Shape, square

Description automatically generated**TRƯỜNG   
VIỆN**

-----🙞🙜🕮🙞🙜-----

**<<<logo trường>>>**

*Báo cáo môn học: Trí tuệ nhân tạo*

**Đề tài:**

\*\*\*

**Giảng viên hướng dẫn:**

**Người thực hiện:**

Phạm Văn MSV:

*Huế, 2022*

Contents

[**Lời mở đầu** 3](#_Toc106146005)

[**Tóm tắt báo cáo** 3](#_Toc106146006)

[**Mở đầu** 4](#_Toc106146007)

[**1.** **Lý do chọn đề tài** 4](#_Toc106146008)

[**2.** **Mục đích và phạm vi nghiên cứu** 5](#_Toc106146009)

[**a.** **Mục đích** 5](#_Toc106146010)

[**b.** **Phạm vi đề tài** 5](#_Toc106146011)

[**CHƯƠNG 1: GIỚI THIỆU** 6](#_Toc106146012)

[**1.** **Bài toán phát hiện khuôn mặt (Face detection)** 6](#_Toc106146013)

[**2.** **Những khó khăn của hệ thống nhận dạng khuôn mặt** 8](#_Toc106146014)

[**CHƯƠNG 2: CƠ SỞ LÝ THUYẾT** 11](#_Toc106146015)

[**2.** **Thuật toán Facial landmarks** 12](#_Toc106146016)

[**3.** **Ước lượng tư thế đầu** 13](#_Toc106146017)

[a) Ước lượng tư thế đầu là gì? 13](#_Toc106146018)

[b) Biểu diễn về mặt toán học 13](#_Toc106146019)

[c. Thuật toán ước lượng tư thế 15](#_Toc106146020)

[**4.** **Phương pháp xác định góc quay** 17](#_Toc106146021)

[a. Nhắc lại kiến thức hình học 17](#_Toc106146022)

[b. Thuật toán 17](#_Toc106146023)

[**5.** **Thư viện OpenCV** 18](#_Toc106146026)

[**CHƯƠNG 3: CHƯƠNG TRÌNH ĐỀ XUẤT VÀ CÀI ĐẶT** 21](#_Toc106146053)

[**1.** **Chuẩn bị** 21](#_Toc106146054)

[**2.** **Mô tả bài toán** 21](#_Toc106146055)

[**4.** **Chạy thức nghiệm chương trình** 23](#_Toc106146056)

[**PHẦN 4: KẾT LUẬN** 24](#_Toc106146057)

[**TÀI LIỆU THAM KHẢO** 25](#_Toc106146058)

# **Lời mở đầu**

Đề tài này được thực hiện tại Viện …. Trường,,,,,

Em xin chân thành cảm ơn Thầy…. đã tận tình hướng dẫn em trong suốt thời gian thực hiện đề tài

# **Tóm tắt báo cáo**

Bài báo cáo gồm 04 phần:

+ Phần 1: giới thiệu về đề tài

+ Phần 2: Cơ sở lý thuyết

+ Phần 3: Chương trình

+ Phần 4: Kết luận

# **Mở đầu**

1. **Lý do chọn đề tài**

Với sự phát triển không ngừng của khoa học và công nghệ, đặc biệt là các thiết bị hỗ trợ công nghệ xử lý ảnh ngày càng hiện đại và được sử dụng phổ biến trong đời sống con người đã làm cho lượng thông tin thu được bằng hình ảnh ngày càng tăng và trở nên phổ biến. Theo đó, lĩnh vực xử lý ảnh cũng được chú trọng phát triển, ứng dụng rộng rãi trong đời sống xã hội hiện đại. Không chỉ dừng lại ở việc chỉnh sửa, tăng chất lượng hình ảnh mà với công nghệ xử lý ảnh hiện nay chúng ta có thể giải quyết các bài toán nhận dạng chữ viết, nhận dạng dấu vân tay, đặc biệt là nhận dạng khuôn mặt…

Một trong những bài toán được nhiều người quan tâm nhất trong lĩnh vực xử lý ảnh hiện nay đó là nhận dạng khuôn mặt. Như chúng ta đã biết, khuôn mặt đóng một vai trò quan trọng trong quá trình giao tiếp giữa người với người, nó mang một lượng thông tin giàu có, chẳng hạn như từ khuôn mặt chúng ta có thể xác định giới tính, tuổi tác, trạng thái cảm xúc, đặc biệt là xác định mối quan hệ với đối tượng. Do đó, bài toán nhận dạng khuôn mặt đóng vài trò quan trọng trong nhiều lĩnh vực của con người như các hệ thống giám sát, tìm kiếm thông tin,…đặc biệt là hệ thống an ninh bảo mật.

