

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

KHOA CÔNG NGHỆ THÔNG TIN I



BÁO CÁO ĐỀ TÀI
XÂY DỰNG HỆ CSDL PHÂN LOẠI ẢNH MẶT NGƯỜI
THEO GIỚI TÍNH

<i>Môn học</i>	Hệ cơ sở dữ liệu đa phương tiện	
<i>Nhóm lớp:</i>	D19-04	
<i>Nhóm thực hiện:</i>	Nhóm 1	
<i>Sinh viên thực hiện:</i>	Trần Khánh Hưng	B19DCCN331
	Đỗ Nhật Minh	B19DCCN
	Nguyễn Danh Bình	B19DCCN

Hà Nội, năm 2023

Mục lục

I.	Kiến thức chuyên môn.....	3
a.	Ảnh chân dung.....	3
i.	Khái niệm.....	3
ii.	Đặc điểm.....	3
iii.	Điểm khác biệt so với các loại ảnh khác.....	3
b.	Trích rút đặc trưng.....	4
i.	Khái niệm đặc trưng ảnh.....	4
c.	Phân loại hình ảnh.....	4
i.	Phân loại hình ảnh là gì?.....	4
ii.	Các kỹ thuật phân loại ảnh.....	5
II.	Bài toán phân loại giới tính theo ảnh mặt người.....	7
a.	Dữ liệu sử dụng trong hệ thống.....	7
b.	Trích rút đặc trưng.....	8
i.	Đưa ảnh về ảnh xám và chuẩn hóa kích thước, cân bằng mức sáng trong ảnh.....	9
ii.	Đặc trưng HOG.....	9
iii.	Đặc trưng tỉ lệ màu tóc trong ảnh:.....	11
iv.	Lưu trữ vector đặc trưng.....	12
c.	So sánh đặc trưng.....	12
III.	Xây dựng hệ thống.....	14
a.	Tổng quan sơ đồ vận hành hệ thống.....	14
b.	Mã nguồn.....	15
IV.	Kết quả và thực nghiệm.....	15

I. Kiến thức chuyên môn

a. Ảnh chân dung

i. Khái niệm

Ảnh chân dung là một dạng nhiếp ảnh được sử dụng rất nhiều trong cuộc sống mỗi ngày kết hợp với nhiều mục tiêu khác nhau. Đối với ảnh chân dung, trọng tâm thường là khuôn mặt của đối tượng mục tiêu, cũng có khả năng thêm một phần cơ thể, hoặc nền và bối cảnh, tuy nhiên chỉ là phần phụ giúp thể hiện rõ nét đối tượng mục tiêu hơn. Chân dung truyền thống hay cổ điển sẽ gợi nhớ đến hình ảnh mà gương mặt là yếu tố chiếm ưu thế. Ảnh chân dung dùng để lột tả hình ảnh đặc trưng của một cá nhân. Chủ thể thường được bố trí nhìn chính diện vào máy ảnh. Như với tên gọi là “chụp cận cảnh gương mặt”, hai phần ba hoặc tất cả khuôn mặt đều có thể được sử dụng.

ii. Đặc điểm

1. Tập trung vào gương mặt: Ảnh chân dung tập trung vào việc chụp lại gương mặt và nét vẽ của người được chụp. Thông thường, phần lớn ảnh sẽ chỉ hiển thị gương mặt của người đó, đôi khi có thể bao gồm cả phần cổ và vai.
2. Tạo dấu ấn cá nhân: Ảnh chân dung có mục đích tạo dấu ấn cá nhân, giúp người được chụp có thể tự tin và tỏa sáng trên ảnh. Nhiếp ảnh gia thường sẽ sử dụng các kỹ thuật chụp ảnh để tạo ra bức ảnh chân dung đẹp và nổi bật.
3. Thể hiện tính cách: Ảnh chân dung thường có khả năng thể hiện tính cách và cảm xúc của người được chụp. Vì vậy, nhiếp ảnh gia thường cố gắng giao tiếp và kết nối với người được chụp để có thể tạo ra một bức ảnh chân dung đầy cảm xúc.
4. Chất lượng hình ảnh cao: Ảnh chân dung thường được chụp bởi các thiết bị chụp ảnh chuyên nghiệp hoặc kỹ thuật số với độ phân giải cao, giúp tạo ra những bức ảnh chân dung sắc nét và chi tiết.
5. Sử dụng trong nhiều mục đích: Ảnh chân dung được sử dụng trong nhiều mục đích khác nhau, từ giới thiệu cá nhân, đăng tải lên mạng xã hội, trang web, cho đến sử dụng trong các bài báo, tạp chí và truyền thông khác.

