



(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)



1-0038745

(51)^{2020.01} G06K 9/00

(13) B

(21) 1-2020-05273

(22) 14/09/2020

(45) 26/02/2024 431

(43) 27/06/2022 411

(73) 1. Trường Đại học Công nghệ Thông tin (VN)

Khu phố 6, phường Linh Trung, quận Thủ Đức, thành phố Hồ Chí Minh

2. Đại học Quốc gia thành phố Hồ Chí Minh (VN)

Phường Linh Trung, quận Thủ Đức, thành phố Hồ Chí Minh

(72) Lê Đình Duy (VN); Võ Quốc Hưng (VN); Dương Anh Đức (VN).

(54) PHƯƠNG PHÁP TÌM NGƯỜI TẠI VỊ TRÍ XÁC ĐỊNH TỪ DỮ LIỆU VIDEO

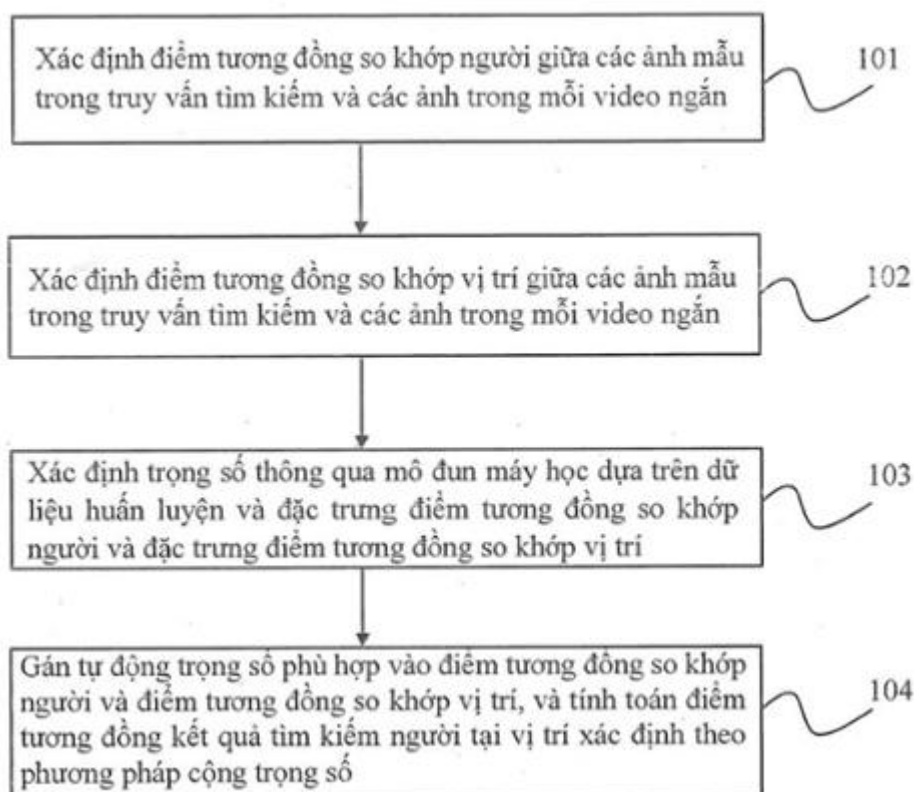
(57) Sáng chế đề cập đến phương pháp tìm người tại vị trí xác định từ dữ liệu video thông qua máy tính bao gồm các bước sau:

xác định điểm tương đồng so khớp người giữa các ảnh mẫu trong truy vấn tìm kiếm và các ảnh trong mỗi video ngắn;

xác định điểm tương đồng so khớp vị trí giữa các ảnh mẫu trong truy vấn tìm kiếm và các ảnh trong mỗi video ngắn;

xác định trọng số thông qua mô đun máy học dựa trên dữ liệu huấn luyện và đặc trưng điểm tương đồng so khớp người và đặc trưng điểm tương đồng so khớp vị trí;

gán tự động trọng số phù hợp vào điểm tương đồng so khớp người và điểm tương đồng so khớp vị trí, và tính toán điểm tương đồng kết quả tìm kiếm người tại vị trí xác định theo phương pháp cộng trọng số.



Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập đến phương pháp tìm người tại vị trí xác định từ dữ liệu video.

Tình trạng kỹ thuật của sáng chế

Hiện nay phương pháp tìm người tại vị trí xác định từ dữ liệu video, với dữ liệu video được phân thành các video ngắn (còn gọi là shot), được thực hiện như sau:

thực hiện tìm kiếm người bằng cách tính điểm tương đồng giữa các ảnh mẫu (ảnh chứa khuôn mặt đối tượng muốn tìm kiếm) trong truy vấn tìm kiếm (search query) và các ảnh trong mỗi shot dựa trên so khớp khuôn mặt (face matching); sau đó các điểm tương đồng so khớp người được sắp xếp theo thứ tự giảm dần tạo thành bảng danh sách kết quả; vị trí càng ở đầu trong bảng danh sách là kết quả có điểm tương đồng càng cao;

thực hiện tìm kiếm vị trí bằng cách tính điểm tương đồng giữa các ảnh mẫu (ảnh chứa vị trí muốn tìm kiếm) trong truy vấn tìm kiếm (search query) và các ảnh trong mỗi shot dựa trên so khớp ảnh (image matching); sau đó các điểm tương đồng so khớp vị trí được sắp xếp theo thứ tự giảm dần tạo thành bảng danh sách kết quả; vị trí càng ở đầu trong bảng danh sách là kết quả có điểm tương đồng càng cao; và

thực hiện kết hợp (fusion) kết quả tìm kiếm người và tìm kiếm vị trí bằng phương pháp cộng theo trọng số các điểm tương đồng so khớp người và so khớp vị trí để trả về bảng danh sách kết quả điểm tương đồng tìm kiếm người tại vị trí xác định.

Việc chọn trọng số phù hợp khi thực hiện kết hợp hai kết quả tìm kiếm riêng biệt ảnh hưởng lớn đến kết quả cuối cùng là người tại vị trí xác định. Hiện nay việc chọn trọng số thường được thực hiện một cách thủ công và cố định, và trọng số được gán cố định giống nhau cho tất cả các kết quả tìm kiếm người hoặc vị trí. Điều này không phù hợp do có trường hợp kết quả tìm kiếm có điểm tương đồng về người rất cao nhưng điểm tương đồng vị trí lại rất thấp, một số trường hợp thì ngược lại, vì thế các trọng số cần được gán một cách phù hợp và linh hoạt hơn thay vì gán cố định.

Bản chất kỹ thuật của sáng chế

Sáng chế đề xuất phương pháp tìm người tại vị trí xác định từ dữ liệu video giúp tăng độ chính xác trong kết quả tìm kiếm, tăng hiệu năng và tốc độ tính toán trong các hệ thống tìm kiếm, bằng cách kết hợp kết quả tìm kiếm người và kết quả tìm kiếm vị trí thông qua phương pháp cộng trọng số, trong đó trọng số được xác định một cách riêng biệt và phù hợp cho từng kết quả tìm kiếm và được gán vào phép cộng trọng số một cách tự động, linh hoạt thông qua mô đun máy học.

Phương pháp tìm người tại vị trí xác định từ tập dữ liệu video thông qua máy tính bao gồm các bước sau:

xác định điểm tương đồng so khớp người giữa các ảnh mẫu trong truy vấn tìm kiếm và các ảnh trong mỗi video ngắn;

xác định điểm tương đồng so khớp vị trí giữa các ảnh mẫu trong truy vấn tìm kiếm và các ảnh trong mỗi video ngắn;

xác định trọng số thông qua mô đun máy học dựa trên dữ liệu huấn luyện và đặc trưng điểm tương đồng so khớp người và đặc trưng điểm tương đồng so khớp vị trí;

gán tự động trọng số phù hợp vào điểm tương đồng so khớp người và điểm tương đồng so khớp vị trí, và tính toán điểm tương đồng kết quả tìm kiếm người tại vị trí xác định theo phương pháp cộng trọng số.

