



(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)



1-0048689

(51)<sup>2021.01</sup> G06K 9/00; G06N 3/00; G06F 16/58

(13) B

(21) 1-2022-05569

(22) 31/08/2022

(45) 25/07/2025 448

(43) 25/11/2022 416A

(73) TẬP ĐOÀN CÔNG NGHIỆP - VIỆN THÔNG QUÂN ĐỘI (VN)

Lô D26 Khu đô thị mới Cầu Giấy, phường Yên Hoà, quận Cầu Giấy, thành phố Hà Nội

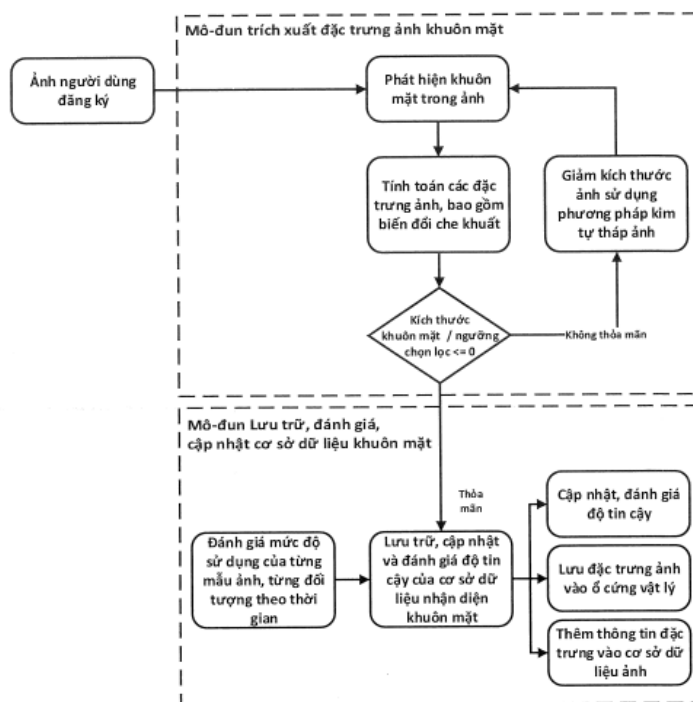
(72) Lê Văn Bằng (VN); Phạm Cao Bằng (VN); Phạm Thị Thu Trang (VN).

(74) Công ty TNHH NACILAW (NACILAW)

(54) PHƯƠNG PHÁP CẤU TRÚC VÀ ĐÁNH GIÁ CƠ SỞ DỮ LIỆU TỰ THÍCH ỨNG  
CHO NHẬN DIỆN KHUÔN MẶT

(21) 1-2022-05569

(57) Sáng chế này đề cập đến phương pháp cấu trúc và đánh giá cơ sở dữ liệu tự thích ứng cho nhận diện khuôn mặt. Phương pháp thực hiện bao gồm phát hiện khuôn mặt trong ảnh đăng ký (để sử dụng hệ thống nhận diện khuôn mặt) của người dùng. Sau đó, tính toán các đặc trưng ảnh, bao gồm biến đổi che khuất. Trong quá trình này, có mười chín trường thông tin được tính toán và phân biệt được lưu trữ vật lý vào đĩa cứng và cơ sở dữ liệu môn-gô. Tiếp theo, đánh giá cơ sở dữ liệu khuôn mặt, xác định ngưỡng tự thích ứng đối với từng mẫu ảnh, cũng như nhằm loại bỏ các mẫu ảnh không tin cậy, đồng thời cập nhật thay thế mẫu ảnh mới phù hợp được lấy từ trong quá trình vận hành hệ thống trong thực tế. Cuối cùng, trong quá trình vận hành thực tế, thống kê tần suất được dùng để phân loại đối với các mẫu ảnh, còn gọi là số lần được kích hoạt của mẫu ảnh, lọc và loại bỏ những mẫu không được hoặc rất ít khi được sử dụng trong quá trình phân loại tại các hệ thống được triển khai thực tế.



Hình 1

### **Lĩnh vực kỹ thuật được đề cập**

Sáng chế đề cập đến phương pháp cấu trúc và đánh giá cơ sở dữ liệu tự thích ứng cho nhận diện khuôn mặt. Cụ thể, phương pháp cấu trúc và đánh giá cơ sở dữ liệu tự thích ứng cho nhận diện khuôn mặt được sử dụng trong các hệ thống trí tuệ nhân tạo và thị giác máy tính sử dụng công nghệ nhận diện khuôn mặt.

### **Tình trạng kỹ thuật của sáng chế**

Trong các hệ thống nhận diện khuôn mặt hiện nay nói riêng và các hệ thống nhận diện đối tượng nói chung, cơ sở dữ liệu thường được cố định từ lúc khởi tạo, quá trình cập nhật thủ công chủ yếu chỉ dựa trên những quan sát bằng mắt thường đối với lịch sử nhận diện tại vị trí triển khai.

Quá trình nhận diện còn kém thông minh, các hệ thống hiện nay chủ yếu tính toán so khớp toàn cục trên toàn bộ cơ sở dữ liệu. Đặc biệt, khi một khuôn mặt bất kỳ bị nhận diện sai thì kết quả đó là duy nhất cho dù thực hiện bao nhiêu lần so khớp. Có nhiều trường hợp, các khuôn mặt có đặc tính khác nhau có khoảng cách giữa các véc-tơ đặc trưng là rất nhỏ, tuy nhiên lại không cùng một nhóm. Ví dụ điển hình như trường hợp nhận diện khuôn mặt đeo khẩu trang, hệ thống tìm kiếm thấy một mẫu tối ưu với khoảng cách nhỏ nhất, nhưng lại không đúng, do các mẫu trong cơ sở dữ liệu chưa được lọc theo đặc tính riêng (ở đây là đeo khẩu trang), dẫn đến tính ổn định hệ thống kém. Khi so khớp toàn cục, số lượng phép tính cũng tăng lên, từ đó làm chậm hiệu quả tìm kiếm, đặc biệt đối với những giải pháp yêu cầu giá rẻ, sử dụng thiết bị có cấu hình thấp.

Ngoài ra, các phương pháp định lượng đánh giá tự động chất lượng cơ sở dữ liệu cũng như từng mẫu ảnh trong đó còn chưa được sử dụng rộng rãi, dẫn đến khó khăn trong quá trình vận hành, những thao tác so sánh và cập nhật thủ công sẽ làm giảm hiệu suất làm việc trong quá trình nghiên cứu, phát triển và triển khai.

### **Bản chất kỹ thuật của sáng chế**

Mục đích của sáng chế là khắc phục các vấn đề kỹ thuật còn tồn tại, như nêu trên, bằng việc đề xuất phương pháp cấu trúc và đánh giá cơ sở dữ liệu tự thích ứng cho nhận diện khuôn mặt. Phương pháp đề cập trong sáng chế có mục đích cung cấp phương pháp cấu trúc cơ sở dữ liệu nhằm tăng cường độ chính xác khi tìm kiếm đối tượng sử dụng đặc trưng khuôn mặt trong ảnh, và cung cấp cách đánh giá, tối ưu và cập nhật cơ sở dữ liệu khuôn mặt nhằm kịp thời cập nhật cũng như tối ưu hiệu quả tìm kiếm.

Trong sáng chế này, phương pháp cấu trúc và đánh giá cơ sở dữ liệu tự thích ứng cho nhận diện khuôn mặt dựa trên một hệ thống bao gồm các bộ phận cấu thành như sau:

Mô-đun trích xuất đặc trưng ảnh khuôn mặt: dùng để trích xuất các đặc trưng về hình học cũng như phân bố bề mặt (dùng để tăng cường hiệu quả tìm kiếm) và đặc trưng khuôn mặt (dùng trong tìm kiếm) trong ảnh.

Mô-đun lưu trữ, đánh giá và cập nhật cơ sở dữ liệu khuôn mặt: dùng để thực hiện quá trình lưu trữ cơ sở dữ liệu bằng, cũng như lưu trữ các mảng ảnh và đặc trưng ảnh phục vụ tối ưu cơ sở dữ liệu và tìm kiếm đối tượng. Đồng thời là công cụ đánh giá các mẫu ảnh trong cơ sở dữ liệu, là tiền đề để xác định các mẫu khả dụng cũng như các mẫu hết hạn sử dụng (do thời gian dài hoặc do sự thay đổi về ngoại hình của đối tượng).

