**TRƯỜNG ĐẠI HỌC XÂY DỰNG HÀ NỘI**

**KHOA CÔNG NGHỆ THÔNG TIN**

**NGÀNH KHOA HỌC MÁY TÍNH**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

A blue and white logo

Description automatically generated

**BÁO CÁO ĐỒ ÁN MÔN XỬ LÝ ẢNH**

**Đề tài: Phát hiện khuôn mặt**

**Giảng viên hướng dẫn: Thái Thị Nguyệt**

**Sinh viên thực hiện: Nguyễn Thu Huyền 0215366**

**Nguyễn Minh Hoàng 0214966**

**Trần Anh Dũng 0212066**

**Hoàng Đình Vinh 0222066**

**Nhóm số: 5**

**Lớp: 66CS1**

**Hà Nội, 20 tháng 12 năm 2023**

# **LỜI NÓI ĐẦU**

Với sự phát triển không ngừng của khoa học và công nghệ, đặc biệt là với những chiếc điện thoại thông minh (smartphone) 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. 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, nhận dạng khuôn mặt…  
 Một trong những bài toán được nhiều người quan tâm nhất của lĩnh vực xử lý ảnh hiện nay đó là nhận dạng khuôn mặt (Face Recognition). Như chúng ta đã biết, khuôn mặt đóng 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, chủng tộc, trạng thái cảm xúc, đặc biệt là xác định mối quan hệ với đối tượng (có quen biết hay không). Do đó, bài toán nhận dạng khuôn mặt đóng vai trò quan trọng trong nhiều lĩnh vực đời sống hằng ngày của con người như các hệ thống giám sát, quản lý vào ra, tìm kiếm thông tin một người nổi tiếng, …đặc biệt là an ninh, bảo mật. Có rất nhiều phương pháp nhận dạng khuôn mặt để nâng cao hiệu suất tuy nhiên dù ít hay nhiều những phương pháp này đang vấp phải những thử thách về độ sáng, hướng nghiêng, kích thước ảnh, hay ảnh hưởng của tham số môi trường.

Nhóm em xin được gửi lời cảm ơn chân thành tới cô Thái Thị Nguyệt đã dành thời gian và tâm huyết hướng dẫn nhóm em trong suốt quá trình thực hiện đồ án Xử lý ảnh. Dưới sự hướng dẫn tận tình của cô, nhóm em đã có cơ hội được tìm hiểu và nghiên cứu về đề tài Phát hiện khuôn mặt. Em và các bạn đã tìm tòi và học hỏi được rất nhiều kiến thức và kỹ năng mới từ cô, giúp nhóm em có thể hoàn thành đồ án một cách tốt nhất.

Nhóm em xin chân thành cảm ơn cô một lần nữa vì sự giúp đoc của cô.

Nhóm 5

Nguyễn Thu Huyền

Nguyễn Minh Hoàng

Trần Anh Dũng

Hoàng Đình Vinh

**Mục lục**

[**I.** **Cơ sở lý thuyết** 5](#_Toc154043218)

[1. Phát hiện vùng da trên không gian màu Ycbcr 5](#_Toc154043219)

[2. Haar cascade 6](#_Toc154043220)

[**II.** **Ý tưởng bài toán** 8](#_Toc154043221)

[1. Đào tạo 1 haar cascade face detection 8](#_Toc154043222)

[2. Đọc ảnh đâu vào chuyển qua không gian YCrBr phát hiện tone da. 10](#_Toc154043223)

[3. Thực hiện bitwise\_and 10](#_Toc154043224)

[**III.** **Thực thi** 11](#_Toc154043225)

1. **Cơ sở lý thuyết**
2. **Phát hiện vùng da trên không gian màu Ycbcr**

Phát hiện vùng da trên không gian màu YcbCr là một kỹ thuật xử lý ảnh được sử dụng để xác định các vùng trong một hình ảnh có khả năng chứa da người. Kỹ thuật này dựa trên những đặc tính khác biệt của da người so với các vật thể khác trong môi trường.

