(12) BẢN MÔ TẢ SÁNG CHẾ THUỘC BẰNG ĐỘC QUYỀN SÁNG CHẾ

(19) **Cộng hòa xã hội chủ nghĩa Việt Nam (VN)**
CỤC SỞ HỮU TRÍ TUỆ

(11) **1-0053629**

(51)[2020.01] **G06N 3/02**; G06K 9/00

(13) **B**

(21) 1-2021-07623

(22) 26/11/2021

(45) 25/11/2025 452

(43) 25/02/2022 407A

(73) TẬP ĐOÀN CÔNG NGHIỆP - VIỄN THÔNG QUÂN ĐỘI (VN)
Lô D26 Khu đô thị mới Cầu Giấy, phường Yên Hoà, quận Cầu Giấy, thành phố Hà Nội

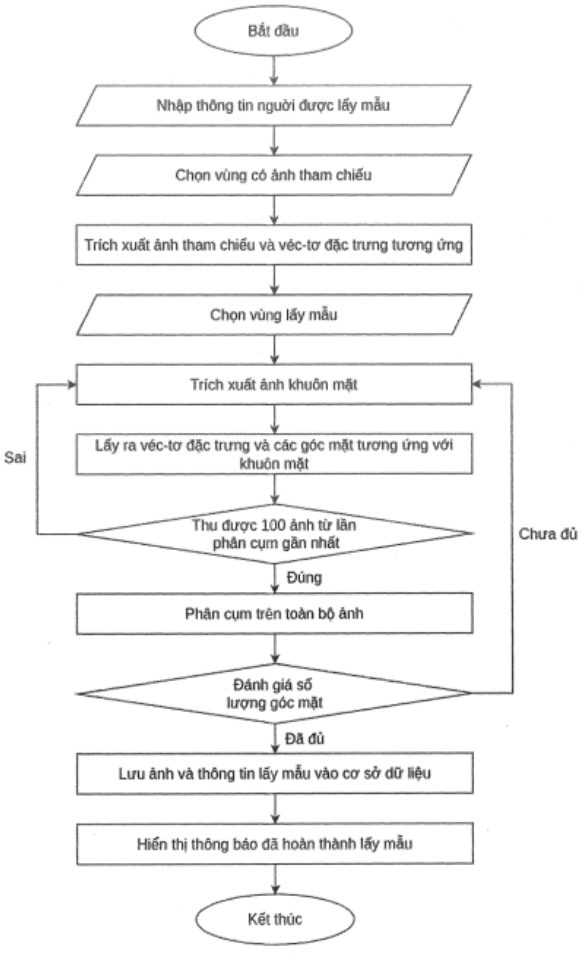
(72) NINH THỊ HƯƠNG (VN); THÁI VĂN CHIẾN (VN); TRẦN TIẾN HẢI (VN).

(74) Công ty TNHH NACILAW (NACILAW)

(54) QUY TRÌNH THU THẬP MẪU DỮ LIỆU BÁN TỰ ĐỘNG CHO HỆ THỐNG NHẬN DIỆN KHUÔN MẶT

(21) 1-2021-07623

(57) Quy trình thu thập mẫu dữ liệu bán tự động cho hệ thống nhận diện khuôn mặt mô tả các công đoạn thu thập dữ liệu khuôn mặt, được thực hiện đơn giản, dễ dàng sử dụng trong thực tế. Quy trình gồm các bước chính: chọn ảnh tham chiếu - ảnh chính diện khuôn mặt của người được lấy mẫu, thay đổi hướng nhìn theo các góc để tăng độ đa dạng cho dữ liệu, lưu trữ tự động dữ liệu ảnh và các thông tin liên quan trong quá trình lấy mẫu vào cơ sở dữ liệu. Nhờ việc tự động phân cụm, đánh giá và lưu trữ, việc thu thập dữ liệu giảm thiểu được thời gian cũng như công sức, trong khi vẫn đảm bảo được độ chính xác cao. Việc sử dụng lưu trữ dữ liệu tập trung trong giúp cho người dùng có được sự tiện lợi trong việc sử dụng trong tương lai. Nhờ sự nhanh chóng và tiện lợi, quy trình có thể được áp dụng trong việc thu thập dữ liệu cho các hệ thống thực tế như giám sát, điểm danh bằng khuôn mặt với số lượng người rất lớn.



Hình 1

Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập đến quy trình thu thập mẫu dữ liệu bán tự động cho hệ thống nhận diện khuôn mặt. Cụ thể, quy trình mô tả cách thu thập dữ liệu một cách bán tự động để huấn luyện cho các thuật toán nhận diện khuôn mặt. Để thực hiện được việc này, quy trình sử dụng các mô hình học sâu để phát hiện, trích xuất đặt trưng từ khuôn mặt và dự đoán hướng mặt. Ngoài ra, quy trình còn sử dụng thuật toán tìm kiếm theo chiều sâu và kĩ thuật lưu trữ dữ liệu tập trung vào các cơ sở dữ liệu có cấu trúc.

Tình trạng kỹ thuật của sáng chế

Trong các bài toán về học máy (machine learning) nói chung và nhận diện khuôn mặt nói riêng, dữ liệu đóng một vai trò cực kì quan trọng. Dữ liệu cho bài toán này đòi hỏi sự đa dạng về mặt phân phối, để các mô hình học sâu (deep learning) có thể học được các đặc tính ẩn của dữ liệu và đưa ra dự đoán chính xác hơn khi sử dụng trong thực tế. Tuy nhiên, việc thu thập dữ liệu cho bài toán nhận diện khuôn mặt đòi hỏi rất nhiều công sức, chủ yếu đến từ việc gán nhãn cho dữ liệu khi số lượng người rất lớn. Ngoài ra, việc đánh giá chất lượng mẫu thu được cũng là một vấn đề quan trọng.