iii. Điểm khác biệt so với các loại ảnh khác

Ảnh chân dung khác với các loại ảnh khác bởi vì nó tập trung vào việc chụp lại gương mặt và nét vẽ của một người nào đó. Trong khi các loại ảnh khác

có thể tập trung vào cảnh quan, đồ vật, hoặc các khung cảnh khác, ảnh chân dung thường chỉ tập trung vào con người. Ngoài ra, ảnh chân dung thường có mục đích để giới thiệu, tạo dấu ấn cá nhân hoặc đại diện cho một người nào đó, trong khi các loại ảnh khác có thể có mục đích khác như ghi lại khoảnh khắc, chia sẻ thông tin hay tạo ra một tác phẩm nghệ thuật.

Để chụp một bức ảnh chân dung đẹp, nhiếp ảnh gia thường phải sử dụng kỹ thuật khác với các loại ảnh khác. Điều này có thể bao gồm cách sắp xếp ánh sáng, cách điều chỉnh độ sâu trường ảnh (depth of field), và cách tạo cảm giác tự nhiên và thoải mái cho người được chụp.

b. Trích rút đặc trưng

i. Khái niệm đặc trưng ảnh

Đặc trưng ảnh (image feature) là các thông tin trừu tượng được trích xuất từ một ảnh và có thể sử dụng để mô tả và phân loại ảnh. Các đặc trưng này có thể được sử dụng để thực hiện các tác vụ như nhận dạng đối tượng, phân loại, gom nhóm và phân tích ảnh.

Các đặc trưng ảnh có thể được trích xuất bằng các phương pháp khác nhau như phân tích cạnh (edge analysis), phân tích màu sắc (color analysis), phân tích hình dạng (shape analysis), phân tích texture (texture analysis), phân tích biên độ sóng (wavelet analysis) và nhiều phương pháp khác. Các đặc trưng này có thể được trích xuất từ các mức độ khác nhau của ảnh, từ các điểm ảnh riêng lẻ cho đến các đối tượng lớn hơn.

Một số ví dụ về đặc trưng ảnh bao gồm:

1. Đặc trưng màu sắc
2. Đặc trưng tần số ảnh
3. Đặc trưng cạnh
4. Đặc trưng cường độ sáng

c. Phân loại hình ảnh

i. Phân loại hình ảnh là gì?

Phân loại hình ảnh (Image classification) hay Nhận dạng hình ảnh (Image recognition) là một trong những tác vụ của thị giác máy tính, ở đó thuật toán xem xét và dán nhãn cho hình ảnh từ một tập danh mục được xác định và đào tạo trước.

Ví dụ, với một tập các hình ảnh, mỗi hình ảnh mô tả một con mèo hoặc một con chó, thuật toán sẽ “quan sát” toàn bộ dữ liệu và dựa trên hình dạng, màu

sắc để hình thành giả thuyết liên quan đến nội dung của ảnh. Kết quả thu được là từ tập dữ liệu ban đầu, các hình ảnh chó/mèo đã được phân loại một cách tự động.

Thực tế, thị giác góp phần tạo nên 80-85% nhận thức của con người về thế giới. Hàng ngày, mỗi người phải thực hiện phân loại trên bất kỳ dữ liệu hình ảnh nào mà chúng ta bắt gặp.

Do đó, mô phỏng nhiệm vụ phân loại với sự trợ giúp của mạng nơ-ron là một trong những ứng dụng đầu tiên của thị giác máy tính mà các nhà nghiên cứu nghĩ đến.

ii. Các kỹ thuật phân loại ảnh

Có nhiều thuật toán khác nhau được ứng dụng trong việc phân loại hình ảnh. Các thuật toán này được chia thành hai nhóm chính là Học có giám sát (supervised learning) và Học không giám sát (unsupervised learning).

1. Phân loại có giám sát

Trong học máy có giám sát, thuật toán được huấn luyện trên một tập hình ảnh đã được dán nhãn. Từ dữ liệu mẫu này, thuật toán có thể trích xuất thông tin, phục vụ phân loại ngay cả những hình ảnh chưa từng nhìn thấy trước đó.

