**ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**



**ĐỐ ÁN MÔN HỌC XỬ LÝ ẢNH VÀ ỨNG DỤNG**

**Counting People Using Image Processing**

**Trần Anh Khoa – 21522233**

**Liêu Xuân Hiền – 21522061**

**TP. HỒ CHÍ MINH, 2024**

**ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

**Trần Anh Khoa – 21522233**

**Liêu Xuân Hiền – 21522061**

**ĐỒ ÁN MÔN HỌC XỬ LÝ ẢNH VÀ ỨNG DỤNG**

**Đếm số người sử dụng xử lý ảnh**

**GIẢNG VIÊN HƯỚNG DẪN**

**TS.Mai Tiến Dũng**

**TP. HỒ CHÍ MINH, 2024**

MỤC LỤC

[Chương 1. Giới thiệu đề tài 1](#_Toc156144595)

[1.1. Giới thiệu 1](#_Toc156144596)

[1.2. Ứng dụng 1](#_Toc156144597)

[1.3. Nghiên cứu liên quan 1](#_Toc156144598)

[Chương 2. Đếm số người ra vào thuần xử lý ảnh 2](#_Toc156144599)

[2.1. Phương pháp trừ nền 2](#_Toc156144600)

[2.1.1. Phương phép trừ nền GMM(Gaussian Mixture Model) 2](#_Toc156144601)

[2.1.2. Các phép biến đổi ảnh 3](#_Toc156144602)

[2.2. Phương pháp Contour 3](#_Toc156144603)

[2.2.1. Contour 3](#_Toc156144604)

[2.2.1. Các đặc trưng của contour 4](#_Toc156144605)

[2.3. Kết quả 5](#_Toc156144606)

[2.4. Đánh giá 5](#_Toc156144607)

[Chương 3. Phương pháp đếm số người ra vào sử dụng mô hình 5](#_Toc156144608)

[3.1. Phát hiện đối tượng sử dụng mô hình MobileNet SSD. 5](#_Toc156144609)

[3.1.1. Depthwise Seperable Convolutions MobileNet 5](#_Toc156144610)

[3.1.2. Mô hình phát hiện đối tượng sử dụng Single multi Shot Detection 7](#_Toc156144611)

[3.2. Thuật toán dùng để theo dõi và dự đoán hướng di chuyển [2] 8](#_Toc156144612)

[3.2.1. Tổng quan 8](#_Toc156144613)

[3.2.2. Sử dụng thuật toán theo dõi tương quan Dlib [3][4] 9](#_Toc156144614)

[3.2.3. Phương pháp theo dõi tâm 9](#_Toc156144615)

[3.3. Kết quả 11](#_Toc156144616)

[3.4. Đánh giá 12](#_Toc156144617)

[Chương 4. Kết luận 12](#_Toc156144618)

[TÀI LIỆU THAM KHẢO 14](#_Toc156144619)

# Giới thiệu đề tài

## Giới thiệu

Bắt đầu từ nhu cầu thực tế, Chúng ta có thể thấy sự cần thiết của việc đếm người thông qua camera và xử lý ảnh trong nhiều lĩnh vực như an ninh, quản lý sự kiện, hay thậm chí trong ngành bán lẻ. Điều này không chỉ giúp tăng cường an toàn mà còn mang lại hiệu suất và sự thuận tiện trong quản lý. Trong nghiên cứu này, chúng tôi sẽ đào sâu vào việc hiểu và áp dụng các thuật toán đếm người, từ những phương pháp cơ bản đến những tiến triển hiện đại. Đồng thời, việc làm quen với OpenCV sẽ giúp chúng ta nhanh chóng triển khai những giải pháp thực tế dựa trên kiến thức lý thuyết đã nắm bắt.