Trong số các tài liệu sáng chế đã được công bố, có một số công trình có nội dung liên quan đến việc thu thập dữ liệu khuôn mặt. Tuy nhiên, những sáng chế liên quan còn một số tồn tại hoặc hạn chế, cụ thể như:

Sáng chế US 8031914 B2 ngày 04/10/2011 của Hoa Kỳ mô tả về phương pháp giúp giảm thiểu thời gian gán nhãn khi dữ liệu lớn. Nội dung sáng chế trình bày về thuật toán phân cụm cho dữ liệu hình ảnh khuôn mặt. Việc gán nhãn sẽ được thực hiện trên cụm gồm các khuôn mặt có độ tương đồng cao thay vì gán nhãn cho từng hình ảnh, Mặc dù giảm thiểu được thời gian, tuy nhiên hiệu quả gán nhãn phụ thuộc vào kết quả của việc phân cụm trong dữ liệu thực tế. Với mỗi cụm, các trường hợp nhiễu (các khuôn mặt trong cùng một cụm không cùng thuộc một người) vẫn có thể xảy ra. Hơn nữa, việc thu thập mẫu thụ động khiến dữ liệu thu được có chất lượng không tốt (hình ảnh bị mờ, độ đa dạng thấp, mất cân bằng dữ liệu giữa mỗi người), dẫn đến việc dự đoán không chính xác của các mô hình học sâu.

Sáng chế CN 106204779 A ngày 31/8/2018 của Trung Quốc mô tả cách thu thập dữ liệu bằng video. Mỗi người được thu dữ liệu video trong khoảng 30s với các hành động khác nhau, dữ liệu hình ảnh khuôn mặt sẽ được trích xuất ra từ video này. Tuy nhiên, giải pháp được đề xuất chưa đề cập đến vấn đề đánh giá sự đa dạng và chất lượng dữ liệu thu được. Hơn nữa, ngoài thông tin khuôn mặt, video có thể chứa thêm nhiều

thông tin không cần thiết khác, dẫn đến thách thức trong việc lưu trữ trong hệ thống nhận diện với số người lớn.

Để khắc phục những hạn chế trên, nhóm tác giả đã đề xuất quy trình thu thập mẫu dữ liệu bán tự động cho hệ thống nhận diện khuôn mặt, không giống với bất kì sáng chế nào đã từng được công bố.

Bản chất kỹ thuật của sáng chế

Mục đích của sáng chế là xây dựng quy trình thu thập mẫu dữ liệu bán tự động cho hệ thống nhận diện khuôn mặt, giúp khắc phục những nhược điểm của các sáng chế trước đó. Quy trình được đề xuất giúp giảm thiểu thời gian và công sức trong việc thu thập dữ liệu trong khi vẫn đảm bảo được chất lượng của dữ liệu, giúp các mô hình học sâu có thể đưa ra dự đoán chính xác hơn trong các ứng dụng thực tế. Ngoài ra, dữ liệu được lưu trữ một cách có hệ thống, tiện lợi cho việc sử dụng trong tương lai. Quy trình được xây dựng trên phần mềm máy tính, dễ dàng trong quá trình cài đặt và sử dụng.

Để đạt được mục đích trên, quy trình được đề xuất trong sáng chế được thực hiện qua các bước sau:

Bước 1: chọn ảnh tham chiếu - ảnh chính diện khuôn mặt của người được lấy mẫu.

Bước 2: thay đổi hướng nhìn theo các góc để tăng độ đa dạng cho dữ liệu. Đánh giá chất lượng dữ liệu khuôn mặt thu được tự động bằng cách đưa qua bộ lọc về hướng mặt và so sánh độ tương đồng với ảnh tham chiếu, quá trình phân cụm dữ liệu thu được theo từng giai đoạn tới khi số lượng ảnh thu được đáp ứng yêu cầu của hệ thống.

Bước 3: lưu trữ tự động dữ liệu ảnh và các thông tin liên quan trong quá trình lấy mẫu vào cơ sở dữ liệu.

Trong đó, quy trình lấy mẫu dữ liệu khuôn mặt bán tự động có các đặc trưng sau:

- Ngoại trừ việc thu thập chỉ một ảnh tham chiếu, tất cả các bước còn lại đều được thực hiện một cách tự động. Điều này giúp giảm thiểu công sức gán nhãn cho những người phát triển hệ thống.
- Việc đánh giá dữ liệu được thực hiện tự động thông qua các mô hình học sâu, vì thế dữ liệu được đánh giá dưới của góc nhìn của máy tính thay vì góc nhìn của con người. Do máy tính "nhìn" hình ảnh dưới dạng ma trận ba chiều, xử lí chúng một cách chi tiết đến từng điểm ảnh, từng góc cạnh, khác với cách nhìn tổng quan của mắt người, cho nên việc đánh giá tự động này vừa giúp giảm sức người, vừa giúp máy tính tự đưa ra những dữ liệu giúp quá trình huấn luyện học máy hiệu quả hơn.
- Các phương pháp phát hiện khuôn mặt trong ảnh, trích xuất đặc trưng và dự đoán hướng mặt của khuôn mặt đều được sử dụng các mô hình học sâu có độ chính xác cũng như tốc độ xử lí cao. Thông số góc từ mô hình dự đoán hướng

mặt được sử dụng cho việc loại bỏ các khuôn mặt có góc nghiêng quá lớn, mất mát nhiều thông tin cho việc nhận dạng. Dữ liệu được phân cụm dựa trên thuật toán tìm kiếm theo chiều sâu với đỉnh gốc là ảnh tham chiếu, giúp đảm bảo được việc máy tính tự xác định được những khuôn mặt trong cùng một cụm là của một người.