Xuyên suốt quá trình đào tạo, đặc điểm của ma trận hình ảnh sẽ được trích xuất dưới dạng dữ liệu quan trọng để đưa vào xử lý. Các đặc điểm này đại diện cho hình ảnh trong

không gian chiều thấp (lower-dimensional feature space) và là cơ sở để thuật toán tiến hành phân loại.

Trong quá trình đánh giá, các đặc điểm của ảnh thử nghiệm được thu thập và tái phân loại với sự hỗ trợ của mạng thần kinh nhân tạo. Hệ thống lúc này đã có thể nhận biết các đặc điểm điển hình của mọi lớp hình ảnh mà nó được đào tạo.

Các phương pháp phân loại phổ biến dựa trên học có giám sát bao gồm:

- Support Vector Machines
- Decision Trees

- K Nearest Neighbors

Các mạng nơ-ron thường được sử dụng để phân loại hình ảnh có giám sát bao gồm AlexNet, ResNet, DenseNet và Inception.

Đối với phân loại có giám sát, việc dán nhãn dữ liệu đóng vai trò quan trọng. Độ chính xác của dữ liệu được dán nhãn quyết định phần lớn hiệu suất của mô hình học máy. Các thuật toán phân loại có giám sát có thể được chia thành hai mục nhỏ hơn dựa trên nhãn dữ liệu.

Phân loại nhãn đơn

Phân loại nhãn đơn (Single-label classification) là tác vụ phổ biến nhất trong phân loại ảnh có giám sát. Theo đó, mỗi hình ảnh được đại diện bởi một nhãn/chú thích (a single label or annotation). Mô hình xuất ra một giá trị hoặc dự đoán duy nhất cho mỗi hình ảnh mà nó xử lý.

Đầu ra từ mô hình là mã hóa One-hot (từng giá trị được biến đổi thành các đặc trưng nhị phân chỉ chứa giá trị 1 hoặc 0). Mã hóa One-hot có độ dài bằng số lớp và giá trị biểu thị xác suất hình ảnh thuộc về lớp này.

Hàm Softmax được sử dụng để đảm bảo các xác suất tổng bằng một và xác suất tối đa được chọn làm đầu ra của mô hình. Mặc dù Softmax không có giá trị về mặt dự đoán, nhưng nó giúp ràng buộc đầu ra giữa 1 và 0, nhờ vậy, có thể đánh giá độ tin cậy của mô hình từ điểm Softmax.

Một số ví dụ về bộ dữ liệu phân loại nhãn đơn bao gồm MNIST, SVHN, ImageNet, v.v.

Phân loại nhãn đơn có thể được xếp vào phân loại đa lớp (Multiclass classification) hoặc phân loại nhị phân (binary classification).

Phân loại đa nhãn

Phân loại đa nhãn là một tác vụ phân loại trong đó mỗi hình ảnh có thể chứa nhiều hơn một nhãn hoặc một số hình ảnh chứa đồng thời tất cả các nhãn.

Phân loại đa nhãn xuất hiện phổ biến trong lĩnh vực xử lý hình ảnh y tế, khi một bệnh nhân có thể được chẩn đoán mắc nhiều bệnh dựa trên dữ liệu chụp X-quang.

2. Phân loại không giám sát

Trong học máy không giám sát, thuật toán chỉ sử dụng dữ liệu thô để đào tạo. Các nhãn phân loại thường không xuất hiện trong kiểu học này và mô hình học bằng cách nhận dạng các mẫu trong tập dữ liệu huấn luyện.

Giống như phân loại có giám sát, các phương pháp dựa trên không giám sát cũng liên quan đến bước trích xuất đặc điểm với các thông tin chi tiết nhất về hình ảnh. Các đặc điểm này sau đó được xử lý bằng các phương pháp phân cụm tham số (Gaussian Mixture Models) và phi tham số (K-means) hoặc các thuật toán học không giám sát khác.

Các thuật toán và kỹ thuật phân loại của thị giác máy tính không chỉ giới hạn trong dữ liệu hình ảnh 2D đơn giản mà còn mở rộng ra dưới dạng Video và ảnh 3D.