## Ứng dụng

Những ứng dụng tiềm năng của việc đếm người và sử dụng OpenCV là vô cùng đa dạng. Trong lĩnh vực an ninh và giám sát, hệ thống này có thể cung cấp một công cụ hiệu quả để theo dõi và đánh giá các hoạt động misicphar cảnh trong thời gian thực. Trong quản lý lưu lượng khách hàng, việc tự động đếm số lượng người có thể giúp tối ưu hóa dịch vụ và kế hoạch phục vụ. Trong quản lý tòa nhà, sự kiện và hội nghị, khả năng đếm người có thể được tích hợp để đảm bảo an toàn và thoải mái cho mọi người tham gia. Trong chăm sóc khách hàng, thông tin về lưu lượng và chuyển động của khách hàng có thể giúp cải thiện trải nghiệm mua sắm hoặc sử dụng dịch vụ. Cuối cùng, việc theo dõi nhân sự là một ứng dụng quan trọng trong quản lý nhân sự và tăng cường hiệu suất làm việc. Sự kết hợp giữa các thuật toán đếm người và OpenCV mở ra nhiều cơ hội sáng tạo để áp dụng vào các lĩnh vực khác nhau, mang lại lợi ích to lớn cho nhiều tổ chức và doanh nghiệp.

## Nghiên cứu liên quan

Real-Time Object Detection with Yolo [1]

* Tác giả Geetha Priya và cộng sự sử dụng kỹ thuật YOLO để phát hiện đối tượng bằng cách dự đoán các hộp giới hạn bằng mạng convolutional và xác suất lớp cho những hộp này, đồng thời phát hiện hình ảnh nhanh hơn so với các thuật toán khác.

A Foreground-background Segmentation Algorithm for Video Sequences [2]

* Tác giả Wei, Li và Yue cung cấp một hệ thống phát hiện chuyển động foreground-background cho việc theo dõi đối tượng di chuyển từ một máy quay video trong. Các tác giả đạt được điều này dựa trên một hàm mật độ xác suất được gọi là Mô hình Hỗn hợp Gaussian (GMM).

[1] Geethapriya. S, N. Duraimurugan, S.P. Chokkalingam, “Real-Time Object Detection with Yolo”, International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-8, Issue-3S, February 2019.

[2] Wei, Li, Yue. A Foreground-background Segmentation Algorithm for Video Sequences. Jiangnan Univ. China. 2015

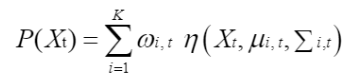
# Đếm số người ra vào thuần xử lý ảnh

## Phương pháp trừ nền

### Phương phép trừ nền GMM(Gaussian Mixture Model)

Gaussian Mixture Model là một thuật toán trừ nền / tiền cảnh dựa trên hỗn hợp Gaussian. Thuật toán được giới thiệu trong bài báo "An improved adaptive background mixture model for real-time tracking with shadow detection" của P. KadewTraKuPong và R. Bowden vào năm 2001. Nó sử dụng một phương pháp để mô hình hóa từng pixel nền bằng hỗn hợp phân phối K Gaussian ( K = 3 đến 5). Trọng lượng của hỗn hợp đại diện cho tỷ lệ thời gian mà những màu đó ở lại trong cảnh. Các màu nền có thể xảy ra là những màu tồn tại lâu hơn và tĩnh hơn.

Công thức:



Trong đó:

Xt : giá trị của pixel tại thời điểm t

XT = {Xt, ..., X(t-T) } : training set tại thời gian t với T là khoảng thời gian

K: số thành phần(component), thường trong khoảng 3 đến 5.

ωi,t:trọng số của Gaussian thứ i, là giá trị không âm và có tổng bằng 1.

µi,t: giá trị trung bình của Gaussian thứ i ở thời gian t

∑i,t : Ma trận hiệp phương sai của Gaussian thứ i tại thời gian t.

### Các phép biến đổi ảnh

Sử dụng các phép biến đổi ảnh để xử lý kết quả sau khi thực hiện trừ nền bằng GMM. Dùng phép nhị phân ảnh để loại bỏ bóng của vật thể. Dùng phép mở ảnh để trơn biên của đối tượng trong ảnh, cũng như loại bỏ những thành phần nhô ra có kích nhỏ trong ảnh. Phép đóng ảnh cũng tương tự làm trơn biên của đối tượng trong ảnh nhưng ngược với phép mở. Phép toán này thường làm hợp nhất các đoạn gãy hẹp, loại bỏ lỗ hỏng nhỏ và làm đầy các khe hở.