Có rất nhiều tác vụ có thể thực hiện trong bài toán này như phát hiện đeo khẩu trang, điểm danh, cảnh báo xâm nhập,… Trong bài báo cáo này, em xin trình bày về một bài toán khá mới trong lĩnh vực này, đó chính là bài toán xác định góc quay, cũng như tọa độ của các điểm trên khuôn mặt. Bài toán này nhìn chung là khá mới so với các bài toán kể trên, tuy vậy tiềm năng ứng dụng của nó là không hề nhỏ. Ví như việc phát hiện gian lận thi cử, định danh khuôn mặt thông qua cử chỉ,…

1. **Mục đích và phạm vi nghiên cứu**
2. **Mục đích**

* Xây dựng một chương trình cơ sở phát hiện khuôn mặt.
* Tìm hiểu về phương pháp xác định khuôn mặt, điểm điểm ảnh trên khuôn mặt, xác định tọa độ của khuôn mặt thông qua 1 việc xác định tọa độ 1 điểm cố định trên khuôn mặt, xác định góc quay của khuôn mặt.
* Tìm hiểu về các phương pháp xử lý ảnh, một số thư viện như OpenCV, từ đó xây dựng chương trình thực nghiệm.

1. **Phạm vi đề tài**

Đồ án tập trung vào nghiên cứu các thành phần chính và các lý thuyết liên quan đến hệ thống phát hiện khuôn mặt, các phương pháp xác định tọa độ, tính góc quay của khuôn mặt. Từ đó cài đặt chương trình và chạy trên máy tính cá nhân bằng ngôn ngữ Python.

# **CHƯƠNG 1: GIỚI THIỆU**

1. **Bài toán phát hiện khuôn mặt (Face detection)**

Phát hiện khuôn mặt là công nghệ máy tính dựa trên trí tuệ nhân tạo (AI) được sử dụng để tìm và xác định khuôn mặt của con người trong hình ảnh kỹ thuật số. Công nghệ phát hiện khuôn mặt có thể được áp dụng cho các lĩnh vực khác nhau - bao gồm an ninh, sinh trắc học, thực thi pháp luật, giải trí và an toàn cá nhân - để cung cấp giám sát và theo dõi mọi người trong thời gian thực

A person with blue hair

Description automatically generated with low confidence

Hình : Bài toán phát hiện khuôn mặt

Phát hiện khuôn mặt đã tiến triển từ các kỹ thuật thị giác máy tính thô sơ đến những tiến bộ trong học máy (ML) đến các mạng thần kinh nhân tạo ngày càng tinh vi (ANN) và các công nghệ liên quan; Kết quả đã được cải thiện hiệu suất liên tục. Bây giờ nó đóng một vai trò quan trọng là bước đầu tiên trong nhiều ứng dụng chính - bao gồm theo dõi khuôn mặt, phân tích khuôn mặt và nhận dạng khuôn mặt. Phát hiện khuôn mặt có ảnh hưởng đáng kể đến cách các hoạt động tuần tự sẽ thực hiện trong ứng dụng.

Trong phân tích khuôn mặt, phát hiện khuôn mặt giúp xác định phần nào của hình ảnh hoặc video nên được tập trung vào để xác định tuổi, giới tính và cảm xúc bằng cách sử dụng biểu cảm khuôn mặt. Trong một hệ thống nhận dạng khuôn mặt - ánh xạ các tính năng khuôn mặt của một cá nhân về mặt toán học và lưu trữ dữ liệu dưới dạng dữ liệu phát hiện khuôn mặt là cần thiết cho các thuật toán phân biệt các phần của hình ảnh hoặc video là cần thiết để tạo khuôn mặt. Sau khi được xác định, FacePrint mới có thể được so sánh với các dấu ấn được lưu trữ để xác định xem có trận đấu nào không.

**Vậy làm thế nào để máy nhận diện được khuôn mặt?**

Các ứng dụng phát hiện khuôn mặt sử dụng các thuật toán và ML để tìm khuôn mặt của con người trong các hình ảnh lớn hơn, thường kết hợp các vật thể phi mặt khác như phong cảnh, tòa nhà và các bộ phận cơ thể người khác như bàn chân hoặc bàn tay. Các thuật toán phát hiện khuôn mặt thường bắt đầu bằng cách tìm kiếm mắt người - một trong những tính năng dễ phát hiện nhất. Thuật toán sau đó có thể cố gắng phát hiện lông mày, miệng, mũi, lỗ mũi và mống mắt. Khi thuật toán kết luận rằng nó đã tìm thấy một vùng khuôn mặt, nó áp dụng các thử nghiệm bổ sung để xác nhận rằng trên thực tế, nó đã phát hiện một khuôn mặt.