Mô tả vắn tắt hình vẽ

Hình 1 thể hiện phương pháp tìm người tại vị trí xác định từ tập dữ liệu video thông qua máy tính theo một phương án của sáng chế.

Mô tả chi tiết sáng chế

Theo Hình 1 thể hiện phương pháp tìm người tại vị trí xác định từ tập dữ liệu video bằng cách kết hợp kết quả tìm kiếm người và kết quả tìm kiếm vị trí thông qua máy tính theo một phương án của sáng chế. Trong đó, đầu vào là một số ảnh có khuôn mặt của người cần tìm và một số ảnh có vị trí (địa điểm) xác định mà người cần tìm xuất hiện, với vị trí xác định và duy nhất. Ví dụ với vị trí là quán Cafe Highland tại Vincom Megamall Thảo Điền, Quận 2, Thành phố Hồ Chí Minh, Việt Nam thì đây là vị trí xác định và duy nhất.

Phương pháp này bao gồm các bước sau:

Bước 101: xác định điểm tương đồng so khớp người giữa các ảnh mẫu trong truy vấn tìm kiếm và các ảnh trong mỗi video ngắn.

Trong bước xác định điểm tương đồng so khớp người, khuôn mặt sẽ được sử dụng như là thông tin chính để xác định đối tượng người quan tâm. Phương pháp sử dụng các mô đun cơ bản, thường thấy trong các phương pháp so khớp người hiện nay bao gồm: mô đun phát hiện khuôn mặt để phát hiện khuôn mặt từ ảnh, mô đun rút trích đặc trưng khuôn mặt và mô đun so khớp đặc trưng dựa trên độ đo tương đồng cosin (cosine); mô đun tăng cường độ phân giải ảnh khuôn mặt (face super-resolution) nhằm tăng cường chất lượng ảnh mẫu, và mô đun phân lớp chất lượng khuôn mặt (good-bad face classifier) nhằm loại bỏ các mặt xấu ảnh hưởng đến kết quả tìm kiếm.

Với đầu vào là một ảnh mẫu, cụ thể là ảnh của đối tượng người mà ta mong muốn tìm kiếm và tập dữ liệu truy vấn các video đã được phân thành các video ngắn (còn gọi là shot), phương pháp trả về danh sách xếp hạng kết quả tìm kiếm người, cụ thể là bảng danh sách kết quả các điểm tương đồng so khớp người được sắp xếp theo thứ tự giảm dần; vị trí càng ở đầu trong bảng danh sách là kết quả có điểm tương đồng càng cao so với ảnh mẫu.

Bước 102: xác định điểm tương đồng so khớp vị trí giữa các ảnh mẫu trong truy vấn tìm kiếm và các ảnh trong mỗi video ngắn.

Phương pháp tìm kiếm vị trí bao gồm các mô đun cơ bản được sử dụng trong các phương pháp thông thường so khớp vị trí như mô đun rút trích đặc trưng vị trí và mô đun so khớp đặc trưng dựa trên độ đo tương đồng cosin (cosine). Với đầu vào là một ảnh mẫu, cụ thể là ảnh chứa vị trí mà tìm kiếm và tập dữ liệu truy vấn video đã được phân thành các video ngắn (còn gọi là shot), phương pháp trả về danh sách xếp hạng kết quả tìm kiếm vị trí, cụ thể là bảng danh sách kết quả các điểm tương đồng so khớp vị trí được sắp xếp theo thứ tự giảm dần; vị trí càng ở đầu trong bảng danh sách là kết quả có điểm tương đồng càng cao so với ảnh mẫu.

Bước 103: xác định trọng số thông qua mô đun máy học dựa trên dữ liệu huấn luyện và đặc trưng điểm tương đồng so khớp người và đặc trưng điểm tương đồng so khớp vị trí. Sau khi thực hiện bước 101 và bước 102, phương pháp có hai danh sách điểm tương đồng so khớp người và điểm tương đồng so khớp vị trí. Trong sáng chế này, phương pháp đã huấn luyện mô đun máy học trên bộ dữ liệu huấn luyện để từ đó

mô đun có khả năng tự động lựa chọn trọng số phù hợp cho điểm tương đồng so khớp người và điểm tương đồng so khớp vị trí, nhằm chuẩn bị cho bước kết hợp và tính toán kết quả điểm tương đồng người tại vị trí xác định.

