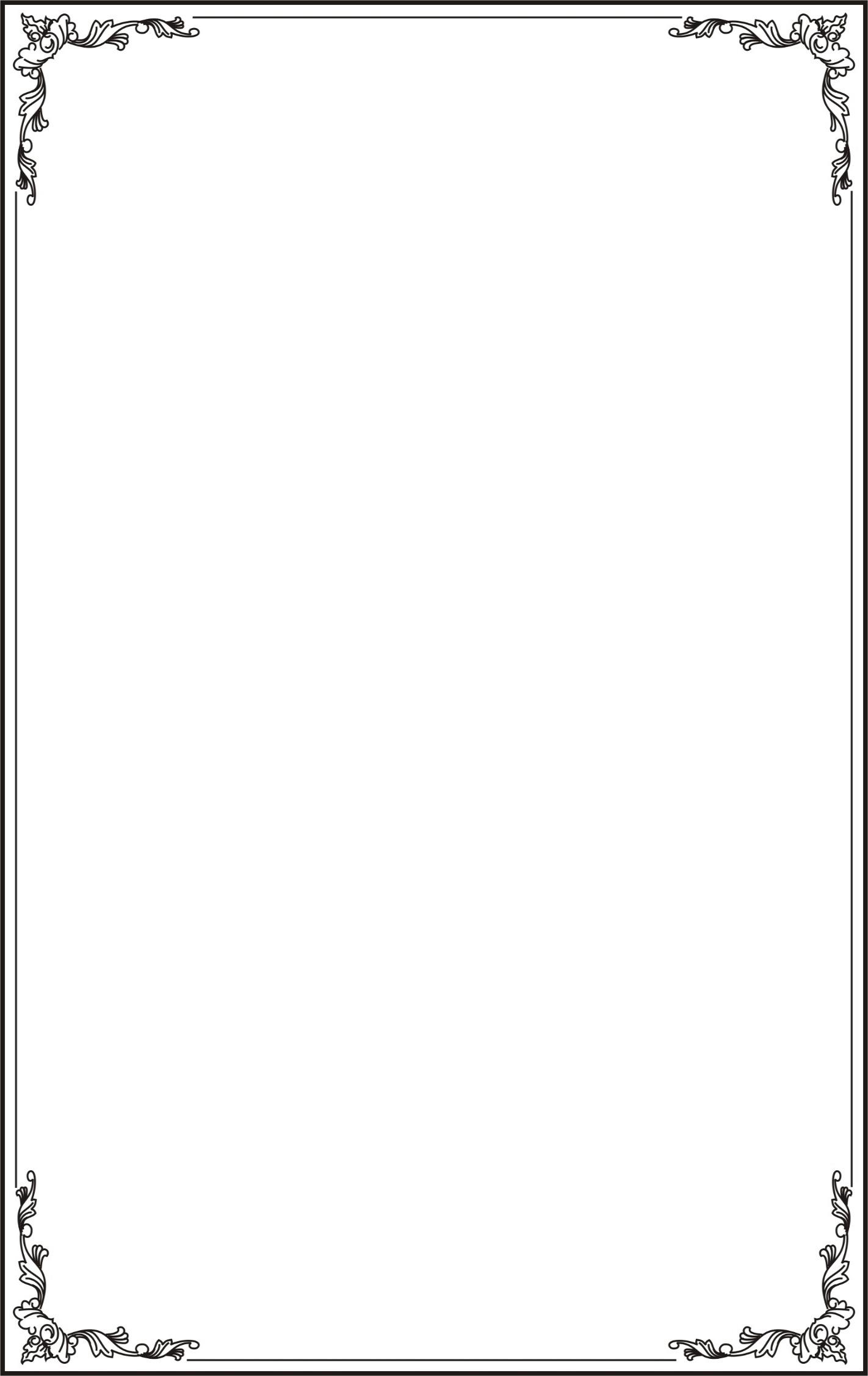
**ĐẠI HỌC QUỐC GIA TP.HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC BÁCH KHOA**

**KHOA KHOA HỌC & KỸ THUẬT MÁY TÍNH**

**🙚🙚🕮🙘🙘**

**ĐỀ CƯƠNG LUẬN VĂN THẠC SĨ**

**GÁN NHÃN ĐỐI TƯỢNG DI CHUYỂN QUA NHIỀU CAMERA**

**(TO ASSIGN LABEL FOR MOVING OBJECTS IN MULTIPLE CAMERAS)**

**GVHD: TS. Nguyễn Thanh Bình**

**---o0o---**

**HVTH : Võ Hoàng An 1670211**

**TP. Hồ Chí Minh – Tháng 6/2018**

# MỤC LỤC

**Nội dung Trang**

[MỤC LỤC 2](#_Toc510985424)

[Chương I. GIỚI THIỆU 1](#_Toc510985425)

[***1.1.*** ***GIỚI THIỆU ĐỀ TÀI*** 1](#_Toc510985426)

[***1.2.*** ***MỤC TIÊU VÀ NỘI DUNG ĐỀ TÀI*** 1](#_Toc510985427)

[***1.2.1.*** ***Mục tiêu đề tài*** 1](#_Toc510985428)

[***1.2.2.*** ***Nội dung đề tài*** 1](#_Toc510985429)

[***1.3.*** ***GIỚI HẠN ĐỀ TÀI*** 1](#_Toc510985430)

[***1.4.*** ***ĐÓNG GÓP CỦA ĐỀ TÀI*** 2](#_Toc510985431)

[***1.4.1.*** ***Đóng góp về mặt khoa học*** 2](#_Toc510985432)

[***1.4.2.*** ***Đóng góp về mặc thực tiễn*** 2](#_Toc510985433)

[***1.5.*** ***PHƯƠNG PHÁP NGHIÊN CỨU*** 2](#_Toc510985434)