Cụ thể, để đạt được mục đích đã nêu ở trên, phương pháp cấu trúc và đánh giá cơ sở dữ liệu tự thích ứng cho nhận diện khuôn mặt bao gồm các bước thực hiện như sau:

Bước 1: phát hiện khuôn mặt trong ảnh đăng ký (để sử dụng hệ thống nhận diện khuôn mặt) của người dùng. Trong quá trình này, xác định tọa độ góc trái trên cùng và tọa độ góc phải dưới cùng của khuôn mặt (trên hệ tọa độ ảnh) với độ tin cậy cao nhằm loại bỏ những trường hợp phát hiện nhầm khuôn mặt mà có thể dẫn đến sai sót trong quá trình nhận diện của hệ thống.;

Bước 2: xác định các đặc trưng ảnh, bao gồm biến đổi che khuất. Trong quá trình này, có mười chín trường thông tin được tính toán và phân biệt được lưu trữ vật lý vào đĩa cứng và cơ sở dữ liệu môn-gô. Do điều kiện đặc thù, ảnh hưởng từ đại dịch cô-vít, dẫn đến

nhu cầu phân loại ảnh khuôn mặt đeo khẩu trang, trong bước này đồng thời thực hiện biến đổi che khuất bằng cách thêm khẩu trang vào khuôn mặt trong ảnh, và tiến hành trích xuất đặc trưng như đối với ảnh gốc ban đầu. Đồng thời kiểm tra kích thước ảnh khuôn mặt phát hiện được trong ảnh, thực hiện thao tác kim tự tháp ảnh, sau đó tiếp tục trích xuất đặc trưng, như một mẫu mới.

Trong mười chín đặc trưng đó, mười tám đặc trưng được tính trong bước 2, và đặc trưng thống kê tần suất thứ mười chín được khởi tạo khi vận hành thực tế hệ thống nhận diện khuôn mặt.

Bước 3: đánh giá cơ sở dữ liệu khuôn mặt; bước này đánh giá cơ sở dữ liệu khuôn mặt để xác định ngưỡng tự thích ứng đối với từng mẫu ảnh, cũng như nhằm loại bỏ các mẫu ảnh không tin cậy, đồng thời cập nhật thay thế mẫu ảnh mới phù hợp được lấy từ trong quá trình vận hành hệ thống trong thực tế;

Bước 4: thống kê tần suất được dùng để phân loại đối với các mẫu ảnh; bước thống kê tần suất được dùng để phân loại đối với các mẫu ảnh còn gọi là số lần được kích hoạt của mẫu ảnh, lọc và loại bỏ những mẫu không được hoặc rất ít khi được sử dụng trong quá trình phân loại tại các hệ thống được triển khai thực tế.

Bước 1, bước 2 được thực hiện trong mô-đun trích xuất đặc trưng ảnh khuôn mặt; bước 3, bước 4 được thực hiện trong mô-đun lưu trữ, đánh giá và cập nhật cơ sở dữ liệu khuôn mặt.

### **Mô tả vắn tắt các hình vẽ**

Hình 1 là hình vẽ thể hiện các bước thực hiện phương pháp cấu trúc và đánh giá cơ sở dữ liệu tự thích ứng cho nhận diện khuôn mặt;

Hình 2 là hình vẽ thể hiện nội dung các đặc trưng được trích xuất khi xử lý ảnh đăng ký của người dùng;

Hình 3 là hình vẽ mô tả góc khuôn mặt trong ảnh, trong đó bao gồm góc theo chiều thẳng đứng (*yaw*), góc theo chiều song song (*pitch*) và góc theo trục vuông góc với ảnh (*roll*);

Hình 4 là hình vẽ quá trình lưu trữ, đánh giá và cập nhật cơ sở dữ liệu (mô-đun lưu trữ, đánh giá và cập nhật cơ sở dữ liệu khuôn mặt);

Hình 5 là hình vẽ mô tả ma trận tương quan được tính từ đặc trưng nhận diện khuôn mặt dựa trên khoảng cách cô-sin tại hệ thống nhận diện diện triển khai ở VTNET;

Hình 6 là hình vẽ mô tả trực quan các chỉ số đánh giá cơ sở dữ liệu tại hệ thống nhận diện diện triển khai ở VTNE;

Hình 7 là hình vẽ mô tả ngưỡng tự thích ứng và sự phân biệt giữa mẫu ảnh tin cậy-không tin cậy.

### **Mô tả chi tiết sáng chế**

Tham chiếu Hình 1, là hình vẽ thể hiện các bước thực hiện phương pháp cấu trúc và đánh giá cơ sở dữ liệu tự thích ứng cho nhận diện khuôn mặt.

Cụ thể, các bước thực hiện phương pháp như sau:

Bước 1: phát hiện khuôn mặt trong ảnh đăng ký (để sử dụng hệ thống nhận diện khuôn mặt) của người dùng;

Bước này sử dụng mô hình học sâu để phát hiện khuôn mặt người trong ảnh được đăng ký bởi người dùng. Đối với các ảnh đăng ký để sử dụng hệ thống nhận diện khuôn mặt, yêu cầu độ tin cậy (có giá trị từ 0 đến 1) cao, ngưỡng phát hiện khuôn mặt được thiết lập lớn hơn 0,95. Sau khi phát hiện khuôn mặt trong ảnh đăng ký, vị trí khuôn mặt được xác định bởi hai điểm là điểm góc trái trên cùng (top-left) và điểm góc phải dưới cùng (bottom-right), theo tọa độ ảnh.

Bước 2: xác định các đặc trưng ảnh, bao gồm biến đổi che khuất;

Tham chiếu Hình 2, quá trình tính toán đặc trưng ảnh được thực hiện qua các bước nhỏ như sau:

B2.1: xác định các đặc trưng định danh

Đặc trưng thứ nhất: số thứ tự mẫu ảnh, là thứ tự xử lý của mẫu ảnh trong cả quá trình khởi tạo cũng như cập nhật cơ sở dữ liệu. Đặc trưng này dùng để bảo đảm tính duy nhất của từng mẫu ảnh.

Đặc trưng thứ hai: tên đối tượng, là tên định danh của người dùng trên hệ thống, mỗi người dùng sẽ được cung cấp một tên duy nhất. Mỗi một đối tượng có thể có nhiều mẫu ảnh được đăng ký.

Đặc trưng thứ ba: trạng thái sử dụng của mẫu ảnh, bao gồm hai giá trị “đang sử dụng” biểu thị mẫu ảnh này đang được sử dụng để so khớp tìm kiếm khi nhận diện khuôn mặt và “hết hạn” biểu thị mẫu ảnh này vẫn tồn tại trong cơ sở dữ liệu nhưng không còn được dùng để so khớp, tìm kiếm, nhằm xác định các mẫu đã bị loại bỏ bởi cơ sở dữ liệu thông qua phương pháp xóa mềm. Trường thông tin này bảo đảm việc loại bỏ mẫu ảnh mà không ảnh hưởng đến cả cơ sở dữ liệu cũng như quá trình tìm kiếm đối tượng.

Đặc trưng thứ tư: mã phân biệt mẫu, xác định bởi tên người được đăng ký, thứ tự đăng ký và thời gian đăng ký. Mã phân biệt này có cấu trúc gồm thời gian đăng ký (ngày-tháng-năm), thứ tự mẫu ảnh và tên đối tượng, tổ hợp lại thành một chuỗi ký tự và được mã hóa thông qua phương pháp mã hóa base64 kết hợp với phương pháp băm dữ liệu SHA256 để tạo ra một mã đại diện cho mẫu ảnh được đăng ký. Đặc trưng này bảo đảm tính duy nhất của mẫu, bất biến trong quá trình lưu trữ cũng như truyền tải, hoặc đồng bộ các cơ sở dữ liệu, cũng là yếu tố kiểm tra chéo đối với đặc trưng thứ nhất và đặc trưng thứ hai.

B2.2: xác định đặc trưng nhận diện khuôn mặt (đặc trưng thứ năm). Bước này sử dụng mô hình học sâu được huấn luyện trên 50 triệu ảnh khuôn mặt với đầu vào là ảnh khuôn mặt (được phát hiện ở bước 1) đã được hiệu chỉnh về kích thước 112 x 112 (đây là kích thước chuẩn của mô hình, định nghĩa lúc huấn luyện mô hình). Đặc trưng nhận diện khuôn mặt này là một véc-tơ 512 chiều, sử dụng để so khớp, tìm kiếm khuôn mặt cho các hệ thống nhận diện.

Nhằm thuận tiện hơn trong quá trình lưu trữ, đặc trưng này được mã hóa thành dạng chuỗi ký tự, thông qua phương pháp mã hóa base64.

