



(12) **BẢN MÔ TẢ GIẢI PHÁP HỮU ÍCH THUỘC BẰNG ĐỘC QUYỀN
GIẢI PHÁP HỮU ÍCH**

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

(11)



2-0003609

(51) **G06T 15/00; G06T 17/00**
2022.01

(13) **Y**

(21) 2-2023-00197

(22) 17/04/2023

(45) 27/05/2024 434

(43) 26/06/2023 423

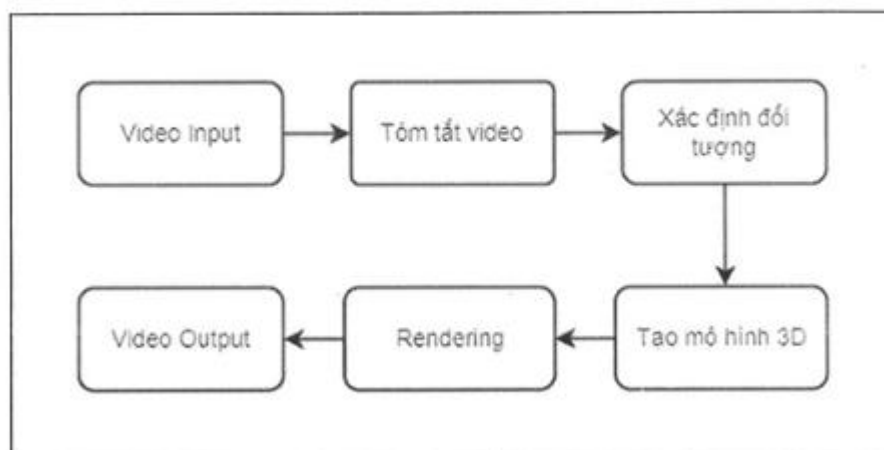
(73) **CÔNG TY CỔ PHẦN GIẢI PHÁP CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN
THÔNG MQ (VN)**

Tầng 4, số 119-121 phố Đại Từ, phường Đại Kim, quận Hoàng Mai, Hà Nội

(72) **LƯƠNG HỒNG QUÂN (VN).**

(54) **PHƯƠNG PHÁP TẠO RA BẢN TÓM TẮT VIDEO SỬ DỤNG THIẾT BỊ TÍNH TOÁN**

(57) Giải pháp hữu ích đề cập đến Phương pháp tạo ra bản tóm tắt video sử dụng thiết bị tính toán thu được các video được ghi lại từ nhiều camera khác nhau, mỗi video bao gồm các khung hình chứa các đối tượng; tóm tắt mỗi video từ đoạn video dài để tạo ra đoạn video ngắn hơn; xác định đối tượng để tìm ra các đối tượng cụ thể xuất hiện trong đoạn video ngắn hơn; tạo mô hình ba chiều (3D) của đối tượng đã gắn nhãn; sử dụng mô hình 3D của đối tượng để xác định đối tượng trong các khung ảnh riêng lẻ của chuỗi video; gắn nhãn đối tượng trong mỗi khung ảnh riêng lẻ; huấn luyện mô hình nhận dạng đối tượng sử dụng dữ liệu là tập các ảnh đã gắn nhãn; kết xuất mô hình 3D của đối tượng để tạo ra bản tóm tắt video chứa ảnh đối tượng được gắn nhãn và sử dụng mô hình nhận dạng đối tượng để xác định đối tượng trong bản tóm tắt video. Mục đích của quá trình tóm lược video là giúp các cơ quan quản lý có thể giám sát cũng như trích xuất đối tượng một cách nhanh chóng và ngắn gọn, từ đó phục vụ cho các mục tiêu điều tra và khảo sát đối tượng.



Lĩnh vực kỹ thuật được đề cập

Giải pháp hữu ích đề cập đến lĩnh vực tóm lược nội dung chính của video, cụ thể là, đề cập đến phương pháp tạo ra bản tóm tắt video sử dụng thiết bị tính toán.

Tình trạng kỹ thuật của giải pháp hữu ích

Tóm lược video là quá trình tóm tắt nội dung chính của một video trong một khoảng thời gian ngắn, thông thường là từ vài giây đến vài phút. Mục đích của tóm lược video là giúp người xem có cái nhìn tổng quan về nội dung của video mà không cần phải xem toàn bộ video. Tóm lược video có thể được thực hiện bằng nhiều cách khác nhau, bao gồm viết chú thích, sử dụng hình ảnh hoặc trình chiếu đoạn clip ngắn từ video gốc. Tóm lược video cũng có thể được sử dụng như một công cụ tiếp thị để thu hút sự chú ý của khán giả và giới thiệu nội dung của video.

Có nhiều phương pháp để tóm lược video, bao gồm:

Sử dụng chú thích: Đây là phương pháp phổ biến nhất để tóm lược video. Người tạo nội dung sẽ viết những câu chú thích ngắn gọn để truyền tải thông điệp chính của video.

Sử dụng hình ảnh: Tóm lược video bằng hình ảnh là cách tạo ra một loạt các hình ảnh tóm tắt nội dung của video. Điều này sẽ giúp khán giả có được cái nhìn tổng quan về nội dung của video một cách nhanh chóng.