- Sử dụng các kỹ thuật tối ưu tốc độ xử lý của các mô hình học sâu như phát hiện khuôn mặt, trích xuất đặc trưng, tính toán góc mặt và thuật toán phân cụm giúp quá trình lấy mẫu tự động đáp ứng thời gian thực.
- Dữ liệu được lưu trữ bao gồm hình ảnh và các thông tin về quá trình lấy mẫu (thông tin người được lấy mẫu, thời gian, địa điểm) trên cơ sở dữ liệu được đặt tại máy chủ, phục vụ cho việc thuận tiện sử dụng, truy vấn dữ liệu trong tương lai.

Với các đặc trưng nêu trên, quy trình này có thể khắc phục được những điểm hạn chế của các phương pháp lấy mẫu trước đó, vừa giảm thiểu sức người trong quá trình chuẩn bị dữ liệu, vừa thu được dữ liệu có độ đa dạng cao dưới góc nhìn của máy tính. Trong triển khai thực tế, việc sử dụng quy trình giúp thời gian lấy mẫu trung bình đạt 0,5phút/người với dữ liệu hình ảnh từ camera 15FPS.

Mô tả vắn tắt hình vẽ

Hình 1 là hình vẽ minh họa quy trình thu thập dữ liệu khuôn mặt.

Mô tả chi tiết sáng chế

Quy trình thu thập dữ liệu khuôn mặt bán tự động với khả năng đọc ảnh từ camera và hiển thị lên màn hình, đồng thời có thể triển khai các mô hình học sâu cho quá trình thu mẫu. Các mô hình học sâu được thiết kế dựa trên mạng nơ-ron tích chập chi tiết như sau:

- Mô hình phát hiện khuôn mặt trong ảnh: có tên là RetinaFace, nhận đầu vào là ảnh và đầu ra là tọa độ điểm trên trái và dưới phải của các khuôn mặt được phát hiện trong ảnh, ngoài ra còn có thông tin về tọa độ của hai mắt, mũi và hai khóe miệng của khuôn mặt này.
- Mô hình trích xuất đặc trưng khuôn mặt: nhận đầu vào là ảnh khuôn mặt với kích thước 112x112, và đầu ra là véc-tơ đặc trưng tương ứng với số chiều là 512. Mô hình sử dụng hàm mất mát ArcFace, với mục đích ánh xạ các điểm dữ liệu lên không gian cầu, và các điểm dữ liệu cùng một lớp sẽ gần nhau và cách xa các điểm thuộc lớp khác theo không gian góc.
- Mô hình dự đoán hướng mặt: nhận đầu vào là ảnh khuôn mặt và đầu ra là giá trị của ba góc biểu thị cho hướng ngang, dọc và nghiêng của khuôn mặt.

Các mô hình được huấn luyện trên các bộ dữ liệu lớn, đạt độ chính xác cao và có khả năng tổng quát hóa khi áp dụng vào các ứng dụng thực tế.

Chi tiết các bước của sáng chế được mô tả như sau:

Bước 1: chọn ảnh tham chiếu - ảnh chính diện khuôn mặt của người được lấy mẫu.

Sau khi nhập thông tin định danh cho đối tượng đang được thu mẫu khuôn mặt, người thực hiện quy trình sẽ thao tác lựa chọn một vùng ảnh xung quanh vùng hiển thị khuôn mặt để xử lý nhằm xác định một ảnh chính diện làm tham chiếu. Mô hình phát hiện khuôn mặt sẽ đưa ra tọa độ hình chữ nhật bao quanh vùng ảnh khuôn mặt. Việc lựa chọn vùng xử lý nhỏ giúp tăng tốc độ xử lý và tránh được các khuôn mặt khác gây nhiễu dữ liệu. Khi hoàn thành lựa chọn ảnh tham chiếu, ảnh sẽ được đưa qua mô hình trích xuất đặc trưng để tạo ra véc-tơ đặc trưng làm dữ liệu tham chiếu.

Bước 2: thu thập dữ liệu tự động.

Sau khi có dữ liệu tham chiếu, đối tượng được lấy mẫu khuôn mặt sẽ được yêu cầu thực hiện thao tác nhìn theo các góc từ trái qua phải. Người thực hiện quy trình sẽ thực hiện thao tác lựa chọn vùng ảnh xung quanh khuôn mặt để xử lý. Dữ liệu khuôn mặt thu thập được đánh giá tự động theo cơ sở lý thuyết sau:

Ảnh khuôn mặt phát hiện được sẽ được điều chỉnh kích thước về 112x112 và đưa qua mô hình trích xuất đặc trưng và mô hình dự đoán hướng mặt, thu được thông tin về véc-tơ đặc trưng và góc xoay ngang tương ứng của mặt.