### B2.3: xác định các đặc trưng mô tả bề mặt

Đặc trưng thứ sáu: độ nhiễu của ảnh, được tính bởi phương pháp giả định nhiễu tuân theo phân bố Gaussian. Trong đó chỉ số nhiễu được ước tính dựa trên độ lệch tuyệt đối trung bình của hệ số chi tiết từ biến đổi sóng nhỏ. Giá trị nhiễu càng nhỏ càng tốt.

Đặc trưng thứ bảy: độ sắc nét của ảnh, được tính bởi phương pháp biến đổi Laplacian. Trong đó, giá trị độ sắc nét là tổng số điểm ảnh trên ảnh khuôn mặt sau khi biến đổi Laplacian có giá trị lớn hơn ngưỡng được thiết lập trước. Độ sắc nét của ảnh càng cao càng tốt.

Đặc trưng thứ tám: độ sáng, được tính bởi giá trị trung bình của kênh sáng của ảnh khuôn mặt. Độ sáng có giá trị từ 0 đến 255 (tương ứng với giá trị lớn nhất của một ảnh 8 bit), trong đó giá trị càng lớn càng sáng, theo trực quan.

Đặc trưng thứ chín: độ tương phản, được tính bởi phương pháp trung bình cục bộ của ảnh khuôn mặt, giá trị độ tương phản càng lớn thì tính phân biệt cục bộ của ảnh càng cao.

Các đặc trưng từ thứ sáu đến đặc trưng thứ tám được dùng để làm tiền đề cho quá trình tối ưu cơ sở dữ liệu trong các điều kiện khắc nghiệt (trời tối, nhiễu) nếu cần. Đồng thời cũng là chỉ số thống kê quan trọng để đánh giá hiệu quả cũng như độ chính xác của hệ thống khi vận hành thực tế.

### B2.4: xác định các đặc trưng hình học

Đặc trưng thứ mười, đặc trưng thứ mười một: lần lượt là điểm góc trái trên cùng và điểm góc phải dưới cùng được trả về trong quá trình phát hiện khuôn mặt. Hai tọa độ này dùng để hỗ trợ phát hiện các đối tượng nổi bật trong ảnh đăng ký (thông thường ở chính giữa), và xác định mức độ được kích hoạt trong hình ảnh thực tế nơi vận hành hệ thống nhận diện khuôn mặt.

Đặc trưng thứ mười hai, đặc trưng thứ mười ba, đặc trưng thứ mười bốn: là các góc khuôn mặt so với trục tạo ảnh (trục pháp tuyến của mặt cầu tạo ảnh), được tính bởi mô hình

hồi quy được huấn luyện trên tập dữ liệu 2000 ảnh với phạm vi góc từ -90 độ đến 90 độ. Mô hình này có đầu vào là ảnh khuôn mặt, và đầu ra là ba chỉ số góc khuôn mặt, bao gồm góc theo chiều song song, góc theo chiều thẳng đứng và góc theo chiều vuông góc. Ý nghĩa trực quan của ba góc này được mô tả như trong Hình 3. Các đặc trưng này là một tham chiếu quan trọng trong quá trình nhận diện khuôn mặt của hệ thống.

Đặc trưng thứ mười lăm: số lần kim tự tháp ảnh, có nghĩa là số lần giảm kích thước của ảnh khuôn mặt (ngang và cao mỗi chiều giảm một nửa giá trị cho mỗi lần kim tự tháp ảnh). Chỉ số này xác định mức độ kích thước ảnh so với ngưỡng cơ bản (được hệ thống thiết lập từ trước). Chỉ số này bảo đảm cho ảnh nhận diện và ảnh được nhận diện trên cùng một mức độ kích thước, tăng độ tin cậy tìm kiếm đối tượng.

Mỗi một lần thực hiện kim tự tháp ảnh, các chỉ tiêu đều được tính toán lại và lưu trữ trong cơ sở dữ liệu như một mẫu mới.

B2.5: lưu trữ các thông tin ảnh được xử lý.

Đặc trưng thứ mười sáu: là ảnh được biến đổi che khuất (thêm khẩu trang) từ ảnh khuôn mặt ban đầu, sử dụng mô hình học sâu (mạng đối nghịch, Generative Adversarial Networks - GAN). Để thuận tiện trong việc lưu trữ, ảnh được gắn khẩu trang được mã hóa thành chuỗi ký tự thông qua phương pháp mã hóa base64.

Đặc trưng thứ mười tám: là ảnh khuôn mặt, được mã hóa bởi base64.

Đặc trưng thứ mười bảy: là giá trị mô tả trạng thái khuôn mặt, bao gồm “khuôn mặt” và “khẩu trang”.

Ảnh khuôn mặt khi đăng ký và thực hiện kim tự tháp ảnh đều tiến hành biến đổi che khuất (thêm khẩu trang), sau đó tính toán lại các chỉ số như trên, và được coi như một mẫu mới. Khi tìm kiếm, ảnh khuôn mặt có đeo khẩu trang sẽ chỉ so khớp với các mẫu có khẩu trang.

Như quá trình mô tả ở trên, tổng số mẫu mới trong cơ sở dữ liệu được tính toán từ một ảnh đăng ký = *số lần kim tự tháp* x 2.

Bước 3: đánh giá cơ sở dữ liệu khuôn mặt. Tham chiếu Hình 4, quá trình đánh giá được thực hiện qua các bước nhỏ như sau:

B3.1: các thông tin đặc trưng thứ nhất, thứ hai, thứ ba, thứ tư, thứ năm, thứ mười sáu, thứ mười bảy và thứ mười tám được lưu vào dữ liệu định dạng \*.parquet. Dạng dữ liệu này thuận tiện trong việc phân mảnh, và ghi nối, khi cập nhật mẫu mới của đối tượng nhất định thì mức độ biến động nhỏ và thao tác đơn giản.

Các thông tin đặc trưng không bao gồm thứ năm, thứ mười sáu, thứ mười bảy và thứ mười tám được lưu vào trong cơ sở dữ liệu môn-gô. Khi hệ thống tiến hành nhận diện, thông tin từ pa-quét (parquet) và môn-gô (mongo) được đồng bộ và xử lý dưới dạng chuỗi dữ liệu (dataframe).

B3.2: tính toán ma trận tương quan từ các đặc trưng nhận diện (đặc trưng thứ năm) của từng mẫu ảnh với nhau. Một cơ sở dữ liệu có  $N$  mẫu sẽ tạo ra một ma trận tương quan có kích thước  $N \times N$ , là một ma trận vuông, dương và đối xứng. Điểm có vị trí  $(i, j)$  trong ma trận tương quan là giá trị khoảng cách cô-sin của véc-tơ thứ  $i$  và thứ  $j$ . Những mẫu cùng nhóm sẽ có giá trị thấp hơn so với những mẫu không cùng nhóm.

B3.3: đánh giá toàn cục, các chỉ số được tính toán từ ma trận tương quan (B3.2), với một ngưỡng nhất định, tại vị trí  $(i, j)$  có giá trị nhỏ hơn ngưỡng thì mẫu thứ  $i$  và thứ  $j$  được coi là cùng nhóm, ngược lại là khác nhóm. Chỉ số tính toán bao gồm chỉ số chấp nhận nhầm (False Acceptance Rate, FAR) là chỉ số đánh giá tỉ lệ các mẫu khác nhóm (cùng của một người đăng ký) nhưng được xác định là của người không cùng nhóm; chỉ số loại bỏ nhầm (False Rejection Rate, FRR) là chỉ số đánh giá tỉ lệ các mẫu cùng nhóm nhưng không được xác định là giống nhau.

Công thức tính như sau:

$$FAR = \frac{FP}{TN + FP}$$

$$FRR = \frac{FN}{TP + FN}$$

Trong đó, TN, TP, FN, FP lần lượt là chỉ số âm thật, dương thật, âm giả và dương giả khi phân loại, tương ứng với một ngưỡng phân loại nhất định (có giá trị từ 0 đến 1). Đối với một cơ sở dữ liệu, FAR và FRR càng thấp càng tốt (đơn vị là phần trăm, %).