## Phương pháp Contour

### Contour

Contour đơn giản là một đường cong kết nối các điểm liên tục theo đường biên, chia sẻ cùng màu sắc hoặc giá trị cường độ. Được sử dụng chủ yếu trong việc phân tích hình dạng, nhận diện và phát hiện vật thể, contour giúp chúng ta hiểu rõ hơn về cấu trúc và đặc điểm của hình ảnh. [1]

Để đạt độ chính xác tốt nhất, việc chuyển đổi hình ảnh sang dạng nhị phân, chỉ chứa hai màu đen và trắng, trước khi áp dụng thuật toán contour là quan trọng. Bạn có thể thực hiện điều này bằng cách sử dụng threshold hoặc thuật toán Canny. [1]

Lưu ý rằng hàm findContour và drawContour sẽ thay đổi hình ảnh gốc, nếu bạn muốn giữ lại hình ảnh gốc sau khi tìm contour, hãy lưu nó vào một biến khác.

Trong OpenCV, việc tìm contours giống như việc "rải" màu trắng lên vật thể từ nền màu đen. Do đó, quan trọng để đảm bảo đối tượng cần tìm có màu trắng và nền là màu đen trong quá trình xử lý ảnh.

A person standing in a room with a black background and a red outline

Description automatically generated

### Các đặc trưng của contour

**Moment**

Cx  là tọa độ x của tâm trọng tâm và Cy là tọa độ y của tâm trọng tâm của vùng hoặc contour trong ảnh. Công thức này giúp xác định vị trí trung tâm của vật thể trong không gian ảnh dựa trên phân bố của các pixel.

Công thức tính Centroid:

Cx = , Cy =

Giải thích công thức:

M10 là moment bậc 10, thể hiện tổng giá trị pixel nhân với tọa độ x.

M01 là moment bậc 01, thể hiện tổng giá trị pixel nhân với tọa độ y.

M00 là moment bậc 00, là tổng giá trị pixel không có tọa độ (chỉ là tổng giá trị pixel).

## Kết quả



**Kết quả chạy trên khung hình**

## Đánh giá

Ưu điểm của phương pháp này nằm ở sự đơn giản và tính dễ triển khai, giúp nhanh chóng tích hợp vào các ứng dụng thực tế. Tuy nhiên, nhược điểm xuất phát từ sự phụ thuộc đối với điều kiện ánh sáng và nền, cũng như khả năng chồng chéo giữa các đối tượng, làm giảm độ chính xác của hệ thống. Ngoài ra, việc mô hình còn dựa vào kích thước và hình dạng của bounding box, có thể dẫn đến những hiểu lầm khi nhận diện đối tượng, đặc biệt là trong trường hợp đối tượng không phải là người.

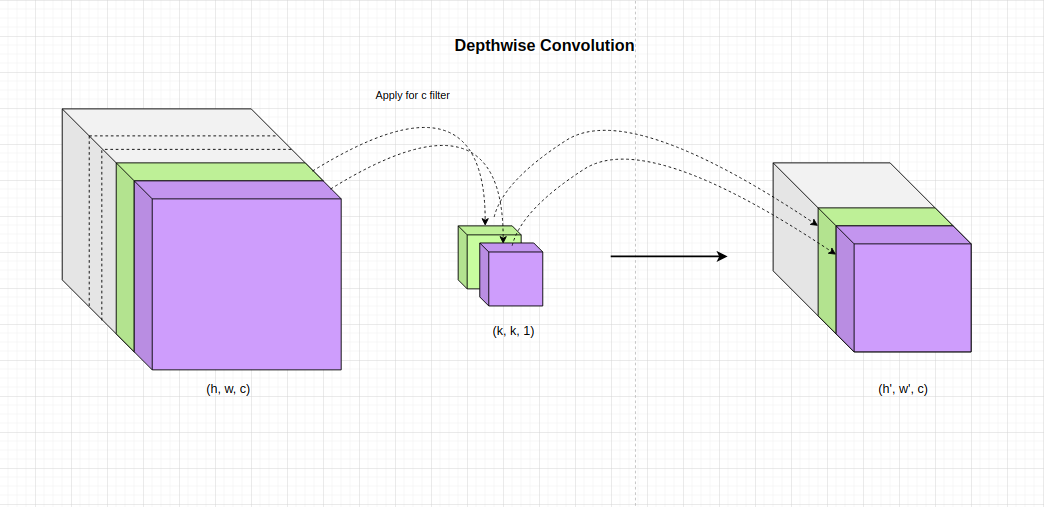
# Phương pháp đếm số người ra vào sử dụng mô hình

## Phát hiện đối tượng sử dụng mô hình MobileNet SSD.

### Depthwise Seperable Convolutions MobileNet

a) Tích chập chiều sâu (Depthwise Convolutions)