Đồ thị vô hướng G = (V, E) thể hiện về việc liên kết giữa các điểm dữ liệu tương ứng là các ảnh khuôn mặt, với V là tập ảnh, E là tập các cạnh. Xét cặp đỉnh u và v thuộc tập

V, tương ứng với hai ảnh trong tập dữ liệu khuôn mặt thu được. Cặp đỉnh u và v được xem là hai ảnh khuôn mặt thuộc cùng một người nếu như có độ tương đồng cao, có giá trị lớn hơn ngưỡng *threshold* . Độ tương đồng của hai ảnh được tính dựa trên khoảng cách góc giữa hai véc-tơ đặc trưng tương ứng, với công thức như sau:

$$cosine_similarity(u, v) = \frac{\langle feat(u), feat(v) \rangle}{\|feat(u)\| * \|feat(v)\|}$$

Với *feat(u), feat(v)* lần lượt là véc-tơ đặc trưng khuôn mặt tương ứng với đầu vào là ảnh u, v. Trong phương án thực hiện sáng chế, các véc-tơ đặc trưng được chuẩn hóa, ví dụ *feat(u)* được chuẩn hóa thành $f(u) = feat(u)/\|u\|$.Vì thế độ tương đồng giữa hai ảnh u và v lúc này được tính bằng tích vô hướng giữa hai véc-tơ đặc trưng đã được chuẩn hóa:

$$cosine_similarity(u, v) = \langle f(u), f(v) \rangle$$

Với tất các cặp đỉnh (u, v), nếu *cosine_similarity(u, v) >= threshold*, ta xây dựng cạnh giữa hai đỉnh này. Đồ thị được xây dựng có tập đỉnh là tập dữ liệu ảnh thu thập được, và cạnh giữa hai đỉnh thể hiện rằng hai ảnh khuôn mặt tương ứng với hai đỉnh đó thuộc cùng một người. Sau khi xây dựng được đồ thị, từ đỉnh gốc là ảnh tham chiếu, tiến hành tìm kiếm theo chiều sâu để tìm ra được đồ thị con liên thông gồm các ảnh được máy tính xem là cùng một người. Mô tả chi tiết về việc tìm kiếm này như sau:

- Xây dựng mảng num_neighbors với num_neighbors[u] là số lượng đỉnh kề với u.
- Duyệt đồ thị từ đỉnh gốc (đỉnh ứng với ảnh tham chiếu).
- Với đỉnh u đang được duyệt. Sau đó, xét tất cả các đỉnh v kề với u, nếu num_neighbors[v] < MIN_SAMPLE, loại bỏ đỉnh v. Nếu không, duyệt tiếp đỉnh v. Quá trình kết thúc khi tất cả các đỉnh có thể đi đến được từ đỉnh gốc đã được duyệt. Những đỉnh được đánh dấu đã thăm tương ứng với những hình ảnh được giữ lại.

Mục đích của quá trình này là thu thập tự động những hình ảnh chất lượng tốt, loại bỏ đi những ảnh kém chất lượng dưới góc nhìn của máy tính. Quá trình này còn loại bỏ được những ảnh nhiễu, ví dụ như có người khác vô tình xuất hiện trong vùng phát hiện trong lúc thu thập. Những ảnh với số lượng ảnh có độ tương đồng cao với nó nhỏ hơn ngưỡng (MIN_SAMPLE) sẽ được loại bỏ để tránh các trường hợp nhiễu. Trong đề xuất sáng chế, nhóm tác giả đặt ngưỡng tương đồng giữa hai ảnh *threshold* là 0,65 và ngưỡng số lượng hàng xóm của một đỉnh MIN_SAMPLE là N/100 với N là tổng số lượng ảnh đang được duyệt. Giá trị 0,65 của *threshold* được nhóm tác giả lựa chọn trong quá trình thực nghiệm khi đánh giá về độ tương đồng giữa hai ảnh thuộc cùng một người và giữa hai ảnh thuộc hai người khác nhau. Giá trị này là giá trị tối ưu nhất trên một tập dữ liệu nhỏ, nếu thấp hơn 0,65 sẽ khiến máy tính nhận nhầm hai ảnh thuộc hai người khác nhau là cùng một người, và lớn hơn 0,65 sẽ tăng tỉ lệ nhận nhầm hai ảnh thuộc cùng một người là hai người khác nhau.

Để tự động đánh giá và đảm bảo độ đa dạng của mẫu dữ liệu đã được phân cụm, phương pháp được sử dụng là tính số lượng các khoảng góc mặt. Góc ngang với giá trị trong đoạn [-50, 50] được chia làm năm khoảng:

- Góc ngang trái: các giá trị trong nửa khoảng [-50, -40);
- Góc ¾ trái: các giá trị trong nửa khoảng [-40, -20);
- Góc chính diện: các giá trị trong đoạn [-20, 20];
- Góc ¾ phải: các giá trị trong nửa khoảng (20, 40];
- Góc ngang phải: các giá trị trong nửa khoảng (40, -50].

Theo một phương án thực hiện sáng chế, tập dữ liệu được gọi là đủ đa dạng nếu số lượng ảnh thu được tại góc chính diện lớn hơn hoặc bằng 30, góc ¾ trái và ¾ phải có số lượng ảnh thu được lớn hơn hoặc bằng 25, và các khoảng góc ngang trái và ngang phải có số lượng ảnh thu được lớn hơn hoặc bằng 5. Các ảnh có góc ngang ngoài khoảng này được loại bỏ. Các số lượng trên được sử dụng để nhằm đảm bảo chất lượng dữ liệu, và tối thiểu hóa thời gian lấy mẫu cũng như giảm thiểu không gian lưu trữ và thời gian khi xử lí dữ liệu trong tương lai (huấn luyện cho các mô hình trí tuệ nhân tạo, tìm kiếm, truy vấn).