Không gian màu Ycbcr là một không gian màu phổ biến trong xử lý ảnh, được sử dụng trong nhiều ứng dụng như truyền hình, video …. Trong không gian màu này, các giá trị RGB của mỗi pixel được chuyển đổi thành các giá trị Y, Cb, Cr như sau:

Y = 0.299R + 0.587G + 0.114B

Cb = -0.1687R – 0.3313G + 0.5000B

Cr = 0.5000R – 0.4187G – 0.0813B

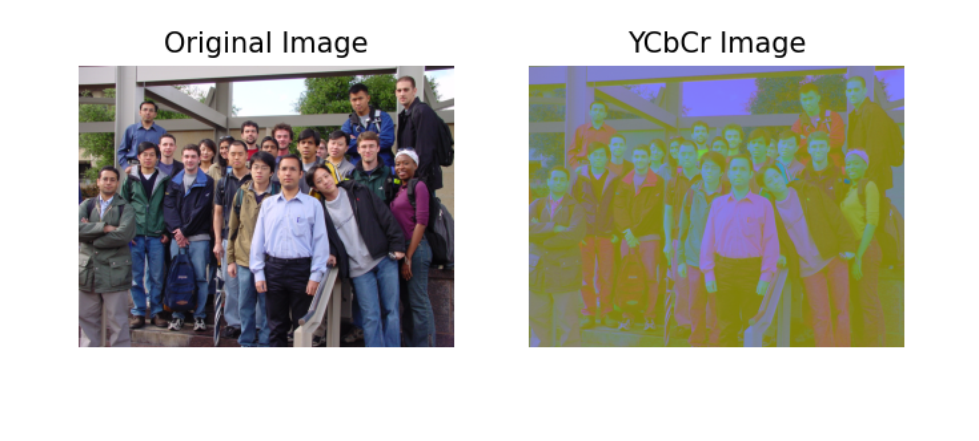
A screen shot of a diagram

Description automatically generated

Các đặc tính khác biệt của da người trong không gian màu YCbCr bao gồm :

* Giá trị Y cao: Da người có giá trị Y cao hơn đáng kể so với các vật thể khác trong môi trường
* Giá trị Cb thấp: Da người có giá trị Cb thấp hơn đáng kể so với các vật thể có màu xanh lá cây hoặc xanh dương.
* Giá trị Cr thấp: Da người có giá trị Cr thấp hơn đáng kể so với các vật thể có màu đỏ hoặc tím.

Qua việc kiểm chọn kĩ lưỡng các ngưỡng màu da người từ không gian YCbCr ta xác định được lower bound (giới hạn dưới) là [0, 133, 77] và upper bound (giới hạn trên) là [255, 173, 127] của màu da trong không gian màu YCbCr.



1. **Haar cascade**

Haar cascade là một thuật toán phát hiện đối tượng được giới thiệu bởi Paul Viola và Michael Jones vào năm 2001. Nó hoạt động bằng cách sử dụng các đặc trưng Harr để xác định các đối tượng trong hình ảnh.

Các đặc trưng Haar là các tính toán toán học đơn giản được thực hiện trên các vùng nhỏ của hình ảnh. Chúng có thể được sử dụng để phát hiện các đặc điểm đặc trưng của khuôn mặt, chẳng hạn như mắt, mũi, miệng, vùng sáng và tối …



Haar Cascade hoạt động để xử lý hình ảnh và phát hiện khuôn mặt bằng cách sử dụng các đặc điểm hình chữ nhật như hình trên.