Cụ thể, với đầu vào là bộ dữ liệu huấn luyện, và đặc trưng dùng để huấn luyện thì mô đun máy học, cụ thể phương pháp sử dụng là mô đun máy học Learn-to-rank, sẽ tự động học và xác định trọng số cho điểm tương đồng so khớp người và điểm tương đồng so khớp vị trí của truy vấn tìm kiếm mới đưa vào, dựa trên bộ dữ liệu huấn luyện và các đặc trưng so khớp người và đặc trưng so khớp vị trí.

Trước tiên, huấn luyện mô đun máy học learn-to-rank trên bộ dữ liệu huấn luyện, cụ thể là danh sách kết quả tìm kiếm được xếp hạng theo thứ tự ưu tiên từ cao đến thấp của một số truy vấn tìm kiếm đã thực hiện trước và tiến hành gán nhãn cho các kết quả tìm kiếm này. Cụ thể phương pháp xây dựng bộ dữ liệu huấn luyện từ 10 danh sách kết quả tìm kiếm của 10 truy vấn tìm kiếm đầu tiên (9219-9228) trong số 30 truy vấn tìm kiếm của cuộc thi TRECVID 2020 (TREC Video Retrieval Evaluation). Bộ dữ liệu mẫu huấn luyện càng lớn thì mô đun máy học càng đưa ra kết quả trọng số càng chính xác.

Về loại đặc trưng đầu vào cho việc huấn luyện bằng mô đun máy học, phương pháp sử dụng đặc trưng điểm tương đồng so khớp người và đặc trưng điểm tương đồng so khớp vị trí. Cụ thể, phương pháp sử dụng đặc trưng điểm tương đồng so khớp người được tính theo các công thức sau:

$$per_sim_1(query, shot_i) = \frac{1}{N} \sum_{k=1}^N \max_{l=1..M} (\cosine(feats_k^{query}, feats_l^{shot_i}))$$

$$per_sim_2(query, shot_i) = \max_{k=1..N} (\max_{l=1..M} (\cosine(feats_k^{query}, feats_l^{shot_i})))$$

$$per_sim_3(query, shot_i) = \min_{k=1..N} (\max_{l=1..M} (\cosine(feats_k^{query}, feats_l^{shot_i})))$$

Trong đó: per_sim_1 : giá trị trung bình đặc trưng điểm tương đồng so khớp người khi so khớp ảnh mẫu trong truy vấn tìm kiếm (query) và ảnh trong video ngắn thứ i của tập dữ liệu video truy vấn;

per_sim_2 : giá trị lớn nhất (max) đặc trưng điểm tương đồng so khớp người khi so khớp ảnh mẫu trong truy vấn tìm kiếm (query) và ảnh trong video ngắn thứ i của tập dữ liệu video truy vấn;

per_sim_3 : giá trị nhỏ nhất (min) đặc trưng điểm tương đồng so khớp người khi so khớp ảnh mẫu trong truy vấn tìm kiếm (query) và ảnh trong video ngắn thứ i của tập dữ liệu video truy vấn;

N : là số lượng khuôn mặt mẫu (khuôn mặt của đối tượng cần tìm kiếm) trong câu truy vấn; do trong câu truy vấn tìm kiếm có thể có nhiều khuôn mặt mẫu để mô tả cùng một đối tượng, nên mỗi đối tượng có nhiều khuôn mặt mẫu;

M : là số lượng khuôn mặt được phát hiện từ các ảnh (frames) trích xuất từ mỗi video ngắn thứ i trong tập dữ liệu video truy vấn; do trong mỗi video ngắn có thể có nhiều ảnh, mỗi ảnh có thể có nhiều khuôn mặt, nên M chính là số lượng khuôn mặt được phát hiện từ nhiều ảnh (frames) trích xuất từ mỗi video ngắn có trong tập dữ liệu video truy vấn;

$shot_i$: là video ngắn thứ i trong tập dữ liệu video truy vấn;

$feat_k^{query}$: đặc trưng của khuôn mặt thứ k trong truy vấn tìm kiếm;

$feat_l^{shot_i}$: đặc trưng của khuôn mặt thứ l của video ngắn thứ i .