[***1.6.*** ***CẤU TRÚC LUẬN VĂN*** 2](#_Toc510985435)

[Chương II. CƠ SỞ LÝ THUYẾT VÀ CÁC NGHIÊN CỨU LIÊN QUAN 3](#_Toc510985436)

[***2.1*** ***CƠ SỞ LÝ THUYẾT*** 3](#_Toc510985439)

[**2.1.1** **Cắt frame từ đoạn video** 3](#_Toc510985440)

[**2.1.2** **Phát hiện đối tượng di chuyển** 3](#_Toc510985441)

[**2.1.3** **Rút trích đặc trưng của đối tượng** 4](#_Toc510985442)

[**2.1.4** **Gán nhãn đối tượng** 5](#_Toc510985443)

[**2.2** ***CÁC NGHIÊN CỨU LIÊN QUAN*** 5](#_Toc510985444)

[**2.2.1.** **Nhiều camera theo dõi vùng không gian trùng lắp** 6](#_Toc510985449)

[**2.2.2.** **Nhiều camera theo dõi vùng không gian không trùng lắp** 7](#_Toc510985450)

[Chương III. PHƯƠNG PHÁP ĐỀ XUẤT GÁN NHÃN ĐỐI TƯỢNG DI CHUYỂN QUA NHIỀU CAMERA 9](#_Toc510985451)

[***3.1.*** ***MÔ TẢ BÀI TOÁN*** 9](#_Toc510985455)

[***3.2.*** ***PHƯƠNG PHÁP ĐỀ XUẤT*** 9](#_Toc510985456)

[**3.2.1.** **Chuẩn bị dữ liệu** 9](#_Toc510985457)

[**3.2.2.** **Phát hiện đối tượng** 9](#_Toc510985458)

[**3.2.3.** **Rút trích đặc trưng đối tượng** 10](#_Toc510985459)

[**3.2.4.** **Gán nhãn cho đối tượng** 11](#_Toc510985460)

[**3.2.5.** **Dữ liệu đầu ra** 12](#_Toc510985461)

[***3.3.*** ***PHƯƠNG PHÁP ĐÁNH GIÁ*** 12](#_Toc510985462)

[Chương IV. THÍ NGHIỆM VÀ ĐÁNH GIÁ KẾT QUẢ 13](#_Toc510985463)

[**4.1.** **TẬP DỮ LIỆU ĐÁNH GIÁ** 13](#_Toc510985464)

[**4.2.** **KẾT QUẢ THÍ NGHIỆM** 13](#_Toc510985465)

[Chương IV. KẾT LUẬN 13](#_Toc510985466)

[**5.1.** **KẾT QUẢ ĐẠT ĐƯỢC** 13](#_Toc510985467)

[**5.2.** **ƯU ĐIỂM VÀ NHƯỢC ĐIỂM CỦA PHƯƠNG PHÁP ĐỀ XUẤT** 13](#_Toc510985468)

[**5.3.** **HƯỚNG MỞ RỘNG** 13](#_Toc510985469)

[TÀI LIỆU THAM KHẢO 14](#_Toc510985470)

# Chương I. GIỚI THIỆU

## ***GIỚI THIỆU ĐỀ TÀI***

## ***MỤC TIÊU VÀ NỘI DUNG ĐỀ TÀI***

* + 1. ***Mục tiêu đề tài***

Chúng ta có thể thấy được tầm quan trọng của thị giác máy tính trong việc thay thế mắt người to lớn như thế nào. Từ việc giám sát, thống kê đến việc vận hành tự động của máy móc, thiết bị, dây chuyền sản xuất, các thiết bị tự động … Chính vì thế mà ngày càng có nhiều công trình nghiên cứu nhằm cải thiện và phát triển thị giác máy tính. Với đề tài này, tôi đặt mục tiêu xây dựng một hệ thống có thể phát hiện đối tượng di chuyển thông qua hai camera quan sát. Một hệ thống như vậy rất có ích cho việc giám sát và truy vết đối tượng, có thể được áp dụng phổ biến trong các hệ thống giám sát sân bay, nhà ga giúp cho việc tìm người bị thất lạc nhanh hơn, ít tốn nhân lực hơn và không gây hoang mang cho những người khác, hay cũng có thể áp dụng cho việc giữ an ninh đối với các cơ quan, tổ chức và các nơi công cộng. Khi có bất kì đối tượng trộm cắp, cướp bóc hay thậm chí là đối tượng tình nghi có khả năng gây ra một hành động phạm tội nào đó, bằng việc phân tích thêm hành vi của đối tượng, ta cũng có thể phát hiện sớm và đưa ra cảnh báo cho các cá nhân, tổ chức có liên quan đến việc đảm bảo an ninh của khu vực đó. Mục tiêu của đề tài này vẫn chỉ ở mức làm sao có thể phát hiện đối tượng di chuyển qua nhiều camera nhằm truy vết đối tượng.

* + 1. ***Nội dung đề tài***

Như đã giới thiệu trong chương mục tiêu của đề tài, tôi muốn xây dựng một hệ thống có thể phát hiện đối tượng di chuyển qua nhiều camera nhằm truy vết đối tượng. Đối với hệ thống này, tôi sẽ:

- Sử dụng hai camera có thể trùng lắp (hai camera được đặt ở vị trí mà tồn tại khoảng tầm nhìn cả hai camera đều thấy được) hoặc không trùng lắp (hai camera được đặt ở hai vị trí khác nhau và không có tồn tại khoảng tầm nhìn trùng lắp giữa hai camera) để giám sát các đối tượng di chuyển trong nhà.

- Với hai camera này, tôi sẽ có được hai đoạn video giám sát và từ đó tôi sẽ sử dụng phương pháp đề xuất trong chương III để gán nhãn cho từng đối tượng di chuyển trong các camera đó.

