TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP

KHOA CÔNG NGHỆ THÔNG TIN

-----------------------------------------------------

ĐỒ ÁN MÔN HỌC

ĐỀ TÀI

Friendly Robot



Lớp: ĐHKHMT11A

Sinh viên thực hiện:

Trần Nguyễn Minh Thông 15073331

Đào Minh Sơn 15036781

Lê Quốc Cương 15092581

Giảng viên hướng dẫn: Tôn Long PhướcLời cảm ơn

Nhóm em xin gửi lời cảm ơn đối với thầy Tôn Long Phước đã tạo điều kiện cho nhóm em có cơ hội làm việc với nhau và đã nhiệt tình hướng dẫn để chúng em hoàn thành tốt bài làm về phần đồ án.

Trên thực tế không có sự thành công nào mà không gắn liền với những sự hỗ trợ, giúp đỡ dù ít hay nhiều, dù trực tiếp hay gián tiếp của người khác.

Trong suốt thời gian từ khi bắt đầu học đến nay, chúng em đã nhận được rất nhiều sự quan tâm, giúp đỡ của thầy cùng với tri thức và tâm huyết của mình thầy đã truyền đạt vốn kiến thức quý báu cho chúng em trong suốt thời gian học tập.

Trong quá trình làm bài báo cáo đồ án, khó tránh khỏi sai sót, rất mong thầy bỏ qua. Đồng thời do trình độ lý luận cũng như kinh nghiệm thực tiễn còn hạn chế nên bài báo cáo không thể tránh khỏi những thiếu sót, chúng em rất mong nhận được ý kiến đóng góp thầy để chúng em học thêm được nhiều kinh nghiệm và sẽ hoàn thành tốt hơn các bài báo cáo sắp tới.

# MỤC LỤC

[MỞ ĐẦU 4](#_Toc496237378)

[CƠ SỞ LÝ THUYẾT 6](#_Toc496237379)

[1. Microsoft Cognititive Service 6](#_Toc496237380)

[a. Vision API : Là nhóm API liên quan đến xử lý hình ảnh, trong nhóm Vision API này hiện Microsoft cung cấp 4 API sau: 6](#_Toc496237381)

[2. OpenCV 7](#_Toc496237386)

[a. Xác định khuôn mặt người (Face Detection): 7](#_Toc496237387)

[b. Face preprocessing: 8](#_Toc496237388)

[c. Collect and learn faces: 8](#_Toc496237389)

[d. Face recognition 8](#_Toc496237390)

[3. Nhận diện sự vật (Microsoft Computer Vision) 8](#_Toc496237391)

[4. FACE API 9](#_Toc496237392)

[5. Nhận diện danh tính người 10](#_Toc496237393)

[6. Tương tác phần cứng với Raspbery 11](#_Toc496237394)

[a. Kiến thức phần cứng 11](#_Toc496237395)

[b. Kiến thức về ngôn ngữ lập trình 11](#_Toc496237396)

[7. Web Socket 12](#_Toc496237397)

[a. Khái niệm: 12](#_Toc496237398)

[b. Ưu điểm 12](#_Toc496237399)

[c. Nhược điểm 13](#_Toc496237400)

[8. Web Front-End 13](#_Toc496237401)

[GIỚI HẠN ĐỀ TÀI 14](#_Toc496237402)

[1. Phần cứng 14](#_Toc496237403)

[2. Phần mềm 14](#_Toc496237404)

[3. Khả năng hiện tại 14](#_Toc496237405)

[TÀI LIỆU THAM KHẢO 15](#_Toc496237406)

# MỞ ĐẦU

Với cuộc cách mạng “**Công Nghiệp 4**” đang rộ lên ở thời gian gần đây thì việc sử dụng các thiết bị hiện đại đi kèm với giải thuật thông minh đã và đang được sự quan tâm phát triển trên khắp toàn thế giới. Những sản phẩm từ cuộc cách mạng CN4 này không những giúp hỗ trợ con người trong sản xuất, sinh hoạt, làm việc hằng ngày mà ngày càng đang tiến sâu hơn tới các lĩnh vực như giải trí, thư giãn.. Từ đó giúp con người có thể phát triển toàn diện từ thể chất đến tinh thần và nhất là giúp chúng ta có nhiều thời gian hơn để ở bên gia đình và bạn bè – nơi chứa những ý nghĩa lớn lao nhất của cuộc sống.

Tuy nhiên việc viết ra các thuật toán máy học, việc rèn luyện mô hình máy học… không phải dễ dàng và ai cũng có đủ kiến thức để lập trình nó. Với lí do đó Microsoft đã cho ra đời **Microsoft Cognitive Services**, một bộ dịch vụ giúp lập trình viên tích hợp những công nghệ tân tiến như AI vào trong các ứng dụng một cách dễ dàng. Sử dụng sức mạnh của machine learning vào ứng dụng chỉ với vài dòng code. Microsoft Cognititive Service ngày càng phát triển và được đánh giá cao vì các tính năng hữu ích của nó.

Một số những ứng dụng của Microsoft Cognitive Services hiện tại có thể nghĩ đến như nhận diện khuôn mặt trên điện thoại, nhận diện chủ nhân trong ngôi nhà, nhận diện danh tính của người ra vào một công ty hoặc tổ chức nào đó, tìm kiếm và chuẩn đoán một số đồ vật để phục vụ trong ngành cảnh sát…

Từ những dịch vụ hỗ trợ của Microsoft cũng như từ quan sát xã hội, con người… quan sát những điều gì là thiết yếu và cần thiết cho con người. Nhóm đã nảy ra ý tưởng sẽ làm một con robot có khả năng trợ giúp con người trong những công việc đơn giản hàng ngày như một người giúp việc, một người quản gia hay đặc biệt là cả một người bạn giúp tâm sự, giải tỏa những nỗi buồn, những áp lực trong cuộc sống hàng ngày!

Và với đề tài này nhóm em hiện tại đã hiện thực được với các dịch vụ của Microsoft là **Face API** (Có tích hợp **Emotion**) để nhận diện khuôn mặt, ngũ quan, phụ kiện, danh tính và cả cảm xúc cùng với **Computer Vision API** để nhận diện đồ vật, con người, khung cảnh... Hệ thống bao gồm chương trình xử lý chính được cài đặt chạy trên máy tính nhúng **Raspberry Pi 3 Model B**, cùng với các **cảm biến** và chương trình hiển thị trên nền web, kết nối với nhau thông qua mạng **wifi hoặc LAN**. Ngoài ra trên chương trình web còn được tích hợp **Voice Recognition** để nhận tín hiệu lời nói của con người và truyền xuống Raspi để xử lý, trong thời gian tới khi có đủ thiết bị sẽ tích hợp vào luôn bên dưới Raspi để có thể giao tiếp với người bằng giọng nói.

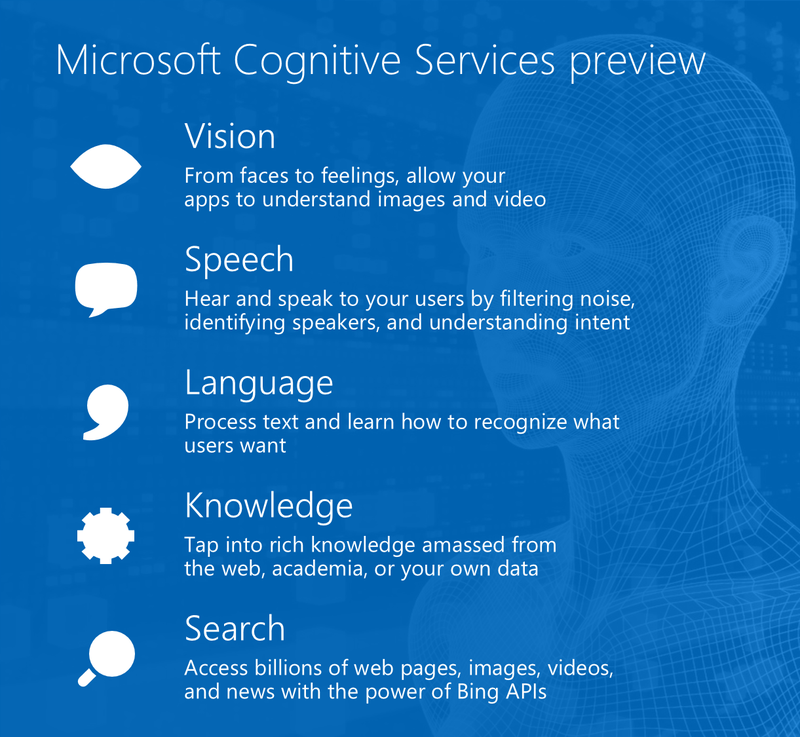
# PHẦN MỘT: CƠ SỞ LÝ THUYẾT

## Microsoft Cognititive Service

Bao gồm một bộ các API ứng dụng trí tuệ nhân tạo thông minh, cho phép lập trình viên ở mọi cấp độ từ những bạn sinh viên viết ứng dụng đầu tiên của mình hay những lập trình viên chuyên nghiệp làm việc cho những công ty, tổ chức lớn đều có thể tạo ra được những ứng thông minh hơn một cách dễ dàng.

Các API của Cognitive Services được viết dưới dạng REST API do vậy lập trình viên có thể tích hợp các API này trên nhiều nền tảng khác nhau như iOS, Android, hay Windows, chỉ cần có kết nối Internet.

Tính đến thời điểm viết bài viết này, Microsoft Cognitive Services bao gồm 21 API được chia thành 5 nhóm: Vision, Speech, Language, Knowledge và Search. Hãy cùng tìm hiểu 5 nhóm API này là gì?