Sử dụng trình chiếu đoạn clip ngắn: Đây là phương pháp tạo ra một đoạn clip ngắn từ video gốc để truyền tải thông điệp chính của video.

Sử dụng phần mềm tóm tắt video: Có rất nhiều phần mềm tóm tắt video miễn phí hoặc trả phí trên thị trường để giúp người dùng tóm tắt nội dung của video một cách nhanh chóng và dễ dàng.

Sử dụng giọng nói off: Tóm lược video bằng giọng nói off là phương pháp dùng giọng nói để truyền tải thông điệp chính của video. Phương pháp này thường được sử dụng trong các video hướng dẫn hoặc đào tạo.

Tùy thuộc vào mục đích và tính chất của video, người tạo nội dung có thể lựa chọn một hoặc nhiều phương pháp tóm lược video để giúp khán giả hiểu được nội dung của video một cách nhanh chóng và hiệu quả.

Trong lĩnh vực tóm tắt video, có nhiều kỹ thuật và công nghệ được sử dụng nhằm giúp tự động hóa quá trình tóm tắt video và tăng cường hiệu quả của quá trình này. Một số lĩnh vực kỹ thuật chính được đề cập trong lĩnh vực tóm tắt video gồm:

Xử lý ngôn ngữ tự nhiên (Natural Language Processing - NLP): Kỹ thuật này được sử dụng để hiểu và phân tích nội dung của video, từ đó tạo ra các câu tóm tắt mô tả bối cảnh, nội dung chính và các thông tin quan trọng khác.

Học máy (Machine Learning): Kỹ thuật này được sử dụng để phân tích nội dung của video, nhận dạng các đối tượng, hình ảnh và âm thanh trong video, và tạo ra các đặc trưng cần thiết cho quá trình tóm tắt.

Trí tuệ nhân tạo (Artificial Intelligence - AI): Kỹ thuật này được sử dụng để tăng cường khả năng tương tác giữa hệ thống tóm tắt video và người dùng, đồng thời cải thiện chất lượng của quá trình tóm tắt.

Kỹ thuật tiền xử lý (Pre-processing): Kỹ thuật này được sử dụng để tinh chỉnh và cải thiện nội dung video trước khi bắt đầu quá trình tóm tắt, bao gồm cắt ghép video, lọc nhiễu và tăng cường chất lượng hình ảnh và âm thanh.

Kỹ thuật tối ưu hóa (Optimization): Kỹ thuật này được sử dụng để cải thiện hiệu suất của hệ thống tóm tắt video, tăng tốc quá trình tóm tắt và cải thiện độ chính xác của kết quả tóm tắt.

Tuy nhiên, vẫn còn nhiều thách thức và hạn chế khi áp dụng các kỹ thuật này vào thực tiễn, bao gồm độ phức tạp của quá trình tóm tắt, khả năng tương tác với người dùng và độ chính xác của kết quả tóm tắt.

Trong các kỹ thuật đã đề cập tại phần trước, để xử lý video các chuyên gia thường dùng các thuật toán để nén và xử lý các vấn đề tại video được nén lại. Việc xử lý video nén là công việc dễ nảy sinh sai sót nếu như không có một thuật toán nén video hiệu quả. Bên cạnh đó, việc theo dõi video là công việc cần dành nhiều thời gian vì thời gian video là không thay đổi sau quá trình nén. Nguyên nhân là do video chỉ giảm dung lượng của video trong suốt quá trình nén.

Mặt khác để phục vụ cho mục đích xem lại các camera, mục tiêu chính là xem lại các nội dung chính từ video nhằm phục vụ cho một mục đích cụ thể như điều tra, giám sát hành vi của đối tượng. Chính vì nguyên nhân này, vậy nên nhiệm vụ tóm tắt video được thực hiện bằng phương pháp trích xuất các ảnh theo yêu cầu từ video. Ngoài ra, việc lưu lại hành vi và các đặc tính của đối tượng là cần thiết cho bài toán tra cứu đối tượng dựa trên mô tả.

Bản chất kỹ thuật của giải pháp hữu ích

Giải pháp hữu ích đề xuất phương pháp tạo ra bản tóm tắt video sử dụng dữ liệu được thu thập từ video giám sát hằng ngày để tóm lược lại những hoạt động, những sự kiện xảy ra hằng ngày có trong video đó. Thực chất, video giám sát được phát theo thời gian thực là quá trình diễn ra một chuỗi các sự kiện một cách tuần tự. Tuy nhiên, quá trình quan sát hết một video hoàn chỉnh dựa theo thời gian là hết sức tốn thời gian và công sức. Hệ thống tóm lược video theo giải pháp hữu ích giúp giảm dung lượng của video mà vẫn nắm bắt được những nội dung chính có trong video đó. Mục đích của quá trình tóm lược video là giúp các cơ quan quản lý có thể giám sát cũng như trích xuất đối tượng một cách nhanh chóng và ngắn gọn, từ đó phục vụ cho các mục tiêu điều tra và khảo sát đối tượng.