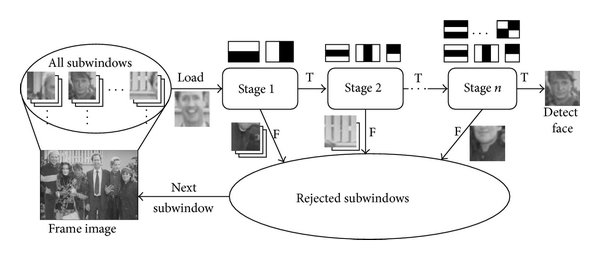
* Sử dụng hướng làm cửa sổ trượt (sliding window): Dùng một window có kích cỡ nhất định và di chuyển dọc teo ảnh. Với mỗi vị trí, classifler sẽ kiểm tra xem vị trí đó của ảnh có giống với đặc trưng của khuôn mặt hay không.
* Phát hiện đặc trưng: Dùng những đặc trưng cơ bản như cạnh, đường thẳng để nhận dạng đối tượng. Nó tìm kiếm các mẫu trong ảnh, ví dụ như các vùng tương phản sáng và tối.
* Quét ở nhiều tỷ lệ (Multiple scales scanning): Nó quét ảnh ở nhiều tỷ lệ khác nhau để phát hiện các đối tượng có kích thước khác nhau. Điều này giúp phát hiện các khuôn mặt có thể xuất hiện lớn hơn hoặc nhỏ hơn do khoảng cách khác nhau với máy ảnh.

.

Trong hình minh họa bên dưới, đặc điểm hình chữ nhật đầu tiên đang tính toán cho sự khác biệt về cường độ giữa vùng mắt và vùng má trên khuôn mặt, Và đặc điểm hình chữ nhật thứ hai là đo sự chênh lệch về cường độ giữa hai vùng mắt và sống mũi. Bộ lọc Haar chỉ có thể nhìn cụ thể vào một vùng trong cửa số để nhận diện.



Các bước nhận diện khuôn mặt của Haar Cascade



* **Bước 1:** Hình ảnh (đã được gửi đến bộ phân loại) được **chia thành các phần nhỏ**(hoặc các cửa sổ con như trong hình minh họa).
* **Bước 2:** Chúng ta đặt N không có bộ dò theo cách xếp tầng trong đó mỗi bộ phát hiện sự kết hợp của các loại **đặc trưng khác nhau** từ các hình ảnh (ví dụ: đường thẳng, cạnh, hình tròn, hình vuông) được truyền qua. Giả sử khi việc trích xuất đối tượng địa lý được thực hiện, mỗi phần phụ được gán một **giá trị tin cậy**.
* **Bước 3:** Hình ảnh (hoặc hình ảnh phụ) có độ tin cậy cao nhất được phát hiện dưới dạng khuôn mặt và được gửi đến **bộ tích lũy** trong khi phần còn lại bị từ chối. Do đó, Cascade tìm nạp khung hình / hình ảnh tiếp theo nếu còn lại và bắt đầu lại quá trình.

1. **Ý tưởng bài toán**

## **1. Đào tạo 1 haar cascade face detection** Dựa trên một tập dataset bao gồm:

* Positive Images: chứa những hình ảnh chứa khuôn mặt. Các hình ảnh này nên có thể được cắt chỉ chứa khu vực khuôn mặt và nên có kích thước và hướng khác nhau.
* Negative Images: chứa những hình ảnh không chứa khuôn mặt. Những hình ảnh này là những hình ảnh ngẫu nhiên từ nhiều nguồn và không chứa bất kỳ khuôn mặt nào.

Lưu ý: Số lượng Negative Images nhiều hơn đáng kể so với Positive Images

* Annotation files: Đối với mỗi positive image tương ứng một tệp chú thích (annotation file) để chỉ định vị trí của khuôn mặt trong hình ảnh. Điều này có thể được thực hiện bằng cách tạo một tệp văn bản cho mỗi hình ảnh tích cực và chỉ định tọa độ của khu vực khuôn mặt (ví dụ: tọa độ bounding box).

Ta có thể tiến hành đào tạo, mục đích tạo ra file model.xml chứa thông tin cần thiết để áp dụng bộ phân loại cascade đã được huấn luyện cho các hình ảnh mới và phát hiện khuôn mặt dựa trên các mẫu và đặc trưng đã học.