## ***GIỚI HẠN ĐỀ TÀI***

Trong vấn đề nhận dạng đối tượng di chuyển qua nhiều camera thông qua việc gán nhãn đối với đối tượng, chúng ta có rất nhiều hướng để phát triển mà mở rộng. Với đề tài này của mình, tôi muốn xoáy quanh việc giải quyết mục tiêu mà mình đặt ra. Đó là:

- Chỉ hiện thực trên hai camera được thiết kế theo nội dung của đề tài.

- Từ hai đoạn video thu được từ hai camera, tôi có thể gán nhãn thành công các đối tượng di chuyển trong mỗi đoạn video, đồng thời có thể đảm bảo được tính nhất quán, chính xác của dữ liệu.

- Đảm bảo tính nhất quán trong việc gán nhãn đối tượng khi đối tượng đó di chuyển qua hai camera. Có nghĩa là khi một đối tượng di chuyển từ camera này sang camera khác, hệ thống mà tôi đề xuất phải đảm bảo gán nhãn hai đối tượng di chuyển thu được trên hai camera đó là giống nhau.

- Thực hiện đo đạt kết quả của hệ thống đối với cả hai trường hơp đặt camera trùng lắp và không trùng lắp, từ đó so sánh đâu là cách đặt camera cho độ chính xác cao hơn.

* 1. ***ĐÓNG GÓP CỦA ĐỀ TÀI***
     1. ***Đóng góp về mặt khoa học***
     2. ***Đóng góp về mặc thực tiễn***
  2. ***PHƯƠNG PHÁP NGHIÊN CỨU***
  3. ***CẤU TRÚC LUẬN VĂN***

Bài báo cáo sẽ được chia làm năm chương chính: chương một sẽ giới thiệu sơ qua về đề tài, mục tiêu và nội dung cũng như những giới hạn của nó trong khi hiện thực; chương hai sẽ giới thiệu cơ sở lý thuyết và các nghiên cứu liên quan đến đề tài; chương ba là chương chính mô tả về những yêu cầu của bài toán, phương pháp mà tôi đề xuất để giải quyết các bài toán đó và phần phương pháp đánh giá để xác định được phương pháp đề xuất này hiệu quả hay không trên các ngữ cảnh mà tôi đã đặt ra trong đề tài này; chương bốn giới thiệu về tập dữ liệu mà tôi sử dụng để hiện thực phương pháp mà mình đề xuất, từ đó thực hiện thí nghiệm trên tập dữ liệu này để thu được kết quả và đưa ra nhận xét, đánh giá đối với phương pháp được đề xuất. Từ những kết quả thu được, tôi sẽ đưa ra nhận xét, đánh giá ở chương năm. Kết luận về ưu nhược điểm cũng như những nguyên nhân dẫn đến các ưu nhược điểm này, đồng thời có những bình luận về chúng để đưa ra hướng mở rộng cho phương pháp nhằm cải tiến phương pháp đề xuất để thu được kết quả tốt hơn.

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



## ***CƠ SỞ LÝ THUYẾT***

* + 1. **Cắt frame từ đoạn video**

Đối với thị giác máy tính, chúng ta không thể xử lý trực tiếp trên dữ liệu là một đoạn video được, mà chúng ta cần phải chuyển các đoạn video này sang tập các frame ảnh. Đoạn video cơ bản được tạo thành từ một tập hợp các frame ảnh liên tiếp nhau. các frame này là các ảnh chụp được tại một thời điểm cụ thể. Do đó mà việc xử lý video sẽ được ánh xạ sang việc xử lý từng frame ảnh. Việc chuyển từ video sang frame ảnh ta có thể sử dụng các chương trình bổ trợ như matlab 7.12.0, FFMPEG, OpenCV, EmguCV…Tất cả các chương trình này được áp dụng rất phổ biến trong lĩnh vực xử lý ảnh và yêu cầu cụ thể ở đây là chuyển video sang các frame ảnh.

FFMPEG là một thư viện có nhiều tiện ích hỗ trợ cho việc xử lý video. Tính năng nổi bật nhật là khả năng encode/decode nhiều định dạng video khác nhau, giúp chuyển video qua từ định dạng này sang định dạng khác. Ngoài ra, FFMPEG còn hỗ trợ cắt đoạn video để chuyển sang dạng frame và từ frame chuyển sang các file ảnh.

OpenCV có lẽ là một bộ thư viện khá phổ biến đối với hầu hết những người làm việc trong lĩnh vực xử lý ảnh. OpenCV là một thư viện mã nguồn mở cung cấp các interface C/C++, Java, Python và hỗ trợ cho Windows, Linux, Mac OS, iOS và cả Android. Các đặc trưng nổi bậc có thể kể đến rút trích đặc trưng thông qua các giải thuật như PCA…, phát hiện đối tượng như khuôn mặt, người, xe hơi, xử lý đoạn video và chuyển sang các frame…

EmguCV cơ bản là một OpenCV nhưng nó được tạo ra để hỗ trợ phát triển trên ngôn ngữ C#, vì vậy nên nó có đầu đủ các tính năng nổi bật của OpenCV.

* + 1. **Phát hiện đối tượng di chuyển**

Có rất nhiều phương pháp đã được đưa ra để phát hiện đối tượng di chuyển trong một đoạn video. Ta có thể chia ra thành các lớp: Point detectors, background subtraction, segmentation, supervisioned learning. Đơn giản nhất và cổ điển nhất là background subtraction. Tuy nhiên, phương pháp đem lại hiệu quả cao nhất và có tính khoa học lại phải kể đến phương pháp supervisioned learning.