Đầu tiên phương pháp tính độ tương đồng (similarity score) thông qua hàm cosin (cosine) giữa một khuôn mặt mẫu (khuôn mặt của đối tượng muốn tìm kiếm) thứ k trong truy vấn tìm kiếm (query) với một khuôn mặt thứ l được phát hiện từ ảnh (frames) trích xuất từ video ngắn thứ i trong tập dữ liệu video truy vấn; và tương tự phương pháp tính độ tương đồng cho khuôn mặt mẫu thứ k nêu trên với tất cả M khuôn mặt được phát hiện từ các ảnh (frames) trích xuất từ video ngắn thứ i có trong tập dữ liệu video truy vấn; kết quả của bước này là một tập hợp các giá trị độ tương đồng giữa một khuôn mặt mẫu trong truy vấn tìm kiếm (query) với các khuôn mặt M được phát hiện từ ảnh (frames) trích xuất từ video ngắn thứ i trong tập dữ liệu video truy vấn; tiếp theo, phương pháp tính giá trị lớn nhất trên tập hợp nêu trên và điểm số độ tương đồng lớn nhất được giữ lại để thực hiện cho các bước tính toán sau;

Tiếp theo, trong một truy vấn tìm kiếm có thể có nhiều khuôn mặt mẫu N trong ảnh để minh họa cho một đối tượng người, phương pháp tiếp tục tính toán lại nhiều điểm số độ tương đồng lớn nhất ứng với từng khuôn mặt mẫu N và tạo thành một tập hợp điểm số độ tương đồng lớn nhất cho nhiều khuôn mặt mẫu N. Các giá trị per_sim_1 , per_sim_2 , per_sim_3 chính là các kết quả lần lượt khi lấy giá trị trung bình, lấy giá trị lớn nhất và lấy giá trị nhỏ nhất của tập hợp này. Kết quả thu được là ba giá trị điểm số độ tương đồng trung bình, lớn nhất và nhỏ nhất của đặc trưng điểm tương đồng so khớp người.

Phương pháp sử dụng đặc trưng điểm tương đồng so khớp vị trí được tính theo các công thức sau:

$$loc_sim_1(query, shot_i) = \max_{k=1..N}(\max_{l=1..M}(\cosine(feats_k^{query}, feats_l^{shot_i})))$$

$$loc_sim_2(query, shot_i) = \min_{k=1..N}(\min_{l=1..M}(\cosine(feats_k^{query}, feats_l^{shot_i})))$$

$$loc_sim_3(query, shot_i) = \frac{1}{N*M} \sum_{k=1}^N \sum_{l=1}^M (\cosine(feats_k^{query}, feats_l^{shot_i}))$$

Trong đó : loc_sim_1 : giá trị lớn nhất đặc trưng điểm tương đồng so khớp vị trí khi so khớp ảnh mẫu trong truy vấn tìm kiếm (query) và ảnh trong video ngắn thứ i của tập dữ liệu video truy vấn;

loc_sim_2 : giá trị nhỏ nhất đặc trưng điểm tương đồng so khớp vị trí khi so khớp ảnh mẫu trong truy vấn tìm kiếm (query) và ảnh trong video ngắn thứ i của tập dữ liệu video truy vấn;

loc_sim_3 : giá trị trung bình đặc trưng điểm tương đồng so khớp vị trí khi so khớp ảnh mẫu trong truy vấn tìm kiếm (query) và ảnh trong video ngắn thứ i của tập dữ liệu video truy vấn;

N: là số lượng ảnh mẫu của vị trí truy vấn; do trong truy vấn có thể có nhiều ảnh mẫu (frames) vị trí để minh họa cho cùng một vị trí nên mỗi vị trí có nhiều ảnh mẫu;

M: là số lượng ảnh (frames) trích xuất từ mỗi video ngắn thứ i trong tập dữ liệu video truy vấn; do trong mỗi video ngắn có thể có nhiều ảnh, nên sẽ có nhiều ảnh (frames) M trích xuất từ mỗi video ngắn có trong tập dữ liệu video truy vấn;

$shot_i$: là video ngắn thứ i trong dữ liệu video truy vấn;

$feat_k^{query}$: đặc trưng của vị trí thứ k trong truy vấn tìm kiếm;

$feat_l^{shot_i}$: đặc trưng của vị trí thứ l của video ngắn thứ i.