Các phương pháp được sử dụng trong phát hiện khuôn mặt có thể dựa trên kiến ​​thức, dựa trên tính năng, khớp mẫu hoặc dựa trên ngoại hình. Mỗi người đều có những ưu điểm cũng như những khuyết điểm:

* Các phương pháp dựa trên kiến ​​thức, hoặc dựa trên quy tắc, mô tả một khuôn mặt dựa trên các quy tắc. Thách thức của phương pháp này là khó khăn trong việc đưa ra các quy tắc được xác định rõ ràng.
* Các phương pháp bất biến có tính năng - sử dụng các tính năng như mắt hoặc mũi của một người để phát hiện khuôn mặt - có thể bị ảnh hưởng tiêu cực bởi tiếng ồn và ánh sáng.
* Các phương pháp phù hợp với mẫu dựa trên việc so sánh hình ảnh với các mẫu hoặc tính năng khuôn mặt tiêu chuẩn đã được lưu trữ trước đó và tương quan hai để phát hiện khuôn mặt. Thật không may, các phương pháp này không giải quyết các biến thể trong tư thế, quy mô và hình dạng.
* Phương pháp dựa trên xuất hiện sử dụng phân tích thống kê và học máy để tìm các đặc điểm có liên quan của hình ảnh khuôn mặt. Phương pháp này, cũng được sử dụng trong trích xuất tính năng để nhận dạng khuôn mặt, được chia thành các phương pháp phụ.

Một số kỹ thuật cụ thể hơn được sử dụng trong phát hiện khuôn mặt bao gồm:

* Loại bỏ nền. Ví dụ, nếu một hình ảnh có nền đơn giản, màu đơn sắc hoặc nền tĩnh được xác định trước, thì việc loại bỏ nền có thể giúp tiết lộ ranh giới khuôn mặt.
* Trong hình ảnh màu, đôi khi màu da có thể được sử dụng để tìm mặt; Tuy nhiên, điều này có thể không hoạt động với tất cả các phức tạp.
* Sử dụng chuyển động để tìm khuôn mặt là một lựa chọn khác. Trong video thời gian thực, một khuôn mặt hầu như luôn luôn di chuyển, vì vậy người dùng phương pháp này phải tính toán khu vực di chuyển. Một nhược điểm của phương pháp này là nguy cơ nhầm lẫn với các đối tượng khác di chuyển trong nền.
* Một sự kết hợp của các chiến lược được liệt kê ở trên có thể cung cấp một phương pháp phát hiện khuôn mặt toàn diện.

Phát hiện khuôn mặt trong hình ảnh có thể phức tạp do sự thay đổi của các yếu tố như tư thế, biểu hiện, vị trí và định hướng, màu da và giá trị pixel, sự hiện diện của kính hoặc tóc trên khuôn mặt, và sự khác biệt về tăng camera, điều kiện ánh sáng và độ phân giải hình ảnh. Những năm gần đây đã mang lại những tiến bộ trong việc phát hiện khuôn mặt bằng cách sử dụng học tập sâu, điều này thể hiện lợi thế của việc vượt trội so với các phương pháp thị giác máy tính truyền thống.

1. **Những khó khăn của hệ thống nhận dạng khuôn mặt**

Bài toán nhận dạng mặt người là một bài toán khó nên những nghiên cứu hiện tại vẫn chưa đạt được những kết quả mong muốn. Chính vì thế, vấn đề này vẫn đang được nhiều tổ chức trên thế giới quan tâm nghiên cứu. Khó khăn của bài toán nhận dạng mặt người có thể kể đến như sau:

* Tư thế chụp, góc chụp: Ảnh chụp khuôn mặt có thể thay đổi rất nhiều bởi vì góc chụp giữa camera và khuôn mặt. Chẳng hạn như: chụp thẳng, chụp chéo bên trái hay chụp chéo bên phải, chụp từ trên xuống, chụp từ dưới lên,… Với các tư thế khác nhau, các thành phần trên khuôn mặt như mắt, mũi, miệng có thể bị khuất một phần hoặc thậm chí khuất hết.
* Thiếu một số thành phần của khuôn mặt: Các đặc trưng như: râu mép, râu hàm, mắt kính,… có thể xuất hiện hoặc không. Vấn đề này làm cho bài toán càng trở nên khó hơn rất nhiều.
* Biểu cảm của khuôn mặt: Biểu cảm của khuôn mặt con người có thể làm ảnh hưởng đáng kể lên các thông số của khuôn mặt. Chẳng hạn, cùng một khuôn mặt một người, nhưng có thể sẽ rất khác khi họ cười hoặc sợ hãi,…
* Góc khuất: Khuôn mặt có thể bị che khuất bởi các đối tượng khác hoặc các khuôn mặt khác.
* Hướng của ảnh (pose variations): Trong ảnh, khuôn mặt có thể biến đổi rất nhiều với các góc quay khác nhau của trục camera. Chẳng hạn chụp với trục máy ảnh nghiêng làm cho khuôn mặt bị nghiêng so với trục của ảnh.
* Điều kiện ảnh: Ảnh được chụp trong các điều kiện khác nhau như là chiếu sáng, về tính chất camera (máy kỹ thuật số, máy hồng ngoại, v.v…), ảnh có chất lượng thấp ảnh hưởng rất nhiều đến chất lượng ảnh khuôn mặt.