II. Bài toán phân loại giới tính theo ảnh mặt người

a. Dữ liệu sử dụng trong hệ thống

Bộ dữ liệu ảnh chân dung được nhóm thu thập từ 120 người, trong đó bao gồm 60 giới tính nam và 60 giới tính nữ. Bộ dữ liệu đã được nhóm thực hiện công việc lọc để đảm bảo các bức ảnh về cùng 1 giới tính nam hoặc nữ đều có cùng bố cục, có cùng màu da, góc chụp và kiểu tóc (ngắn đối với nam, ngang vai hoặc dài qua vai đối với nữ).

Dữ liệu được chia ra thành 2 folder: Training và Test

./data

./Training

./Nam (bao gồm 50 ảnh)

./Nu (bao gồm 50 ảnh)

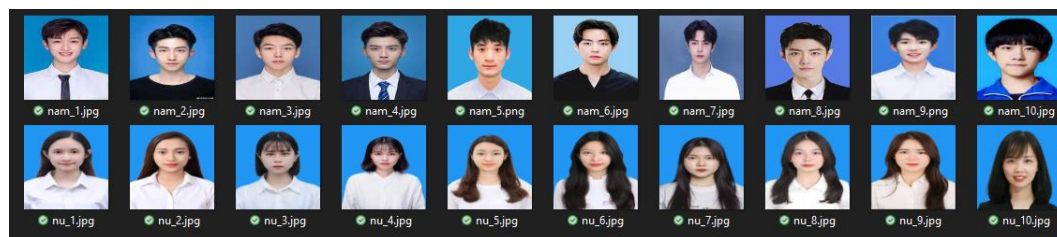
./Test (bao gồm 20 ảnh)



Dữ liệu trong folder “Nam”



Dữ liệu trong folder “Nu”



Dữ liệu trong folder “Test”

b. Trích rút đặc trưng

Từ bộ data đã có, nhóm tiến hành trích rút đặc trưng tiêu biểu của ảnh. Các đặc trưng nhóm chọn là đặc trưng HOG và đặc trưng tỉ lệ màu tóc trong ảnh.

i. Đưa ảnh về ảnh xám và chuẩn hóa kích thước, cân bằng mức sáng trong ảnh.

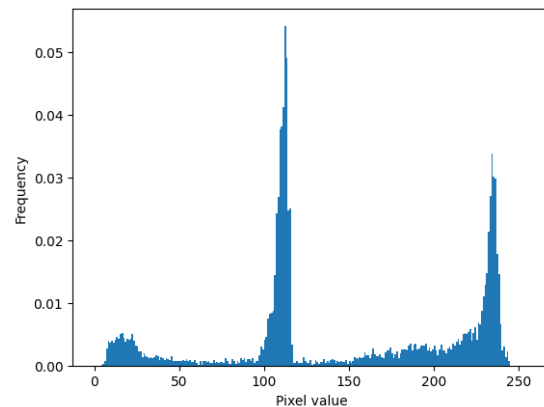
Với đề bài “phân loại giới tính nam nữ”, nhóm quyết định đưa ảnh về dạng ảnh xám để dễ dàng hơn trong quá trình trích rút đặc trưng ảnh mà vẫn đảm bảo không làm ảnh hưởng hay có ảnh hưởng ít đến mục tiêu đề ra của bài toán. Sau đó, ảnh được chuẩn hóa kích thước với chiều cao bằng 128, chiều rộng bằng 96 để đảm bảo số lượng đặc trưng trích xuất là như nhau.



Ảnh ban đầu



Ảnh đa mức xám



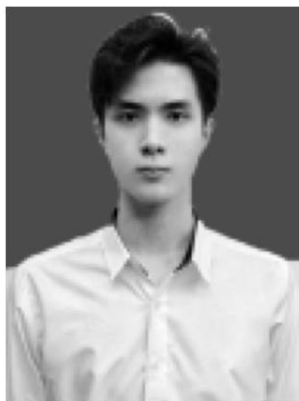
Ảnh và histogram sau khi thu nhỏ

ii. Đặc trưng HOG.