### Vision API

Là nhóm API liên quan đến xử lý hình ảnh cũng là những API thông dụng và là phần API nhóm em sử dụng trong đề tài, trong nhóm Vision API này hiện Microsoft cung cấp 4 API sau:

1. Computer Vision API: API này cho phép trích xuất những thông tin có giá trị từ bức ảnh của bạn với khả năng xác định được kiểu đối tượng trong ảnh (là bánh mỳ, con chó, con mèo hay cây cối, …) hay nếu là người thì API này cũng xác định được giới tính của nhân vật trong ảnh. Ngoài ra, API này cũng hỗ trợ nhận diện được những nhận vật nổi tiếng hay trích xuất chữ có trong bức hình của bạn.
2. **Face API:** Cái tên nói lên tất cả, đây là API cho phép phát hiện khuôn mặt có trong bức hình của bạn. Ngoài ra, API này cũng trả về các thuộc tính của khuôn mặt như tuổi, giới tính, độ rạng ngời của nụ cười hay thậm chí là chiều dài của tóc, … Ngoài phát hiện khuôn mặt, Face API còn cho phép so sánh 2 khuôn mặt có phải là của cùng một người hay không.
3. **Emotion API:** API này cho phép xác định tâm trạng của người có trong bức hình xem họ đang vui, đang buồn hay đang giận dữ.
4. Video API: API này là một tập hợp các thuật toán xử lý video tân tiến của Microsoft. Với Video API, các nhà phát triển có thể tích hợp các tính năng chỉnh sửa video bao gồm chống rung, phát hiện khuôn mặt người, phát hiện chuyển động hay tạo video thumbnail.

### Speech API

Là nhóm API liên quan đến xử lý âm thanh với công nghệ xử lý đến từ Bing. Trong nhóm Speech API này hiện Microsoft cung cấp 3 API:

* 1. Bing Speech API:API này cho phép trích xuất một tập tin âm thanh sang dạng chữ, chuyển đổi định dạng chữ sang âm thanh (tức là đọc chữ) hay đoán ý của một câu nói.
  2. Custom Recognition Intelligent Service (CRIS): CRIS cho phép bạn có thể tùy biến language model và acoustic model sao cho phù hợp với ứng dụng hoặc người dùng của bạn.
  3. Speaker Recognition API: Với những thuật toán nhận dạng giọng nói tân tiến của Microsoft, API này cho phép nhận dạng giọng nói của người nói trong một tập tin âm thanh. API này bao gồm 2 thành phần: speaker verification và speaker identification tạm dịch tương ứng là xác nhận người nói và xác định người nói.
  4. ***Speaker Verification****:* Cho phép xác nhận và xác thực người dùng bằng giọng nói của họ. Lập trình viên chỉ cần cho người dùng đọc một đoạn văn bản có sẵn để lưu lại dữ liệu giọng nói của họ (enrollment) rồi ở mỗi lần cần xác thực, người dùng chỉ cần đọc lại chính xác đoạn văn bản đã được dùng để lấy dữ liệu giọng nói của mình (bước enrollment) để xác thực
  5. ***Speaker Identification***: Có thể xác định được người đang nói trong một tập tin âm thanh dựa trên một tập dữ liệu các người nói tiềm năng. Tính năng này cũng có thể được sử dụng để xác thực người dùng bằng giọng nói. Tuy nhiên thay vì phải đọc chính xác một đoạn văn bản cố định như Speaker Verification thì khi sử dụng tính năng Speaker Identification này, người dùng có thể đọc một đoạn văn bản bất kỳ, API sẽ phân tích và đối chiếu với tập dữ liệu giọng nói của người dùng để so sánh và xác thực.

### Language API

Là nhóm API liên quan đến xử lý ngôn ngữ. Nhóm Language API này hiện được Microsoft cung cấp 5 API:

1. Bing Spell Check API: API này cho phép phát hiện và sửa các lỗi chính tả có trong một đoạn văn bản mà bạn cung cấp. API còn có khả năng phát hiện từ lóng, sửa lỗi tên riêng hay sửa các từ đồng âm, …
2. Web Language Model API: API này giúp hỗ trợ xử lý ngôn ngữ tự nhiên, với khả năng chèn khoảng cách vào 1 đoạn văn bản được viết liền nhau như hashtag hay đường dẫn.
3. Linguistic Analysis API: The Linguistic Analysis API giúp bạn hiểu sâu hơn văn bản của mình. API này sẽ giúp phân tích cú pháp của ngôn ngữ tự nhiên để dễ dàng xác định được các thực thể (danh từ) hay các hành động (động từ) có trong văn bản. Việc xử lý văn bản này có thể hữu ích cho các công việc phân tích như phân tích tâm lý.
4. Language Understanding Intelligent Service (LUIS): LUIS cho phép lập trình viên xây dựng các model hiểu được ngôn ngữ tự nhiên cũng như hiểu được các câu lệnh riêng được thiết kế riêng cho ứng dụng của bạn. Ví dụ: Bạn có thể nói “bật đèn trong phòng ngủ”, gửi câu lệnh đó tới LUIS model, và thay vì trả lại chính xác các câu từ có trong câu lệnh trên, LUIS sẽ trả về dữ liệu chứa: thông tin hành động là “bật” vị trí là “phòng ngủ” và đối tượng hướng đến là “bóng đèn”, từ đó ứng dụng có thể dễ dàng xử lý được câu lệnh của bạn.
5. Text Analytics API: API này giúp xác định các ẩn ý, từ khóa, chủ đề hay ngôn ngữ được sử dụng có trong một đoạn văn bản.

### Knowledge API

Là nhóm API liên quan đến tri thức. Nhóm Knowledge API này hiện bao gồm 4 API:

1. Academic Knowledge API:API này cho phép lập trình viên xây dựng những giải pháp tìm kiếm tài liệu học thuật với tính năng Interpret, trả về kết quả gợi ý cho từ khóa mà người dùng nhập vào dựa vào nguồn dữ liệu phong phú từ hệ thống [Microsoft Academic Graph](https://www.microsoft.com/en-us/research/project/microsoft-academic-graph/) (MAG).
2. Knowledge Exploration Service API: API này cho phép lập trình viên xây dựng những giải pháp tìm kiếm sử dụng ngôn ngữ tự nhiên bằng cách dịch ngôn ngữ tự nhiên mà người dùng nhập vào sang các biểu thức truy vấn có cấu trúc mà máy tính có thể dễ dàng hiểu và xử lý được.
3. Entity Linking Intelligence Service API:  Với một đoạn văn bản, Entity Linking Intelligence Service sẽ nhận dạng và xác định từng thực thể (entity) có trong đoạn văn dựa vào ngữ cảnh của đoạn văn đó và sẽ liên kết những entity này tới Wikipedia. Lấy ví dụ rằng bạn có một đoạn văn bản trong đó chứa từ cloud, từ cloud này có thể hiểu sang thành “Cloud Computing” (điện toán đám mây) hay “Cloud” (đám mây trên trời), dựa vào ngữ cảnh mà API này sẽ xác định được rằng từ cloud có ý nghĩa như thế nào.
4. Recommendations API: API này cho phép xây dựng các giải pháp khuyến nghị cho người dùng. Chẳng hạn như bạn xây dựng một ứng dụng bán hàng, sử dụng API này cho phép bạn dễ dàng xây dựng ra các tính năng khuyến nghị mua hàng như “Các sản phẩm được bán chạy”, “Các sản phẩm được mua cùng” hay “Những sản phẩm hàng đầu trong mặt hàng Đồ gia dụng” chẳng hạn, từ đó sẽ khuyến khích người dùng mua nhiều hơn.

### Search API

Search API cũng cấp các giải pháp tìm kiếm sử dụng sức mạnh của Bing.com cũng như từ nhiều đối tác như AOL, Apple, Amazon, Yahoo, …. Nhóm Search API bao gồm 5 API cho các tác vụ tìm kiếm trang web, hình ảnh, video hay tin tức.

1. Bing Web Search API: Đây là API chủ lực của gói Search API. Chỉ với một cú pháp lệnh gọi đến API này, lập trình viên có thể lấy được các kết quả trả về cho trang web, hình ảnh, video hay tin tức tương ứng. Nó khá tương tự với việc bạn tìm kiếm trên các công cụ tìm kiếm như Bing.com hay Google.com. Ngoài ra, lập trình viên cũng sẽ nhận được những tính năng mạnh mẽ từ công cụ tìm kiếm Bing Search như ranking kết quả tìm kiếm, phân loại kết quả tìm kiếm theo vùng, …
2. Bing Autosuggest API: API này cho phép lập trình viên có thể xây dựng tính năng đề xuất các từ khóa tìm kiếm liên quan kể cả khi từ khóa tìm kiếm chưa được điền đầy đủ. Ví dụ nếu người dùng gõ từ khóa tìm kiếm là “Thời tiết tại H”, API sẽ trả về danh sách các từ khóa gợi ý như “Thời tiết tại Hà Nội”, “Thời tiết tại Hồ Chí Minh” hay “Thời tiết tại Hà Giang” chẳng hạn.
3. Bing Image Search API: API này cho phép lập trình viên có thể tìm kiếm các hình ảnh tương ứng với từ khóa nhập vào. Ngoài trả về đường dẫn của hình ảnh, API này cũng trả về các metadata hữu ích như kích thước ảnh, màu chủ đạo của ảnh, …
4. Bing Video Search API: API này cho phép lập trình viên có thể tìm kiếm các video tương ứng với từ khóa nhập vào. Ngoài trả về đường dẫn của video, API này cũng trả về các metadata hữu ích khác như tên nhà sản xuất, định dạng mã hóa, ảnh thumbnail, …
5. Bing News Search API: API này cho phép lập trình viên có thể tìm kiếm các tin tức, bài báo tương ứng với từ khóa nhập vào. Ngoài ra, API cũng trả về các metadata hữu ích khác như thể loại, thông tin nhà xuất bản, ngày xuất bản, …

## OpenCV

OpenCV (Open Source Computer Vision) là một thư viện điện toán phổ biến của Intel bắt đầu vào năm 1999. Thư viện nền tảng cho phép tập trung vào quá trình xử lý hình ảnh thời gian thực và bao gồm các triển khai bằng máy tính mới nhất. Trong năm 2008 Willow Garage tiếp nhận hỗ trợ và OpenCV 2.3.1 giờ đây có giao diện lập trình cho C, C ++, Python và Android. OpenCV được phát hành dưới giấy phép BSD do đó nó được sử dụng trong các dự án học tập và các sản phẩm thương mại như nhau.OpenCV 2.4 bây giờ đi kèm với lớp FaceRecognizer rất mạnh mẽ trong việc nhận diện khuôn mặt. OpenCV là bộ thư viện dành cho thị giác máy tính. Xử lí khuôn mặt gồm 2 bước nhận diện khuôn mặt và Phát hiện khuôn mặt .