Trong quá trình đánh giá, ngưỡng phân loại được xác định từ 0 đến 1 với bước là 0,01. Mỗi một cơ sở dữ liệu được đánh giá trên 100 ngưỡng đó, vị trí ngưỡng tối ưu được xác định bởi vị trí có FAR và FRR gần bằng nhau nhất. Giá trị FAR và FRR tốt được xác định là nhỏ hơn hoặc bằng 0,1% (giá trị được tính toán từ thực nghiệm thống kê). Với một cơ sở dữ liệu bất kỳ, nếu FAR và FRR lớn hơn giá trị 0,1% thì cần phải tiến hành lọc hoặc thay thế mẫu ở trong đó. Trong đó các mẫu dương giả và âm giả sẽ được loại bỏ khỏi cơ sở dữ liệu, bằng cách thiết lập đặc trưng thứ ba bởi giá trị “hết hạn”.

B3.4: đánh giá cục bộ, tính toán giá trị ngưỡng tự thích ứng ( $thr_{accept}$ ) và ngưỡng phân biệt ( $thr_{reject}$ ) của từng mẫu trong cơ sở dữ liệu. Đối với một mẫu  $i$  bất kỳ trong cơ sở dữ liệu, trong ma trận tương quan  $M$ , và thứ tự (đặc trưng thứ nhất) của các mẫu cùng nhóm là từ  $n$  đến  $m$  thì công thức tính ngưỡng như sau:

$$thr_{accept} = \max(M(i, j) | j \in [n, m])$$

$$thr_{reject} = \min(M(i, j) | j \notin [n, m])$$

Các mẫu đạt yêu cầu là các mẫu  $thr_{accept} < thr_{reject} + \zeta$  trong đó  $\zeta$  là chỉ số đệm, lớn hơn 0. Những mẫu không đạt yêu cầu sẽ bị loại bỏ và được thay thế bởi các mẫu ảnh thu được từ quá trình vận hành thực địa.

Bước 4: thống kê tần suất được dùng để phân loại đối với các mẫu ảnh. Cụ thể được thực hiện qua các bước nhỏ như sau:

B4.1: thực hiện phân loại đối với một khuôn mặt bất kỳ, trong quá trình vận hành hệ thống nhận diện khuôn mặt. Trong đó, các đặc trưng thứ sáu, thứ bảy, thứ tám, thứ chín, thứ mười hai, thứ mười ba, thứ mười bốn, thứ mười lăm, thứ mười bảy được sử dụng như yếu tố hỗ trợ tìm kiếm. Hay nói cách khác, việc so khớp giữa khuôn mặt bất kỳ  $v$  với cơ sở

dữ liệu  $D_{face}$  là sự tổ hợp giữa nhiều yếu tố, công thức lựa chọn các mẫu trong  $D_{face}$  để so khớp như sau:

$$d_{face} = \begin{cases} D_i | D_{if} \in [v_f \mp \delta_f], f = 6 \sim 9, 12 \sim 15 \\ D_{if} = v_f, f = 17 \end{cases}$$

Sau khi chọn lọc, quá trình so khớp giữa véc-tơ  $v_{f5}$  của mẫu cần nhận diện và  $d_{face}$  được tính như sau:

$$score = \frac{dot(v_{f5}, d_{face})}{\|v_{f5}\| \|d_{face}\|}$$

$score$  là một chuỗi khoảng cách cô-sin từ mẫu cần nhận diện đến các mẫu đã được chọn lọc ở trong cơ sở dữ liệu, mẫu khớp nhất là mẫu có giá trị nhỏ nhất, hay:

$$matched = argmin(score)$$

B4.2: xác định kết quả so khớp được chấp nhận hoặc không. Sau khi so khớp, nếu mẫu khớp nhất  $matched$  có khoảng cách nhỏ hơn ngưỡng tự thích ứng  $thr_{accept}$  thì được coi là kết quả đúng, ngược lại kết quả nhận diện sẽ không được hệ thống công nhận. Như vậy, mỗi một mẫu ảnh trong cơ sở dữ liệu sẽ có ngưỡng so khớp riêng, tự thích ứng với phân bố toàn cục của cơ sở dữ liệu.

B4.3: tích lũy tần suất so khớp thành công của mẫu ảnh trong cơ sở dữ liệu (xác định bởi B4.2). Mỗi lần so khớp thành công, mẫu ảnh tương ứng sẽ được cộng một lượt tần suất. Thống kê trên toàn bộ mẫu trong cơ sở dữ liệu và cập nhật thành trường thông tin thứ mười chín lưu trữ trong môn-gô. Trường dữ liệu này chỉ được kích hoạt khi bắt đầu vận hành thực tế.

Trong quá trình triển khai hệ thống nhận diện, sau mỗi chu kỳ T, các mẫu không được sử dụng để so khớp sẽ được thay thế bởi các mẫu thu thập được tại thực địa.

### Ví dụ thực hiện sáng chế

Nhóm nghiên cứu đã tiến hành thử nghiệm và đánh giá hiệu quả của phương pháp được đề cập trong sáng chế này với nhiều dạng dữ liệu khác nhau. Trong đó, ví dụ về quá

trình phân tích hệ thống nhận diện khuôn mặt triển khai tại Tổng công ty mạng lưới Viettel (VTNET) được mô tả như trong Hình 5, Hình 6 và Hình 7.

Dữ liệu bao gồm ảnh của 750 cán bộ nhân viên tại địa điểm thử nghiệm, sau khi xử lý, trích xuất đặc trưng tạo ra cơ sở dữ liệu hơn 6000 mẫu. Hình 5 mô tả ma trận tương quan của các đặc trưng nhận diện khuôn mặt, trong đó Hình 5A thể hiện toàn bộ ma trận, Hình 5B thể hiện ví dụ về một số nhóm đối tượng. Trong đó, màu càng tối thì có giá trị càng nhỏ, và càng có khả năng thuộc cùng một nhóm đối tượng.

Hình 6 mô tả các giá trị FAR và FRR khi phân tích khả năng nhận diện trên 100 ngưỡng từ 0 đến 1. Giá trị tối ưu FAR, FRR thu được lần lượt là 0,1% và 0,11%. Đồng thời, kết quả phân tích ngưỡng tự thích ứng của tất cả các mẫu được mô tả như trong Hình 7, trong đó hình bên trái mô tả ngưỡng tự thích ứng (xanh lam), và ngưỡng phân biệt (đỏ), hình bên phải mô tả sự chênh lệch giữa hai ngưỡng trên. Chúng ta có thể thấy có nhiều mẫu giá trị ngưỡng phân biệt thấp hơn ngưỡng tự thích ứng, cho thấy những mẫu này không thể dùng để nhận diện, cần loại bỏ và thay thế.

Quá trình thống kê tần suất và đưa ra quyết định cập nhật cơ sở dữ liệu cũng đang được thử nghiệm và cho kết quả tốt.

### **Lợi ích (hiệu quả) có thể đạt được bởi sáng chế**

Phương pháp được đề cập trong sáng chế đã được sử dụng trong các sản phẩm sử dụng công nghệ nhận diện khuôn mặt, chủ yếu dùng để phân tích cơ sở dữ liệu và đưa ra quyết định cập nhật, dùng trong giải pháp tòa nhà thông minh, an ninh căn hộ của trung tâm camera, thuộc tổng công ty công nghiệp công nghệ cao Viettel.

Phương pháp đề cập đưa ra cách sử dụng các đặc trưng hỗ trợ để giảm thiểu thời gian so khớp với độ ổn định cao hơn. Những mẫu tương đồng về độ sáng, độ nét, kích thước hoặc góc khuôn mặt sẽ được so khớp với nhau, tránh những trường hợp tìm kiếm toàn cục, chỉ đưa ra được kết quả duy nhất trong mọi lần tìm kiếm, tăng cường khả năng sửa sai, tự thích ứng của hệ thống.

Phương pháp còn đưa ra cách bố trí lưu trữ thông tin tối ưu, giảm thiểu thời gian khởi tạo cơ sở dữ liệu (với cơ sở dữ liệu VTNET, thời gian khởi tạo chưa đến 70 mili giây), và cách đồng bộ dữ liệu khi lưu trữ phân mảnh (parquet và môn-gô).

Ngoài ra, phương pháp còn đưa ra cách tính toán và sử dụng ngưỡng cho riêng mỗi mẫu ở trong cơ sở dữ liệu, bảo đảm độ tin cậy cao nhất. Đồng thời thống kê tần suất sử dụng, là căn cứ để cập nhật dữ liệu.