Trích xuất đặc trưng HOG (Histogram of Oriented Gradients) là một kỹ thuật phân tích ảnh để trích xuất các đặc trưng của vùng ảnh cục bộ. Nhận thấy đặc trưng tóc là 1 trong những đặc trưng quan trọng và trích xuất đặc trưng HOG trong trường hợp này có thể đưa ra trích xuất tốt. (Như hình bên dưới)



Ảnh và phân phối HOG tương ứng



Ảnh và phân phối HOG tương ứng

Các bước thực hiện thuật toán HOG:

- 1) Tính toán gradient: HOG tính toán gradient của ảnh bằng cách sử dụng các bộ lọc tích chập để tìm ra độ lớn và hướng của gradient tại mỗi điểm ảnh.
- 2) Phân bố hướng gradient vào các bin: Sau khi tính toán gradient, HOG sẽ phân bố các hướng gradient vào các bin để tạo thành một histogram. Số lượng bin và khoảng cách giữa các bin được xác định trước.
- 3) Chuẩn hóa: HOG thực hiện các bước chuẩn hóa để đảm bảo rằng các vector đặc trưng có cùng tổng bằng một. Các bước chuẩn hóa này bao gồm chuẩn hóa L1 và chuẩn hóa L2.
- 4) Kết hợp các cell để tạo thành các block: HOG chia ảnh thành các cell và kết hợp các cell để tạo thành các block. Các block được trượt qua ảnh và mỗi block tạo ra một vector đặc trưng.
- 5) Thực hiện phân loại: Cuối cùng, vector đặc trưng được sử dụng để thực hiện phân loại ảnh, ví dụ như phân loại khuôn mặt.

Ưu điểm:

- 1) Khả năng trích rút đặc trưng cục bộ: HOG có khả năng trích rút đặc trưng của các phần của ảnh, chứ không chỉ toàn bộ ảnh. Điều này giúp HOG phân biệt các đối tượng chính xác hơn.
- 2) Khả năng xử lý các biến đổi hình học: HOG có khả năng xử lý các biến đổi hình học như xoay và co giãn của đối tượng.
- 3) Hiệu quả tính toán: HOG là một kỹ thuật trích rút đặc trưng đơn giản và hiệu quả về mặt tính toán.

Nhược điểm:

- 1) Nhạy cảm với ánh sáng và độ sáng của ảnh: Khi ánh sáng và độ sáng của ảnh thay đổi, đặc trưng HOG có thể bị ảnh hưởng và không hoạt động hiệu quả.
- 2) Nhược điểm khi phát hiện đối tượng có kích thước khác nhau: Nếu đối tượng cần được phát hiện có kích thước khác nhau, HOG có thể không phát hiện được các đối tượng nhỏ hoặc lớn hơn kích thước được huấn luyện.
- 3) Khả năng xử lý nhiễu hạn chế: HOG có thể không hoạt động hiệu quả khi phải xử lý ảnh nhiễu hoặc ảnh chất lượng kém.

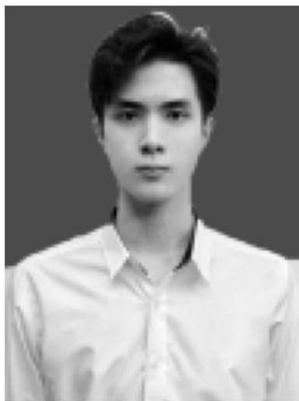
iii. Đặc trưng tỉ lệ màu tóc trong ảnh:

Để tăng khả năng phân biệt tóc theo giới tính nam nữ như trên, nhóm thực hiện trích rút đặc trưng dựa trên tỉ lệ màu tóc trên bức ảnh. Các pixel trong ảnh được so sánh với ngưỡng (threshold = 51) để từ đó tính ra được tỉ lệ màu tóc trong ảnh theo công thức:

$$ratio_h = \frac{sum(pixels \leq 51)}{sum(total_pixels)}$$



Ảnh và Ảnh nhị phân thể hiện tỉ lệ tóc với ngưỡng $K = 51$



Ảnh và Ảnh nhị phân thể hiện tỉ lệ tóc với ngưỡng $K = 51$

iv. Lưu trữ vector đặc trưng.

Mỗi vector đặc trưng của mỗi ảnh được lưu trong 1 file riêng với tên tương ứng với tên của ảnh.

Bảng dưới đây mô tả file lưu trữ đặc trưng (ảnh dạng bảng mang tính chất minh họa)

Đặc trưng	Đặc trưng cường độ theo hướng	Đặc trưng cường độ màu
Các giá trị	$X_1, X_2, \dots, X_{k-1}, X_k$	$X_{k+1}, X_{k+2}, \dots, X_n$

c. So sánh đặc trưng

KNN (K-Nearest Neighbors) là một trong những thuật toán học có giám sát đơn giản nhất được sử dụng nhiều trong khai phá dữ liệu và học máy. Ý tưởng của thuật toán này là nó không học một điều gì từ tập dữ liệu học (nên

KNN được xếp vào loại lazy learning), mọi tính toán được thực hiện khi nó cần dự đoán nhãn mới của dữ liệu mới.