Chúng ta sẽ chia khối input tensor3D thành những lát cắt ma trận theo độ sâu. Thực hiện phép tích chập trên từng lát cắt như hình minh họa bên dưới:



Mỗi một kênh sẽ áp dụng bộ lọc khác nhau và hoàn toàn không chia sẻ tham số. Điều này có ba tác dụng chính cho mô hình:

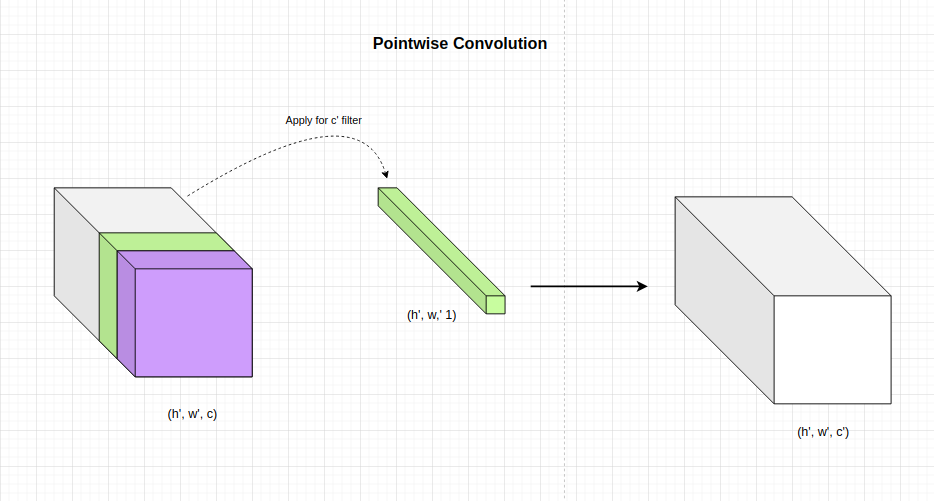
**\* Nhận diện đặc trưng**: Quá trình học và nhận diện đặc trưng sẽ được tách biệt theo từng bộ lọc. Nếu đặc trưng các kênh là khác xa nhau thì sử dụng các bộ lọc riêng cho kênh sẽ chuyên biệt hơn trong việc phát hiện các đặc trưng. Chẳng hạn như đầu vào là 3 keeng RGB thì mỗi kênh áp dụng một bộ lọc khác nhay chuyên biệt.

\* **Giảm thiểu khối lượng tính toán**: Để tạo ra một điểm pixel trên output thì tích chập thông thường cần sử dụng k×k×c phép tính trong khi tích chập chiều sâu tách biệt chỉ cần k×k phép tính.

**\* Giảm thiếu số lượng tham số** : Ở tích chập chiều sâu cần sử dụng c×k×k tham số. Số lượng này ít hơn gấp c’ lần so với tích chập chiều sâu thông thường.

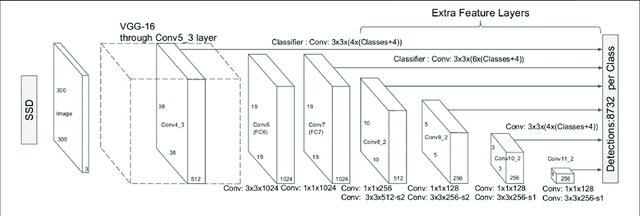
b) Tích chập điểm (Point-wise Convolutions)

Có tác dụng thay đổi độ sâu của output bước trên từ c sang c’. Chúng ta sẽ áp dụng c’ bộ lọc kích thước 1x1xc. Như vật kích thước width và height không thay đổi mà chỉ độ sâu thay đổi.



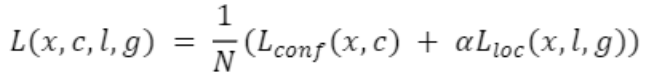
### 3.1.2. Mô hình phát hiện đối tượng sử dụng Single multi Shot Detection

Kiến trúc của MobileNet SSD được xây dựng trên kiến trúc MobileNet. SSD có thể nhận diện vật thể với các lớp khác nhau.