Phương pháp supervisioned learning gọi theo tiếng việt là học có giám sát, đây là một phương pháp khá là phổ biến trong thời gian gần đây và có khuynh hướng sẽ trở thành phương pháp chính hỗ trợ không chỉ trong việc phát hiện đối tượng mà còn cả trong các mặt ứng dụng khác của thị giác máy tính nói riêng và trí tuệ nhân tạo nói chung. Phương pháp này mang tính khoa học ở chỗ nó mô phỏng quá trình học của con người để áp dụng cho máy tính. Con người thông qua quá trình sống, học tập và làm việc mà ghi nhận được những kiến thức khoa học thì ở đây máy tính cũng được học từ những kiến thức khoa học mà con người thu nhận và truyền tải cho nó để biến nó trở thành tri thức mà máy có thể hiểu và thực thi.

Một cách cụ thể hơn bằng ngôn ngữ tự nhiên, con người học được cách nhận dạng một chiếc xe máy bằng việc tiếp nhận vô số các hình ảnh, âm thanh cũng như các tính chất của xe máy để từ đó lọc ra được những đặc trưng mà một chiếc xe máy có thể có và như thế ta học được cách nhận dạng đối tượng nào là xe máy. Còn đối với máy tính, ta cũng cung cấp tập dữ liệu gọi là tập huấn luyện. Tập dữ liệu này cũng tương tự là những hình ảnh, âm thanh, các tính chất của một chiếc xe máy. Với tập dữ liệu này, chúng ta sẽ xây dựng phương pháp để máy tính tính toán, xây dựng mẫu đại diện cho đối tượng xe máy. Cuối cùng máy tính có thể đưa ra quyết tịnh một đối tượng có phải là xe máy hay không dựa trên mức độ tương đồng của đối tượng đó với mẫu đã xây dựng.

* + 1. **Rút trích đặc trưng của đối tượng**

Đối với các phương pháp rút trích đặc trưng của đối tượng, mục tiêu là tìm ra được các đặc trưng thể hiện sự khác biệt giữa đối tượng này với các đối tượng khác. Một trong số đó có thể kể đến như:

1. **Màu**

Màu là một trong những đặc trưng quan trọng nhất giúp cho việc phân biệt đối tượng. Màu rất dễ để phân tích thông qua frame ảnh cũng như ý tưởng khá đơn giản. Chất lượng của đặc trưng màu phụ thuộc lớn vào không gian màu sử dụng để biểu diễn đối tượng. Một số phương pháp rút trích đặc trưng màu của đối tượng phổ biến là moment màu, moment màu mờ, biểu đồ màu…Chính vì có nhiều phương pháp rút trích đặc trưng màu mà nó trở nên rất phổ biến trong việc nhận diện đối tượng trong ảnh.

1. **Hình dáng**

Hình ảnh cũng khá là quan trọng trong rút trích đặc trưng đối tượng. Khi một đối tượng quay lưng lại hoặc thay đổi trang phục thì đặc trưng màu sắc không còn có thể giúp phân biệt các đối tượng được nữa, nhưng hình dáng của đối tượng thì vẫn sẽ không thay đổi nhiều, một người béo thì không thể gầy ngay trong khi di chuyển qua camera được và ngược lại. Một đặc trưng về hình dáng tốt là một đặc trưng mà không bị ảnh hưởng bởi sự thay đổi hình dáng do quá trình di chuyển, xoay hay thu phóng kích thước của đối tượng. Phương pháp hiệu quả nhất là sử dụng moment bất biến của hình dáng. Phướng pháp này sẽ giúp rút trích ra các vector đặc trưng không thay đổi dựa trên hình dáng của đối tượng.

1. **Kết cấu**

Kết cấu là một trong những đặc trưng quan trọng trong việc phân tích ảnh cho nhiều ứng dụng. Kết cấu phân tích trực quan hình ảnh thông qua các thuật ngữ về độ thô, mịn, mượt…như là một hàm chỉ sự biến đổi trong không gian về cường độ của điểm ảnh. Đặc trưng kết cấu có thể được rút trích thông qua các phương pháp như Gabor Filter, Haar Wavelet Decomposition và Wavelet GLCM…

1. **Vị trí**

Vị trí được xem là một đặc trưng tức thời của đối tượng, nó thể hiện vị trí của đối tượng trong không gian tại một thời điểm xác định. Trong một khoảng thời gian dài thì đặc trưng này không có tác dụng trong việc nhận dạng đối tượng, nhưng xét về một khoảng thời gian ngắn, nó lại trở thành một đặc trưng rất có giá trị và có độ tin cậy cao trong lĩnh vực thị giác máy tính. Nguyên nhân bởi vì trong một khoảng thời gian ngắn, một đối tượng thông thường không thể di chuyển một khoảng không gian lớn được. Do đó, dựa vào vị trí hiện tại của đối tượng, ta có thể biết được đối tượng này là đối tượng nào trong tập các đối tượng quan sát được, hỗ trợ mạnh mẽ cho bài toán theo dõi đối tượng.

* + 1. **Gán nhãn đối tượng**

Gán nhãn thực chất là một quá trình nhận dạng đối tượng thông qua các đặc trưng sinh trắc của đối tượng đó. Có hai lớp gán nhãn chính mà hầu hết mọi người đều biết đó là phân lớp và phân cụm.

Phân lớp là quá trình học có giám sát. Đây là lớp giải thuật mà đầu vào dữ liệu huấn luyện đã được cho biết trước nhãn của nó. Quá trình huấn luyện được thiết lập để xây dụng được một mô hình tốt đủ để khi một mẫu thử mới chưa được gán nhãn, thông qua mô hình này có thể xác định được nhãn của nó.

Phân cụm là quá trình học không có giám sát. Học không có giám sát là khi mình xây dựng mô hình dựa trên dữ liệu huấn luyện mà dữ liệu này không được cho biết trước nhãn của nó. Sau quá trình huấn luyện ta sẽ được một mô hình mà khi một mẫu thử mới tới, thông qua mô hình huấn luyện này mình sẽ biết nó thuộc nhãn nào trong tất cả các nhãn đã được xây dựng trong tập huấn luyện.