* Các tham số của cascade: thông tin về các tham số đã được sử dụng trong quá trình huấn luyện, chẳng hạn như số giai đoạn, kích thước tối thiểu và tối đa của đối tượng, và hệ số tỉ lệ.
* Các đặc trưng dạng Haar: Mô hình đã được huấn luyện dựa trên các đặc trưng dạng Haar, đó là các bộ lọc hình chữ nhật được sử dụng để phát hiện mẫu trong hình ảnh. File .xml chứa thông tin về các đặc trưng dạng Haar cụ thể đã được chọn và huấn luyện để phát hiện khuôn mặt.
* Các giai đoạn phân loại: Bộ phân loại cascade đã được huấn luyện bao gồm nhiều giai đoạn, mỗi giai đoạn bao gồm một tập hợp các bộ phân loại yếu. File .xml chứa thông tin về số lượng giai đoạn và các bộ phân loại yếu được sử dụng trong mỗi giai đoạn.
* Ngưỡng và trọng số: chứa các giá trị ngưỡng và trọng số cho mỗi bộ phân loại yếu. Các giá trị này được sử dụng để xác định xem một vùng cụ thể trong hình ảnh có chứa khuôn mặt hay không.
* Vị trí và kích thước của đặc trưng: lưu trữ các vị trí và kích thước của các đặc trưng dạng Haar đã được chọn, được sử dụng để tính toán các giá trị đặc trưng ở các tỉ lệ và vị trí khác nhau trong hình ảnh.

Code huấn luyện


Ta có thể sử dụng file “face\_classifier\_DA.xml” vừa huấn luyện để thực hiện tác vụ phát hiện khuôn mặt dự trên dữ liệu mới.

## **2. Đọc ảnh đầu vào chuyển qua không gian YCrBr phát hiện tone da.**

Từ khoảng lower bound và upper bound, ta xác định một mask skin\_mask và một ảnh skin, cụ thể như sau:

* những pixel nằm trong khoảng được quy định bằng lower bound và upper bound trong không gian Ycbcr được đặt là màu trắng (255)
* những pixel không nằm trong khoảng trên được đặt là đen (0).
* Ta được output là một mask với những pixel có thể là da, có màu trắng và những pixel không phải da người màu đen.

A group of people posing for a photo

Description automatically generated

## **3. Thực hiện bitwise\_and**

Từ mask trên, có thể thực hiện phép bitwise\_and giữa ảnh gốc đầu vào và skin\_mask này để giữ lại toàn bộ pixel có màu trắng, loại được đa số những vùng không phải da.

* Ta có ảnh skin sau:

A group of people posing for a photo

Description automatically generated

Sau đó chuyển sang dạng xám (grayscale). Do thuật toán phát hiện khuôn mặt yêu cầu và cũng vì việc chuyển sang dạng xám giảm số lượng tính toán và giảm thời gian xử lý đáng kể.

Chúng ta sau đó sử dụng file "face\_classifier\_DA.xml" đã huấn luyện và bắt đầu thực hiện tác vụ phát hiện khuôn mặt trong ảnh xám. Lưu ý khi sử dụng thư viện OpenCV có thể chỉ rõ các tham số trong quá trình thực hiện phát hiện khuôn mặt nhằm đạt kết quả tốt hơn, ví dụ như:

* scaleFactor: tham số này bù cho việc gương mặt xa camera sẽ nhỏ lại, tuỳ việc chọn giá trị để thực hiện phát hiện trên nhiều kích cỡ ảnh, mặt khác nhau từ đầu vào.
* minSize / maxSize: tham số này định nghĩa kích cỡ tối thiểu hoặc tối đa mà một gương mặt có thể nhận. Ví dụ minSize = (30, 30) nghĩa là bất kì vật thể nào được phát hiện mà nhỏ hơn 30x30 đều bị loại.

1. **Thực thi**

Code phát hiện khuôn mặt

A computer screen shot of a computer code

Description automatically generated

A white background with black text

Description automatically generated

Giải thích:

- Khởi tạo Phân loại Cascade khuôn mặt

Xử dụng file face\_classifier\_DA.xml là kết quả của việc huấn luyện model.