Lớp (nhãn) của một đối tượng dữ liệu mới có thể dự đoán các lớp (nhãn) của k hàng xóm gần nó nhất.

Các bước trong KNN:

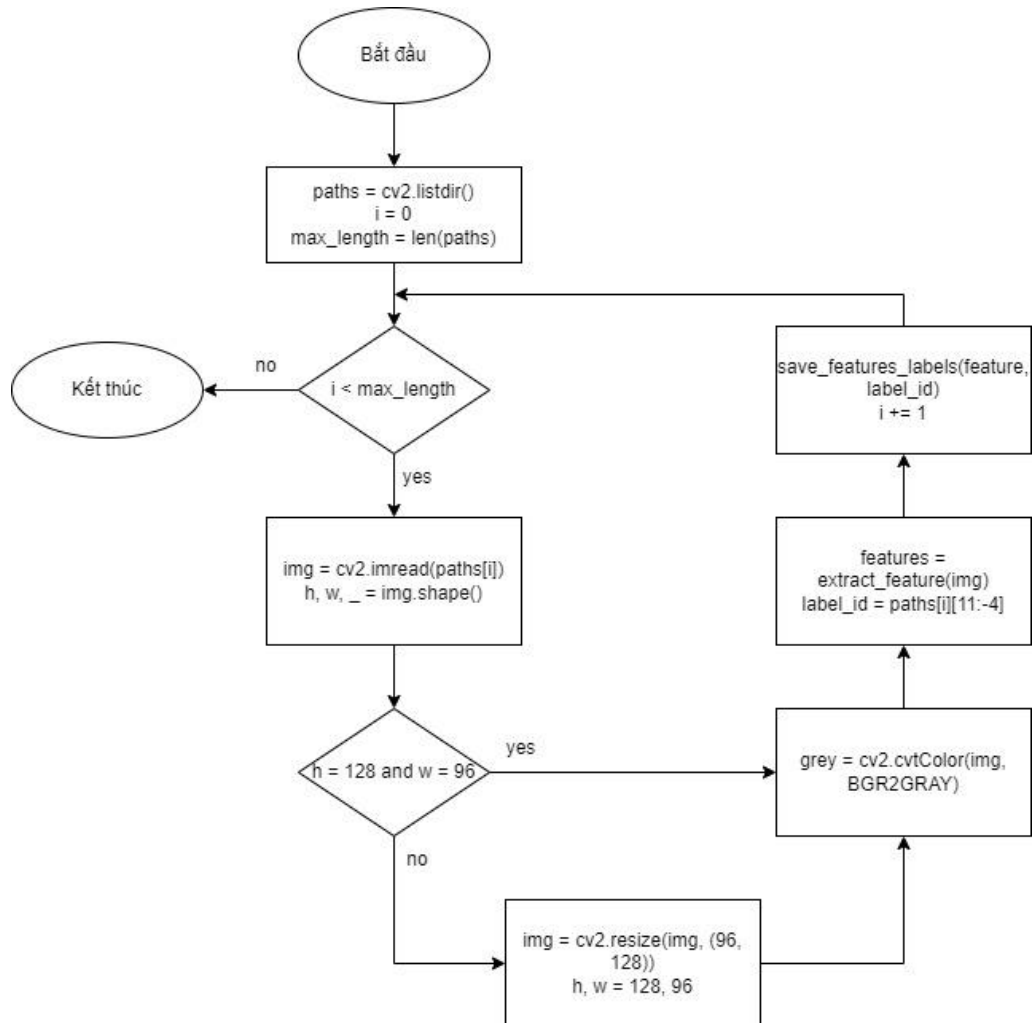
- 1) Chuẩn bị dữ liệu: Bước này đã được xử lý thông qua trích rút đặc trưng, với mỗi ảnh có 1 vector thuộc tính (đặc trưng trích xuất được) riêng và nhãn của nó.
- 2) Chọn giá trị của K: Đây là bước quan trọng trong thuật toán KNN. K là số lượng điểm láng giềng gần nhất mà chúng ta muốn tìm kiếm. Đối với bài toán đề ra, nhóm chọn $K = 3$.
- 3) Tính toán khoảng cách: Thuật toán KNN sử dụng khoảng cách Euclidean để tính toán khoảng cách giữa các điểm dữ liệu. Công thức tính khoảng cách Euclidian:

$$d(p, q) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2}$$

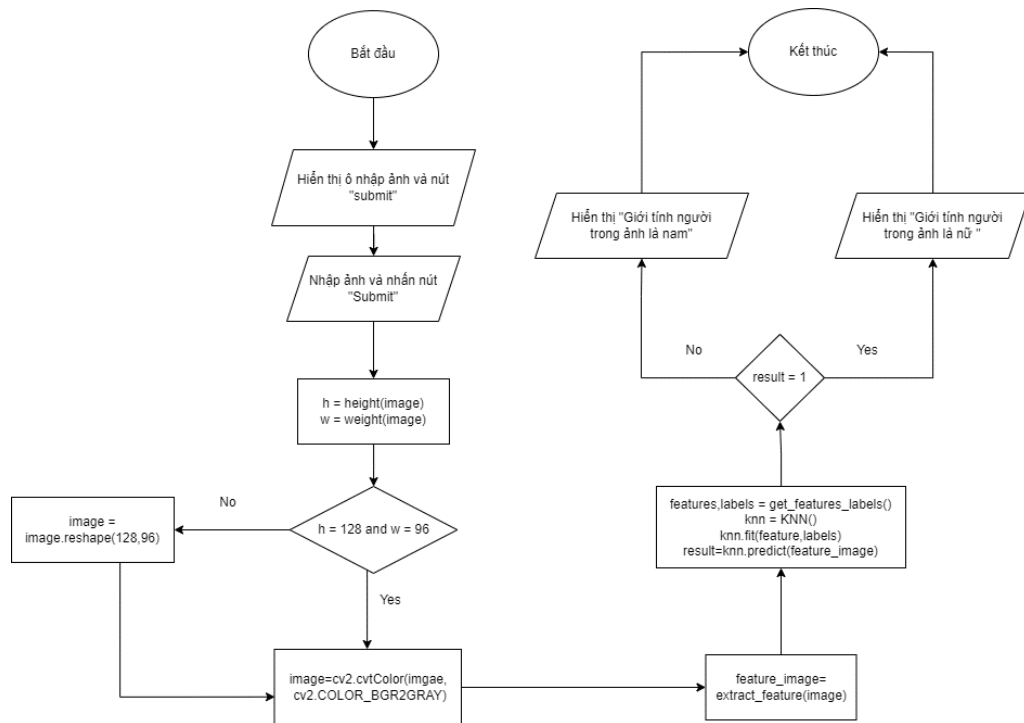
- 4) Tìm K điểm láng giềng gần nhất: Sau khi tính khoảng cách, chúng ta sẽ tìm ra K điểm láng giềng gần nhất với điểm dữ liệu mới. Điều này được thực hiện bằng cách sắp xếp tất cả các điểm dữ liệu trong tập huấn luyện theo khoảng cách Euclidean tăng dần và chọn ra K điểm dữ liệu gần nhất.
- 5) Lựa chọn lớp: Sau khi tìm được K điểm láng giềng gần nhất, chúng ta cần xác định lớp của điểm dữ liệu mới bằng cách lựa chọn lớp xuất hiện nhiều nhất trong K điểm láng giềng đó.
- 6) Đưa ra dự đoán: Cuối cùng, sau khi xác định được lớp của điểm dữ liệu mới, chúng ta sẽ sử dụng lớp đó để đưa ra dự đoán cho điểm dữ liệu mới.

III. Xây dựng hệ thống

a. Tổng quan sơ đồ vận hành hệ thống



Sơ đồ khởi tạo các đặc trưng ảnh



Sơ đồ dự đoán giới tính của hệ thống

b. Mã nguồn

[DPT/Code at main · Hungtran-pro/DPT \(github.com\)](https://github.com/Hungtran-pro/DPT)

IV. Kết quả và thực nghiệm

Mô hình	Precision	Recall	F1_Score
Đặc trưng HOG	0.333	0.2	0.25
Đặc trưng tỉ lệ màu tóc	0.454	0.5	0.47
Sử dụng cả 2 đặc trưng	0.333	0.2	0.25