Quá trình thu thập, phân cụm và đánh giá kết thúc khi đạt đủ số lượng yêu cầu về góc mặt. Do quá trình thực hiện phân cụm tốn khá nhiều thời gian, cho nên để đảm bảo khả năng xử lí thời gian thực, theo một phương án thực hiện sáng chế, quá trình này chỉ thực hiện sau khi nhận 100 ảnh so với lần phân cụm trước.

Bước 3: lưu trữ dữ liệu hình ảnh và thông tin lấy mẫu.

Sau khi quá trình thu thập dữ liệu tự động kết thúc, dữ liệu hình ảnh và thông tin lấy mẫu được lưu trữ vào hệ thống máy chủ nhằm mục đích thuận tiện cho việc sử dụng trong tương lai. Dữ liệu ảnh được lưu vào cơ sở dữ liệu *MinIO*. Các thông tin về người được lấy mẫu (họ tên, mã định danh, địa chỉ email, số điện thoại, ngày sinh, giới tính, các ghi chú khác…) và thông tin ảnh khuôn thu thập được (thời gian lấy mẫu, địa điểm, liên kết của ảnh tại *MinIO*, tọa độ của mặt trong ảnh gốc, kích thước ảnh, tọa độ các điểm mắt, mũi, miệng, véc-tơ đặc trưng, các góc mặt) được lưu trữ tại *PosgreSQL*. Thông tin về người được lấy mẫu và ảnh tương ứng với người đó được liên kết với nhau để dễ dàng truy vấn. Sau khi lưu trữ thành công, màn hình sẽ hiển thị thông báo đã hoàn thành lấy mẫu.

Dữ liệu được lưu trữ tập trung trên hệ thống máy chủ, giúp cho dữ liệu có tính thống nhất, khả năng quản lý cao và có thể chia sẻ dữ liệu dễ dàng. Người dùng được cấp quyền truy cập vào máy chủ có khả năng truy vấn, tải dữ liệu từ xa thông qua đường truyền mạng.

Trong quy trình này, người lấy mẫu chỉ cần thực hiện chọn ảnh tham chiếu và chọn vùng để phát hiện người. Việc thu thập và đánh giá cũng như lưu trữ lượng dữ liệu lớn được thực hiện tự động với tốc độ xử lí cao. Điều này vừa đảm bảo được độ đa dạng cho dữ liệu, thời gian thực hiện lấy mẫu cũng như giảm thiểu tối đa công sức gán nhãn.

Ví dụ thực hiện sáng chế

Phần sau đây đề cập đến ví dụ về thực hiện quy trình lấy mẫu và đánh giá trên hệ thống nhận diện khuôn mặt, có mục đích làm rõ mà không áp đặt bất kì giới hạn nào lên sáng chế đề xuất.

Quy trình thu thập dữ liệu được sử dụng lên camera chất lượng 4K với tốc độ năm khung hình trên giây đặt trong tòa nhà. Các mô hình học sâu được thực hiện trên máy

tính cấu hình cao với card đồ họa Quadro P4000. Số lượng người được lấy mẫu là gần 500 người.

Thời gian lấy mẫu trung bình cho một người là khoảng một phút, và công đoạn chọn ảnh tham chiếu và vùng của người lấy mẫu chỉ mất 10 giây. Trung bình thu được khoảng 100 ảnh đối với mỗi người với các góc mặt khác nhau. Mô hình nhận diện khuôn mặt sử dụng dữ liệu thu được để huấn luyện và đạt độ chính xác 99,99% trên bộ dữ liệu bao gồm hơn 250000 ảnh thường ngày của 455 người (để đảm bảo đánh giá khách quan, bộ dữ liệu này được đánh nhãn cho từng ảnh thu được thực tế bằng camera của từng người trong vòng năm ngày, vì thế bộ dữ liệu có thể không đủ người và ảnh không được đa dạng). Khi so sánh về thời gian thực hiện, việc tự gán nhãn mất năm ngày cho việc thu thập dữ liệu thường ngày của nhân viên trong tòa nhà từ 84 camera, và năm người mất hai tuần cho việc gán nhãn 250000 ảnh khuôn mặt trích xuất được. Tuy nhiên dữ liệu thu được không đủ người do camera được đặt không đủ để bao phủ tất cả các góc, dữ liệu thu được cũng không đủ độ đa dạng.

Hiệu quả kỹ thuật đạt được

Quy trình lấy mẫu khuôn mặt bán tự động được đề xuất trong sáng chế đã giải quyết được hai vấn đề cần thiết trong bài toán nhận diện khuôn mặt sử dụng các mô hình học sâu, đó là xây dựng bộ dữ liệu đa dạng, tổng quát và giảm thời gian thu thập và gán nhãn dữ liệu. Quy trình được thiết kế đơn giản và đóng gói thành phần mềm để dễ dàng sử dụng. Vì thế, quy trình có thể được áp dụng rộng rãi trong thực tế, khi số lượng người lên đến hàng trăm, hàng ngàn người. Ngoài ra, quy trình sử dụng các thuật toán học sâu hiện đại nhất với độ chính xác cao và thời gian xử lí thấp trong các bài toán phát hiện khuôn mặt, trích xuất đặc trưng cho mặt và dự đoán hướng mặt. Nhờ tốc độ xử lí các thuật toán cao và quá trình thu thập và đánh giá tự động, việc lấy mẫu được diễn ra nhanh chóng và không cần sự can thiệp của con người. Dữ liệu thu được tuy ít nhưng vẫn đảm bảo được sự đa dạng và tổng quát được các trường hợp về góc mặt. Nhờ đó, độ chính xác trong quá trình nhận diện khuôn mặt được tăng lên so với các cách thu thập không có đánh giá về chất lượng dữ liệu trước đó, trong khi thời gian và công sức thực hiện thu mẫu giảm đáng kể.