Phương pháp tạo ra bản tóm tắt video sử dụng thiết bị tính toán, phương pháp bao gồm các bước:

thu được các video được ghi lại từ nhiều camera khác nhau, mỗi video bao gồm các khung hình chứa các đối tượng;

tóm tắt mỗi video từ đoạn video dài để tạo ra đoạn video ngắn hơn;
 xác định đối tượng để tìm ra các đối tượng cụ thể xuất hiện trong đoạn video ngắn hơn;
 tạo mô hình ba chiều (3D) của đối tượng đã gán nhãn;
 sử dụng mô hình 3D của đối tượng để xác định đối tượng trong các khung ảnh riêng lẻ của chuỗi video;
 gán nhãn đối tượng trong mỗi khung ảnh riêng lẻ;
 huấn luyện mô hình nhận dạng đối tượng sử dụng dữ liệu là tập các ảnh đã gán nhãn;
 kết xuất mô hình 3D của đối tượng để tạo ra bản tóm tắt video chứa ảnh đối tượng được gán nhãn và sử dụng mô hình nhận dạng đối tượng để xác định đối tượng trong bản tóm tắt video;
 trong đó,
 bước tóm tắt mỗi video từ đoạn video dài để tạo ra đoạn video ngắn hơn bao gồm các bước:
 phân tích video: đoạn video sẽ được phân tích để xác định các thông tin quan trọng, bao gồm các cảnh, sự kiện và đối tượng trong video;
 chọn thông tin quan trọng: Sau khi phân tích video, các thông tin quan trọng sẽ được chọn để tạo ra phiên bản tóm tắt; việc này có thể bao gồm chọn các cảnh, đối tượng hoặc hành động quan trọng nhất trong video;
 tạo đoạn video ngắn hơn: các thông tin được chọn sẽ được sắp xếp và tạo thành đoạn video ngắn hơn;
 bước xác định đối tượng để tìm ra các đối tượng cụ thể xuất hiện trong đoạn video ngắn hơn bao gồm:
 chuẩn bị dữ liệu: các đoạn video được chia thành các khung hình (frames) và xác định các thông tin như độ phân giải, tốc độ khung hình, và định dạng video;
 xác định đối tượng trong khung hình: bước này sẽ tìm kiếm các đối tượng trong từng khung hình và trích xuất các đặc trưng của đối tượng;
 xác định đối tượng qua các khung hình: bước này sẽ theo dõi các đối tượng qua các khung hình khác nhau để tìm ra các đối tượng tương ứng với nhau trong các khung hình khác nhau;
 phân loại và gán nhãn các đối tượng tương ứng;
 bước tạo mô hình ba chiều (3D) của đối tượng với nhãn tương ứng bao gồm:
 xây dựng mô hình 3D: các đối tượng được tái tạo lại dưới dạng các mô hình 3D, bao gồm các thông tin về hình dạng, kích thước, màu sắc và vị trí của chúng;
 tích hợp dữ liệu của đối tượng: dữ liệu được tích hợp bao gồm các thông tin về vị trí, hướng, tốc độ và hành động của các đối tượng trong đoạn video;
 lắp ráp mô hình 3D để tạo thành một mô hình 3D toàn cảnh của đoạn video;
 tối ưu hóa mô hình 3D;
 bước kết xuất mô hình 3D của đối tượng để tạo ra bản tóm tắt video bao gồm:
 thiết lập các thông số của mô hình: thiết lập các thông số cho mô hình 3D, bao gồm ánh sáng, vật liệu, cảnh quan và camera;

tạo ra các khung hình ảnh của mô hình 3D của đối tượng đã được gán nhãn với các góc khác nhau;

tạo ra các bản tóm tắt video các đối tượng trong mô hình 3D với nhãn được gán.

Mô tả vắn tắt các hình vẽ

Hình 1 là hình vẽ minh họa tổng quan phương pháp trích xuất và lưu trữ đối tượng;

Hình 2 là hình vẽ minh họa tổng quan phương pháp tạo ra bản tóm tắt video theo một phương án của giải pháp hữu ích;

Hình 3 là hình vẽ minh họa mô hình hệ thống theo một phương án của giải pháp hữu ích;

Hình 4 là hình vẽ minh họa mô hình sau khi được huấn luyện;

Hình 5 là hình vẽ minh họa sự phát hiện xe cộ tại một quốc lộ tại Việt Nam;

Hình 7 là hình vẽ minh họa sự phát hiện xe cộ tại một quốc lộ tại nước ngoài;

Hình 8 là hình vẽ minh họa nội dung video được trích xuất với đặc tính xe màu đỏ.

Mô tả chi tiết giải pháp hữu ích

Các mô tả dưới đây về hệ thống tóm tắt video và phương pháp lưu đặc tính của vật thể trong video sẽ được giải thích theo các hình vẽ đính kèm, từ đó giải pháp hữu ích sẽ được hiểu một cách rõ ràng và rành mạch. Các mô tả dưới đây có đưa ra những ví dụ khi thực hiện giải pháp hữu ích này và không giới hạn phạm vi kỹ thuật của giải pháp hữu ích.

Trong trường hợp giải pháp hữu ích thuộc phạm vi của lĩnh vực camera giao thông xe ô tô và con người là hai đối tượng phổ biến nhất trong video. Chính vì vậy để hệ thống được nhỏ gọn, giải pháp đi sâu vào bài toán phát hiện đối tượng cũng như trích xuất đặc tính của đối tượng xe và con người, sau đó thực hiện lưu kết quả và hiển thị khi cần thiết. Các trường hợp khác tiến hành tương tự.