### **Yêu cầu bảo hộ**

1. Phương pháp cấu trúc và đánh giá cơ sở dữ liệu tự thích ứng cho nhận diện khuôn mặt, bao gồm:

bước 1: phát hiện khuôn mặt trong ảnh đăng ký (để sử dụng hệ thống nhận diện khuôn mặt) của người dùng; bước này sử dụng mô hình học sâu để phát hiện khuôn mặt người trong ảnh được đăng ký bởi người dùng; đối với các ảnh đăng ký để sử dụng hệ thống nhận diện khuôn mặt, yêu cầu độ tin cậy (có giá trị từ 0 đến 1) cao, ngưỡng phát hiện khuôn mặt được thiết lập lớn hơn 0,95; sau khi phát hiện khuôn mặt trong ảnh đăng ký, vị trí khuôn mặt được xác định bởi hai điểm là điểm góc trái trên cùng (top-left) và điểm góc phải dưới cùng (bottom-right), theo tọa độ ảnh;

bước 2: xác định các đặc trưng ảnh, bao gồm biến đổi che khuất; các đặc trưng ảnh bao gồm đặc trưng định danh, đặc trưng nhận diện khuôn mặt, đặc trưng mô tả bề mặt, đặc trưng hình học; cụ thể:

xác định các đặc trưng định danh, bao gồm:

đặc trưng thứ nhất: số thứ tự mẫu ảnh, là thứ tự xử lý của mẫu ảnh trong cả quá trình khởi tạo cũng như cập nhật cơ sở dữ liệu; đặc trưng này dùng để bảo đảm tính duy nhất của từng mẫu ảnh;

đặc trưng thứ hai: tên đối tượng, là tên định danh của người dùng trên hệ thống, mỗi người dùng sẽ được cung cấp một tên duy nhất; mỗi một đối tượng có thể có nhiều mẫu ảnh được đăng ký;

đặc trưng thứ ba: trạng thái sử dụng của mẫu ảnh, bao gồm hai giá trị “đang sử dụng” biểu thị mẫu ảnh này đang được sử dụng để so khớp tìm kiếm khi nhận diện khuôn mặt và “hết hạn” biểu thị mẫu ảnh này vẫn tồn tại trong cơ sở dữ liệu nhưng không còn được dùng để so khớp, tìm kiếm, nhằm xác định các mẫu đã bị loại bỏ bởi cơ sở dữ liệu thông qua phương pháp xóa mềm; trường thông tin này bảo đảm việc loại bỏ mẫu ảnh mà không ảnh hưởng đến cả cơ sở dữ liệu cũng như quá trình tìm kiếm đối tượng;

đặc trưng thứ tư: mã phân biệt mẫu, xác định bởi tên người được đăng ký, thứ tự đăng ký và thời gian đăng ký; mã phân biệt này có cấu trúc gồm thời gian đăng ký (ngày-tháng-năm), thứ tự mẫu ảnh và tên đối tượng, tổ hợp lại thành một chuỗi ký tự và được mã hóa thông qua phương pháp mã hóa base64 kết hợp với phương pháp băm dữ liệu SHA256 để tạo ra một mã đại diện cho mẫu ảnh được đăng ký; đặc trưng này bảo đảm tính duy nhất của mẫu, bất biến trong quá trình lưu trữ cũng như truyền tải, hoặc đồng bộ các cơ sở dữ liệu, cũng là yếu tố kiểm tra chéo đối với đặc trưng thứ nhất và đặc trưng thứ hai;

xác định đặc trưng nhận diện khuôn mặt:

đặc trưng thứ năm: đặc trưng nhận diện khuôn mặt; bước này sử dụng mô hình học sâu được huấn luyện trên 50 triệu ảnh khuôn mặt với đầu vào là ảnh khuôn mặt (được phát hiện ở bước 1) đã được hiệu chỉnh về kích thước 112 x 112 (đây là kích thước chuẩn của mô hình, định nghĩa lúc huấn luyện mô hình); đặc trưng nhận diện khuôn mặt này là một véc-tơ 512 chiều, sử dụng để so khớp, tìm kiếm khuôn mặt cho các hệ thống nhận diện; Nhằm thuận tiện hơn trong quá trình lưu trữ, đặc trưng này được mã hóa thành dạng chuỗi ký tự, thông qua phương pháp mã hóa base64.

xác định các đặc trưng mô tả bề mặt:

đặc trưng thứ sáu: độ nhiễu của ảnh, được tính bởi phương pháp giả định nhiễu tuân theo phân bố gaussian; trong đó chỉ số nhiễu được ước tính dựa trên độ lệch tuyệt đối trung bình của hệ số chi tiết từ biến đổi sóng nhỏ; giá trị nhiễu càng nhỏ càng tốt;

đặc trưng thứ bảy: độ sắc nét của ảnh, được tính bởi phương pháp biến đổi laplacian; trong đó, giá trị độ sắc nét là tổng số điểm ảnh trên ảnh khuôn mặt sau khi biến đổi laplacian có giá trị lớn hơn ngưỡng được thiết lập trước; độ sắc nét của ảnh càng cao càng tốt;

đặc trưng thứ tám: độ sáng, được tính bởi giá trị trung bình của kênh sáng của ảnh khuôn mặt; độ sáng có giá trị từ 0 đến 255 (tương ứng với giá trị lớn nhất của một ảnh 8 bit), trong đó giá trị càng lớn càng sáng, theo trực quan;

đặc trưng thứ chín: độ tương phản, được tính bởi phương pháp trung bình căn bậc ba của ảnh khuôn mặt, giá trị độ tương phản càng lớn thì tính phân biệt cục bộ của ảnh càng cao;

các đặc trưng từ thứ sáu đến đặc trưng thứ tám được dùng để làm tiền đề cho quá trình tối ưu cơ sở dữ liệu trong các điều kiện khắc nghiệt (trời tối, nhiều) nếu cần; đồng thời cũng là chỉ số thống kê quan trọng để đánh giá hiệu quả cũng như độ chính xác của hệ thống khi vận hành thực tế;

xác định các đặc trưng hình học:

đặc trưng thứ mười, đặc trưng thứ mười một: lần lượt là điểm góc trái trên cùng và điểm góc phải dưới cùng được trả về trong quá trình phát hiện khuôn mặt; hai tọa độ này dùng để hỗ trợ phát hiện các đối tượng nổi bật trong ảnh đăng ký (thông thường ở chính giữa), và xác định mức độ được kích hoạt trong hình ảnh thực tế nơi vận hành hệ thống nhận diện khuôn mặt;

đặc trưng thứ mười hai, đặc trưng thứ mười ba, đặc trưng thứ mười bốn: là các góc khuôn mặt so với trục tạo ảnh (trục pháp tuyến của mặt cầu tạo ảnh), được tính bởi mô hình hồi quy được huấn luyện trên tập dữ liệu 2000 ảnh với phạm vi góc từ -90 độ đến 90 độ; mô hình này có đầu vào là ảnh khuôn mặt, và đầu ra là ba chỉ số góc khuôn mặt, bao gồm góc theo chiều song song, góc theo chiều thẳng đứng và góc theo chiều vuông góc;

đặc trưng thứ mười lăm: số lần kim tự tháp ảnh, có nghĩa là số lần giảm kích thước của ảnh khuôn mặt (ngang và cao mỗi chiều giảm một nửa giá trị cho mỗi lần kim tự tháp ảnh); chỉ số này xác định mức độ kích thước ảnh so với ngưỡng cơ bản (được hệ thống thiết lập từ trước); chỉ số này bảo

đảm cho ảnh nhận diện và ảnh được nhận diện trên cùng một mức độ kích thước, tăng độ tin cậy tìm kiếm đối tượng;

lưu trữ các thông tin ảnh được xử lý.

