh này với khung ảnh hiện tại sẽ là đầu vào của quá trình xử lý tiếp theo

Trong sơ đồ xử lý này, khối xử lý mô hình ảnh nền được coi là khối xử lý quan trọng nhất. Khối này sẽ thực hiện tích hợp các phương pháp xử lý khác nhau như đã trình bày ở trên nhằm đưa ra một mô hình ảnh nền phục vụ cho bước xử lý tiếp theo. Cụ thể các mô hình được sử dụng trong khối xử lý này là:

- Tương hợp mô hình trừ ảnh nền
- Tương hợp mô hình pha trộn Gausse

- Mô hình chèn lệch tạm thời

Các mô hình này sẽ được thực hiện dưới dạng các thư viện xử lý, tích hợp vào trong hệ thống. Khi gặp các điều kiện quan sát khác nhau, hệ thống có thể linh động sử dụng các thư viện xử lý này để đưa ra hiệu quả cao nhất. Dưới đây là một số cơ sở lý thuyết về các mô hình được đề cập ở trên.

1.2.1.1.1. Mô hình trừ ảnh nền thích hợp

Thuật toán trừ ảnh nền thích hợp được thực hiện như sau:

Gọi $I_n(x)$ là cường độ của điểm ảnh tại vị trí x và tại thời điểm khung hình thứ n trong chuỗi các khung hình video $I(n \in [0, 255])$. Gọi $B_n(x)$ là giá trị cường độ của ảnh nền tương ứng cho vị trí x được ước lượng qua chuỗi hình ảnh từ I_0 đến I_{n-1} . Theo phương pháp trừ ảnh thì điểm ảnh tại vị trí x trong khung hình video hiện tại thuộc lớp ảnh nổi khi nó thỏa mãn :

$$|I_n(x) - B_n(x)| > T_n(x) \quad (1.1)$$

Trong công thức này $T_n(x)$ biểu diễn một ngưỡng được ước lượng qua các khung hình từ I_0 đến I_{n-1} . Công thức trên được sử dụng để tạo ra bản đồ các vùng ảnh nổi lên. Biểu diễn các điểm ảnh bởi một ma trận trong đó các điểm ảnh nổi lên sẽ có giá trị là một và các điểm ảnh nền sẽ có giá trị là 0 cho mỗi điểm ảnh. Điểm ảnh nền $B_n(x)$ ban đầu được khởi tạo với khung hình đầu tiên là $B_0 = I_0$. Do sự thay đổi của khung cảnh thu được nên khung ảnh nền và ngưỡng liên tục được cập nhật theo công thức sau:

$$B_{n+1}(x) = \begin{cases} \alpha B_n(x) + (1 - \alpha) I_n(x), & x \in BG \\ \beta B_n(x) + (1 - \beta) I_n(x), & x \in FG \end{cases} \quad (1.2)$$

$$T_{n+1}(x) = \begin{cases} \alpha T_n(x) + (1 - \alpha) I_n(x), & x \in BG \\ T_n(x), & x \in FG \end{cases} \quad (1.3)$$

Trong đó ký hiệu BG là vùng ảnh nền và FG là vùng ảnh nổi lên trên. α và β thuộc khoảng $[0.0, 1.0]$, γ là số điểm ảnh khác nhau từ khung hình được so sánh với ảnh nền.