Đầu tiên phương pháp tính độ tương đồng (similarity score) thông qua hàm cosin (cosine) giữa một ảnh mẫu vị trí thứ k trong truy vấn với một ảnh vị trí thứ l trong video ngắn thứ i và tương tự phương pháp tính độ tương đồng cho ảnh mẫu vị trí nêu trên với tất cả M ảnh (frames) trích xuất từ video ngắn thứ i có trong tập dữ liệu video truy vấn; kết quả của bước này là một tập hợp các giá trị độ tương đồng giữa một ảnh mẫu vị trí trong truy vấn tìm kiếm (query) với số lượng ảnh M được trích xuất từ video ngắn thứ i trong tập dữ liệu video truy vấn; tiếp theo, phương pháp tính giá trị lớn nhất, giá trị nhỏ nhất và giá trị trung bình của tập hợp nêu trên và điểm số độ tương đồng lớn nhất, nhỏ nhất và trung bình được giữ lại để thực hiện cho các bước tính toán sau;

Tiếp theo, trong một truy vấn tìm kiếm (query) có thể có nhiều ảnh mẫu vị trí N để minh họa cho cùng một vị trí, phương pháp tiếp tục tính toán lại nhiều điểm số độ tương đồng lớn nhất, nhỏ nhất, trung bình ứng với nhiều ảnh mẫu vị trí N và tạo thành một tập hợp điểm số độ tương đồng lớn nhất, nhỏ nhất và trung bình cho tất cả ảnh mẫu vị trí N. Các giá trị loc_sim_1 , loc_sim_2 , loc_sim_3 chính là các kết quả lần lượt khi lấy giá trị lớn nhất, lấy giá trị nhỏ nhất và lấy giá trị trung bình của từng tập hợp tương ứng ở trên cho N ảnh mẫu vị trí. Kết quả thu được là ba giá trị điểm số độ tương đồng trung bình, lớn nhất và nhỏ nhất của đặc trưng điểm tương đồng so khớp vị trí.

Với các giá trị đặc trưng điểm tương đồng so khớp người và vị trí tìm được, phương pháp huấn luyện mô đun máy học trên bộ dữ liệu huấn luyện để tính ra các trọng số tương ứng cho từng giá trị đặc trưng điểm tương đồng so khớp người và vị trí, từ đó mô đun máy học có khả năng tự động lựa chọn trọng số phù hợp cho các truy vấn tìm kiếm mới.

Bước 104: gán tự động trọng số phù hợp vào điểm tương đồng so khớp người và điểm tương đồng so khớp vị trí, và tính toán điểm tương đồng kết quả tìm kiếm người tại vị trí xác định theo phương pháp cộng trọng số. Các trọng số được xác định ở bước 103 được gán tự động vào điểm tương đồng so khớp người và điểm tương đồng so

khớp vị trí đã xác định ở bước 101, 102 và tính toán điểm tương đồng kết quả tìm kiếm người tại vị trí xác định theo phương pháp cộng trọng số. Cụ thể công thức cộng trọng số như sau:

$$final_score_{shotx} = \theta_1 \times a + \theta_2 \times b + \dots + \theta_n \times a_n + \theta_m \times b_m$$

Trong đó $final_score_{shotx}$: điểm tương đồng về người tại vị trí xác định trong video ngắn thứ x;

$a_1, \dots, a_n$ điểm tương đồng so khớp người;

$b_1, \dots, b_m$ điểm tương đồng so khớp vị trí;

$\theta_1, \theta_2, \dots, \theta_n, \theta_m$ trọng số được xác định dựa trên mô đun máy học ở bước 103.