### **Xác định khuôn mặt người (Face Detection)**:

Là một kỹ thuật máy tính để xác định các vị trí và các kích thước của các khuôn mặt người trong các ảnh bất kỳ (ảnh kỹ thuật số). Kỹ thuật này nhận biết các đặc trưng của khuôn mặt và bỏ qua những thứ khác, như: tòa nhà, cây cối, cơ thể, …

Để Phát hiện khuôn mặt ta thực thực hiện bằng cách sử dụng OpenCV Sử dụng file xml để dò tìm khuôn mặt . Nhận diện đối tượng sử dụng LBP Classifir Chuyển về ảnh xám Nén ảnh lại Cân bằng sang tối Phát hiện khuôn mặt Khi chúng ta đã chuyển ảnh về ảnh xám, nén ảnh, cân bằng sáng tối thì lúc này đã sẽ sử dụng hàm để xác định khuôn mặt.

### **Face** preprocessing:

Là quá trình xử lý để khuôn mặt trong ảnh chi tiết và sắc nét hơn, giúp giảm độ sai lệch nhận dạng của hệ thống .

Ta cần chỉnh độ sáng tối, hướng để sao cho sự sai khác là ít nhất, nhưng đối với sự tin cậy trong thực thế ,ta cần kĩ thuật tinh vi khác để đảm bảo sự chính. Sắc tối đa (ví dụ: phát hiện đôi mắt, mũi, miệng và lông mày). Phát hiện mắt nó rất hữu ích trong việc tiền xử lý Khi khuôn mặt và mắt đã được phát hiện ta thực hiên sự tiền sử lý.

Chuyển đổi hình học và cắt xén ảnh: Bao gồm loại bỏ cằm, phần trán, căn chỉnh 2 bên, bỏ phần tai và phần nền của ảnh.

Phần chia biểu đồ cân bằng cho bên trái và phải: Là quá trình tiêu chuẩn hóa độ cân bằng sáng, độ tương phản của 2 bên trái và phải.

**Smoothing:** Là quá trình làm giảm nhiễu của ảnh

**Elliptical mask:** Loại bỏ tóc còn lại và hình nền trên khuôn mặt

### Collect and learn faces:

Là quá trình lưu giữ nhiều khuôn mặt đã được xử lý và sau đó sẽ xử lý để nhận dạng ra chúng .

Collecting faces and learning from them Sau khi thu thập đầy đủ khuôn mặt, ta phải đào đạo cho hệ thống để hệ thống nhận dạng được khuôn mặt bằng cách sử dụng thuật toán PCA(Principal Component Analysis (PCA) Thuật toán nhận dạng khuôn mặt sau đó sẽ tìm hiểu làm thế nào để phân biệt giữa các khuôn mặt của những người khác nhau. Điều này được gọi là giai đoạn đào tạo và các khuôn mặt thu thập được gọi là tập huấn luyện, và sau đó đã sẽ lưu dữ liệu này vào file để dùng để nhận diện người trước máy ảnh hoặc camera.

### Face recognition

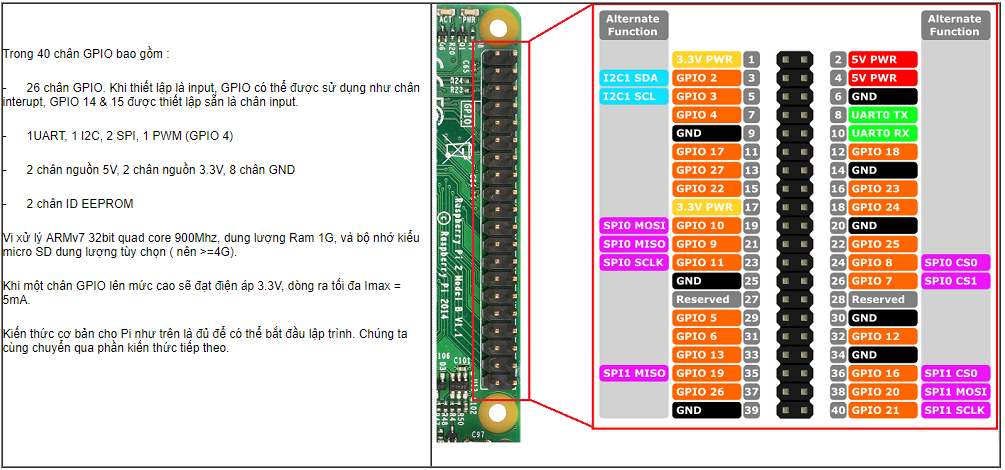
Là quá trình nhận diện khuôn mặt so với ảnh tập hợp.

Khi đã có đầy đủ dữ liệu từ bước 3 lúc này ta sẽ xử lý nhận dạng người từ khuôn mặt của họ Opencv hỗ trợ chúng ta hàm để nhận biết người trong ảnh.

## Tương tác phần cứng với Raspbery

1. Kiến thức phần cứng

Cùng xem GPIO mapping của Raspberry Pi:



1. Kiến thức về ngôn ngữ lập trình

Lập trình trên Pi có nhiều sự lựa chọn. Có thể lập trình trực tiếp từ bash-shell của linux, hoặc lập trình với C đơn thuần, ngoài ra còn có python, perl hay Ruby. Bạn nên chọn lựa một bộ thư viện thay vì chỉ lập trình với ngôn ngữ đơn thuần, vì đơn giản bạn đặt gạch xây nhà nhanh hơn là làm từng viên gạch cho ngôi nhà của mình. Thư viện sẽ giúp bạn bỏ qua lượng công việc vừa phức tạp và tốn công sức như gán địa chỉ của chân GPIO hay làm việc với thanh ghi ..v.v. Bạn có thể tập trung hơn vào xây dựng ứng dụng của mình.

Thư viện cho Pi cũng đa dạng không kém. Một thư viện tốt khi nó cung cấp nhiều hàm xử lý linh hoạt, hỗ trợ nhiều giao tiếp và tốc độ của thư viện nhanh ( tức là nó mất không quá nhiều lần gọi lệnh hay hàm trung gian để có thực hiện mong muốn của bạn ). Vấn đề tốc độ chỉ đáng quan tâm khi bạn thực sự làm việc với yêu cầu vi xử lý thực hiện lệnh nhanh chóng (giả như PWM). Bạn có thể xem qua [Benchmarking](http://codeandlife.com/2012/07/03/benchmarking-raspberry-pi-gpio-speed/) cho các thư việc của Raspberry Pi.

Trong dự án này, nhóm sẽ giới thiệu lập trình ngôn ngữ sử dụng rộng rãi trên Pi là Python. Với bộ thư viện tương ứng là  [RPiGPIO](https://pypi.python.org/pypi/RPi.GPIO).

## Web Socket

### Khái niệm:

WebSoket là công nghệ hỗ trợ giao tiếp hai chiều giữa client và server bằng cách sử dụng một TCP socket để tạo một kết nối hiệu quả và ít tốn kém. Mặc dù được thiết kế để chuyên sử dụng cho các ứng dụng web, lập trình viên vẫn có thể đưa chúng vào bất kì loại ứng dụng nào.

WebSockets mới xuất hiện trong HTML5, là một kỹ thuật Reverse Ajax. WebSockets cho phép các kênh giao tiếp song song hai chiều và hiện đã được hỗ trợ trong nhiều trình duyệt (Firefox, Google Chrome và Safari). Kết nối được mở thông qua một HTTP request (yêu cầu HTTP), được gọi là liên kết WebSockets với những header đặc biệt. Kết nối được duy trì để bạn có thể viết và nhận dữ liệu bằng JavaScript như khi bạn đang sử dụng một TCP socket đơn thuần.

Dữ liệu truyền tải thông qua giao thức HTTP (thường dùng với kĩ thuật Ajax) chứa nhiều dữ liệu không cần thiết trong phần header. Một header request/response của HTTP có kích thước khoảng 871 byte, trong khi với WebSocket, kích thước này chỉ là 2 byte (sau khi đã kết nối).  
Vậy giả sử bạn làm một ứng dụng game có thể tới 10,000 người chơi đăng nhập cùng lúc, và mỗi giây họ sẽ gửi/nhận dữ liệu từ server. Hãy so sánh lượng dữ liệu header mà giao thức HTTP và WebSocket trong mỗi giây:

HTTP: 871 x 10,000 = 8,710,000 bytes = 69,680,000 bits per second (66 Mbps)

WebSocket: 2 x 10,000 = 20,000 bytes = 160,000 bits per second (0.153 Kbps)

Như bạn thấy chỉ riêng phần header thôi cũng đã chiếm một phần lưu lượng đáng kể với giao thức HTTP truyền thống.

### Ưu điểm

WebSockets cung cấp khả năng giao tiếp hai chiều mạnh mẽ, có độ trễ thấp và dễ xử lý lỗi. Không cần phải có nhiều kết nối như phương pháp Comet long-polling và cũng không có những nhược điểm như Comet streaming. API cũng rất dễ sử dụng trực tiếp mà không cần bất kỳ các tầng bổ sung nào, so với Comet, thường đòi hỏi một thư viện tốt để xử lý kết nối lại, thời gian chờ timeout, các Ajax request (yêu cầu Ajax), các tin báo nhận và các dạng truyền tải tùy chọn khác nhau (Ajax long-polling và jsonp polling).