Cũng giống như hầu hết các kiến trúc object detection khác, đầu vào của SSD là tọa độ bounding box của vật thể(hay còn gọi là offsets của bounding box) và nhãn của vật thể chứa trong bounding box. Điểm đặc biệt làm nên tốc độ của SSD là mô hình sử dụng một mang nơ-ron duy nhất. Cách tiếp cận của nó dựa trên việc nhận diện vật thể trong các features map có độ phân giải khác nhau. Mô hình sẽ tạo ra một lưới cá ô vuông gọi là grid cells trên các feature map, mỗi ô được gọi là một cell và từ tâm của mỗi cell xác định một tập hợp các boxes mặc định( default boxes) để dự đoán khung hình có khả năng bao quanh vật thể. Tại thời điểm dự báo mạng nơ-ron sẽ trả về 2 giá trị đó là: phân phối xác xuất của nhãn của vật thể chứa trong bounding box và một tọa độ gọi là offsets của bounding box. Quá trình huấn luyện cũng là quá trình tinh chỉnh xác xuất nhãn và bounding box về đúng với các giá trị groundtruth input của mô hình.

Hàm mất mát của thuật toán SSD có công thức:



Trong đó N là số lượng matched default boxes. Nếu N=0, loss có giá trị mặc định là 0.

## Thuật toán dùng để theo dõi và dự đoán hướng di chuyển [2]

### Tổng quan

Trong quá trình phát triển hệ thống đếm người, giai đoạn phát hiện đối tượng được thực hiện một lần sau mỗi khoảng khung hình, tạo ra sự linh hoạt trong việc xử lý dữ liệu. Việc này không chỉ giảm áp lực tính toán mà còn giúp tối ưu hóa tốc độ so với việc liên tục chạy mô hình phát hiện trong thời gian thực.

Tốc độ nhanh hơn trong giai đoạn này không chỉ mang lại sự hiệu quả mà còn giúp tăng cường khả năng xử lý dữ liệu, đặc biệt là trong các kịch bản mà mô hình thực hiện đối tượng chỉ một lần. Điều này làm cho quá trình đếm người trở nên nhẹ nhàng và linh hoạt hơn, đồng thời giảm độ trễ khi áp dụng các biện pháp phân tích.

Tuy nhiên, như mọi lợi ích đều đi kèm với nhược điểm. Trong tình huống đối tượng biến mất hoặc bị che khuất bởi các đối tượng khác, sự linh hoạt của giai đoạn này cần được kết hợp với các chiến lược xử lý tình huống linh hoạt và độ tin cậy cao. Mặc dù tốc độ là một ưu điểm, nhưng việc đảm bảo độ chính xác và tính đồng nhất trong quá trình đếm vẫn là thách thức đối với các hệ thống đếm người sử dụng mô hình này.

### Sử dụng thuật toán theo dõi tương quan Dlib [3][4]

Theo dõi tương quan của Dlib là một công cụ mạnh mẽ được sử dụng trong thị giác máy tính để theo dõi đối tượng, được xây dựng dựa trên phương pháp DCF, một phương pháp xuất sắc trong các tình huống theo dõi thời gian thực. Tracker hoạt động bằng cách tạo ra một bộ lọc tương quan cho đối tượng cần theo dõi, sau đó cập nhật bộ lọc này một cách hiệu quả qua các khung hình tiếp theo. Cơ chế lọc thích ứng này giúp tracker theo dõi đối tượng một cách mạnh mẽ, ngay cả khi có sự thay đổi về tỉ lệ, xoay hoặc che khuất. Được thiết kế để đáp ứng yêu cầu theo dõi thời gian thực, tracker này tận dụng các thuật toán tối ưu và xử lý đa nhiệm để mang lại tốc độ theo dõi đỉnh cao.

Phương pháp: cung cấp tọa độ hộp giới hạn ban đầu của đối tượng. Kế đến, thuật toán sẽ làm nhiệm vụ cập nhật mô hình nội tại của mình, giữ cho quá trình theo dõi diễn ra mượt mà và chính xác qua từng khung hình.

### Phương pháp theo dõi tâm

**1) Theo dõi đối tượng:**

Đối tượng được theo dõi và vị trí của chúng được xác định.

**2) Kiểm tra sự tồn tại của đối tượng:**

Kiểm tra xem nó có tồn tại trong danh sách object đang theo dõi không. Việc kiểm tra sự tồn tại giúp tránh nhầm lẫn và đồng thời cập nhật thông tin về đối tượng một cách chính xác.