Tham chiếu tới hình 1, giải pháp hữu ích đề xuất phương pháp trích xuất và lưu trữ đối tượng, Các bước thực hiện cho phương pháp trích xuất và lưu trữ đối tượng:

Phân tích và xác định đối tượng trong video: Đầu tiên, phải sử dụng các phương pháp nhận dạng đối tượng, phân đoạn hình ảnh hoặc các thuật toán tương tự để phân tích và xác định các đối tượng trong đoạn video đầu vào. Điều này có thể được thực hiện bằng cách sử dụng các phần mềm hoặc thư viện có sẵn như OpenCV, TensorFlow, hoặc PyTorch.

Trích xuất các đối tượng từ video: Sau khi xác định các đối tượng, tiếp theo là trích xuất chúng từ đoạn video ban đầu. Quá trình này có thể được thực hiện bằng cách sử dụng các phương pháp chuyên dụng như phân đoạn hình ảnh, phát hiện cạnh hoặc phát hiện vùng tương phản.

Xác định tính năng của các đối tượng: Các tính năng của các đối tượng, bao gồm thông tin về hình dạng, màu sắc và động tác, được xác định bằng cách sử dụng các thuật toán phân tích hình ảnh hoặc các phương pháp tương tự. Các tính năng này cũng có thể được lưu trữ dưới dạng các thuộc tính của đối tượng.

Lưu trữ các đối tượng: Các đối tượng trích xuất được lưu trữ dưới dạng các tệp định dạng phù hợp như OBJ, FBX hoặc STL. Các tệp này được lưu trữ trên đĩa cứng hoặc trong cơ sở dữ liệu để có thể truy xuất và sử dụng lại sau này.

Cập nhật các đối tượng: Trong quá trình sử dụng, các đối tượng có thể cần được cập nhật để đảm bảo tính toàn vẹn của dữ liệu. Các cập nhật này bao gồm việc thêm mới các đối tượng, cập nhật tính năng của các đối tượng hiện có và xóa các đối tượng không cần thiết.

Phân loại các đối tượng: Các đối tượng lưu trữ được phân loại và gán nhãn để dễ dàng tìm kiếm và truy cập sau này. Điều này giúp cho việc sử dụng các đối tượng trở nên dễ dàng hơn.

Phân tích và xác định đối tượng

Bước phân tích và xác định đối tượng trong video là bước quan trọng trong quá trình trích xuất thông tin và tóm tắt video. Bước này tập trung vào việc phân tích và xác định các đối tượng xuất hiện trong video và mối quan hệ giữa chúng.

Các bước chi tiết của phương pháp phân tích và xác định đối tượng trong video là:

Tiền xử lý video: Video được tiền xử lý bằng cách chuyển đổi định dạng, cắt bỏ phần không cần thiết, thay đổi độ phân giải và thời lượng để tối ưu hóa quá trình xử lý.

Trích xuất khung hình (frame extraction): Khung hình của video được trích xuất để tạo thành các ảnh tĩnh riêng lẻ cho mỗi khung hình. Các ảnh này được sử dụng để phân tích và xác định đối tượng.

Phát hiện đối tượng (object detection): Các đối tượng trong video được phát hiện bằng cách sử dụng các thuật toán phát hiện đối tượng như YOLO (You Only Look Once) hoặc Faster R-CNN (Region-based Convolutional Neural Network). Các thuật toán này sử dụng mạng nơ-ron để xác định đối tượng, vị trí và kích thước của chúng trong từng khung hình.

Phân loại đối tượng (object classification): Sau khi phát hiện được các đối tượng, chúng được phân loại thành các loại khác nhau. Các thuật toán phân loại đối tượng như AlexNet, VGG (Visual Geometry Group), hoặc ResNet (Residual Network) được sử dụng để đưa ra dự đoán về loại đối tượng và cung cấp các thông tin thêm về đối tượng như màu sắc, hình dạng, v.v.

Theo dõi đối tượng (object tracking): Để đảm bảo rằng cùng một đối tượng được theo dõi liên tục trong các khung hình, các thuật toán theo dõi đối tượng như SORT (Simple Online and Realtime Tracking) hoặc DeepSORT được sử dụng. Các thuật toán này sử dụng các phương pháp tính toán để theo dõi đối tượng và cập nhật thông tin về vị trí và hướng di chuyển của chúng.

Xác định mối quan hệ giữa các đối tượng (object relation): Các thuật toán xác định mối quan hệ giữa các đối tượng được sử dụng để phân tích

Trích xuất đối tượng

Bước trích xuất các đối tượng từ video sau bước phân tích và xác định đối tượng trong video là quá trình lấy ra các đối tượng đã được xác định từ video gốc để sử dụng cho mục đích tiếp theo trong quá trình xử lý video.

Các bước chính để trích xuất các đối tượng từ video bao gồm:

Chọn các khung hình quan trọng: Một số khung hình trong video có thể không chứa thông tin quan trọng về đối tượng, vì vậy chỉ chọn các khung hình quan trọng để tối ưu hóa việc trích xuất đối tượng và giảm thiểu tài nguyên tính toán. Các khung hình quan trọng có thể được xác định bằng các phương pháp như phát hiện sự thay đổi đột ngột trong video, phát hiện sự xuất hiện hoặc biến mất của đối tượng, hoặc các phương pháp khác tùy thuộc vào loại đối tượng cần trích xuất.