- Đọc và xử lý ảnh



* Đọc ảnh từ đường dẫn tới ảnh
* Chuyển đổi ảnh sang không gian màu Ycbcr

- Tạo ra mark từ khoảng mặt định sẵn, tạo ảnh da mặt

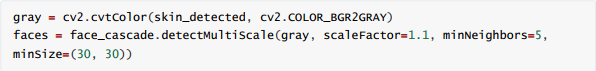
A screenshot of a computer code

Description automatically generated

* lower\_bound và upper\_bound là giới hạn dưới và giới hạn trên cho màu da trong không gian màu YCbCr.
* cv2.inRange() tạo ra một mặt nạ (skin\_mask) xác định các vùng màu da trong ảnh YCbCr. Các pixel nằm trong phạm vi được xác định bởi lower\_bound và upper\_bound được đặt  
  thành màu trắng (255), các pixel nằm ngoài phạm vi này được đặt thành màu đen (0).
* cv2.bitwise\_and() lấy ảnh gốc (image) và chỉ giữ lại những pixel mà mặt nạ là màu trắng, hiệu quả loại bỏ các vùng không phải là màu da.

🡪 ta có ảnh chỉ còn giữ lại những gì quy định là màu da (thuận tiện cho việc phát hiện khuôn mặt)

- Phát hiện khuôn mặt



* Chuyển đổi ảnh đã phát hiện da sang ảnh xám (gray) như yêu cầu của thuật toán phát hiện  
  khuôn mặt, vì nó đơn giản hóa tính toán và giảm thời gian xử lý.
* face\_cascade.detectMultiScale(): Phát hiện khuôn mặt trong ảnh xám bằng cách sử dụng bộ  
  phân loại Haar cascade.
  + scaleFactor: Tham số này bù đắp cho việc một khuôn mặt xa hơn từ máy ảnh sẽ xuất hiện nhỏ hơn. Nó chỉ định mức độ giảm kích thước ảnh tại mỗi tỷ lệ ảnh. Ví dụ, giá trị 1.1 có nghĩa là giảm kích thước đi 10% ở mỗi bước. Giá trị nhỏ hơn có thể làm tăng thời gian phát hiện nhưng có thể tăng độ chính xác.
  + minNeighbors: Đây là tham số chỉ định số lượng hàng xóm (hình chữ nhật ứng cử viên) mà mỗi BB ứng cử viên nên có để giữ lại nó. Giá trị cao sẽ dẫn đến ít phát hiện hơn nhưng chất lượng cao hơn. Giá trị thấp có thể dẫn đến nhiều phát hiện hơn nhưng chất lượng thấp hơn.
  + minSize: Tham số này xác định kích thước đối tượng tối thiểu có thể có. Đối tượng nhỏ hơn kích thước này sẽ bị bỏ qua. Nó giúp loại bỏ các đối tượng nhỏ hoặc dương tính sai. Đối với phát hiện khuôn mặt, chỉ định (30, 30) có nghĩa là bất kỳ đối tượng nhỏ hơn cửa sổ 30x30 được phát hiện sẽ bị bỏ qua.
  + maxSize: Tham số này xác định kích thước đối tượng tối đa có thể có. Đối tượng lớn hơn kích thước này sẽ bị bỏ qua. Nó giúp loại bỏ các đối tượng lớn hoặc giảm các phát hiện sai. Đối với phát hiện khuôn mặt, chỉ định (200, 200) có nghĩa là bất kỳ đối tượng lớn hơn cửa sổ 200x200 sẽ bị bỏ qua.

- Vẽ bounding box cho những gương mặt được phát hiện



Lặp qua các khuôn mặt đã được phát hiện và vẽ các hình chữ nhật xung quanh chúng trên ảnh gốc (image).

**Kết quả**

A group of people standing on a railing

Description automatically generated