Việc phân lớp và phân cụm thường đòi hỏi phải có một lượng lớn dữ liệu để có thể huấn luyện được một mô hình tốt, phục vụ cho việc gán nhãn. Dữ liệu càng lớn, quá trình huấn luyện càng tốn thời gian nhưng tính chính xác của mô hình được xây dựng được lại càng cao.

## ***CÁC NGHIÊN CỨU LIÊN QUAN***

Từ khi hình thành, thị giác máy tính phát triển rất nhanh và dần phổ biến mà ngày nay hầu hết các lĩnh vực đều cần đến như truy vết tội phạm, phát hiện người vi phạm giao thông, tính toán mật độ xe để điều tiết giao thông…tất cả đều xoay quanh bài toán cốt yếu nhất, đó là bài toán nhận dạng đối tượng.

Tiziana [4] đã nghiên cứu phát hiện các đối tượng di chuyển qua hệ thống camera giám sát thế hệ thứ ba. Tác giả giới thiệu rằng ở hệ thống giám sát thế hệ thứ ba này, hệ thống cần phải cung cấp đầy đủ các dịch vụ kĩ thuật số từ cấp độ của cảm biến, đến việc đưa ra các thông tin đã được hình ảnh hóa hộ trợ cho việc thực hiện thao tác đơn giản hơn, nghĩa là người sử dụng có thể thấy được các thông tin đã được trực quan hóa.

Shipra [1] lại làm một cuộc khảo sát rất chi tiết về các kĩ thuật trong xử lý ảnh để giải quyết bài toán truy vết đối tượng trong camera quan sát. Tác giả bắt đầu từ việc giới thiệu làm thế nào để biểu diễn được đối tượng thông qua các đặc trưng quan trọng của nó, từ đó tìm hiểu các phương pháp phát hiện đối tượng thông các đặc trưng biểu diễn đối tượng bao gồm: background subtraction, temporal differencing hay optical flow đến các phương pháp truy vết đối tượng dựa trên đặc trưng đối tượng, dựa trên hình dáng viền và bên trong của đối tượng…Bên cạnh đó việc truy vết dựa trên màu sắc của đối tượng cũng là một đặc trưng được tác giả đặc biệt quan tâm bởi sự đơn giản và dễ rút trích, nhận dạng của nó.

Còn đối với Shantaiya [9], tác giả nghiên cứu về các phương pháp truy vết đối tượng thông qua việc tính toán trạng thái của đối tượng thông qua đặc trưng di chuyển của đối tượng. Trạng thái của đối tượng mới sẽ được dự đoán dựa trên các thông số của trạng thái hiện tại, từ đó có thể giải quyết được bài toán về phủ lấp của đối tượng. Trong khi sử dụng các phương pháp truy vết dựa trên đặc trưng bên ngoài của đối tượng như hình dáng, màu sắc, ta không thể nhận dạng được đối tượng khi đối tượng bị che khuất bởi một đối tượng khác hoặc bơi hình nền. Tuy nhiên, với các phương pháp tính toán trạng thái của đối tượng để dự đoán vị trí của nó trong frame ảnh tiếp theo như thế này, ta có thể dễ dàng ước lượng được vị trí của đối tượng ngay cả trong trường hợp đối tượng bị che khuất. Tác giả nghiên cứu hai phương pháp truy vết khá nổi tiếng là: Kalman filter và Optical flow. Từ việc đánh giá điểm mạnh, điểm yếu của cả hai phương pháp này mà tác giả đã cải tiến giải thuật optical flow để đạt hiệu quả cao hơn, nâng cao cả độ chính xác trong việc truy vết cũng như giảm thời gian tính toán.

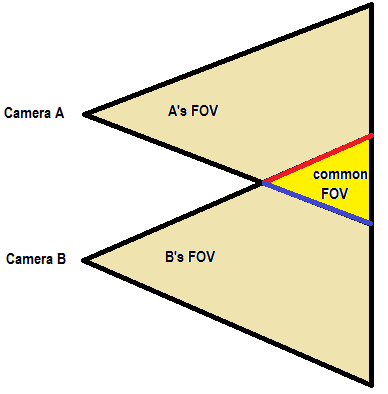
Trong môi trường thực tiễn việc nhận dạng đối tượng thông qua camera không chỉ đơn giản là nhận dạng trên duy nhất một camera, vì phạm vi theo dõi của một camera rất giới hạn nên không thể đáp ứng được yêu cầu thiết yếu của thực tiễn khi mà việc theo dõi đối tượng có thể là xuyên xuốt qua một vùng không gian rộng lớn từ không gian giữa hai căn phòng đến không gian giữa các căn nhà hay thậm chí là không gian giữa các tuyến đường…, Ergys [5] đã làm một cuộc nghiên cứu về vấn đề độ hiệu quả và tập dữ liệu cho nhiều đối tượng truy vết qua nhiều camera. Trong nghiên cứu này, Ristani định nghĩa những biện pháp mới để tính toán hiệu suất theo dõi của nhiều đối tượng trong nhiều camera (MTMC). Tác giả cũng giới thiệu về tập dữ liệu đã được hiệu chỉnh lớn nhất cho tới thời điểm thực hiện nghiên cứu đó để so sánh các phương pháp theo dõi MTMC.

Đối với hệ thống theo dõi MTMC, có hai cách thiết lập camera đó là thiết lập nhiều camera theo dõi vùng không gian trùng lắp và nhiều camera theo dõi vùng không gian không trùng lắp.