### Nhược điểm

* Nó là một đặc tả mới của HTML5, nên nó vẫn chưa được tất cả các trình duyệt hỗ trợ.
* Không có phạm vi yêu cầu nào. Do WebSocket là một TCP socket chứ không phải là HTTP request, nên không dễ sử dụng các dịch vụ có phạm vi-yêu cầu, như SessionInViewFilter của Hibernate. Hibernate là một framework kinh điển cung cấp một bộ lọc xung quanh một HTTP request. Khi bắt đầu một request, nó sẽ thiết lập một contest (chứa các transaction và liên kết JDBC) được ràng buộc với luồng request. Khi request đó kết thúc, bộ lọc hủy bỏ contest này.

## Web Front-End

Phần front-end của một trang web là phần tương tác với người dùng. Tất cả mọi thứ bạn nhìn thấy khi điều hướng trên Internet, từ các font chữ, màu sắc cho tới các menu xổ xuống và các thanh trượt, là một sự kết hợp của HTML, CSS, và JavaScript được điều khiển bởi trình duyệt máy tính của bạn.

Dùng để tạo giao diện hiển thị kết quả .

Sử dụng HTML, CSS, Javascript hỗ trợ cho việc hiển thị

Làm cho việc hiển thị đơn giản , dễ nhìn hơn

Hỗ trợ cho việc xem các lỗi và sửa chữa

Ứng dụng sử dụng websocket để giao tiếp với Raspberry Pi. Qua đó nhận dữ liệu hình ảnh từ camera thu được cũng như dữ liệu phân tích bằng OpenCV hay các Microsoft Cognitive Service.

## Voice Recognition

Sự nhận dạng tiếng người bằng máy tính và biến đổi những từ đã nhận dạng đó thành văn bản hoặc các chỉ lệnh mà máy tính có thể đọc được. Máy tính và con người có chung một đặc tính không may : nói tốt hơn nhiều so với nghe hiểu. Trong hầu hết các hệ thống nghiên cứu tiên tiến, máy tính chỉ có thể nhận biết được vào khoảng một hai trăm từ, và ngay cả khả năng này cũng chỉ đạt được sau khi người nói đã huấn luyện cho hệ thống nhận dạng biết về các mẫu tiếng của mình. Sự nhận dạng tiếng nói liên quan đến một số khả năng nhận dạng mẫu cực kỳ phức tạp trong bộ não người mà hiện nay vẫn chưa biết rõ. Tuy nhiên, với tất cả các phương án đầu vào có sẵn, những người không có điều kiện đầu tư thời gian để học cách gõ bàn phím cũng có thể có một số lệnh nói có thể dùng được như là một cụ chung cho mọi người. Do đó, việc nhận dạng tiếng nói đối với nhiều người được coi như là một phương pháp tín hiệu vào trong tương lai.

## Speech to text

Text-to-Speech (TTS) khả năng cho máy tính đề cập đến khả năng phát lại văn bản trong một giọng nói. Bài viết này mô tả cách cấu hình và sử dụng text-to-speech trong Windows XP và Windows Vista. TTS là khả năng của hệ điều hành để phát lại in văn bản dưới dạng lời nói. Trình điều khiển nội bộ, được gọi là một công cụ TTS, nhận dạng văn bản và sử dụng một giọng nói tổng hợp chọn nhiều nói trước tạo ra, nói đoạn văn. Công cụ TTS được cài đặt với hệ điều hành. Công cụ bổ sung cũng có sẵn thông qua nhà sản xuất bên thứ ba. Các công cụ thường sử dụng các biệt ngữ hoặc từ; Ví dụ: một từ chuyên y hoặc Pháp ngữ. Họ cũng có thể sử dụng giọng nói khác nhau cho phép cho khu vực dấu như Anh hoặc nói ngôn ngữ khác nhau hoàn toàn như tiếng Đức, tiếng Pháp hoặc tiếng Nga. Tab Text-to-Speech cụ nói trong bảng điều khiển trình bày các tùy chọn cho mỗi chương trình TTS. Xem các chủ đề trợ giúp cá nhân để được trợ giúp cụ thể. Ngoài các tuỳ chọn chung, mỗi chương trình có một bộ tính năng cụ thể khác nhau. Do đó, không phải tất cả các hộp thoại kết quả sẽ giống nhau.

# PHẦN HAI : HƯỚNG TIẾP CẬN VÀ PHƯƠNG PHÁP

## Khái niệm

Xử lý ảnh là một phân ngành trong xử lý số tín hiệu với tín hiệu xử lý là ảnh. Đây là một phân ngành khoa học mới rất phát triển trong những năm gần đây. Xử lý ảnh gồm 4 lĩnh vực chính: xử lý nâng cao chất lựong ảnh, nhận dạng ảnh, nén ảnh và truy vấn ảnh. Sự phát triển của xử lý ảnh đem lại rất nhiều lợi ích cho cuộc sống của con người.

Ngày nay xử lý ảnh đã được áp dụng rất rộng rãi trong đời sống như: photoshop, nén ảnh, nén video, nhận dạng biển số xe, nhận dạng khuôn mặt, nhận dạng chữ viết, xử lý ảnh thiên văn, ảnh y tế,....

**Một số khái niệm cơ bản:**

* Ảnh và điểm ảnh: Điểm ảnh được xem như là dấu hiệu hay cường độ sáng tại 1 toạ độ trong không gian của đối tượng và ảnh được xem như là 1 tập hợp các điểm ảnh
* Mức xám, màu: Là số các giá trị có thể có của các điểm ảnh của ảnh.

## Các vấn đề cơ bản trong sử lý ảnh

### ****Chỉnh mức xám****

Nhằm khắc phục tính không đồng đều của hệ thống gây ra. Thông thường có 2 hướng tiếp cận:

* Giảm số mức xám: Thực hiện bằng cách nhóm các mức xám gần nhau thành một bó. Trường hợp chỉ có 2 mức xám thì chính là chuyển về ảnh đen trắng. Ứng dụng: In ảnh màu ra máy in đen trắng.
* Tăng số mức xám: Thực hiện nội suy ra các mức xám trung gian bằng kỹ thuật nội suy. Kỹ thuật này nhằm tăng cường độ mịn cho ảnh

### ****Trích chọn đặc điểm****

Các đặc điểm của đối tượng được trích chọn tuỳ theo mục đích nhận dạng trong quá trình xử lý ảnh. Có thể nêu ra một số đặc điểm của ảnh sau đây:

* Đặc điểm không gian: Phân bố mức xám, phân bố xác suất, biên độ, điểm uốn v.v..
* Đặc điểm biến đổi: Các đặc điểm loại này được trích chọn bằng việc thực hiện lọc vùng (zonal filtering). Các bộ vùng được gọi là “mặt nạ đặc điểm” (feature mask) thường là các khe hẹp với hình dạng khác nhau (chữ nhật, tam giác, cung trònv.v..)
* Đặc điểm biên và đường biên: Đặc trưng cho đường biên của đối tượng và do vậy rất hữu ích trong việc trích trọn các thuộc tính bất biến được dùng khi nhận dạng đối tượng. Các đặc điểm này có thể được trích chọn nhờ toán tử gradient, toán tử la bàn, toán tử Laplace, toán tử “chéo không” (zero crossing) v.v.. Việt trích chọn hiệu quả các đặc điểm giúp cho việc nhận dạng các đối tượng ảnh chính xác, với tốc độ tính toán cao và dung lượng nhớ lưu trữ giảm xuống.

### Nhận dạng

* Nhận dạng tự động (automatic recognition), mô tả đối tượng, phân loại và phân nhóm các mẫu là những vấn đề quan trọng trong thị giác máy, được ứng dụng trong nhiều ngành khoa học khác nhau. Tuy nhiên, một câu hỏi đặt ra là: mẫu (pattern) là gì? Watanabe, một trong những người đi đầu trong lĩnh vực này đã định nghĩa: “Ngược lại với hỗn loạn (chaos), mẫu là một thực thể (entity), được xác định một cách ang áng (vaguely defined) và có thể gán cho nó một tên gọi nào đó”.
* Ví dụ mẫu có thể là ảnh của vân tay, ảnh của một vật nào đó được chụp, một chữ viết, cơ thể người hoặc một ký đồ tín hiệu tiếng nói. Khi biết một mẫu nào đó, để nhận dạng hoặc phân loại mẫu đó có thể:  
   Hoặc phân loại có mẫu (supervised classification), chẳng hạn phân tích phân biệt (discriminant analyis), trong đó mẫu đầu vào được định danh như một thành phần của một lớp đã xác định.  
   Hoặc phân loại không có mẫu (unsupervised classification hay clustering) trong đó các mẫu được gán vào các lớp khác nhau dựa trên một tiêu chuẩn đồng dạng nào đó. Các lớp này cho đến thời điểm phân loại vẫn chưa biết hay chưa được định danh.
* Hệ thống nhận dạng tự động bao gồm ba khâu tương ứng với ba giai đoạn chủ yếu sau đây:  
  1. Thu nhận dữ liệu và tiền xử lý.  
  2. Biểu diễn dữ liệu.  
  3. Nhận dạng, ra quyết định.
* Bốn cách tiếp cận khác nhau trong lý thuyết nhận dạng là:  
  1. Đối sánh mẫu dựa trên các đặc trưng được trích chọn.  
  2. Phân loại thống kê.  
  3. Đối sánh cấu trúc.

**4. Phân loại dựa trên mạng nơ-ron nhân tạo.**

Trong các ứng dụng rõ ràng là không thể chỉ dùng có một cách tiếp cận đơn lẻ để phân loại “tối ưu” do vậy cần sử dụng cùng một lúc nhiều phương pháp và cách tiếp cận khác nhau. Do vậy, các phương thức phân loại tổ hợp hay được sử dụng khi nhận dạng và nay đã có những kết quả có triển vọng dựa trên thiết kế các hệ thống lai (hybrid system) bao gồm nhiều mô hình kết hợp.

Việc giải quyết bài toán nhận dạng trong những ứng dụng mới, nảy sinh trong cuộc sống không chỉ tạo ra những thách thức về thuật giải, mà còn đặt ra những yêu cầu về tốc độ tính toán. Đặc điểm chung của tất cả những ứng dụng đó là những đặc điểm đặc trưng cần thiết thường là nhiều, không thể do chuyên gia đề xuất, mà phải được trích chọn dựa trên các thủ tục phân tích dữ liệu.