Quá trình trên sẽ được thực hiện tương tự cho toàn bộ các video ngắn có trong tập dữ liệu video truy vấn. Và kết quả tìm kiếm người tại vị trí xác định sẽ là những video ngắn kết quả có điểm số tương đồng về người tại vị trí xác định được sắp xếp theo danh sách giảm dần từ điểm cao đến thấp.

YÊU CẦU BẢO HỘ

1. Phương pháp tìm người tại vị trí xác định từ dữ liệu video thông qua máy tính bao gồm các bước sau:

xác định điểm tương đồng so khớp người giữa các ảnh mẫu trong truy vấn tìm kiếm và các ảnh trong mỗi video ngắn;

xác định điểm tương đồng so khớp vị trí giữa các ảnh mẫu trong truy vấn tìm kiếm và các ảnh trong mỗi video ngắn;

xác định trọng số thông qua mô đun máy học dựa trên dữ liệu huấn luyện và đặc trưng điểm tương đồng so khớp người và đặc trưng điểm tương đồng so khớp vị trí;

gán tự động trọng số phù hợp vào điểm tương đồng so khớp người và điểm tương đồng so khớp vị trí, và tính toán điểm tương đồng kết quả tìm kiếm người tại vị trí xác định theo phương pháp cộng trọng số.

2. Phương pháp tìm người tại vị trí xác định từ tập dữ liệu video theo điểm 1, trong đó đặc trưng điểm tương đồng so khớp người được tính theo các công thức sau:

$$per_sim_1(query, shot_i) = \frac{1}{N} \sum_{k=1}^N \max_{l=1..M} (cosine(feats_k^{query}, feats_l^{shot_i}))$$

$$per_sim_2(query, shot_i) = \max_{k=1..N} (\max_{l=1..M} (cosine(feats_k^{query}, feats_l^{shot_i})))$$

$$per_sim_3(query, shot_i) = \min_{k=1..N} (\max_{l=1..M} (cosine(feats_k^{query}, feats_l^{shot_i})))$$

trong đó :

per_sim_1 : giá trị trung bình đặc trưng điểm tương đồng so khớp người khi so khớp ảnh mẫu trong truy vấn tìm kiếm (query) và ảnh trong video ngắn thứ i của tập dữ liệu video truy vấn;

per_sim_2 : giá trị lớn nhất (max) đặc trưng điểm tương đồng so khớp người khi so khớp ảnh mẫu trong truy vấn tìm kiếm (query) và ảnh trong video ngắn thứ i của tập dữ liệu video truy vấn;

per_sim_3 : giá trị nhỏ nhất (min) đặc trưng điểm tương đồng so khớp người khi so khớp ảnh mẫu trong truy vấn tìm kiếm (query) và ảnh trong video ngắn thứ i của tập dữ liệu video truy vấn;

N : là số lượng khuôn mặt mẫu (khuôn mặt của đối tượng cần tìm kiếm) trong câu truy vấn;

M : là số lượng khuôn mặt được phát hiện từ các ảnh (frames) trích xuất từ mỗi video ngắn thứ i trong tập dữ liệu video truy vấn;

$shot_i$: là video ngắn thứ i trong dữ liệu video truy vấn;

$feat_k^{query}$: đặc trưng của khuôn mặt thứ k trong truy vấn tìm kiếm;

$feat_l^{shot_i}$: đặc trưng của khuôn mặt thứ l của video ngắn thứ i .

3. Phương pháp tìm người tại vị trí xác định theo điểm 1, trong đó đặc trưng điểm tương đồng so khớp vị trí được tính theo công thức sau:

$$loc_sim_1(query, shot_i) = \max_{k=1..N}(\max_{l=1..M}(\cosine(feats_k^{query}, feat_l^{shot_i})))$$

$$loc_sim_2(query, shot_i) = \min_{k=1..N}(\min_{l=1..M}(\cosine(feats_k^{query}, feat_l^{shot_i})))$$

$$loc_sim_3(query, shot_i) = \frac{1}{N*M} \sum_{k=1}^N \sum_{l=1}^M (\cosine(feats_k^{query}, feat_l^{shot_i}))$$