Việc lưu trữ dữ liệu trong quy trình được đề xuất giúp làm tiện lợi cho quá trình truy vấn trong tương lai. Nhờ việc lưu trữ tập trung trên một hệ thống máy chủ, dữ liệu được lưu trữ một cách thống nhất, dễ dàng quản lí và nhiều người dùng có thể truy cập để sử dụng. Hơn nữa, nhờ việc lưu trữ này, nhiều người có thể được lấy mẫu tại cùng một thời điểm tại các vị trí camera khác nhau mà không xảy ra xung đột.

Yêu cầu bảo hộ

1. Quy trình lấy mẫu bán tự động cho hệ thống nhận diện khuôn mặt bao gồm ba bước:

bước 1: chọn ảnh tham chiếu - ảnh chính diện khuôn mặt của người được lấy mẫu; sau khi nhập thông tin định danh cho đối tượng đang được thu mẫu khuôn mặt, người thực hiện quy trình sẽ thao tác lựa chọn một vùng ảnh xung quanh vùng hiển thị khuôn mặt để xử lý nhằm xác định một ảnh chính diện làm tham chiếu; mô hình phát hiện khuôn mặt sau đó sẽ đưa ra tọa độ hình chữ nhật bao quanh vùng ảnh khuôn mặt;

bước 2: thay đổi hướng nhìn theo các góc để tăng độ đa dạng cho dữ liệu; đánh giá chất lượng dữ liệu khuôn mặt thu được tự động bằng cách đưa qua bộ lọc về hướng mặt và so sánh độ tương đồng với ảnh tham chiếu, quá trình phân cụm dữ liệu thu được theo từng giai đoạn tới khi số lượng ảnh thu được đáp ứng yêu cầu của hệ thống;

trong đó:

sau khi có dữ liệu tham chiếu, đối tượng được lấy mẫu khuôn mặt sẽ được yêu cầu thực hiện thao tác nhìn theo các góc từ trái qua phải; người thực hiện quy trình sẽ thực hiện thao tác lựa chọn vùng ảnh xung quanh khuôn mặt để xử lý; dữ liệu khuôn mặt thu thập được đánh giá tự động theo cơ sở sau:

ảnh khuôn mặt phát hiện được sẽ được điều chỉnh kích thước về 112x112 và đưa qua mô hình trích xuất đặc trưng và mô hình dự đoán hướng mặt, thu được thông tin về véc-tơ đặc trưng và góc xoay ngang tương ứng của mặt;

đồ thị vô hướng G = (V, E) thể hiện về việc liên kết giữa các điểm dữ liệu tương ứng là các ảnh khuôn mặt, với V là tập ảnh, E là tập các cạnh; xét cặp đỉnh u và v thuộc tập V, tương ứng với hai ảnh trong tập dữ liệu khuôn mặt thu được; cặp đỉnh u và v được xem là hai ảnh khuôn mặt thuộc cùng một người nếu như có độ tương đồng cao, có giá trị lớn hơn ngưỡng *threshold*; độ tương đồng của hai ảnh được tính dựa trên khoảng cách góc giữa hai véc-tơ đặc trưng tương ứng, với công thức như sau:

$$cosine_similarity(u,v) = \frac{\langle feat(u), feat(v)\rangle}{\|feat(u)\| * \|feat(v)\|}$$

với *feat(u), feat(v)* lần lượt là véc-tơ đặc trưng khuôn mặt tương ứng với đầu vào là ảnh u, v; trong phương án thực hiện sáng chế, các véc-tơ đặc trưng được chuẩn hóa, ví dụ *feat(u)* được chuẩn hóa thành $f(u) = feat(u)/\|u\|$; vì thế độ tương đồng giữa hai ảnh u và v lúc này được tính bằng tích vô hướng giữa hai véc-tơ đặc trưng đã được chuẩn hóa:

$$cosine_similarity(u,v) = \langle f(u), f(v)\rangle$$

với tất các cặp đỉnh (u, v), nếu *cosine_similarity(u, v) >= threshold*, ta xây dựng cạnh giữa hai đỉnh này; đồ thị được xây dựng có tập đỉnh là tập dữ liệu ảnh thu thập được, và cạnh giữa hai đỉnh thể hiện rằng hai ảnh khuôn mặt tương ứng với hai đỉnh đó thuộc cùng một người; sau khi xây dựng được đồ thị, từ đỉnh gốc là ảnh tham chiếu, tiến hành tìm kiếm theo chiều sâu để tìm ra được đồ thị con liên thông gồm các ảnh được máy tính xem là cùng một người; mô tả chi tiết về việc tìm kiếm này như sau:

xây dựng mảng num_neighbors với num_neighbors[u] là số lượng đỉnh kề với u;

duyệt đồ thị từ đỉnh gốc (đỉnh ứng với ảnh tham chiếu);

với đỉnh u đang được duyệt. Sau đó, xét tất cả các đỉnh v kề với u, nếu num_neighbors[v] < MIN_SAMPLE, loại bỏ đỉnh v; nếu không, duyệt tiếp đỉnh v; quá trình kết thúc khi tất cả các đỉnh có thể đi đến được từ đỉnh gốc đã được duyệt; những đỉnh được đánh dấu đã thăm tương ứng với những hình ảnh được giữ lại;

những ảnh với số lượng ảnh có độ tương đồng cao với nó nhỏ hơn ngưỡng (MIN_SAMPLE) sẽ được loại bỏ để tránh các trường hợp nhiễu.