1. **Nén ảnh**

Nhằm giảm thiểu không gian lưu trữ. Thường được tiến hành theo cả hai cách khuynh hướng là nén có bảo toàn và không bảo toàn thông tin. Nén không bảo toàn thì thường có khả năng nén cao hơn nhưng khả năng phục hồi thì kém hơn. Trên cơ sởhaikhuynhhướng,có4cáchtiếpcậncơbảntrongnénảnh:

• Nén ảnh thống kê: Kỹ thuật nén này dựa vào việc thống kê tần xuất xuất hiện của giá trị các điểm ảnh, trên cơ sở đó mà có chiến lược mã hóa thích hợp. Một ví dụ điểnhìnhchokỹthuậtmãhóanàylà\*.TIF

• Nén ảnh không gian: Kỹ thuật này dựa vào vị trí không gian của các điểm ảnh để tiến hành mã hóa. Kỹ thuật lợi dụng sự giống nhau của các điểm ảnh trong các vùng gần nhau.Ví dụ cho kỹ thuật này là mã nén\*.PCX

• Nén ảnh sử dụng phép biến đổi: Đây là kỹ thuật tiếp cận theo hướng nén không bảo toàn và do vậy, kỹ thuật thướng nến hiệu quả hơn. \*.JPG chính là tiếp cận theo kỹ thuật nén này.

• Nén ảnh Fractal: Sử dụng tính chất Fractal của các đối tượng ảnh, thể hiện sự lặp lại của các chi tiết. Kỹ thuật nén sẽ tính toán để chỉ cần lưu trữ phần gốc ảnh và quy luật sinh ra ảnh theo nguyên lý Fractal

## Bài toán phát hiện người trong ảnh bằng Viral Jone

### Giới thiệu về nhận diện khuôn mặt

Trong những năm gần đây, có rất nhiều công trình nghiên cứu về bài toán xác định khuôn mặt người từ ảnh đen trắng , xám đến ảnh màu. Ban đầu chỉ có thể xác định ảnh chỉ có một khuôn mặt nhìn thẳng và đầu thẳng đứng trong ảnh , nhưng không đáp ứng được nhu cầu ngày càng cao trong cuộc sống, khoa học. Vì thế đã có những cải tiến về bài toán nhận diện khuôn mặt trong những môi trường phức tạp hơn , có nhù khuôn mặt trong ảnh hơn, nhiều tư thế hơn.

### Mục đích của việc phát hiện

Phát hiện khuôn mặt người là một kỹ thuật để xác định vị trí và kích thước khuôn mặt người có trong bức ảnh bất kì. Kỹ thuật này chỉ nhận biết được các đặc trưng của khuôn mặt và bỏ qua những đặc trưng khác.

### Các phương pháp nhận diện khuôn mặt người

Dựa vào tính chất của các phương pháp xác định mặt người trên ảnh, các phương pháp này được chia thành bốn loại chính, tương ứng với bốn hướng tiếp cận khác nhau. Ngoài ra cũng có rất nhiều nghiên cứu mà phương pháp xác định mặt người không chỉ dựa vào một hướng mà có liên quan đến nhiều hướng.

 Hướng tiếp cận dựa trên tri thức: Dựa vào các thuật toán, mã hóa các đặc trưng và quan hệ giữa các đặc trưng của khuôn mặt thành các luật. Đây là hướng tiếp cận theo kiểu top-down

Hướng tiếp cận dựa trên đặc trưng không thay đổi: Xây dựng các thuật toán để tìm các đặc trưng của khuôn mặt mà các đặc trưng này không thay đổi khi tư thế khuôn mặt hay vị trí đặt camera thay đổi.

 Hướng tiếp cận dựa trên so sánh khớp mẫu: Dùng các mẫu chuẩn của khuôn mặt (các mẫu này đã được chọn và lưu trữ) để mô tả các khuôn mặt hay các đặc trưng của khuôn mặt. Phương 7 pháp này có thể dùng để xác định vị trí hay dò tìm trên khuôn mặt ảnh

 Hướng tiếp cận dựa trên diện mạo: Trái với hướng tiếp cận dựa trên khuôn mẫu, các mô hình hay các mẫu sẽ được học từ một tập ảnh huấn luyện mà thể hiện tính chất tiêu biểu của sự xuất hiện của mặt người trong ảnh. Sau đó hệ thống sẽ xác định mặt người. Phương pháp này còn được biết đến với tên gọi theo tiếp cận theo các phương pháp học máy.

### Khó khăn trong việc xử lí

Khuôn mặt trong ảnh có thể có những hướng nhìn khác nhau như: nhìn thẳng, nhìn nghiêng hay nhìn lên nhìn xuống. Cùng trong một ảnh nhưng sẽ có nhiều hướng nhìn khác nhau của khuôn mặt vì vậy việc nhận dạng sẽ rất khó khăn.

Trong một khuôn mặt không chỉ có những đặc trưng là khuôn mặt mà còn có một số chi tiết không phải là đặc trưng của khuôn mặt nên việc nhận dạng khuôn mặt cũng có thể sai.

Một số khuôn mặt bị che khuất bởi các đối tượng khác cũng gây cản trở cho việc nhận dạng mặt người.

Sự biểu cảm của khuôn mặt: biểu cảm của khuôn mặt có thể làm thay đổi đáng kể các đặc trưng và thông số của khuôn mặt. Như cùng một người nhưng khi cười, tức giận hay sợ hãi cũng dẫn đến sự khác biệt của khuôn mặt.

Giới hạn về số ảnh cần thiết cho tập huấn luyện, tập các ảnh khuôn mặt huấn luyện không thể bao quát được tất cả các biến đổi có thể có trên khuôn mặt của một người cần nhận dạng trong thế giới thực.

## Phát hiện khuôn mặt bằng thuật toán Viral Jone

### Đặc trưng Haar-like

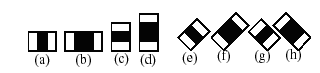
Do Viola và Jones công bố, gồm 4 đặc trưng cơ bản để xác định khuôn mặt người. Mỗi đặc trưng Haar–like là sự kết hợp của hai hay ba hình chữ nhật "trắng" hay "đen" như trong hình sau:

[http://2.bp.blogspot.com/-vfvQEoj5TH4/TzI4gDb7c0I/AAAAAAAAAeo/SKzDh2HPRjo/s320/Haar-like-0.jpg](http://2.bp.blogspot.com/-vfvQEoj5TH4/TzI4gDb7c0I/AAAAAAAAAeo/SKzDh2HPRjo/s1600/Haar-like-0.jpg)

Hình 1: 4 đặt trưng Haar-like cơ bản

Để sử dụng các đặt trưng này vào việc xác định khuôn mặt người, 4 đặt trưng Haar-like cơ bản được mở rộng ra, và được chia làm 3 tập đặc trưng như sau:

* 1. Đặc trưng cạnh (edge features):
     1. [http://1.bp.blogspot.com/-VfVkuegsiwU/TzI414grUcI/AAAAAAAAAew/yWbLYpIyYeQ/s1600/image002.png](http://1.bp.blogspot.com/-VfVkuegsiwU/TzI414grUcI/AAAAAAAAAew/yWbLYpIyYeQ/s1600/image002.png)
  2. Đặc trưng đường (line features):

[](http://2.bp.blogspot.com/-l-KNt0P3waU/TzI5EbCYq9I/AAAAAAAAAe4/vMFZjftJHpY/s1600/Haarlike-2.png)

* 1. Đặc trưng xung quanh tâm (center-surround features):

[http://4.bp.blogspot.com/-J3I13L81GDw/TzI5PBONVRI/AAAAAAAAAfA/lcj_rZ2k8e4/s1600/Haarlike-3.png](http://4.bp.blogspot.com/-J3I13L81GDw/TzI5PBONVRI/AAAAAAAAAfA/lcj_rZ2k8e4/s1600/Haarlike-3.png)

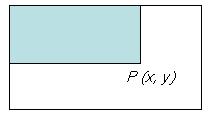
Hình 2,3,4: Các đặc trưng mở rộng của các đặc trưng Haar-like cơ sở

Dùng các đặc trưng trên, ta có thể tính được giá trị của đặc trưng Haar-like là sự chênh lệch giữa tổng của các pixel của các vùng đen và các vùng trắng như trong công thức sau:

f(x) = Tổngvùng đen(các mức xám của pixel) - Tổngvùng trắng(các mức xám của pixel)

Sử dụng giá trị này, so sánh với các giá trị của các giá trị pixel thô, các đặc trưng Haar-like có thể tăng/giảm sự thay đổi in-class/out-of-class (bên trong hay bên ngoài lớp khuôn mặt người), do đó sẽ làm cho bộ phân loại dễ hơn.

Như vậy ta có thể thấy rằng, để tính các giá trị của đặc trưng Haar-like, ta phải tính tổng của các vùng pixel trên ảnh. Nhưng để tính toán các giá trị của các đặc trưng Haar-like cho tất cả các vị trí trên ảnh đòi hỏi chi phí tính toán khá lớn, không đáp ứng được cho các ứng dụng đòi hỏi tính run-time. Do đó Viola và Jones đưa ra một khái niệm gọi là Integral Image, là một mảng 2 chiều với kích thước bằng với kích của ảnh cần tính các đặc trưng Haar-like, với mỗi phần tử của mảng này được tính bằng cách tính tổng của điểm ảnh phía trên (dòng-1) và bên trái (cột-1) của nó. Bắt đầu từ vị trí trên, bên trái đến vị trí dưới, phải của ảnh, việc tính toán này đơn thuần chỉ đựa trên phép cộng số nguyên đơn giản, do đó tốc độ thực hiện rất nhanh.