trong đó:

loc_sim_1 : giá trị lớn nhất đặc trưng điểm tương đồng so khớp vị trí khi so khớp ảnh mẫu trong truy vấn tìm kiếm (query) và ảnh trong video ngắn thứ i của tập dữ liệu video truy vấn;

loc_sim_2 : giá trị nhỏ nhất đặc trưng điểm tương đồng so khớp vị trí khi so khớp ảnh mẫu trong truy vấn tìm kiếm (query) và ảnh trong video ngắn thứ i của tập dữ liệu video truy vấn;

loc_sim_3 : giá trị trung bình đặc trưng điểm tương đồng so khớp vị trí khi so khớp ảnh mẫu trong truy vấn tìm kiếm (query) và ảnh trong video ngắn thứ i của tập dữ liệu video truy vấn;

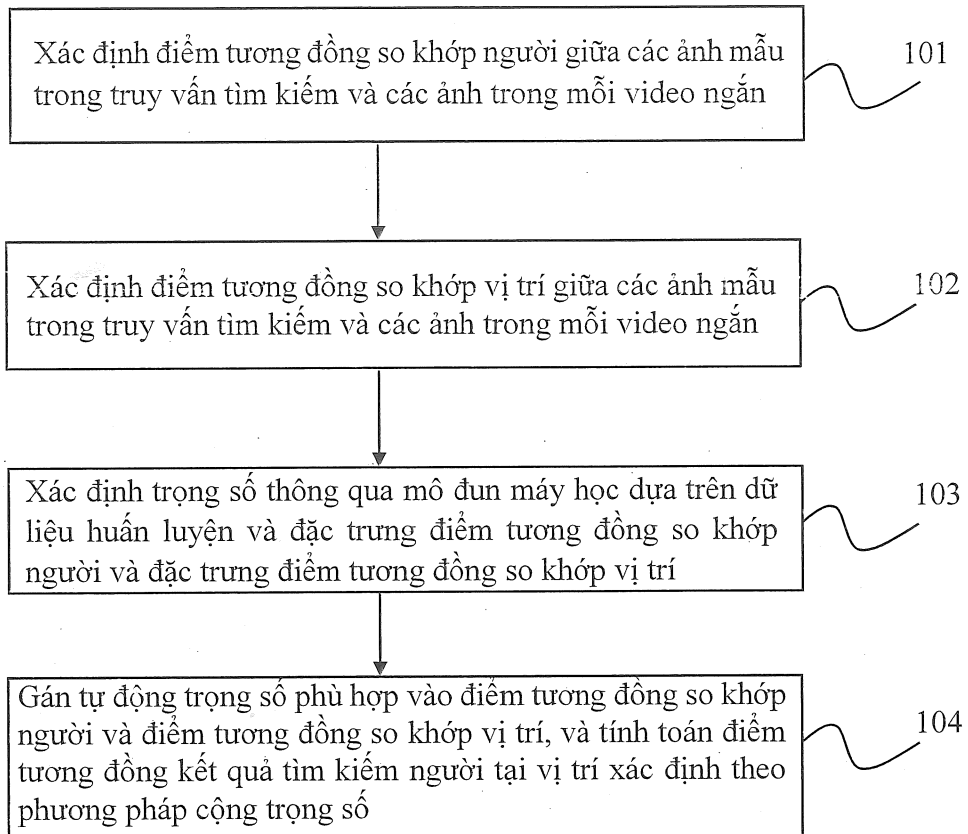
N : là số lượng ảnh mẫu của vị trí truy vấn;

M : là số lượng ảnh (frames) trích xuất từ mỗi video ngắn thứ i trong tập dữ liệu video truy vấn;

shot_i : là video ngắn thứ i trong dữ liệu video truy vấn;

$\text{feat}_k^{\text{query}}$: đặc trưng của vị trí thứ k trong truy vấn tìm kiếm;

$\text{feat}_l^{\text{shot}_i}$: đặc trưng của vị trí thứ l của video ngắn thứ i .



Hình 1