1. **Các ứng dụng nhận diện khuôn mặt**

Bài toán phát hiện mặt người trong ảnh là một bài toán phức tạp nhưng lại hấp dẫn nhiều người nghiên cứu về nó, có lẽ bởi tính ứng dụng thực tế của bài toán là rất lớn và đa dạng trong cuộc sống hiện đại. Có thể kể ra một số ứng dụng thông dụng nhất của bài toán nhận diện mặt người:

A picture containing text, clipart, vector graphics

Description automatically generated

Hình : Nhân dạng khuôn mặt có rất nhiều ứng dụng

* Hệ thống giao tiếp thông minh giữa người và máy: con người có thể xây dựng những hệ thống thông giao tiếp giữa người và máy tính thông qua việc nhận diện khuôn mặt, biểu cảm trên khuôn mặt người để dự đoán, nhận biết trạng thái tâm lý hiện thời của người đó. Một ngôi nhà thông minh trong tương lai có thể nhận biết được chủ nhân của nó thông qua nhận biết khuôn mặt, dáng người, giao tiếp qua giọng nói, vân tay…
* Nhận dạng tội phạm: Hệ thống có thể nhận diện ra một khuôn mặt ngay tức thì và đối chiếu với hàng triệu bản ghi có sẵn trong cơ sở dữ liệu để chỉ ra đó có thể là một tội phạm đang bị truy tìm hay không, hoặc đó có thể là một nhân vật nào đó đặc biệt cần quan tâm,....
* Giải trí: trong hầu hết các máy ảnh hiện đại ngày nay đều có chức năng tự động nhận diện mặt người để có thể lấy độ nét, điều chỉnh ánh sáng cho phù hợp với khung cảnh xung quanh. Trên một số trang web cũng đã áp dụng công nghệ tự động nhận diện mặt người và so sánh với kho dữ liệu khổng lồ của mình để đưa ra những lời chào, tính năng thông minh nhất cho người sử dụng.
* Hệ thống quan sát, theo dõi và bảo vệ: Hệ thống camera sẽ xác định đâu là con người và theo dõi người đó…
* Điều khiển ra vào các cơ quan, văn phòng: cho phép nhân viên ra vào các khu vực quan trọng mà không cần phải đăng nhập hay dùng thẻ. Nếu kết hợp với sử dụng vân tay hay hốc mắt thì sẽ đem lại kết quả chính xác cao.
* Tổ chức tìm kiếm liên quan đến con người thông qua khuôn mặt trên nhiều hệ cơ sở dữ liệu lớn.
* Phân tích cảm xúc của khuôn mặt người

# **CHƯƠNG 2: CƠ SỞ LÝ THUYẾT**

1. **Tổng quan về xử lý ảnh**
2. Khái niệm xử lý ảnh

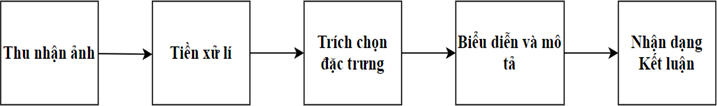
Xử lí ảnh là quá trình biến đổi từ một ảnh ban đầu sang một ảnh mới với các đặc tính và tuân theo ý muốn của người sử dụng. Xử lí ảnh có thể bao gồm quá trình phân tích, phân lớp các đối tượng, làm tăng chất lượng, phân đoạn và tách cạnh, gán nhãn cho vùng hay quá trình biên dịch các thông tin hình ảnh.

**Mục đích của xử lí ảnh:**

* Biến đổi làm tăng chất lượng ảnh
* Tự động nhận dạng ảnh, đánh giá nội dung ảnh
* Kiểm tra, phân loại ảnh
* Nhận biết và đánh giá ảnh từ đó có thể phân tích ảnh thành những phần có ý nghĩa để phân biệt giữa đối tượng này với đối tượng khác từ đó áp dụng thực tiễn vào các ứng dụng trong đời sống

1. Quá trình của xử lí ảnh

Quá trình xử lí ảnh là thực hiện các kỹ thuật xử lí ảnh đầu vào nhằm cho ra kết quả mong muốn. Quá trình xử lí ảnh gồm các bước sau:



Hình : Quá trình xử lí ảnh

**Sơ đồ này gồm các phần sau:**

* Thu nhận ảnh: Tiếp thu ảnh đầu vào có thể từ nhiều nguồn khác nhau từ camera , video, tệp thư mục ảnh…
* Tiền xử lí: Sau khi thu nhận ảnh, ảnh được xử lí có thể là tăng chất lượng ảnh như xử lí nhiễu, phân vùng, tăng độ tương phản ,… để làm ảnh được rõ hơn nét hơn.
* Trích chọn đặc trưng: Sau khi tiền xử lí thu được ảnh đã phù hợp, giai đoạn này trích chọn ra đặc trưng của bức ảnh gắn với việc tách các đặc tính của ảnh dưới dạng các thông tin phù hợp với yêu cầu của cơ sở dữ liệu.
* Biểu diễn và mô tả: Sau khi tìm ra các đặc trưng của bức ảnh , phân loại và chọn liên kết các đặc trưng với nhau tạo thành một khối dễ dàng nhận dạng và đưa ra kết quả.
* Nhận dạng/ kết luận: Sau khi có các dữ liệu cần thiết sẽ tiến hành nhận dạng phân tích và đưa ra kết quả.