Phân đoạn đối tượng: Phân đoạn đối tượng là quá trình tách đối tượng khỏi nền bằng cách xác định vùng ảnh chứa đối tượng trong một khung hình. Có nhiều kỹ thuật phân đoạn đối tượng như phân đoạn đối tượng dựa trên đường biên (contour-based segmentation), phân đoạn đối tượng dựa trên màu sắc (color-based segmentation), phân đoạn đối tượng dựa trên phân loại vật thể (object classification-based segmentation) và phân đoạn đối tượng dựa trên mạng neural (neural network-based segmentation).

Loại bỏ nhiễu và cắt ảnh: Sau khi đối tượng được phân đoạn, quá trình loại bỏ nhiễu và cắt ảnh được thực hiện để loại bỏ các đối tượng nhỏ hơn hoặc các phần nhiễu nhỏ khác trong ảnh. Điều này giúp giảm thiểu việc lưu trữ các tệp không cần thiết và giúp tăng tốc độ xử lý.

Xác định tính năng của các đối tượng

Bước Xác định tính năng của các đối tượng là một bước quan trọng trong quá trình phân tích video. Bước này nhằm mục đích đưa ra các thông tin liên quan đến đặc điểm, tính chất và hành vi của các đối tượng có trong video.

Các kỹ thuật cơ bản để xác định tính năng của các đối tượng gồm:

Phân tích hình dạng: Ứng dụng các phương pháp xử lý ảnh để phân tích hình dạng của các đối tượng, bao gồm kích thước, hình dạng, tỉ lệ khung hình, độ sắc nét, v.v. Thông tin này có thể được sử dụng để phát hiện các đối tượng không bình thường hoặc đối tượng có hành vi bất thường.

Nhận dạng: Sử dụng các thuật toán nhận dạng đối tượng để xác định danh tính của các đối tượng trong video. Các kỹ thuật nhận dạng đối tượng bao gồm nhận dạng khuôn mặt, nhận dạng biển số xe, nhận dạng ký tự, v.v.

Phân tích hành vi: Sử dụng các thuật toán phân tích hành vi để xác định hành vi của các đối tượng trong video. Các kỹ thuật phân tích hành vi bao gồm theo dõi chuyển động, phát hiện hành vi bất thường, phát hiện cử chỉ và phân tích ngôn ngữ cơ thể.

Phân tích âm thanh: Phân tích các tín hiệu âm thanh trong video để xác định các thông tin âm thanh liên quan đến các đối tượng trong video. Các kỹ thuật phân tích âm thanh bao gồm phát hiện tiếng nói, phân tích âm thanh nền và phân tích âm thanh môi trường.

Tổng quan, bước Xác định tính năng của các đối tượng sau bước Trích xuất các đối tượng từ video là một bước rất quan trọng để phân tích video và đưa ra các quyết định cần thiết. Các kỹ thuật xác định tính năng có thể được sử dụng độc lập hoặc kết hợp với nhau để đạt được hiệu quả tốt nhất trong phân tích video.

Lưu trữ dữ liệu

Bước lưu trữ các đối tượng sau bước xác định tính năng của các đối tượng là quá trình lưu trữ và quản lý thông tin liên quan đến các đối tượng đã được phân tích từ video. Bước này rất quan trọng để đảm bảo tính toàn vẹn và sẵn sàng truy xuất các đối tượng đã được phân tích sau này.

Một số kỹ thuật cơ bản để lưu trữ các đối tượng sau bước xác định tính năng của các đối tượng gồm:

Lưu trữ trên đĩa cứng: Các đối tượng và thông tin về tính năng của chúng có thể được lưu trữ trên đĩa cứng. Điều này có thể được thực hiện bằng cách lưu trữ các tệp tin video và các tệp tin liên quan khác trong một thư mục trên đĩa cứng hoặc sử dụng cơ sở dữ liệu quan hệ để lưu trữ thông tin về các đối tượng và tính năng của chúng.

Lưu trữ trên đám mây (Cloud): Các đối tượng và thông tin về tính năng của chúng có thể được lưu trữ trên đám mây. Điều này cho phép truy cập từ xa đến các tệp tin và thông tin và giúp cho quản lý tài nguyên dữ liệu dễ dàng hơn. Bên cạnh đó, việc lưu trữ trên đám mây cũng giúp giảm chi phí lưu trữ và bảo trì.

Lưu trữ trong cơ sở dữ liệu phi quan hệ: Các đối tượng và thông tin về tính năng của chúng có thể được lưu trữ trong các cơ sở dữ liệu phi quan hệ như MongoDB hoặc Cassandra. Các cơ sở dữ liệu phi quan hệ cho phép lưu trữ dữ liệu theo định dạng tài liệu (document) và cho phép truy xuất tài liệu dựa trên các trường tùy ý.