[](http://2.bp.blogspot.com/-n-BAaDMOCvo/TzI5fkMRsFI/AAAAAAAAAfI/PlaaeVry_Fw/s1600/IntegralImag-1-new.gif.jpg)

[http://2.bp.blogspot.com/-_sc_-nfrj1s/TzI5hpdtXFI/AAAAAAAAAfQ/LAntPDhSXCI/s1600/IntegralImag-2-new.gif.jpg](http://2.bp.blogspot.com/-_sc_-nfrj1s/TzI5hpdtXFI/AAAAAAAAAfQ/LAntPDhSXCI/s1600/IntegralImag-2-new.gif.jpg)

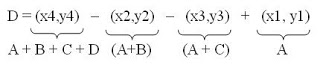
Hình 5: Cách tính Integral Image của ảnh

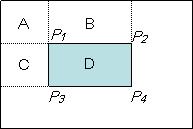
Sau khi đã tính được Integral Image, việc tính tổng các giá trị mức xám của một vùng bất kỳ nào đó trên ảnh thực hiện rất đơn giản theo cách sau:

Giả sử ta cần tính tổng các giá trị mức xám của vùng D như trong hình 4, ta có thể tính như sau:

D = A + B + C + D – (A+B) – (A+C) + A

Với A + B + C + D chính là giá trị tại điểm P4 trên Integral Image, tương tự như vậy A+B là giá trị tại điểm P2, A+C là giá trị tại điểm P3, và A là giá trị tại điểm P1. Vậy ta có thể viết lại biểu thức tính D ở trên như sau:

[](http://1.bp.blogspot.com/-ofY9IDi0KD8/TzI5w8Sh8uI/AAAAAAAAAfY/nWPGqJmTo8s/s1600/IntegralImag-3.gif.jpg)

[](http://4.bp.blogspot.com/-LfhdqoYrXm4/TzI5xRcjMMI/AAAAAAAAAfc/75EyC1f1U18/s1600/sumgraylevel.gif.jpg)

Hình 6: Ví dụ cách tính nhanh các giá trị mức xám của vùng D trên ảnh

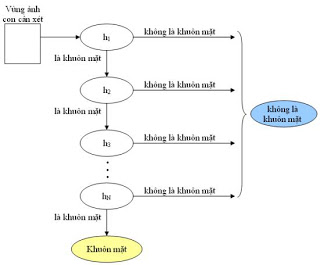
Tiếp theo, để chọn các đặc trưng Haar-like dùng cho việc thiết lập ngưỡng, Viola và Jones sử dụng một phương pháp máy học được gọi là AdaBoost. AdaBoost sẽ kết hợp các bộ phân loại yếu để tạo thành một bộ phân loại mạnh. Với bộ phân loại yếu chỉ cho ra câu trả lời chính xác chỉ hơn viện đoán một cách ngẫn nhiên một chút, còn bộ phân loại mạnh có thể đưa ra câu trả lời chính xác trên 60%.

1. **AdaBoost**

AdaBoost là một bộ phân loại mạnh phi tuyến phức dựa trên hướng tiếp cận boosting được Freund và Schapire đưa ra vào năm 1995 [2]. Adaboost cũng hoạt động trên nguyên tắc kết hợp tuyến tính các weak classifiers để hình thành một strong classifier.

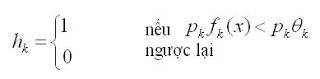
Là một cải tiến của tiếp cận boosting, AdaBoost sử dụng thêm khái niệm trọng số (weight) để đánh dấu các mẫu khó nhận dạng. Trong quá trình huấn luyện, cứ mỗi weak classifiers được xây dựng, thuật toán sẽ tiến hành cập nhật lại trọng số để chuẩn bị cho việc xây dựng weak classifier kế tiếp: tăng trọng số của các mẫu bị nhận dạng sai và giảm trọng số của các mẫu được nhận dạng đúng bởi weak classifier vừa xây dựng. Bằng cách này weak classifer sau có thể tập trung vào các mẫu mà các weak classifiers trước nó làm chưa tốt. Sau cùng, các weak classifers sẽ được kết hợp tùy theo mức độ tốt của chúng để tạo nên strong classifier.

Viola và Jones dùng AdaBoost kết hợp các bộ phân loại yếu sử dụng các đặc trưng Haar-like theo mô hình phân tầng (cascade) như sau:

[](http://2.bp.blogspot.com/-p9jOLhjphGU/TzI59Yb21LI/AAAAAAAAAfo/bAfX1oVLDHs/s1600/AdaBoost-0.gif.jpg)

Hình 5: Mô hình phân tần kết hợp các bộ phân loại yếu để xác định khôn mặt

Trong đó, hk là các bộ phân loại yếu, được biểu diễn như sau:

[](http://4.bp.blogspot.com/-bR-KWQLm71A/TzI6JzAt0yI/AAAAAAAAAfw/ktXSzndxG74/s1600/AdaBoost-1.gif.jpg)

x: cửa sổ con cần xét

Ok: ngưỡng (O = teta)

fk: giá trị của đặc trưng Haar-like

pk: hệ số quyết định chiều của phương trình

AdaBoost sẽ kết hợp các bộ phân loại yếu thành bộ phân loại mạnh như sau:

H(x) = sign(a1h1(x) +a2h2(x) + ... + anhn(x)) (a = alpha)

Với: at >= 0 là hệ số chuẩn hoá cho các bộ phân loại yếu

[](http://4.bp.blogspot.com/-sI2m1ndrF74/TzI6VFVv6xI/AAAAAAAAAf4/HisDFfvPDsY/s1600/AdaBoost-2.jpg)

Hình 6: Kết hợp các bộ phân loại yếu thành bộ phân loại mạnh

 Mỗi bộ phân loại yếu sẽ quyết định kết quả cho một đặc trưng haarlike, được xác định ngưỡng đủ nhỏ sao cho có thể vượt được tất cả các bộ dữ liệu mẫu trong tập dữ liệu huấn luyện (số lượng ảnh khuôn mặt trong tập huấn luyện lớn). Trong quá trình xác định khuôn mặt người, mỗi vùng ảnh con sẽ được kiểm tra với các đặc trưng trong chuỗi các đặc trưng haarlike, nếu có một đặc trưng haarlike nào cho ra kết quả là khuôn mặt người thì các đặc trưng khác không cần xét nữa. Thứ tự xét các đặc trưng trong chuỗi các đặc trưng haarlike sẽ được dựa vào trọng số (weight) của đặc trưng đó do Adaboost quyết định dựa vào số lần và thứ tự xuất hiện của các đặc trưng haarlike.

**Thuật toán Adaboost:**

Cho một tập huấn luyện gồm N mẫu có đánh dấu (x1, y1), (x2, y2), , (xn, yn). Với xk X = (x1k, x2k, ,xmk) là vector đặc trưng và yk (-

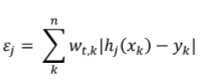
Khởi tạo trọng số ban đầu cho tất cả các mẫu: với m là số mẫu đúng (ứng với object và y= 1) và l là số mẫu sai (ứng với background và y= -1).

W1,k=

Xây dựng T weak classifier

Lặp t = 1….T

Với mỗi đặc trưng trong vector đặc trưng, xây dựng một weak classifier hj với ngưỡng θj và lỗi εj.



Chọn ra hj với εj nhỏ nhất, ta được ht:

ht : X → {1,-1}

Cập nhật lại trọng số:



Trong đó: =ln()

 Zt: hệ số dùng để đưa Wt+1 về đoạn [0, 1] (normalization factor)

Strong classifier được xây dựng:

H(X) = dấu()

 Quá trình huấn luyện bộ phân loại được thực hiện bằng một vòng lặp mà ở mỗi bước lặp, thuật toán sẽ chọn ra weak classifier ht thực hiện việc phân loại với lỗi nhỏ nhất (do đó sẽ là bộ phân loại tốt nhất) để bổ sung vào strong classifier. Mỗi khi chọn được một bộ phân loại ht Adaboost sẽ tính giá trị theo công thức ở trên. cũng được chọn trên nguyên tắc làm giảm thiểu giá trị lỗi .

Hệ số .

Trong công thức của bộ phân loại H(x):

H(x) = dấu ()

Ta thấy tất cả bộ phân loại ht đều có đóng góp vào kết quả phân loại của H(x), và mức độ đóng góp của chúng phụ thuộc vào giá trị tương ứng: ht với càng lớn thì nó càng có vai trò quan trọng trong H(x).

Trong công thức : =ln()

Dễ thấy giá trị . Bởi vì ht được chọn với tiêu chí đạt nhỏ nhất, do đó nó sẽ bảo đảm giá trị lớn nhất.

Sau khi tính được giá trị , AdaBoost tiến hành cập nahtaj lại trọng số của các mẫu : tang trọng số các mẫu mà ht phân loại sai, giảm trọng số các mẫu mà ht phân loại đúng. Bằng cách này, trọng số các mẫu phản ánh đucợ độ khó nhận dạng của mẫu dố và ht+1 sẽ ưu tiên học cách phân loại các mẫu này.

Vòng lặp xây dựng strong classifier sẽ dừng lại sau T lần lặp. Trong thực tế cài đặt, người ta ít sử dụng giá trị T vì không cố công thức nào đảm bảo tính được giá trị T tối ưu cho quá trình huấn luyện. Thay vào đó người ta sử dụng giá trị max false positive hay max false alarm (tỉ lệ nhận dạng sai tối đa các mẫu background). Tỉ lệ này của bộ phân loại cần xây dựng không được phép vượt quá giá trị này. Khi đó, qua các lần lặp, false alarm của strong classifier xây dựng được (tại lần thứ t) sẽ giảm dần, và vòng lặp kết thúc khi tỉ lệ này thấp hơn max false alarm.

1. **Cascade of Classifiers**

Ta thấy quá trình huấn luyện, bộ phân loại phải duyệt qua tất cả các đặc trưng của các mẫu trong tập training. Việc này tốn rất nhiều thời gian. Tuy nhiên, trong các mẫu đưa vào, không phải mẫu nào cũng thuộc loại khó nhận dạng, có những mẫu background rất dễ nhận ra (ta gọi đây là

những mẫu background đơn giản). Đối với những mẫu này, ta chỉ cần xét một hay vài đặc trưng đơn giản là có thế nhận diện được chứ không cần xét tất cả các đặc trưng. Nhưng đối với các bộ phân loại thông thường thì cho dù mẫu cần nhân dạng là dễ hay khó thì nó vẫn sẽ xét tất cả các đặc trưng mà nó rút ra được trong quá trình học. Do đó, chúng tốn thời gian xử lý một cách không cần thiết.