2. 2. 1. **Nhiều camera theo dõi vùng không gian trùng lắp**

Thông thường trong hệ thống có nhiều camera quan sát, mỗi camera đơn phải có khả năng theo dõi đối tượng một cách độc lập với nhau. Các camera sẽ tương tác với nhau, chia sẻ thông tin giúp theo dõi đối tượng trong một khoảng thời gian dài hơn, không gian rộng hơn.

Với vùng không gian trùng lắp, mối quan hệ giữa vùng quan sát được của các camera sẽ được sử dụng như một dạng thông tin quan trọng. Chuyển tất cả các thông tin từ các camera trùng lắp lại thành một hệ thống không gian 3D là một cách tiếp cận khá phổ biến trong trường hợp này. Lalita [2] sử dụng một khái niệm gọi là góc nhìn (Field of view) của các camera khác nhau để kết nối các camera trùng lắp với nhau thành một thộng thống đồng nhất. FOV (Field of View) là vùng không gian mà mỗi camera theo dõi được và khi hai camera có vùng không gian phủ lắp thì có nghĩa là sẽ có một khoảng FOV nào đó mà cả hai camera đều thấy được cùng một hình ảnh (Hình 2.1).



Hình 2.1: Vùng không gian trùng lắp giữa hai camera.

Đường màu xanh được gọi là cạnh của FOV của camera A trong camera B và đường màu đỏ là cạnh của FOV của camera B trong camera A. Việc xác định được hai cạnh này cực kì quan trọng vì nó giúp cho hệ thống có thể kết nối FOV của các camera lại với nhau. Trường hợp một đối tượng đang nằm trong vùng giao nhau (Common FOV), ta chỉ cần tính tổng khoảng cách từ đối tượng đó đến hai cạnh (đỏ và xanh), nếu tổng khoảng canh bằng đúng khoảng cách từ cạnh đỏ đến cạnh xanh thì đối tượng đó trong hai camera là một. Đây chính là cách kết nối thông tin chia sẻ được từ các camera theo dõi vùng không gian trùng lắp.

* + 1. **Nhiều camera theo dõi vùng không gian không trùng lắp**

Trong việc theo dõi đối tượng trên một vùng không gian rộng, không phải lúc nào tất cả các camera đều có vùng không gian trùng lắp. Ví dụ trong trường hợp hai camera được đặt ở hai phòng sát bên nhau, FOV của cả hai camera sẽ là không gian căn phòng tương ứng với nơi đặt camera đó và do vậy ta không có vùng common FOV. Khi đó, việc theo dõi đối tượng thông qua các camera này là vô cùng khó khăn, thách thức. Chúng ta không có mối liên hệ giữa các camera, các đối tượng khi đi qua vùng không trùng lắp sẽ bị mất dấu, do đó mà việc so trùng dựa trên đặc trưng sinh trắc được sử dụng phổ biến để giải quyết thách thức này. Các thông tin thường được dùng để so trùng đối tượng thường bao gồm các đặc trưng về ngoại hình như màu sắc, hình dánh, kết cấu của đối tượng, các dấu hiệu về không thời gian hoặc là kết hợp cả đặc trưng ngoại hình và dấu hiệu về không thời gian của đối tượng.

Michael [7] đã sử dụng đặc trưng về điểm ảnh(point features) cấu tạo nên vị trí của các đối tượng để theo dõi chúng. Ở nghiên cứu này, tác giả cũng hiệu chỉnh giải thuật “multi-camera parallel tracking and mapping” (MCPTAM) để tăng khả năng theo dõi đối tượng trong thời gian thực.

Xiaojing [6] đề xuất sử dụng mô hình tập tham khảo dựa trên ngoại hình (reference set based appearance model)[8] để cải thiện việc theo dõi nhiều đối tượng trong mạng lưới các camera không có vùng không gian quan sát trùng lắm. Theo tác giả, việc sử dụng mạng lưới camera như thế rất dễ làm thay đổi ngoại hình của đối tượng như là điều kiện chiếu sáng, dáng đi, đứng của đối tượng cũng như chất lượng của ảnh trong camera, điều này rất dễ làm cho cùng một đối tượng nhưng trong các camera khác nhau lại bị nhận diện là hai đối tượng. Chính điều này mà tác giả đã không thực hiện so trùng đối tượng trực tiếp dựa trên các đặc trưng ngoại hình mà so trùng gián tiếp dựa trên tập tham khảo.

# Chương III. PHƯƠNG PHÁP ĐỀ XUẤT GÁN NHÃN ĐỐI TƯỢNG DI CHUYỂN QUA NHIỀU CAMERA



## ***MÔ TẢ BÀI TOÁN***

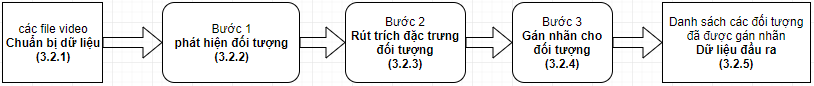
- Phát hiện đối tượng di chuyển trong từng camera.

- Gán nhãn cho các đối tượng này và đảm bảo tính nhất quán trong quá trình gán nhãn khi mà các đối tượng có thể di chuyển theo nhiều hướng khác nhau, hình dạng, kích thức sẽ thay đổi theo góc nhìn, vị trí xa gần đối với camera.

- Đảm bảo tính nhất quán trong việc gián nhãn đối tượng di chuyển qua hai camera. Hai camera này có thể được thiết lập một trong hai trường hợp: có vùng trùng lắp và không có vùng trùng lắp.

## ***PHƯƠNG PHÁP ĐỀ XUẤT***

Với các bài toán trên tôi không thể giải quyết thông qua một bước duy nhất, mà phải chia ra từng phần nhỏ để giải. Sơ đồ khối sau giúp cho tôi có thể giới thiệu khái quát về phương pháp mà tôi đề xuất để giải quyết tuần tự các yêu cầu đối với nội dung đề tài đã đặt ra.