Lưu trữ bằng phần mềm quản lý dữ liệu: Các đối tượng và thông tin về tính năng của chúng có thể được lưu trữ bằng phần mềm quản lý dữ liệu như Elasticsearch hoặc Apache Solr. Các phần mềm này cho phép lưu trữ và truy xuất các tài liệu dựa trên các trường tùy ý và hỗ trợ các tính năng như tìm kiếm toàn văn (full-text search) và lọc dữ liệu.

Cập nhật đối tượng

Bước cập nhật đối tượng sau bước lưu trữ các đối tượng là một trong những bước quan trọng trong việc xử lý dữ liệu. Sau khi các đối tượng đã được lưu trữ vào cơ sở dữ liệu, các thông tin của chúng có thể được thay đổi và cần phải được cập nhật lại để đảm bảo tính toàn vẹn và chính xác của dữ liệu.

Một số kỹ thuật cập nhật đối tượng sau bước lưu trữ các đối tượng bao gồm:

Sử dụng ngôn ngữ truy vấn cơ sở dữ liệu: Kỹ thuật này sử dụng các câu lệnh SQL hoặc các ngôn ngữ truy vấn cơ sở dữ liệu khác để cập nhật thông tin của các đối tượng. Các câu lệnh này bao gồm các phép cập nhật, thêm mới, xóa và truy vấn dữ liệu.

Sử dụng hệ thống cache: Các hệ thống cache như Redis hay Memcached được sử dụng để cung cấp dữ liệu nhanh hơn. Khi các thông tin của các đối tượng bị thay đổi, các giá trị trong cache được cập nhật và được đồng bộ với cơ sở dữ liệu chính.

Sử dụng hệ thống message queue: Hệ thống message queue như RabbitMQ hay Apache Kafka được sử dụng để đưa ra các thông báo khi có sự thay đổi về thông tin của các đối tượng.

Khi có sự thay đổi xảy ra, các thông tin này sẽ được đưa vào message queue và được xử lý bởi các worker để cập nhật dữ liệu trong cơ sở dữ liệu.

Sử dụng hệ thống event sourcing: Hệ thống event sourcing như Apache Kafka hay AWS Kinesis được sử dụng để theo dõi các sự kiện về sự thay đổi của các đối tượng. Khi có sự thay đổi xảy ra, các sự kiện này được lưu trữ vào event log và sử dụng để cập nhật dữ liệu của các đối tượng.

Phân loại đối tượng

Bước phân loại đối tượng sau bước cập nhật đối tượng là một trong những bước quan trọng trong quá trình xử lý dữ liệu. Việc phân loại đối tượng giúp đưa ra những quyết định đúng đắn dựa trên dữ liệu và giúp cho quá trình quản lý dữ liệu dễ dàng hơn.

Một số kỹ thuật phân loại đối tượng sau bước cập nhật đối tượng bao gồm:

Sử dụng thuật toán học máy: Thuật toán học máy như Decision Tree, Random Forest, Support Vector Machine hay Neural Network được sử dụng để phân loại các đối tượng dựa trên các đặc trưng của chúng. Các đặc trưng này có thể được trích xuất từ các thuộc tính của đối tượng, hoặc từ các thông tin bổ sung như hình ảnh, video hay văn bản.

Sử dụng kỹ thuật khai phá dữ liệu: Kỹ thuật khai phá dữ liệu như Clustering, Association Rule hay Pattern Mining được sử dụng để phân loại các đối tượng dựa trên sự tương đồng giữa chúng. Các đối tượng được phân vào các nhóm tương đồng nhau dựa trên các thuộc tính của chúng.

Sử dụng ngôn ngữ lập trình: Ngôn ngữ lập trình như Python hay R được sử dụng để phân loại các đối tượng dựa trên các quy tắc logic được lập trình trước. Các đối tượng sẽ được xử lý và đưa ra quyết định dựa trên quy tắc được lập trình sẵn.

Sử dụng hệ thống hỗ trợ quyết định: Hệ thống hỗ trợ quyết định như Decision Support System hay Expert System được sử dụng để phân loại các đối tượng dựa trên các quy tắc logic và kiến thức chuyên môn. Các đối tượng sẽ được xử lý và đưa ra quyết định dựa trên các kiến thức được lập trình vào hệ thống.

Hình 2 là hình vẽ minh họa tổng quan phương pháp tạo ra bản tóm tắt video theo một phương án của giải pháp hữu ích.

Các bước để thực hiện phương pháp này bao gồm:

thu được các video được ghi lại từ nhiều camera khác nhau, mỗi video bao gồm các khung hình chứa các đối tượng;

tóm tắt mỗi video từ đoạn video dài để tạo ra đoạn video ngắn hơn;

xác định đối tượng để tìm ra các đối tượng cụ thể xuất hiện trong đoạn video ngắn hơn;

tạo mô hình ba chiều (3D) của đối tượng đã gán nhãn;

sử dụng mô hình 3D của đối tượng để xác định đối tượng trong các khung ảnh riêng lẻ của chuỗi video;

gán nhãn đối tượng trong mỗi khung ảnh riêng lẻ;

huấn luyện mô hình nhận dạng đối tượng sử dụng dữ liệu là tập các ảnh đã gán nhãn;

kết xuất mô hình 3D của đối tượng để tạo ra bản tóm tắt video chứa ảnh đối tượng được gán nhãn và sử dụng mô hình nhận dạng đối tượng để xác định đối tượng trong bản tóm tắt video; trong đó,

Tóm tắt video: Bước đầu tiên trong phương pháp này là trích xuất thông tin chính từ đoạn video nguồn bằng các kỹ thuật tóm tắt video. Điều này có thể bao gồm phân tích nội dung để xác định các cảnh quan trọng và đối tượng xuất hiện trong video, phát hiện đối tượng bằng cách sử dụng các thuật toán như phát hiện đối tượng và theo dõi đối tượng, và phân loại đối tượng bằng cách sử dụng các mô hình học sâu để xác định loại đối tượng.