A screenshot of a computer

Description automatically generated

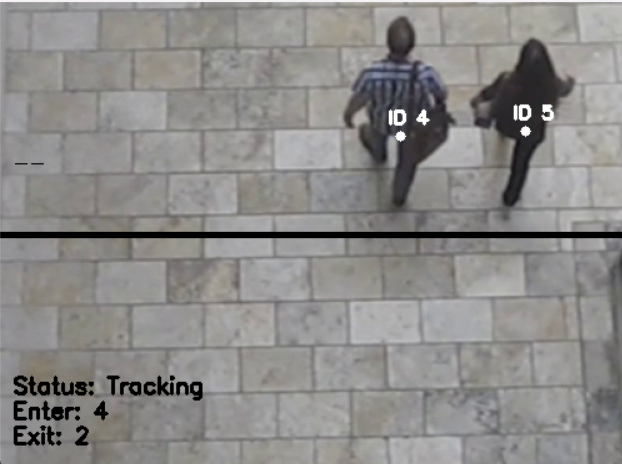
**3) Nhận biết hướng di chuyển:**

Từ sự thay đổi của vị trí tâm, hệ thống có thể dự đoán và nhận biết hướng di chuyển của đối tượng. Điều này mở ra khả năng phân tích sự tương tác và dự đoán động lực đằng sau hành vi của đối tượng.

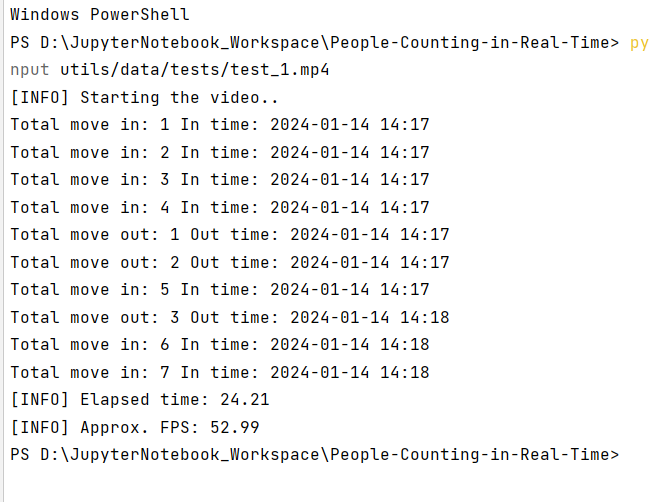
**4) Đếm đối tượng:**

Đối với quá trình đếm, ngoài việc ghi lại số liệu còn phải phân loại và theo dõi các biểu hiện phức tạp như sự tăng tốc, giảm tốc, hoặc về việc bến đổi về hình dạng của đối tượng.

## Kết quả



**Kết quả chạy trực tiếp trên từng khung hình**



**Kết quả hiển thị**

## Đánh giá

Tổng cộng, quy trình này bao gồm tiền xử lý hình ảnh, phát hiện đối tượng bằng mạng nơ-ron và theo dõi đối tượng bằng cách sử dụng phương pháp dựa trên trọng tâm để phát hiện và đếm số người trong các tình huống đông người. CentroidTracker là một thuật toán theo dõi đối tượng đơn giản nhưng hiệu quả, sử dụng khoảng cách Euclidean giữa trọng tâm đối tượng để theo dõi chuyển động của chúng qua các khung.

Thuật toán duy trì một tập hợp các trọng tâm đối tượng đã phát hiện trước đó và so khớp chúng với tập hợp hiện tại các trọng tâm đã phát hiện bằng cách sử dụng thuật toán Hungary, tìm ra sự gán định tối ưu của trọng tâm đối tượng dựa trên việc giảm thiểu tổng khoảng cách Euclidean của chúng.

Các tọa độ hộp chứa được trả về bởi mô hình MobileNet SSD được sử dụng để tạo ra các trọng tâm đối tượng, sau đó được theo dõi bằng mô-đun CentroidTracker. Số người trong phòng được xác định bằng cách đếm số ID đối tượng duy nhất đã được theo dõi theo thời gian.

Sau khi các trọng tâm đối tượng đã được so khớp, CentroidTracker cập nhật vị trí của các đối tượng đã theo dõi và thêm bất kỳ đối tượng mới nào được phát hiện vào tập hợp. Điều này cho phép thuật toán theo dõi các đối tượng qua các khung hình ngay cả khi chúng tạm thời bị che khuất hoặc ra khỏi tầm nhìn.