để tự động đánh giá và đảm bảo độ đa dạng của mẫu dữ liệu đã được phân cụm, phương pháp được sử dụng là tính số lượng các khoảng góc mặt; góc ngang với giá trị trong đoạn [-50, 50] được chia làm năm khoảng:

góc ngang trái: các giá trị trong nửa khoảng [-50, -40);

góc ¾ trái: các giá trị trong nửa khoảng [-40, -20);

góc chính diện: các giá trị trong đoạn [-20, 20];

góc ¾ phải: các giá trị trong nửa khoảng (20, 40];

góc ngang phải: các giá trị trong nửa khoảng (40, -50];

tập dữ liệu được gọi là đủ đa dạng nếu số lượng ảnh thu được tại góc chính diện lớn hơn hoặc bằng 30, góc ¾ trái và ¾ phải có số lượng ảnh thu được lớn hơn hoặc bằng 25, và các khoảng góc ngang trái và ngang phải có số lượng ảnh thu được lớn hơn hoặc bằng 5; các ảnh có góc ngang ngoài khoảng này được loại bỏ; các tham số về số lượng ảnh nêu trên được sử dụng để nhằm đảm bảo chất lượng dữ liệu, và tối thiểu hóa thời gian lấy mẫu cũng như giảm thiểu không gian lưu trữ và thời gian khi xử lí dữ liệu trong tương lai (huấn luyện cho các mô hình trí tuệ nhân tạo, tìm kiếm, truy vấn);

quá trình thu thập, phân cụm và đánh giá kết thúc khi đạt đủ số lượng yêu cầu về góc mặt; do quá trình thực hiện phân cụm tốn khá nhiều thời gian, cho nên

để đảm bảo khả năng xử lí thời gian thực, theo một phương án thực hiện sáng chế, quá trình này chỉ thực hiện sau khi nhận 100 ảnh so với lần phân cụm trước.

bước 3: lưu trữ tự động dữ liệu ảnh và các thông tin liên quan trong quá trình lấy mẫu vào cơ sở dữ liệu; sau khi quá trình thu thập dữ liệu tự động kết thúc, dữ liệu hình ảnh và thông tin lấy mẫu được lưu trữ vào hệ thống máy chủ nhằm mục đích thuận tiện cho việc sử dụng trong tương lai.

2. Quy trình lấy mẫu bán tự động cho hệ thống nhận diện khuôn mặt theo điểm 1, trong đó tại bước 1:

người thực hiện quy trình thao tác lựa chọn một vùng ảnh xung quanh vùng hiển thị khuôn mặt của người được lấy mẫu, việc lựa chọn vùng xử lý nhỏ giúp tăng tốc độ xử lý và tránh được các khuôn mặt khác gây nhiễu dữ liệu; sau đó mô hình phát hiện khuôn mặt sẽ đưa ra tọa độ hình chữ nhật bao quanh khuôn mặt; khi hoàn thành lựa chọn ảnh tham chiếu, ảnh sẽ được đưa qua mô hình trích xuất đặc trưng để tạo ra véc-tơ đặc trưng làm dữ liệu tham chiếu.

3. Quy trình lấy mẫu bán tự động cho hệ thống nhận diện khuôn mặt theo điểm 1, trong đó tại bước 3:

dữ liệu ảnh được lưu vào cơ sở dữ liệu MinIO; các thông tin về người được lấy mẫu (họ tên, mã định danh, địa chỉ email, số điện thoại, ngày sinh, giới tính, các ghi chú khác…) và thông tin ảnh khuôn thu thập được (thời gian lấy mẫu, địa điểm, liên kết của ảnh tại MinIO, tọa độ của mặt trong ảnh gốc, kích thước ảnh, tọa độ các điểm mắt, mũi, miệng, véc-tơ đặc trưng, các góc mặt) được lưu trữ tại PosgreSQL; thông tin về người được lấy mẫu và ảnh tương ứng với người đó được liên kết với nhau để dễ dàng truy vấn; sau khi lưu trữ thành công, màn hình sẽ hiển thị thông báo đã hoàn thành lấy mẫu;