Xác định đối tượng: Sau khi có được thông tin từ bước trên, đối tượng muốn tạo hình ảnh hoặc video mới được chọn từ đoạn video nguồn. Đối tượng này có thể được chọn bằng tay hoặc thông qua các thuật toán phát hiện đối tượng và theo dõi đối tượng.

Tạo mô hình 3D: Tiếp theo, một mô hình 3D của đối tượng được tạo ra bằng cách sử dụng các phần mềm đồ họa máy tính. Để tạo mô hình này, các thông tin đối tượng được sử dụng để tạo ra các hình dạng và chi tiết của đối tượng trong môi trường 3D.

Kết xuất (Rendering): Cuối cùng, đối tượng được đưa vào môi trường 3D và quá trình rendering được thực hiện để tạo ra hình ảnh hoặc video mới. Trong quá trình này, các thông tin về ánh sáng, vật liệu và camera được sử dụng để tạo ra hình ảnh hoặc video mới với đối tượng được hiển thị trong một môi trường 3D.

Tóm tắt video

Bước tóm tắt video là quá trình trích xuất thông tin chính từ một đoạn video dài và phức tạp để tạo ra một phiên bản ngắn gọn hơn và dễ hiểu hơn. Quá trình này thường bao gồm các bước sau:

Phân tích video: Đầu tiên, đoạn video sẽ được phân tích để xác định các thông tin quan trọng, bao gồm các cảnh, sự kiện và đối tượng trong video. Các phương pháp phân tích video có thể bao gồm:

Phân tích cảnh: Xác định các cảnh quan trọng trong video, chẳng hạn như các cảnh diễn biến nhanh hoặc quan trọng đối với nội dung của video.

Phân tích đối tượng: Xác định các đối tượng xuất hiện trong video, phân loại chúng và xác định đối tượng nào quan trọng nhất đối với nội dung của video.

Phân tích hành động: Xác định các hành động và sự kiện quan trọng trong video, chẳng hạn như chuyển động của đối tượng hay các sự kiện quan trọng xảy ra trong video.

Chọn thông tin quan trọng: Sau khi phân tích video, các thông tin quan trọng sẽ được chọn để tạo ra phiên bản tóm tắt. Điều này có thể bao gồm chọn các cảnh, đối tượng hoặc hành động quan trọng nhất trong video.

Tạo phiên bản tóm tắt: Cuối cùng, các thông tin được chọn sẽ được sắp xếp và tạo thành phiên bản tóm tắt của video. Các phương pháp tạo phiên bản tóm tắt video bao gồm:

Cắt ghép video: Chọn các cảnh, đối tượng hoặc hành động quan trọng nhất và cắt ghép chúng lại thành một phiên bản tóm tắt ngắn gọn.

Tạo chú thích: Sử dụng các chú thích hoặc văn bản để giải thích nội dung của video một cách ngắn gọn và dễ hiểu.

Tổng hợp: Tổng hợp các thông tin quan trọng thành một phiên bản tóm tắt dựa trên một số tiêu chí nhất định.

Bước tóm tắt video là bước quan trọng trong việc xử lý và phân tích video, giúp giảm thiểu thời gian xem video và tăng tính tiện

Xác định đối tượng

Sau bước tóm tắt video, bước xác định đối tượng sẽ tiếp tục được thực hiện để tìm ra các đối tượng cụ thể xuất hiện trong đoạn video. Các bước thực hiện trong quá trình này có thể bao gồm:

Chuẩn bị dữ liệu: Đầu tiên, các đoạn video được chia thành các khung hình (frames) để thuận tiện cho việc xử lý. Ngoài ra, các thông tin khác như độ phân giải, tốc độ khung hình, và định dạng video cũng cần được xác định trước để đảm bảo tính nhất quán của dữ liệu.

Xác định đối tượng trong khung hình: Bước này sẽ tìm kiếm các đối tượng trong từng khung hình. Có thể sử dụng các phương pháp phát hiện đối tượng như Haar Cascades, HOG (Histogram of Oriented Gradients), hoặc các mô hình Deep Learning để phát hiện đối tượng. Một khi các đối tượng đã được phát hiện, các đặc trưng của chúng cũng cần được trích xuất để sử dụng cho các bước tiếp theo.

Xác định đối tượng qua các khung hình: Sau khi các đối tượng được xác định trong từng khung hình, các đối tượng này cần được theo dõi qua các khung hình khác nhau để tìm ra các đối tượng tương ứng với nhau trong các khung hình khác nhau. Các phương pháp theo dõi đối tượng (object tracking) như SORT (Simple Online and Real-time Tracking) hoặc DeepSORT (Deep Learning-based SORT) có thể được sử dụng để thực hiện việc này.