# Kết luận

Tóm lại, hệ thống đếm và theo dõi người có thể cung cấp thông tin giám sát lượng người ra vào một khu vực cụ thể. Những hệ thống này có thể đếm và theo dõi khách khi họ vào và ra khỏi một khu vực, cung cấp báo cáo riêng biệt cho mỗi trường hợp. Dữ liệu này đặc biệt hữu ích đối với những đối tượng kinh doanh muốn hiểu về hành vi của khách hàng, bao gồm thời điểm họ vào, cách họ di chuyển qua không gian và khi họ ra khỏi đó. Hơn nữa, công nghệ đếm người có thể cho thấy khách truy cập đang dành thời gian ở đâu và những sản phẩm nào thu hút nhiều sự chú ý nhất.

Hệ thống đếm người trong lĩnh vực bán lẻ có thể xác định số lượng khách truy cập chính xác và theo dõi tỷ lệ những người đó thực sự mua hàng. Thông tin này quan trọng để hiểu xem chiến dịch tiếp thị có hiệu quả không, thu hút khách truy cập và chuyển đổi họ thành khách hàng. Trong các trung tâm mua sắm, công nghệ đếm người có thể giúp các nhà quản lý hiểu rõ về những khu vực thu hút nhất và làm thế nào lưu lượng thay đổi theo thời gian. Thông tin này có thể được sử dụng để hình thành chiến dịch tiếp thị và lập kế hoạch cho các đợt cao điểm.

Hệ thống đếm người cũng hữu ích trong các triển lãm và sự kiện, nơi chúng có thể giúp các quản lý hiểu về giờ cao điểm của sự kiện và tối ưu hóa phân bổ nhân viên tương ứng. Ngoài ra, hệ thống đếm người có thể cung cấp tổng số lượng khách tham dự sự kiện, đánh giá mức độ thành công của sự kiện và lập kế hoạch cho sự kiện trong tương lai.

Nhận dạng và nhận biết đối tượng là các thành phần quan trọng của hệ thống đếm và theo dõi người. Các công nghệ này quan trọng cho các ứng dụng thời gian thực và có thể được sử dụng để phát hiện bất kỳ loại đối tượng nào trong một không gian vật lý. Bằng cách kết hợp nhận dạng và nhận biết đối tượng với theo dõi đối tượng dựa trên trọng tâm, hệ thống đếm và theo dõi người có thể theo dõi chính xác các chuyển động của người trong video và đếm số lượng người vào và ra một cách thời gian thực. Công nghệ này có thể được áp dụng trong nhiều tình huống, chẳng hạn như trong cửa hàng bán lẻ, sân bay và các không gian công cộng khác nơi theo dõi và theo dõi chuyển động của người là cần thiết.

TÀI LIỆU THAM KHẢO

[1] R. M. Gurav and P. K. Kadbe, "Real time finger tracking and contour detection for gesture recognition using OpenCV," 2015 International Conference on Industrial Instrumentation and Control (ICIC), Pune, India, 2015, pp. 974-977, doi: 10.1109/IIC.2015.7150886.

[2] https://pyimagesearch.com/2018/08/13/opencv-people-counter/

[3] https://pyimagesearch.com/2018/10/22/object-tracking-with-dlib/

[4] G. Gamage, I. Sudasingha, I. Perera and D. Meedeniya, "Reinstating Dlib Correlation Human Trackers Under Occlusions in Human Detection based Tracking," 2018 18th International Conference on Advances in ICT for Emerging Regions (ICTer), Colombo, Sri Lanka, 2018, pp. 92-98, doi: 10.1109/ICTER.2018.8615551.

[5] Shahbaz, Ajmal & Hariyono, Joko & Jo, Kang-Hyun. (2015). Evaluation of background subtraction algorithms for video surveillance. 10.1109/FCV.2015.7103699.

[6] Speed/accuracy trade-offs for modern convolutional object detectors

[7] Lightweight SSD: Real-time Lightweight Single Shot Detector for Mobile Devices

[8] Shubham Mishra, Mrs. Versha Verma, Dr. Nikhat Akhtar, Shivam Chaturvedi, Dr. Yusuf Perwej, "An Intelligent Motion Detection Using OpenCV", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN: 2394-4099, Print ISSN: 2395-1990, Volume 9 Issue 2, pp. 51-63, March-April 2022.