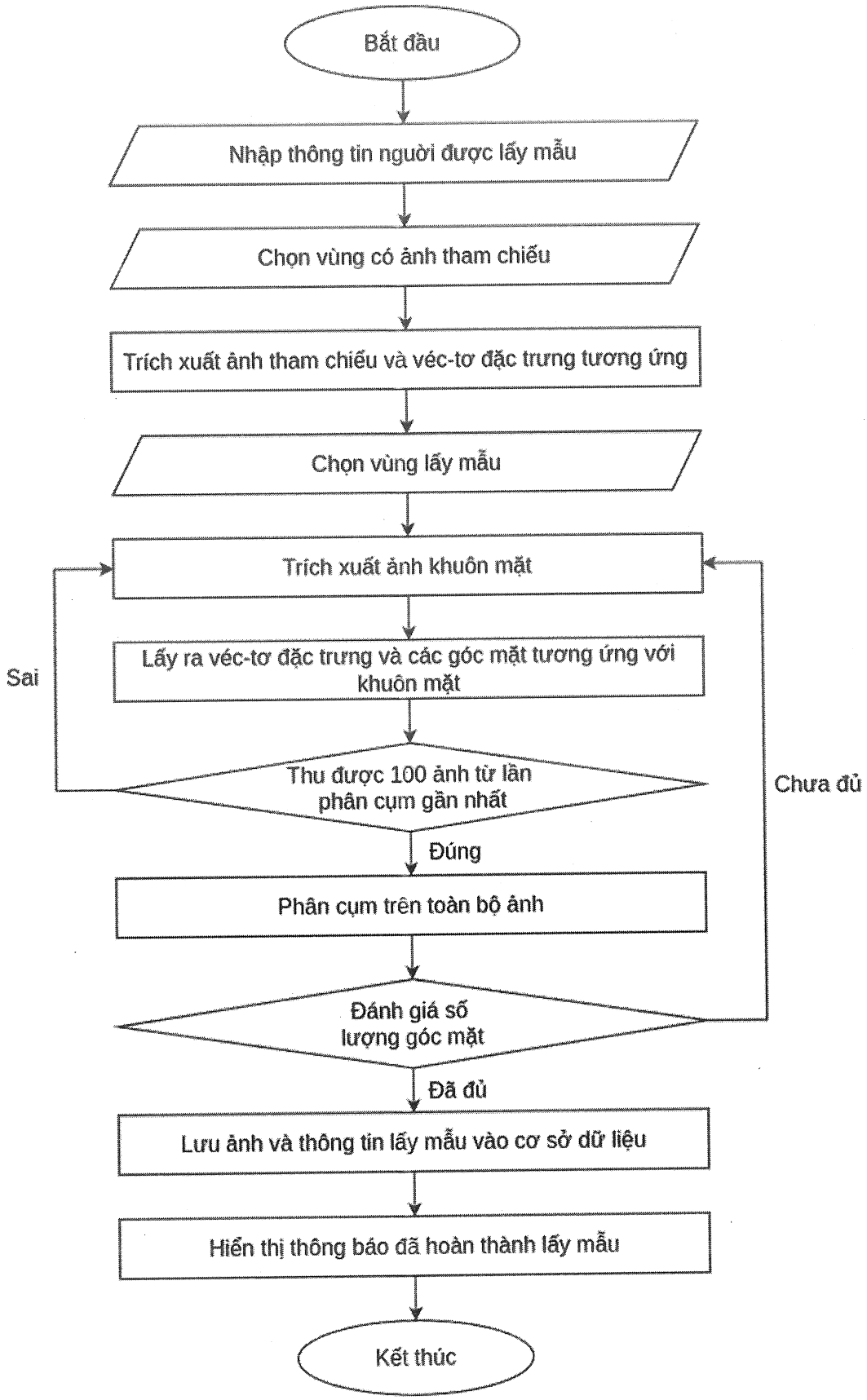
dữ liệu được lưu trữ tập trung trên hệ thống máy chủ, giúp cho dữ liệu có tính thống nhất, khả năng quản lý cao và có thể chia sẻ dữ liệu dễ dàng; người dùng được cấp quyền truy cập vào máy chủ có khả năng truy vấn, tải dữ liệu từ xa thông qua đường truyền mạng;

trong quy trình này, người lấy mẫu chỉ cần thực hiện chọn ảnh tham chiếu và chọn vùng để phát hiện người; việc thu thập và đánh giá cũng như lưu trữ lượng dữ liệu lớn được thực hiện tự động với tốc độ xử lí cao; điều này vừa đảm bảo được độ đa dạng cho dữ liệu, thời gian thực hiện lấy mẫu cũng như giảm thiểu tối đa công sức gán nhãn.

4. Quy trình lấy mẫu bán tự động cho hệ thống nhận diện khuôn mặt theo điểm 1, trong đó quá trình phân cụm và đánh giá độ đa dạng góc mặt được thực hiện sau khi thu thập được 100 ảnh (tính từ lần phân cụm và đánh giá trước đó) để cân bằng giữa thời gian lấy mẫu và khối lượng tính toán của máy tính.

5. Quy trình lấy mẫu bán tự động cho hệ thống nhận diện khuôn mặt theo điểm 1, trong đó ngưỡng để xác định hai ảnh là tương đồng là 0,65, ngưỡng về số lượng ảnh

tương đồng của một đỉnh để quyết định xem ảnh đấy có được chọn hay không là tổng số lượng ảnh / 100.


Bắt đầu
Nhập thông tin người được lấy mẫu
Chọn vùng có ảnh tham chiếu
Trích xuất ảnh tham chiếu và véc-tơ đặc trưng tương ứng
Chọn vùng lấy mẫu
Trích xuất ảnh khuôn mặt
Lấy ra véc-tơ đặc trưng và các góc mặt tương ứng với khuôn mặt
Thu được 100 ảnh từ lần phân cụm gần nhất
Sai
Chưa đủ
Đúng
Phân cụm trên toàn bộ ảnh
Đánh giá số lượng góc mặt
Đã đủ
Lưu ảnh và thông tin lấy mẫu vào cơ sở dữ liệu
Hiển thị thông báo đã hoàn thành lấy mẫu
Kết thúc

Hình 1