Cascade of Classifiers được xây dụng chính là nhằm rút ngắn thời gian xử lý, giảm thiểu false alarm cho bộ phân loại. Cascade tree gồm nhiều stage (hay còn gọi là layer) mỗi stage củacây sẽ là một stage classifier. Một mẫu để được phân loại là đối tượng thì nó cần phải đi qua hết tất cả các stages của cây. Các stage classifiers ở stage sau được huấn luyện bằng những mẫu negative mà stage classifiier trước nó nhận dạng sai tức là nó sẽ tập trung học từ các mẫu background khó hơn, do đó sự kết hợp các stage classifiers này lại sẽ giúp bộ phân loại có false alarm thấp. Với cấu trúc này, những mẫu background dễ nhận điện sẽ bị loại ngay từ những stages đầu tiên, giúp đáp ứng tốt nhất đối với độ phức tạp gia tăng của các mẫu đưa vào, đồng thời giúp rút ngắn thời gian xử lý.

Thuật toán cascade training:

*1. Gọi:*

F là giá trị false alam và d là đó chính xác của weak < classifier ở mỗi stage

: Giá trị max false alarm.

P, N là số lương mẫu positive và negative.

, mà tập positive và negative cho bộ phân lớp ở tàng thứ i.

,: Giá trị false alarm và đó chính xác của cascade trước khi đến tầng

thứ i.

*2. Khởi tạo i=O; Fo=1.0; Do = 1.0*

3. Lặp : While

i = i+1;

Huấn luyện bộ phân loại từ tập và với detection rate d và max

false alarm f. Thêm vào cây phân lớp.

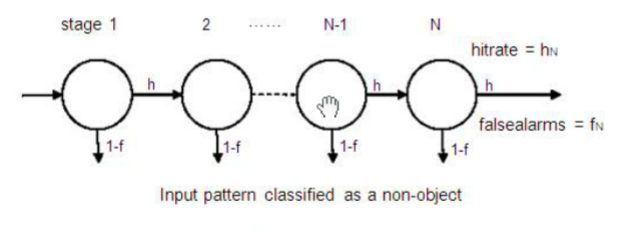
Dùng cây phân lớp hiện có để tính: Duyệt qua N mẫu negative cho đến khi nào tìm đủ n mẫu mà cây phân lớp hiện có phân loại sai.

Nếu

N={số mẫu sai ở tage hiện tại phân loại sai }

P={số mẫu positive mà stage hiện tại phân loại đúng}

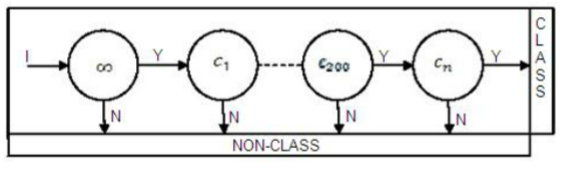
Minh họa thuật toán Cascade training:



Hình trên minh họa sự huấn luyện của một cascade gồm N stages. Ở mỗi stage, weak classifier tương ứng sẽ được huấn luyện sao cho dộ chính xác của nó là h và false alarm bằng f.

1. **Cascade of boosting classifiers**

Cascade of boosting classifiers là mô hình cascade of classifiers với mối classifier được xây dựng bằng adaboost sử dụng haar-like.

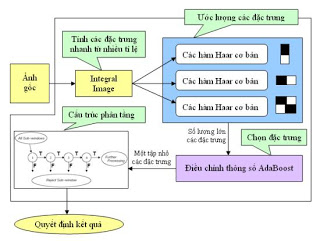


Viola va Jones đã sử dụng rất thành công cascade of boosting classiﬁers cho bài tóan nhận dạng mặt người. Với tập huấn luyện gồm 4196 hình mặt người được đưa về ảnh grayscale kich thuoc 24x24 và 9500 hình background, hai ông đã xây dựng cấu trucs cascade tree gồm 38 stage với tổng cộng 6060 đặc trưng haar-like. Thực nghiệm đã cho thấy classiﬁer 6 stage đầu tiến sử dụng 2 đặc trưng và loại được khoảng 50% mẫu background (không phải mặt người) và có độ chính xác là d= 100%. Classiﬁer ở stage thứ 2 sử dụng 10 đặc trưng loại được 80% mẫu background với độ chính xác vẫn là l00%. Hệ thống này được so sánh với hệ thống của Rowley-Baluja-Kanade (sử dụng mạng neural) , Schneiderman-Kanade (sử dụng phương pháp thống kê) và cho thấy tỉ lệ nhận dạng là ngang nhau, trong khi hệ thống của Viola va Jones chạy nhanh hơn đến 15 lần so với hệ thống của Rowley-Baluja-Kanade và nhanh hơn 600 lần hệ thống của Schneiderman-Kanade.

Lý do mà cấu trúc cascade đạt tốc độ nhận dạng nhanh chính là nhờ nó sớm loại bỏ được các mẫu background đơn giản (thường có số lượng lớn hơn nhìu so với các mẫu chứa object -cac mẫu chứa khuôn mặt cần tiến hành nhận dạng). Bên cạnh đó hệ thống của Viola va Jones cũng

đat được độ chính xác khá cao nhờ vào thuật toán cascade training các bộ nhận dạng được huấn luyện bằng AdaBoost với đặc trưng Haar-like mô tả tốt thông tin đối tượng, cùng với cách tính Integral Image tính nhanh các đặc trưng, không làm giảm tốc độ nhận dạng của hệ thống. Như vậy, mô hình Cascade of Boosted Classifiers thật sự là một cách tiếp cận tốt cả về tốc độ lần khả năng nhận dạng, rất phù hợp với bài nhận dạng mặt người.

1. **Hệ thống xác định vị trí khuôn mặt người**

[](http://4.bp.blogspot.com/-KdYw3fMVH84/TzI6gxosTUI/AAAAAAAAAgA/VC5axVMJ5EQ/s1600/Facedetectionsystem.gif.jpg)

Hình 7: Hệ thống xác định vị trí khuôn mặt người (Face detection system)

Như trong hình trên, từ ảnh gốc ban đầu, ta sẽ được tính Integral Image, là mảng 2 chiều với phần tử (x, y) sẽ được tính bằng tổng của các phần tử (x', y') với x' < x và y' < y, mục đích là để tính nhanh tổng của các giá trị mức xám của một vùng hình chữ nhật bất kỳ trên ảnh gốc. Các vùng ảnh con này sẽ được đưa qua các hàm Haar cơ bản để ước lượng đặc trưng, kết quả ước lượng sẽ được đưa q11ua bộ điều chỉnh AdaBoost để loại bỏ nhanh các đặc trưng không có khả năng là đặc trưng của khuôn mặt người. Chỉ có một tập nhỏ các đặc trưng mà bộ điều chỉnh AdaBoost cho là có khả năng là đặc trưng của khuôn mặt người mới được chuyển sang cho bộ quyết định kết quả (là tập các bộ phân loại yếu có cấu trúc như trong hình 5). Bộ quyết định sẽ tổng hợp kết quả là khuôn mặt người nếu kết quả của các bộ phân loại yếu trả về là khuôn mặt người.

Mỗi bộ phân loại yếu sẽ quyết định kết quả cho một đặc trưng Haar-like, được xác định ngưỡng đủ nhỏ sao cho có thể vượt được tất cả các bộ dữ liệu mẫu trong tập dữ liệu huấn luyện (số lượng ảnh khuôn mặt trong tập huấn luyện có thể rất lớn). Trong quá trình xác định khuôn mặt người, mỗi vùng ảnh con sẽ được kiểm tra với các đặc trưng trong chuỗi các đặc trưng Haar-like, nếu có một đặc trưng Haar-like nào cho ra kết quả là khuôn mặt người thì các đặc trưng khác không cần xét nữa. Thứ tự xét các đặc trưng trong chuỗi các đặc trưng Haar-like sẽ được dựa vào trọng số (weight) của đặc trưng đó do AdaBoost quyết định dựa vào số lần và thứ tự xuất hiện của các đặc trưng Haar-like.

# PHẦN BA : XÂY DỰNG CHƯƠNG TRÌNH

## Đặc tả chương trình

* Hướng xử lý từ Raspi: Chương trình thu thập dữ liệu hình ảnh từ camera và phân tích xử lý sơ bộ bằng Open CV để nhận diện các khuôn mặt xuất hiện trong tầm quay của camera. Dữ liệu ảnh thu được đó đồng thời được gửi đến các dịch vụ của Microsoft để phân tích (Computer Vision API và Face API). Dữ liệu trả về là mô tả về khung cảnh, con người và cảm xúc của những người đang có trong tấm ảnh gửi lên. Từ dữ liệu đó robot sẽ có thể phản ứng lại với cảm xúc của người đối diện. Ngoài ra dữ liệu từ camera và kết quả phân tích còn được gửi lên trên máy tính thông qua websocket để hiển thị trực quan với người dùng; giúp kiểm tra quá trình hoạt động của Raspi.
* Hướng điều khiển từ máy tính ngoài: Chương trình chạy trên web còn có tích hợp Voice Recognition giúp nhận diện giọng nói của người dùng và truyền tới Raspi. Raspi từ đó có thể phản ứng lại theo các điều kiện lập trình trước (chưa có ứng dụng AI). Và để tương tác tốt hơn, trình duyệt web còn sử dụng Text to Speech được tích hợp sẵn trong Chrome để chuyển đổi nội dung phản hồi từ Raspi sang tiếng nói giúp giao tiếp tự nhiên và thân thiện hơn.

## Thiết kế hệ thống

Với các chức năng đã nêu ở trên thì nhìn chung chương trình sẽ có những phần sau đây : Kết nối Raspi, Xử lí đầu vào, Phân tích đầu vào, Đầu ra.