1. **Thuật toán Facial landmarks**

Facial landmarks được sử dụng để xác định vị trí và thể hiện các vùng nổi bật trên khuôn mặt, chẳng hạn như: mắt, lông mày, mũi, miệng, hàm răng,…

Facial landmarks là một tập con của bài toán dự đoán hình dạng. Đưa ra một hình ảnh đầu vào (và thường là ROI chỉ định đối tượng quan tâm), công cụ dự đoán hình dạng cố gắng xác định vị trí các điểm quan tâm chính dọc theo hình dạng.

Mục tiêu của facial landmarks là phát hiện các cấu trúc khuôn mặt quan trọng trên khuôn mặt bằng các phương pháp dự đoán hình dạng.

Chart, scatter chart

Description automatically generated

Hình : 68 tọa độ facial landmark

Việc nhận diện Facial landmarks đi theo 2 bước sau:

* **Bước 1**: Khoanh vùng khuôn mặt trong ảnh.

Có nhiều cách thực hiện bước này, trong đó ta có thể sử dụng một bộ nhận dạng pre-trained HOG + Linear SVM được đào tạo trước dành riêng cho nhiệm vụ phát hiện khuôn mặt.

* **Bước 2**: Phát hiện các cấu trúc khuôn mặt chính trên ROI trên khuôn mặt.

Có nhiều loại facial landmark, nhưng tất cả các phương pháp về cơ bản đều cố gắng xác định vị trí và gắn nhãn các vùng trên khuôn mặt như: Miệng, Lông mày phải, Lông mày trái, Mắt phải, Mắt trái, Mũi, Hàm.

Kazemi và Sullivan (2014) đã đề xuất một phương pháp như sau:

* Một tập hợp đào tạo về các điểm mốc trên khuôn mặt được gắn nhãn trên một hình ảnh. Những hình ảnh này được gắn nhãn thủ công, xác định tọa độ (x, y) cụ thể của các vùng xung quanh mỗi cấu trúc khuôn mặt.
* Xác suất về khoảng cách giữa các cặp pixel đầu vào.

Với dữ liệu đào tạo này, một nhóm các cây hồi quy được đào tạo để ước tính các vị trí facial landmark trực tiếp từ chính cường độ pixel (tức là không có "trích xuất tính năng").

1. **Ước lượng tư thế đầu**

Trong nhiều ứng dụng, chúng ta cần biết làm thế nào đầu bị nghiêng đối với máy ảnh. Ví dụ, trong một ứng dụng thực tế ảo, người ta có thể sử dụng tư thế của đầu để hiển thị góc nhìn đúng của cảnh. Trong hệ thống hỗ trợ lái xe, máy ảnh nhìn vào mặt người lái trong xe có thể sử dụng ước tính tư thế đầu để xem người lái xe có chú ý đến đường hay không. Và tất nhiên người ta có thể sử dụng cử chỉ dựa trên tư thế đầu để kiểm soát một ứng dụng / trò chơi rảnh tay. Ví dụ, nghiêng đầu từ trái sang phải có thể biểu thị không.

1. Ước lượng tư thế đầu là gì?

Trong thị giác máy tính, tư thế của một đối tượng đề cập đến định hướng và vị trí tương đối của nó đối với máy ảnh. Ta có thể thay đổi tư thế bằng cách di chuyển đối tượng đối với máy ảnh hoặc máy ảnh đối với đối tượng.

Vấn đề ước tính tư thế được mô tả trong hướng dẫn này thường được gọi là vấn đề quan điểm-N-điểm hoặc PNP trong biệt ngữ tầm nhìn máy tính. Như chúng ta sẽ thấy trong các phần sau chi tiết hơn, trong vấn đề này, mục tiêu là tìm tư thế của một đối tượng khi chúng ta có một camera được hiệu chỉnh và chúng ta biết các vị trí của N 3D điểm trên đối tượng và các phép chiếu 2D tương ứng trong bức hình.

1. Biểu diễn về mặt toán học

Một đối tượng cứng 3D chỉ có hai loại chuyển động đối với máy ảnh:

* **Dịch chuyển**: Di chuyển máy ảnh từ vị trí 3D hiện tại (*X, Y, Z*) sang vị trí 3D mới được gọi là (*X’, Y’, Z’*). Như ta có thể thấy dịch có 3 tọa độ - ta có thể di chuyển theo hướng X, Y hoặc Z. Dịch được đại diện bởi một vector **t** bằng (*X’ - X, Y’ - Y, Z’ - Z*).
* **Xoay**: Ta cũng có thể xoay camera về trục X, Yvà Z. Do đó, một vòng quay cũng có ba độ tự do. Có nhiều cách thể hiện vòng quay. Ta có thể biểu diễn nó bằng cách sử dụng các góc Euler, ma trận xoay (3x3) hoặc hướng xoay (tức là trục) và góc.