đặc trưng thứ mười sáu: là ảnh được biến đổi che khuất (thêm khẩu trang) từ ảnh khuôn mặt ban đầu, sử dụng mô hình học sâu (mạng đối nghịch, generative adversarial networks - GAN); để thuận tiện trong việc lưu trữ, ảnh được gắn khẩu trang được mã hóa thành chuỗi ký tự thông qua phương pháp mã hóa base64;

đặc trưng thứ mười tám: là ảnh khuôn mặt, được mã hóa bởi base64;

đặc trưng thứ mười bảy: là giá trị mô tả trạng thái khuôn mặt, bao gồm “khuôn mặt” và “khẩu trang”;

ảnh khuôn mặt khi đăng ký và thực hiện kim tự tháp ảnh đều tiến hành biến đổi che khuất (thêm khẩu trang), sau đó xác định lại các chỉ số như trên, và được coi như một mẫu mới; khi tìm kiếm, ảnh khuôn mặt có đeo khẩu trang sẽ chỉ so khớp với các mẫu có khẩu trang; như quá trình mô tả ở trên, tổng số mẫu mới trong cơ sở dữ liệu được tính toán từ một ảnh đăng ký = số lần kim tự tháp  $\times 2$ ;

bước 3: đánh giá cơ sở dữ liệu khuôn mặt; cụ thể được thực hiện qua các bước nhỏ như sau:

các thông tin đặc trưng thứ nhất, thứ hai, thứ ba, thứ tư, thứ năm, thứ mười sáu, thứ mười bảy và thứ mười tám được lưu vào dữ liệu định dạng \*.parquet; dạng dữ liệu này thuận tiện trong việc phân mảnh, và ghi nối, khi cập nhật mẫu mới của đối tượng nhất định thì mức độ biến động nhỏ và thao tác đơn giản; các thông tin đặc trưng không bao gồm đặc trưng thứ năm, thứ mười sáu, thứ mười bảy và thứ mười tám được lưu vào trong cơ sở dữ liệu môn-gô; khi hệ thống tiến hành nhận diện, thông tin từ pa-quét và môn-gô được đồng bộ và xử lý dưới dạng chuỗi dữ liệu (dataframe);

xác định ma trận tương quan từ các đặc trưng nhận diện (đặc trưng thứ năm) của từng mẫu ảnh với nhau; một cơ sở dữ liệu có  $N$  mẫu sẽ tạo ra một ma trận tương quan có kích thước  $N \times N$ , là một ma trận vuông, dương và đối xứng; điểm có vị trí  $(i, j)$  trong ma trận tương quan là giá trị khoảng cách cô-sin của véc-tơ thứ  $i$  và thứ  $j$ ; những mẫu cùng nhóm sẽ có giá trị thấp hơn so với những mẫu không cùng nhóm;

đánh giá toàn cục, các chỉ số được tính toán từ ma trận tương quan, với một ngưỡng nhất định, tại vị trí  $(i, j)$  có giá trị nhỏ hơn ngưỡng thì mẫu thứ  $i$  và thứ  $j$  được coi là cùng nhóm, ngược lại là khác nhóm; chỉ số tính toán bao gồm chỉ số chấp nhận nhầm (false acceptance rate, FAR) là chỉ số đánh giá tỉ lệ các mẫu khác nhóm (cùng của một người đăng ký) nhưng được xác định là của người không cùng nhóm; chỉ số loại bỏ nhầm (false rejection rate, FRR) là chỉ số đánh giá tỉ lệ các mẫu cùng nhóm nhưng không được xác định là giống nhau.

công thức tính như sau:

$$FAR = \frac{FP}{TN + FP}$$

$$FRR = \frac{FN}{TP + FN}$$

trong đó, TN, TP, FN, FP lần lượt là chỉ số âm thật, dương thật, âm giả và dương giả khi phân loại, tương ứng với một ngưỡng phân loại nhất định (có giá trị từ 0 đến 1); đối với một cơ sở dữ liệu, FAR và FRR càng thấp càng tốt (đơn vị là phần trăm, %);

trong quá trình đánh giá, ngưỡng phân loại được xác định từ 0 đến 1 với bước là 0,01; mỗi một cơ sở dữ liệu được đánh giá trên 100 ngưỡng đó, vị trí ngưỡng tối ưu được xác định bởi vị trí có FAR và FRR gần bằng nhau nhất; giá trị FAR và FRR tốt được xác định là nhỏ hơn hoặc bằng 0,1% (giá trị được tính toán từ thực nghiệm thống kê); với một cơ sở dữ liệu bất kỳ, nếu FAR và FRR lớn hơn giá trị 0,1% thì cần phải tiến hành lọc hoặc thay

thể mẫu ở trong đó; trong đó các mẫu dương giả và âm giả sẽ được loại bỏ khỏi cơ sở dữ liệu, bằng cách thiết lập đặc trưng thứ ba bởi giá trị “hết hạn”.

đánh giá cục bộ, tính toán giá trị ngưỡng tự thích ứng,  $thr_{accept}$  và ngưỡng phân biệt,  $thr_{reject}$  của từng mẫu trong cơ sở dữ liệu; đối với một mẫu  $i$  bất kỳ trong cơ sở dữ liệu, trong ma trận tương quan  $M$ , và thứ tự (đặc trưng thứ nhất) của các mẫu cùng nhóm là từ  $n$  đến  $m$  thì công thức tính ngưỡng như sau:

$$thr_{accept} = \max(M(i, j) | j \in [n, m])$$

$$thr_{reject} = \min(M(i, j) | j \notin [n, m])$$

các mẫu đạt yêu cầu là các mẫu  $thr_{accept} < thr_{reject} + \zeta$  trong đó  $\zeta$  là chỉ số đệm, lớn hơn 0; những mẫu không đạt yêu cầu sẽ bị loại bỏ và được thay thế bởi các mẫu ảnh thu được từ quá trình vận hành thực địa;

bước 4: thống kê tần suất được dùng để phân loại đối với các mẫu ảnh; Cụ thể được thực hiện qua các bước nhỏ như sau:

thực hiện phân loại đối với một khuôn mặt bất kỳ, trong quá trình vận hành hệ thống nhận diện khuôn mặt; trong đó, các đặc trưng thứ sáu, thứ bảy, thứ tám, thứ chín, thứ mười hai, thứ mười ba, thứ mười bốn, thứ mười lăm, thứ mười bảy được sử dụng như yếu tố hỗ trợ tìm kiếm; hay nói cách khác, việc so khớp giữa khuôn mặt bất kỳ  $v$  với cơ sở dữ liệu  $D_{face}$  là sự tổ hợp giữa nhiều yếu tố, công thức lựa chọn các mẫu trong  $D_{face}$  để so khớp như sau:

$$d_{face} = \begin{cases} D_i | D_{if} \in [v_f \mp \delta_f], f = 6 \sim 9, 12 \sim 15 \\ D_{if} = v_f, f = 17 \end{cases}$$

sau khi chọn lọc, quá trình so khớp giữa véc-tơ  $v_{f5}$  của mẫu cần nhận diện và  $d_{face}$  được tính như sau:

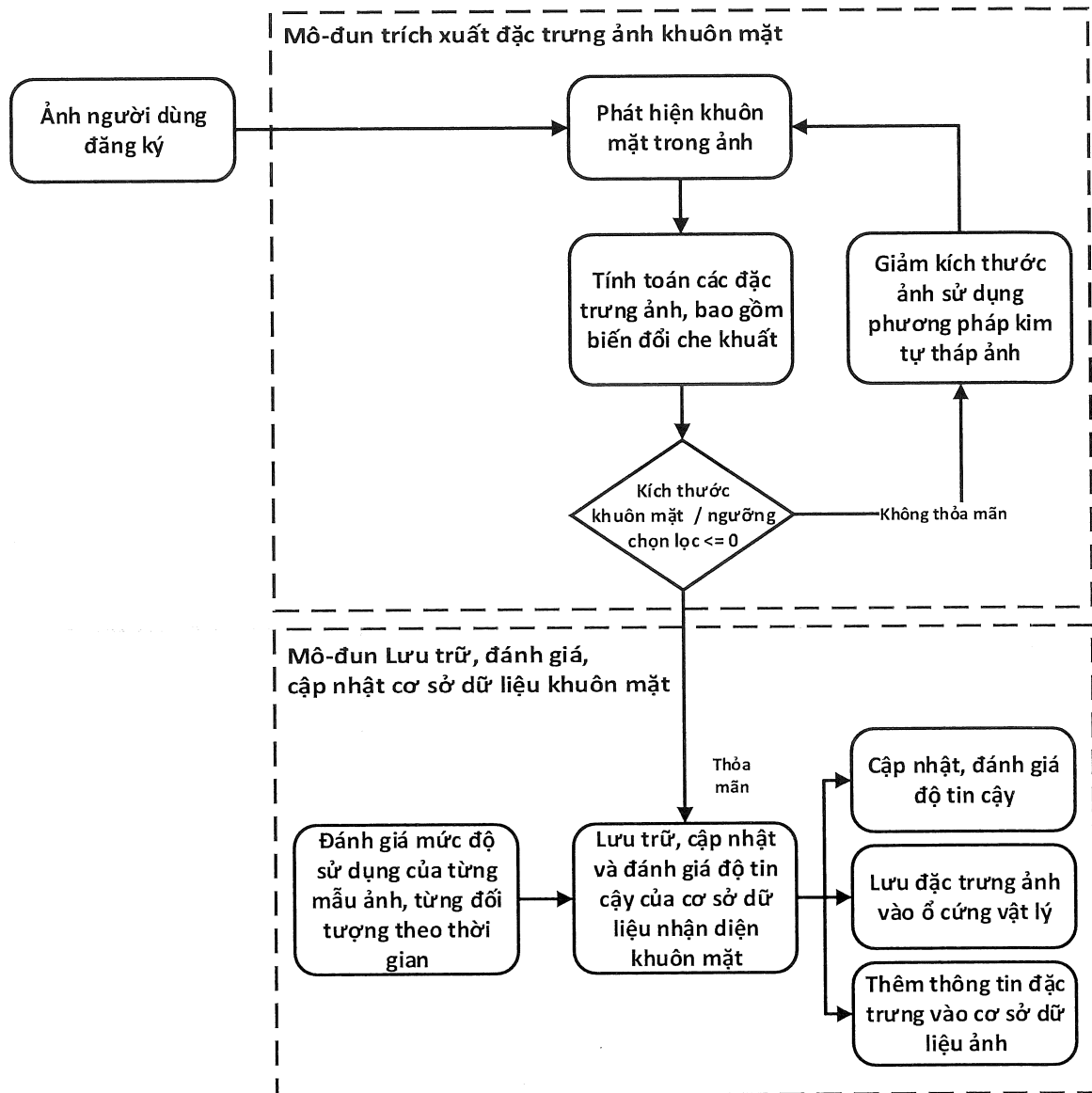
$$score = \frac{dot(v_{f5}, d_{face})}{\|v_{f5}\| \|d_{face}\|}$$