### Kết nối Raspi

Sử dụng chương trình Websocket để kết nối giao tiếp với Raspi

### Xử lí đầu vào

1. Đầu vào của chương trình là một hình ảnh được nhận dạng từ camera của Raspi nên ta phải chuyển thành các ảnh tĩnh và xử lý trên từng ảnh tĩnh. Sau khi đã có ảnh đầu vào thì sẽ chuyển ảnh cho giai đoạn xứ lý tiếp theo, đó là phát hiện và nhận dạng các khuôn mặt có trong ảnh.
2. Phân tích đầu vào

Chúng ta sẽ tiến hành sử dụng:

* OpenCV để nhận diện khuôn mặt bằng thời gian thực
* Microsoft Computer Vision API: phân tích khung cảnh hiện tại, đồng thời nhận diện khuôn mặt, giới tính, tuổi tác người trong khung cảnh
* Microsoft Face API: Phân tích sâu khuôn mặt, hình dáng, ngũ quan, trang sức... và nhận biết cảm xúc, mức độ hạnh phúc... Ngoài ra còn giúp nhận diện danh tính người trong ảnh với dữ liệu rèn luyện trước
* Về phần giọng nói ta sẽ sử dụng Voice Recognition của Google để nhận diện

### Đầu ra

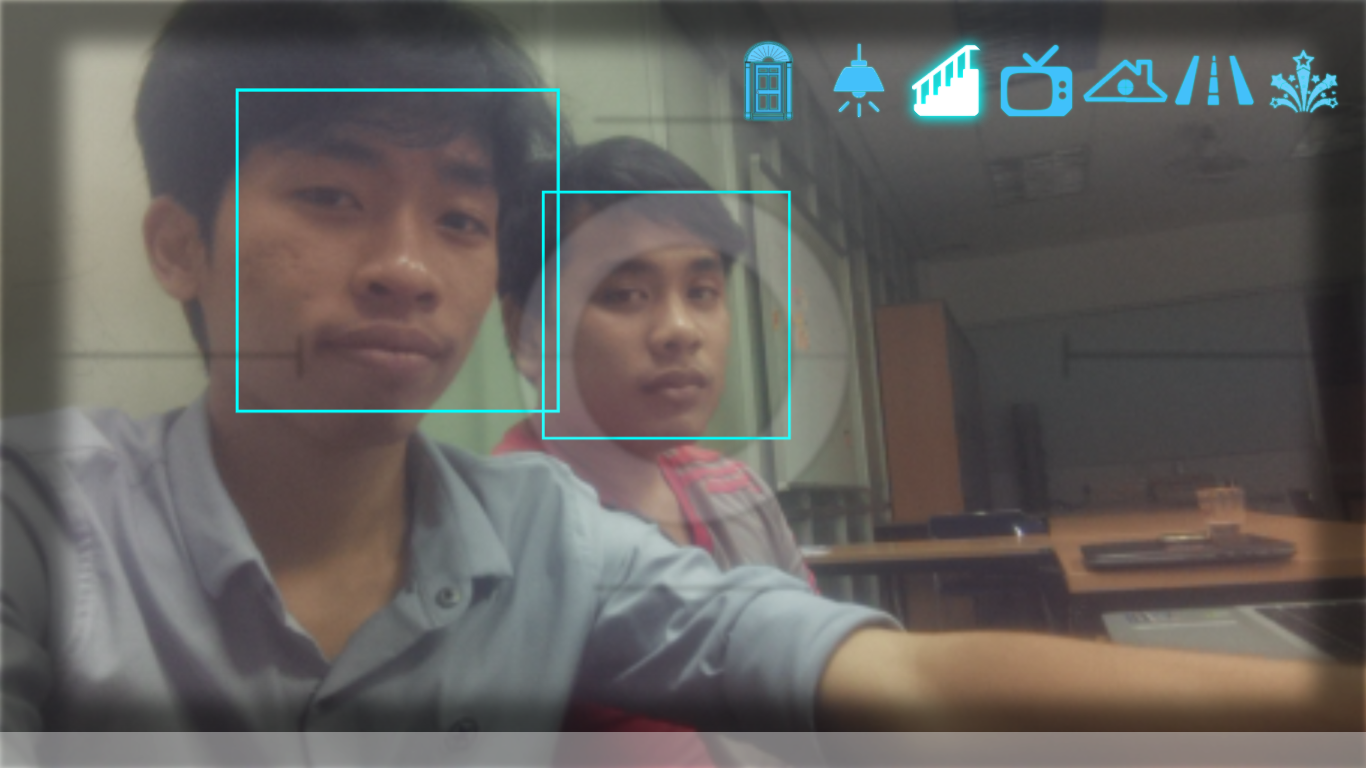
Các khuôn mặt sẽ được tách ra và lưu dưới dạng file. Sau đó trả về trên giao diện của một trang web, dùng các ngôn ngữ lập trình web front-end (HTML, CSS, JS) và một số bộ thư viện hỗ trợ (Bootstrap) để thực hiện. Được viết trên nền tảng tiện ích cho trình duyệt nhân chromium (khả dụng trên chrome, cốc cốc, chromium)

Về phần giọng nói sẽ sử dụng Speech to text của chrome để phát lại nội dung trả lời từ Raspi được lập trình sẵn trên máy.

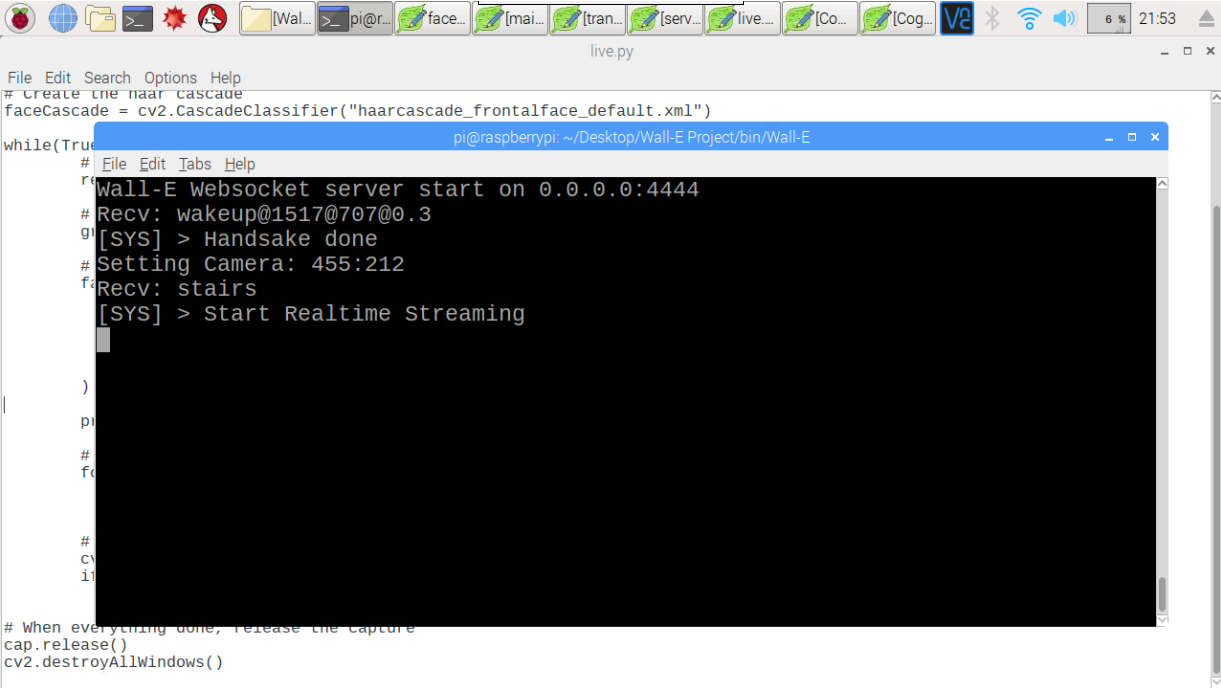
## Một số hình ảnh từ chương trình



Nhận diện khung cảnh với ComputerVision



Nhận diện thời gian thực với OpenCV



Chương trình chính trên Raspi



Trò chuyện bằng giọng nói



Xác định cảm xúc, danh tính từ dữ liệu rèn luyện

## Giới hạn hệ thống

* Chương trình chưa nhận dạng được một cách chính xác hoàn toàn, phần khuôn mặt người nhận diện không được nhanh và chuẩn xác. Có nhiều khi vẫn không nhận diện được khuôn mặt do một số tác động bên ngoài như ánh sang, độ gần xa, góc độ.
* Robot chưa thể làm gì ngoài việc nhận diện và tương tác với con người một cách đơn giản những câu đã được lập trình sẵn trong hệ thống. Chưa thể giao tiếp một cách tự nhiên cũng như giúp đỡ con người trong những công việc đơn giản khác.
* Do phần kinh phí đầu tư thấp nên ngoại hình robot không được đẹp và trang bị thêm những thiết bị tốt hơn nhiều tính năng hơn.

## Hướng phát triển

* Hoàn thiện robot để có thể làm được những công việc như chào người khác, nhận ra được đâu là người quen đâu là người lạ.
* Di chuyển được theo ý của con người hoặc tự động làm những công việc cần thiết như xếp lại giày dép, mở và đóng cửa…
* Có thể giao tiếp được với con người và trò chuyện ở một mức độ đơn giản nào đó như là kể lại những thứ hàng ngày gặp phải hoặc đọc tên được những thứ mà con người đưa ra
* Đầu tư thêm về kinh phí để mua thêm những thiết bị tốt hơn, làm thêm nhiều tính năng cho robot ở mức độ khó hơn.
* Đề tài này có thể áp dụng để tạo ra những con robot có ích cho xã hội sau này

# TÀI LIỆU THAM KHẢO

Face API : <https://azure.microsoft.com/en-us/services/cognitive-services/face/>

Computer Vision API: <https://azure.microsoft.com/en-us/services/cognitive-services/computer-vision/>

OpenCV: <https://opencv.org/>

Web Socket: <https://en.wikipedia.org/wiki/WebSocket>

Raspberry Pi Document: <https://www.raspberrypi.org/>