Vì vậy, ước tính tư thế của một vật thể 3D có nghĩa là tìm 6 tham số - ba để dịch và ba để quay.

A picture containing person, person, indoor, looking

Description automatically generated

Hình : Ước lượng tư thế

Để tính toán tư thế 3D của một đối tượng trong một hình ảnh, ta cần thông tin sau:

* **Tọa độ 2D của một vài điểm**: Ta cần các vị trí 2D (x, y) của một vài điểm trong hình ảnh. Trong trường hợp của một khuôn mặt, ta có thể chọn các góc của mắt, đầu mũi, khóe miệng, v.v. Máy dò mốc khuôn mặt của Dlib cung cấp cho chúng ta nhiều điểm để lựa chọn. Trong hướng dẫn này, chúng ta sẽ sử dụng đầu mũi, cằm, góc trái của mắt trái, góc phải của mắt phải, góc bên trái của miệng và góc phải của miệng.
* **Vị trí 3D của cùng một điểm**: Ta cũng cần vị trí 3D của các điểm tính năng 2D. Ta có thể nghĩ rằng ta cần một mô hình 3D của người trong ảnh để có được vị trí 3D. Lý tưởng nhất là có, nhưng trong thực tế, ta không. Một mô hình 3D chung sẽ đủ. Ta lấy mô hình 3D của đầu từ đâu? Ta thực sự không cần một mô hình 3D đầy đủ. Ta chỉ cần vị trí 3D của một vài điểm trong một số khung tham chiếu tùy ý. Ví dụ, ta có thể sử dụng các điểm 3D sau đây:
* Đầu mũi: (0,0, 0,0, 0,0)
* Cằm: (0,0, -330.0, -65.0)
* Góc trái của mắt trái: (-225.0, 170.0, -135.0)
* Góc phải của mắt phải: (225.0 , 170.0, -135.0)
* Góc bên trái của miệng: (-150.0, -150.0, -125.0)
* Góc phải của miệng: (150.0, -150.0, -125.0)

1. Thuật toán ước lượng tư thế

Có một số thuật toán để ước tính tư thế. Thuật toán đầu tiên được biết đến có từ năm 1841. Nó nằm ngoài phạm vi của bài đăng này để giải thích các chi tiết của các thuật toán này nhưng đây là một ý tưởng chung.

Có ba hệ tọa độ đang đặt ở đây. Các tọa độ 3D của các đặc điểm khuôn mặt khác nhau được hiển thị ở trên là trong tọa độ thế giới. Nếu chúng ta biết vòng quay và dịch (tức là tư thế), chúng ta có thể chuyển đổi các điểm 3D trong tọa độ thế giới thành các điểm 3D trong tọa độ camera. Các điểm 3D trong tọa độ camera có thể được chiếu lên mặt phẳng hình ảnh (tức là hệ tọa độ hình ảnh) bằng cách sử dụng các tham số nội tại của máy ảnh (tiêu cự, trung tâm quang học, v.v.).

Diagram

Description automatically generated

Hình : Thuật toán ước lượng tư thế

Trong hình trên, *o* là tâm của máy ảnh và mặt phẳng được hiển thị trong hình là mặt phẳng hình ảnh. Chúng ta quan tâm đến việc tìm hiểu những phương trình chi phối phép chiếu *p* của điểm ảnh 3D *P* lên mặt phẳng hình ảnh.

Giả sử chúng ta biết vị trí của một điểm 3D *P* với tọa độ (*U, V, W*) trong mặt phẳng. Nếu chúng ta biết phép quay *R* (ma trận 3 × 3) và phép dịch chuyển *t* (một vectơ 3 × 1), của tọa độ phẳng đối với tọa độ camera, chúng ta có thể tính toán vị trí (*X, Y, Z*) của điểm *P* trong hệ tọa độ camera bằng cách sử dụng phương trình sau đây:

Engineering drawing

Description automatically generated with medium confidence

Ở dạng mở rộng, phương trình trên trở thành:

Table

Description automatically generated

**Biến đổi tuyến tính trực tiếp:**

Chúng ta biết nhiều điểm trên mô hình 3D (tức là (*U, V, W*)), nhưng chúng ta không biết (*X, Y, Z*). Chúng ta chỉ biết vị trí của các điểm 2D (tức là (*x,y)*). Trong trường hợp không có biến dạng xuyên tâm, các tọa độ (*x,y*) của điểm *p* trong tọa độ hình ảnh được đưa ra bởi:

A picture containing text, clock

Description automatically generated

ở đâu, và là độ dài tiêu cự theo hướng x và y, và  là trung tâm quang học. Mọi thứ trở nên phức tạp hơn một chút khi có liên quan đến biến dạng xuyên tâm và với mục đích đơn giản, tôi sẽ bỏ nó ra.