$score$  là một chuỗi khoảng cách cô-sin từ mẫu cần nhận diện đến các mẫu đã được chọn lọc ở trong cơ sở dữ liệu, mẫu khớp nhất là mẫu có giá trị nhỏ nhất, hay:

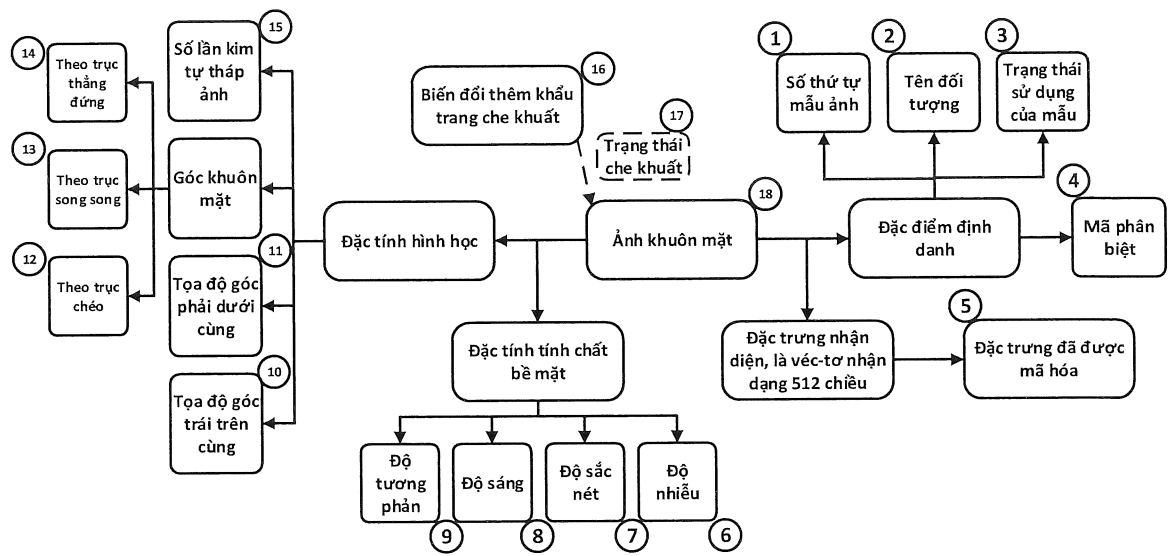
$$matched = \operatorname{argmin}(score)$$

xác định kết quả so khớp được chấp nhận hoặc không; sau khi so khớp, nếu mẫu khớp nhất  $matched$  có khoảng cách nhỏ hơn ngưỡng tự thích ứng  $thr_{accept}$  thì được coi là kết quả đúng, ngược lại kết quả nhận diện sẽ không được hệ thống công nhận; như vậy, mỗi một mẫu ảnh trong cơ sở dữ liệu sẽ có ngưỡng so khớp riêng, tự thích ứng với phân bố toàn cục của cơ sở dữ liệu;

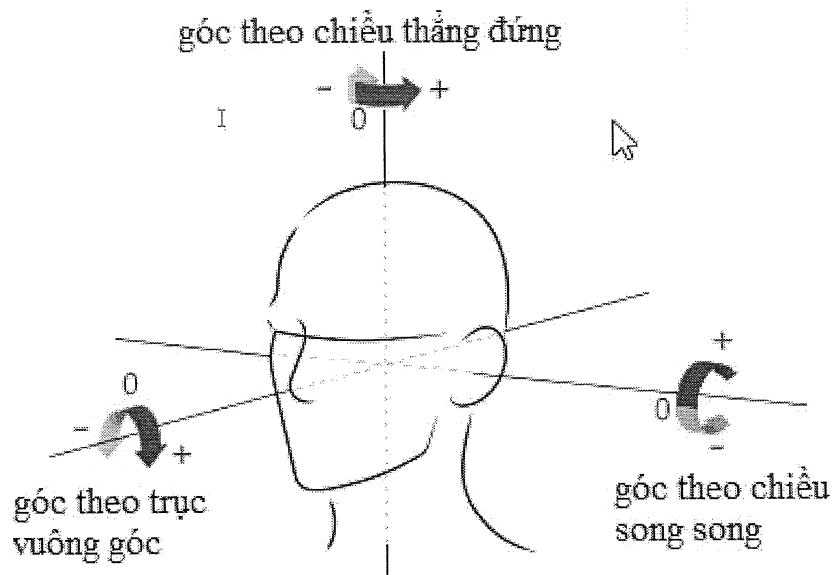
tích lũy tần suất so khớp thành công của mẫu ảnh trong cơ sở dữ liệu; mỗi lần so khớp thành công, mẫu ảnh tương ứng sẽ được cộng một lượt tần suất; thống kê trên toàn bộ mẫu trong cơ sở dữ liệu và cập nhật thành trường thông tin thứ mười chín lưu trữ trong môn-gô; trường dữ liệu này chỉ được kích hoạt khi bắt đầu vận hành thực tế; trong quá trình triển khai hệ thống nhận diện, sau mỗi chu kỳ  $T$ , các mẫu không được sử dụng để so khớp sẽ được thay thế bởi các mẫu thu thập được tại thực địa.