Hình 3.1. Sơ đồ khối quy trình gán nhãn cho các đối tượng di chuyển

* + 1. **Chuẩn bị dữ liệu**

Với một hệ thống camera được thiết kế theo mô tả trong chương giới thiệu đề tài tôi sẽ thu được hai đoạn video định dạng dav tương ứng với từng camera. Tôi sẽ đặt tên file lần lượt là camera1.dav và camera2.dav.

* + 1. **Phát hiện đối tượng**

Đối với bài toán nhận dạng đối tượng di chuyển trong đoạn video sẽ có nhiều cách để giải quyết. Ở đây tôi sử dụng phương pháp mạng nơ ron học sâu để xây dựng mô hình đối tượng. Với tập dự liệu huấn luyện là một danh sách các ảnh về đối tượng di chuyển, qua mạng nơ ron học sau, tôi sẽ tạo ra mô hình phát hiện đối tượng. Từ mô hình phát hiện đối tượng đã được huấn luyện đó, tôi sẽ lần lượt đưa các frame ảnh có chứa đối tượng vào để mạng nơ ron xác định được đâu là đối tượng và vị trí của nó trong frame ảnh. Kết quả của quá trình này sẽ ra một frame có chứa các đối tượng di chuyển đã được xác định.

* + 1. **Rút trích đặc trưng đối tượng**

Đầu ra của quá trình nhận dạng đối tượng giúp ta tìm được vị trí thực của đối tượng trong foreground frame, từ đó ta tiến hành thu thập các đặc trưng nhận dạng của đối tượng bao gồm:

* + - 1. **Màu**

Có rất nhiều phương pháp rút trích đặc trưng cho đối tường từ màu, có thể kể đến đó là moment màu (Moment color), moment màu mờ (Fuzzy color moments), biểu đồ màu (Color histogram). Trong đó, moment màu là phương pháp được ưa chuộng nhất vì thời gian tính toán nhanh hơn so với moment màu mờ và độ chính xác cao hơn so với biểu đồ màu.

Moment màu: Với mỗi đối tượng, tôi lần lượt tính các giá trị Mean, Variance, skewness theo công thức sau:

(3.1)

(3.2)

(3.3)

*Trong đó, M và N là chiều cao và chiều rộng của khung ảnh chứa đối tượng.*

*P[i][j] là giá trị màu tại pixel [i][j].*

Khi đó vector đặc trưng của đối tượng trong dải màu RGB sẽ là:

(3.4)

* + - 1. **Hình dáng**

Để tăng tính chính xác của việc phân lớp, tôi sử dụng thêm phương pháp các moment bất biến của Hu (Hu’s invariant moments HIM) để rút trích đặc trưng hình dáng của đối tượng. Điểm mạnh của phương pháp này là loại bỏ được rào cản về sự thay đổi hình dáng của đối tượng bởi việc xoay, thay đổi kích thước, góc nhìn đối với camera. HIM là đặc trưng rất có lợi đối với ảnh 2 chiều, nếu chúng ta đại diện đối tượng R cho một khung ảnh, moment trung tâm của thứ tự (p + q) của R được định nghĩa là:

(3.5)

Trong đó, (xc,yc) là trung tâm của đối tượng.

Ta chuẩn hóa moment trung tâm theo công thức:

(3.6)

Dựa vào các moment trung tâm đã được chuẩn hóa, Hu giới thiệu bảy moment bất biến.

Bảy moment bất biến này là nhưng đặc trưng cực kì hữu dụng khi mà nó không bị thay đổi cho dù đối tượng trong ảnh có bị thay đổi kích thước, xoăy hoặc di chuyển theo các chiều khác nhau. Điều này giúp cho việc phân lớp đối tượng chính xác hơn trong môi trường thực tế đối với camera quan sát. Đối tượng có thể di chuyển qua lại theo nhiều hướng khác nhau hoặc di chuyển lại gần hay ra xa camera, tất cả những thay đổi đó đều làm cho hình dáng của đối tượng không còn giống như ban đầu nữa và khiến cho việc phân lớp đối tượng gặp khó khăn.

* + 1. **Gán nhãn cho đối tượng**

Quá trình gán nhãn là quá trình giải quyết mục tiêu chính của đề tài. Quá trình gán nhãn sẽ được thực thi tuần tự theo từng frame trong các đoạn video. Các đối tượng vì vậy mà sẽ được gán nhãn theo thứ tự nên trong bước này, tôi áp dụng lớp giải thuật incremental clustering. Lớp giải thuật incremental clustering dựa vào giả thuyết rằng các đối sẽ được truyển tới theo luồng và tại mỗi thời điểm có một đối tượng đến, khi đó ta sẽ gán nhãn đối tượng cho các lớp nhãn đã được xây dựng trước đó. Một đối tượng mới được gán nhãn vào một lớp nhãn sẽ không gây ảnh hưởng đến các phần tử của các lớp nhãn đã được xây dựng. Một trong những giải thuật cơ bản và đơn giản nhất đại diện cho lớp giải thuật này là giải thuật Leader.

Giải thuật leader bao gồm 2 bước chính:

Bước 1: Gán nhãn cho đối tượng đầu tiên O1 là nhãn C1, khởi tạo i = 1 và j = 1. Gọi Li = Oj là leader đại diện cho nhãn Ci.

Bước 2: gán j = j + 1; Để gán nhãn cho đối tượng Oj, tôi lần lượt tính độ sai lệch đặc trưng từ Oj đến lần lượt các leader Li (i = 1,..,n với n là tổng số lớp nhãn). Khi khoảng cách từ Oj đến Li là nhỏ nhất và nhỏ hơn ngưỡng do người dùng định nghĩa thì Oj sẽ được gán nhãn cho lớp Ci hoặc ngược lại thì Oj sẽ được gán một nhãn mới và quá trình gán nhãn cho đối tượng kết thúc.