*s* là một yếu tố quy mô chưa biết. Nó tồn tại trong phương trình do thực tế là trong bất kỳ hình ảnh nào chúng ta không biết độ sâu. Nếu ta cho bất kỳ điểm 3D *P* nào vào trung tâm của máy ảnh, điểm, trong đó tia giao nhau giao nhau, mặt phẳng hình ảnh là hình ảnh của *P*. Lưu ý rằng tất cả các điểm dọc theo tia tham gia trung tâm của máy ảnh và điểm *P* tạo ra cùng một hình ảnh. Nói cách khác, sử dụng phương trình trên ta chỉ có thể tìm được *(X, Y, Z)* với tỉ lệ *s*.

Do đó:

Table

Description automatically generated

Phương trình có dạng trên có thể được giải quyết bằng một số cách có trong đại số cí như phương pháp biến đổi tuyến tính.

1. **Phương pháp xác định góc quay**

Về tổng quan, phương pháp sử dụng các vị trí landmark trên khuôn mặt để nối lại thành một tam giác gồm 3 điểm, ví dụ như mắt trái, mắt phải, và mũi. 3 điểm này hoàn toàn xác định về tọa độ, do đó ta có thể xác định được độ dài 3 cạnh. Từ độ dài 3 cạnh ta hoàn toàn có thể xác định được góc của tam giác. Việc di chuyển mặt sẽ làm tọa độ các điểm thay đổi, kéo theo đó là góc của tam giác đó cũng thay đổi. Ta gọi đó là góc quay của khuôn mặt.

1. Nhắc lại kiến thức hình học

Hai vecto a và b tạo thành 1 góc :

A picture containing clock

Description automatically generated

Hình : Góc tạo thành từ 2 vecto

Ta hoàn toàn có thể tính được góc xem giữa 2 vecto thông qua các kiến thức ở phổ thông:

A picture containing text, device, gauge

Description automatically generated

Hình : Tính góc giữa 2 vecto

1. Thuật toán

Dưới đây là thuật toán xác định góc quay khuôn mặt:

Diagram

Description automatically generated

Hình : Thuật toán xác định góc quay khuôn mặt

Chúng ta xác định định góc quay của khuôn mặt thông qua việc xác định 2 góc  và . Thuật toán cụ thể thuật toán trên xác định 2 góc  và  thông qua việc tính các cạnh của tam giác. Sau khi xác định được các mốc điểm trên khuôn mặt, chúng ta tính độ lớn các vecto  và  và tính góc giữa chúng. Tương tự, đối với 2 vecto  và . Do vậy chúng ta xác định được  và . Sau khi xác định được  và , chúng ta đưa  và  vào thuật toán như đầu vào của thuật toán. Tuy vậy, do ta đã xác định được vị trí của khuôn mặt ở thuật toán trên, nên ta chỉ sử dụng thuật toán này và ý tưởng của nó để xác định góc quay của khuôn mặt.

1. **Thư viện OpenCV**

OpenCV (Open Source Computer Vision Library) là một thư viện phần mềm và thị giác máy tính, mã nguồn mở. OpenCV cung cấp một cơ sở hạ tầng chung cho các ứng dụng thị giác máy tính.

Thư viện có hơn 2500 thuật toán được tối ưu hóa, bao gồm toàn bộ thuật toán máy học và thị giác máy tính cổ điển và hiện đại. OpenCV có hơn 47 nghìn người dùng cộng đồng và số lượt tải xuống ước tính vượt quá 18 triệu lượt.

OpenCV có trên các giao diện C, C++, Python và Java hỗ trợ nhiều hệ điều hành khác nhau Windows, Linux, Mac OS, IOS, Android.

**Ứng dụng của OpenCV:**

OpenCV được sử dụng cho đa dạng mục đích và ứng dụng khác nhau bao gồm:

* Kiểm tra và giám sát tự động
* Robot và xe hơi tự lái
* Phân tích hình ảnh y học
* Tìm kiếm và phục hồi hình ảnh/video
* Phim – cấu trúc 3D từ chuyển động

**Chức năng của OpenCV:**

* Xử lí và hiển thị hình ảnh/video
* Phát hiện các vật thể.
* Điện toán hình ảnh.
* Áp dụng vào học máy và phân cụm.