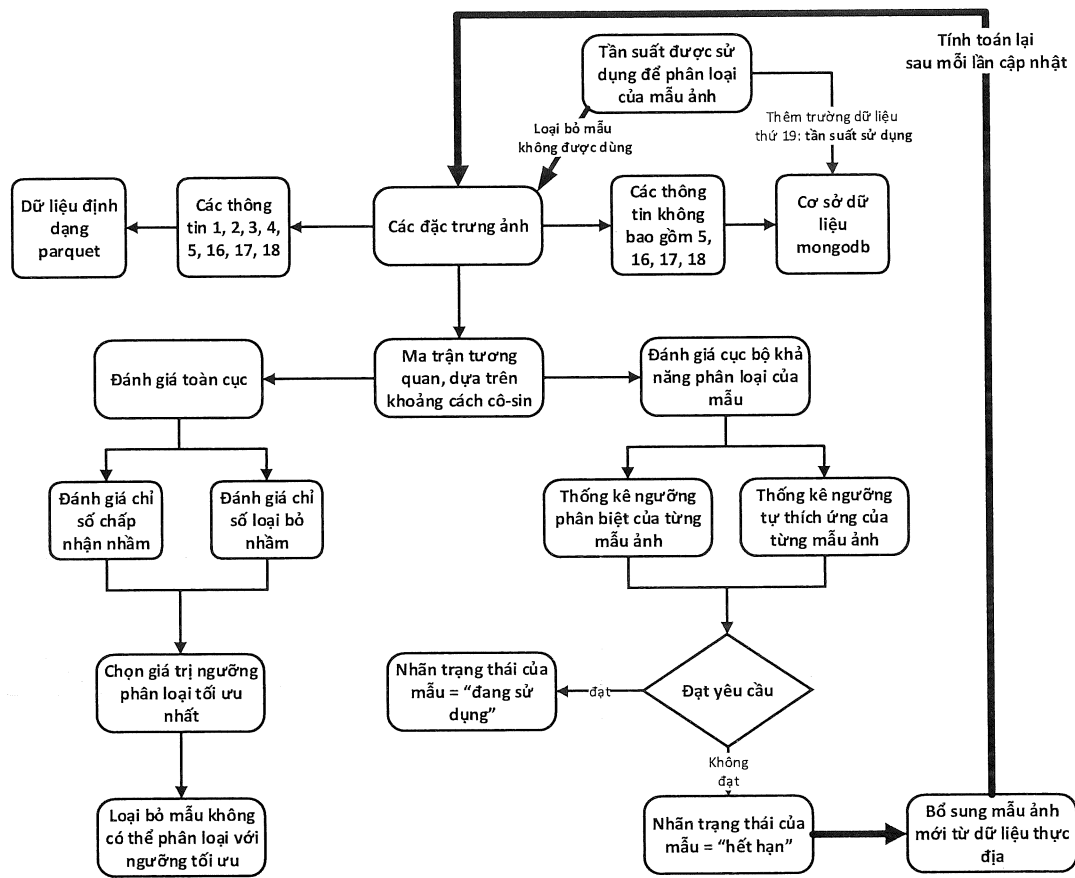
Hình 1



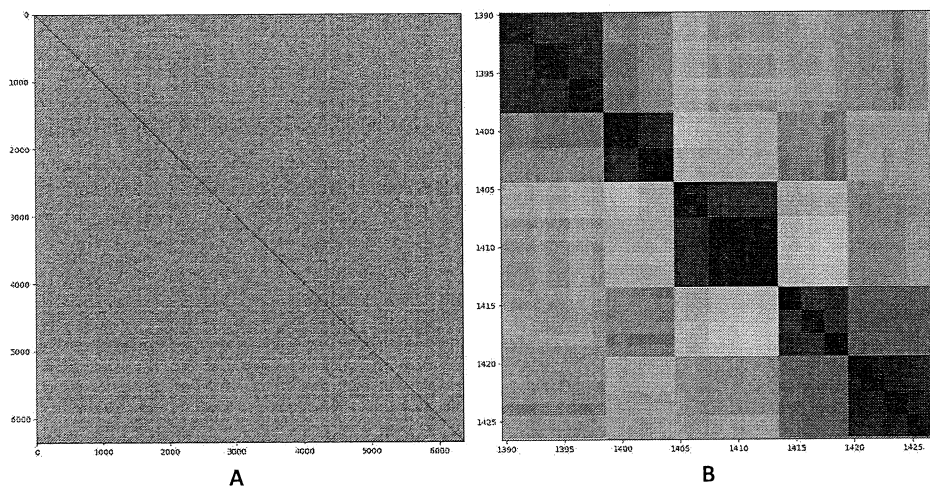
Hình 2



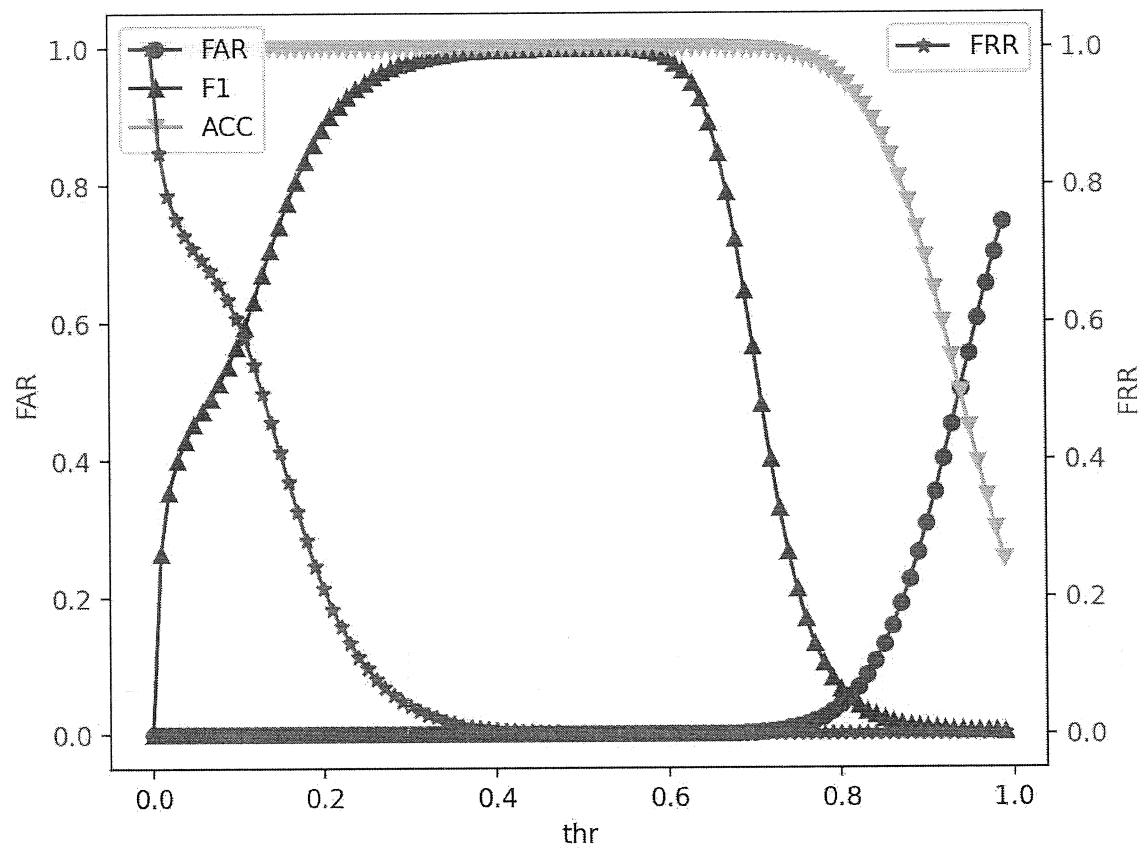
Hình 3



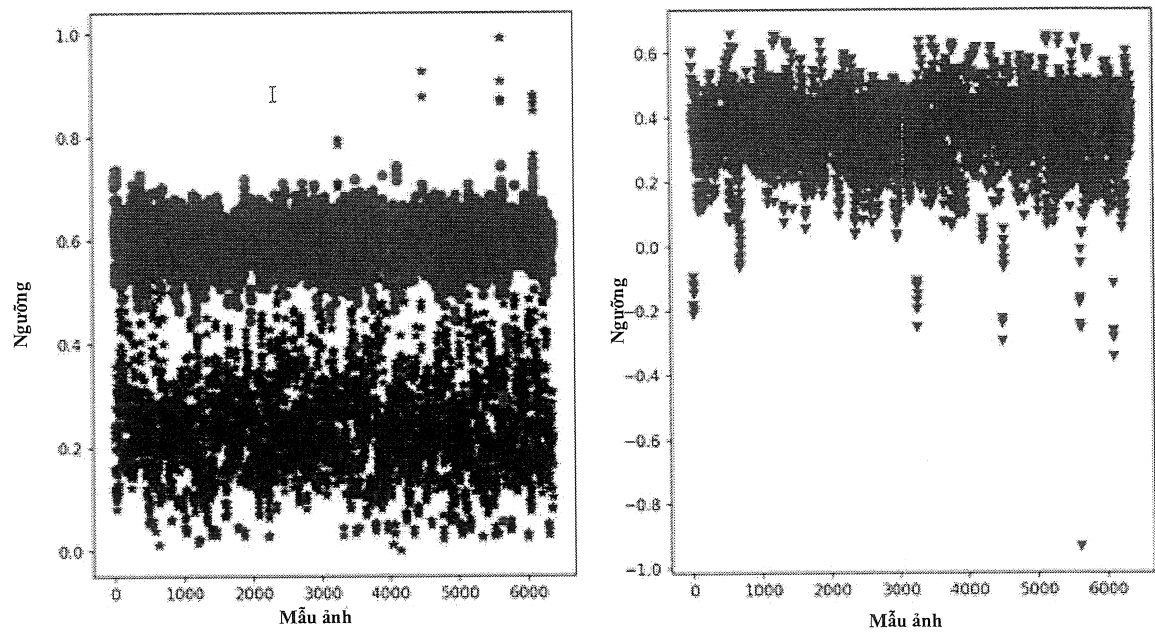
Hình 4



Hình 5



Hình 6



Hình 7