Khi một đối tượng mới cần được gán nhãn tôi sẽ lặp lại bước 2, 3.

Ưu, nhược điểm của giải thuật leader:

- Ưu điểm: đơn giản, dễ hiện thực, độ phức tạp thấp.

- Nhược điểm: kết quả gán nhãn bị phụ thuộc vào thứ tự đối tượng đến gián nhãn.

* + 1. **Dữ liệu đầu ra**

Sau khi gán nhãn xong tôi thu được đối tượng đã được gán nhãn với các đặc trưng của nó. Dữ liệu này tôi sẽ lưu vào kho dữ liệu. Mỗi lần xử lý một frame, tôi sẽ có một danh sách các đối tượng di chuyển trong foreground frame với đầy đủ thông tin về đặc trưng màu sắc, hình dáng và nhãn của nó.

## ***PHƯƠNG PHÁP ĐÁNH GIÁ***

Với bài toán này, tôi sẽ có 2 tiêu chí để đánh giá độ chính xác của việc gán nhãn: tính nhất quán, chính xác của việc gán nhãn trên từng camera và tính nhất quán, chính xác của việc gán nhãn trên hai camera.

Tôi sẽ thực hiện gán nhãn trực tiếp cho các đối tượng di chuyển trong hai đoạn video camera1.mp4 và camera2.mp4. Sau đó, hai đoạn video này sẽ được đưa vào hệ thống mà tôi đã đề xuất để gán nhãn tự động cho các đối tượng. Tôi sẽ đưa ra kết quả dựa trên độ chính xác giữa gán nhãn tự động và kết quả thực tế. Tỉ lệ phần trăm của việc gán nhãn thành công, chính xác là thước đo để đánh giá tính hiệu quả của phương pháp đề xuất.

Tôi cũng thực hiện chạy phương pháp đề xuất của mình trên cả hai hệ thống thiết lập camera trùng lắp và không trùng lắp. Dựa trên việc so sánh kết quả thu được của hai hệ thống này, tôi sẽ đưa ra nhận xét về môi trường vận hành nào tốt hơn đối với phương pháp đề xuất của mình.

# Chương IV. THÍ NGHIỆM VÀ ĐÁNH GIÁ KẾT QUẢ

* 1. **TẬP DỮ LIỆU ĐÁNH GIÁ**

Với mục tiêu đã đặt ra trong chương I, tôi sẽ xây dựng một hệ thống gán nhãn đối tượng di chuyển qua hai camera được thiết lập trong cả hai trường hợp: có vùng không gian trùng lắp và không có vùng không gian trùng lắp. Kết quả dự kiến sau khi xây dựng hệ thống:

- Hệ thống có khả năng phát hiện đối tượng di chuyển trong hai đoạn video.

- Gán nhãn thành công và đảm bảo được tính nhất quán đối với việc gán nhãn. Một đối tượng di chuyển qua nhiều vị trí khác nhau trong đoạn video đều được gán nhãn giống nhau.

- Đảm bảo tính nhất quan trong việc gán nhãn đối với một đối tượng di chuyển qua cả hai camera.

- Tính chính xác của việc gán nhãn qua hai camera trùng lắp cao hơn đối với hệ thống hai camera không trùng lắp.

* 1. **KẾT QUẢ THÍ NGHIỆM**

# Chương IV. KẾT LUẬN

* 1. **KẾT QUẢ ĐẠT ĐƯỢC**
  2. **ƯU ĐIỂM VÀ NHƯỢC ĐIỂM CỦA PHƯƠNG PHÁP ĐỀ XUẤT**
  3. **HƯỚNG MỞ RỘNG**

# TÀI LIỆU THAM KHẢO

1. OJHA, Shipra; SAKHARE, Sachin. Image processing techniques for object tracking in video surveillance-A survey. In: Pervasive Computing (ICPC), 2015 International Conference on. IEEE. pp(1-6), **2015**.
2. GAVIT, Lalita, et al. Object Tracking Using Multiple Cameras. *IJ of Technology Enhancements and Emerging Engineering Reaearch*, vol.2, pp(36-41), **2014**.
3. Wu, Yi, Jongwoo Lim, and Ming-Hsuan Yang. "Object tracking benchmark." *IEEE Transactions on Pattern Analysis and Machine Intelligence* vol.37, pp(1834-1848), **2015**.
4. D'Orazio, Tiziana, and Cataldo Guaragnella. "A survey of automatic event detection in multi-camera third generation surveillance systems." *International Journal of Pattern Recognition and Artificial Intelligence*, vol.29, 1555001, **2015**.
5. Ristani, Ergys, et al. "Performance measures and a data set for multi-target, multi-camera tracking." *European Conference on Computer Vision*. Springer International Publishing, pp(17-35), **2016**.
6. Chen, Xiaojing, Le An, and Bir Bhanu. "Multitarget tracking in nonoverlapping cameras using a reference set." *IEEE Sensors Journal,* vol.15: pp(2692-2704), **2015**.
7. Tribou, Michael J., et al. "Multi-camera parallel tracking and mapping with non-overlapping fields of view." *The International Journal of Robotics Research*, pp(1-21), **2015**.
8. Chen, X., An, L., & Bhanu, B. (2013, October). Reference set based appearance model for tracking across non-overlapping cameras. in Proc. 7th ACM/IEEE Int. Conf. Distrib. Smart Cameras (ICDSC, pp(1–6), **2013**.
9. Shantaiya, Sanjivani, Kesari Verma, and Kamal Mehta. Multiple object tracking using kalman filter and optical flow. *European Journal of Advances in Engineering and Technology*, vol.2, pp(34-39), **2015**.