**Các mô-đun phổ biến của OpenCV:**

OpenCV có cấu trúc mô-đun, bao gồm một số thư viện liên kết tĩnh (static libraries) hoặc thư viện liên kết động (shared libraries). Một số các module phổ biến có sẵn như sau:

* Mô-đun core (Core functionality): Đây là mô-đun chứa các cấu trúc, lớp cơ bản mà OpenCV sẽ sử dụng trong việc lưu trữ và xử lí hình ảnh và các phương thức cơ bản sử dụng cho các mô-đun khác.
* Mô-đun imgproc (Image Processing): Mô-đun xử lí hình ảnh bao gồm cả lọc hình ảnh tuyến tính và phi tuyến (linear and non-linear image filtering), phép biến đổi hình học, chuyển đổi không gian màu, biểu đồ và nhiều cái khác.
* Mô-đun highgui (High-level GUI): Đây là một mô-đun cho phép tương tác với người dùng trên giao diện như hiển thị hình ảnh , quay video.

* Mô-đun feature2d (2D Features Framework): Phát hiện tính nổi bật của bộ phận nhận diện, truy xuất thông số, thông số đối chọi.
* Mô-đun objdetect (Object Detection): Mô-đun cho việc phát hiện các đối tượng và mô phỏng các hàm được định nghĩa sẵn như: Khuôn mặt, cốc, người, xe,…
* Mô-đun video (Video Analysis) : Module phân tích video bao gồm các tính năng ước tính chuyển động, tách nền và các thuật toán theo dõi vật thể.
* Mô-đun calib3d (Camera Calibration and 3D Reconstruction): Thuật toán hình học đa chiều cơ bản, hiệu chuẩn máy ảnh, dự đoán kiểu dáng của đối tượng và các yếu tố tái tạo 3D.

# **CHƯƠNG 3: CHƯƠNG TRÌNH ĐỀ XUẤT VÀ CÀI ĐẶT**

1. **Chuẩn bị**

* Môi trường: Ngôn ngữ lập trình Python
* Thư viện: dlib, numpy, opencv-python,…

1. **Mô tả bài toán**

Sau khi sử dụng công cụ trong thư viện opencv để mở camera, chúng ta tiến hành xác định landmarks của khuôn mặt. Ta phân biệt nhiệm vụ của 2 loại landmarks 2D và 3D như sau:

* Các điểm landmarks 2D sẽ được sử dụng để xác định góc quay của khuôn mặt
* Các điểm landmarks 3D sẽ xác định vị trí của khuôn mặt trong không gian 3D, từ đó xác định chúng ta có đang quay trái, phải, lên, xuống hay không.

1. **Xây dựng chương trình**

Text

Description automatically generatedChương trình được xây dựng theo cấu trúc Class, đi kèm với đó là các hàm xử lý độc lập. Có 3 lớp chính trong chương trình là:

Lớp Timer: lớp này có nhiệm vụ tính thời gian thực hiện từng tác vụ của chương trình để đảm bảo tính tối ưu của chương trình.

Text

Description automatically generated  
Lớp Annotator: lớp này đảm nhận việc hiển thị trên giao diện của cửa sổ real-time.

Trong chương trình, việc ta hiển thị quá nhiều thông tin khiến cho tốc độ xử lý chậm đi đáng kể, đặc biệt là ở chế độ real-time.

Text

Description automatically generated

Lớp HeadposeDetection: Đảm nhận việc xác định tọa độ của khuôn mặt

Ngoài ra, còn một số hàm xử lý riêng biệt, ví như hàm npAngle(a,b,c): khi truyền vào tọa độ 3 điểm nó sẽ trả về góc của tam giác, hay hàm face\_angle(im): truyền vào ảnh sẽ trả về góc quay trái, phải của mặt.

1. **Chạy thức nghiệm chương trình**

Phần này bạn chạy… và chụp ảnh quá trình bạn chạy vào đây nhé

# **PHẦN 4: KẾT LUẬN**

Báo cáo đồ án đã trình bày các kiến thức cơ bản để nhận diện khuôn mặt. Trên cơ sở tìm hiểu về bài toán nhận diện mặt người, bài báo cáo đã trình bày được phương pháp tính góc xoay và xác định được vị trí của khuôn mặt trong không gian 3 chiều. Cùng với đó, tác giả cũng xây dựng được chương trình và thực nghiệm ra kết quả khả thi.

Do thời gian hạn chế nên chương trình vẫn còn nhiều nhược điểm chưa được xử lý hết. Để chương trình được sử dụng vào trong thực tế, cần giải quyết những nhược điểm mà phầm mềm còn gặp phải, thực hiện các ý tưởng mới nhằm nâng cao tốc độ, hiệu suất và độ chính xác của chương trình.

# **TÀI LIỆU THAM KHẢO**

1, Nawaf Alageel, “Face Pose Detection”, June 2021

2, Adrian Rosebrock, “Facial landmarks with dlib, OpenCV, and Python”, April 3, 2017, url=<https://pyimagesearch.com/2017/04/03/facial-landmarks-dlib-opencv-python>

3, Satya Mallick, “Head Pose Estimation using OpenCV and Dlib”, SEPTEMBER 26, 2016, url=” <https://learnopencv.com/head-pose-estimation-using-opencv-and-dlib>”