Phân loại đối tượng: Cuối cùng, các đối tượng cần được phân loại vào các lớp (class) tương ứng. Các mô hình Deep Learning như Convolutional Neural Networks (CNNs) hoặc Recurrent Neural Networks (RNNs) có thể được sử dụng để huấn luyện trên các tập dữ liệu chứa các đối tượng tương ứng với các lớp khác nhau. Sau đó, các mô hình này có thể được sử dụng để phân loại các đối tượng đã được xác định trong đoạn video.

Tóm lại, bước xác định đối tượng sau bước tóm tắt video là quá trình phát hiện và nhận dạng các đối tượng xuất hiện trong đoạn video dựa trên các phương pháp xử lý ảnh và Deep Learning.

Tạo mô hình 3D

Sau khi đã xác định được các đối tượng trong đoạn video và các thông tin liên quan đến chúng, bước tiếp theo là tạo mô hình 3D để mô phỏng chúng. Quá trình này có thể bao gồm các bước sau:

Xây dựng mô hình 3D: Trước hết, các đối tượng được tái tạo lại dưới dạng các mô hình 3D, bao gồm các thông tin về hình dạng, kích thước, màu sắc và vị trí của chúng. Các mô hình 3D

này có thể được tạo ra bằng cách sử dụng các phần mềm đồ họa 3D như Blender, Maya hoặc 3DS Max.

Tích hợp dữ liệu: Sau khi đã có các mô hình 3D của các đối tượng, các thông tin về chúng cần được tích hợp với các dữ liệu được xác định từ các bước trước đó, bao gồm các thông tin về vị trí, hướng, tốc độ và hành động của các đối tượng trong đoạn video. Các phương pháp theo dõi đối tượng (object tracking) như SORT hoặc DeepSORT có thể được sử dụng để theo dõi các đối tượng và tích hợp dữ liệu này vào các mô hình 3D tương ứng.

Lắp ráp mô hình 3D: Sau khi đã có các mô hình 3D của các đối tượng và thông tin liên quan đến chúng, các mô hình này cần được lắp ráp lại để tạo thành một mô hình 3D toàn cảnh của đoạn video. Các phương pháp lắp ráp mô hình 3D như Structure from Motion (SfM) hoặc Multi-View Stereo (MVS) có thể được sử dụng để thực hiện việc này.

Tối ưu hóa mô hình 3D: Cuối cùng, mô hình 3D của đoạn video cần được tối ưu hóa để đảm bảo tính chính xác và chất lượng của nó. Các phương pháp tối ưu hóa mô hình 3D như Bundle Adjustment (BA) hoặc Global Optimization có thể được sử dụng để tối ưu hóa mô hình 3D.

Tóm lại, bước tạo mô hình 3D sau bước xác định đối tượng là quá trình tái tạo các đối tượng trong đoạn video dưới dạng các mô hình 3D và tích hợp các thông tin liên quan đến chúng để tạo thành một mô hình 3D

Kết xuất (Rendering)

Sau khi đã có mô hình 3D của đoạn video, bước tiếp theo là rendering để tạo ra các hình ảnh và video chính xác của các đối tượng trong mô hình 3D. Các bước chi tiết trong quá trình rendering có thể bao gồm:

Thiết lập các thông số của mô hình: Trước khi bắt đầu quá trình rendering, cần thiết lập các thông số cho mô hình 3D, bao gồm ánh sáng, vật liệu, cảnh quan và camera. Các thông số này được thiết lập để đảm bảo rằng hình ảnh được tạo ra phù hợp với mục đích sử dụng.

Tạo ra các hình ảnh của mô hình: Quá trình này bao gồm tạo ra các hình ảnh của mô hình 3D từ các góc quan sát khác nhau. Các hình ảnh này có thể được tạo ra bằng cách sử dụng các phần mềm đồ họa như Blender hoặc Maya.

Áp dụng kỹ thuật rendering: Sau khi đã có các hình ảnh của mô hình 3D, cần áp dụng kỹ thuật rendering để tạo ra các hình ảnh chính xác của các đối tượng trong mô hình 3D. Các phương pháp rendering bao gồm ray tracing, rasterization và path tracing.

Tối ưu hóa quá trình rendering: Rendering có thể là quá trình tốn nhiều thời gian và tài nguyên, vì vậy cần tối ưu quá trình này để tăng tốc độ và tiết kiệm tài nguyên. Các phương pháp tối ưu hóa rendering như distributed rendering, multi-threading và caching có thể được sử dụng để tối ưu hóa quá trình này.

Xuất các đối tượng được render: Sau khi đã tạo ra các hình ảnh và video của các đối tượng trong mô hình 3D, cần xuất chúng ra dưới dạng các tệp định dạng phù hợp để có thể sử dụng được. Các định dạng phổ biến cho đầu ra bao gồm MPEG-4, AVI, MOV và GIF.

Tóm lại, quá trình rendering sau bước tạo mô hình 3D là quá trình tạo ra các hình ảnh và video chính xác của các đối tượng trong mô hình 3D bằng cách áp dụng kỹ thuật rendering và tối ưu quá trình này để tăng tốc độ và tiết kiệm tài nguyên.

Ở bước sử dụng mô hình 3D của đối tượng để xác định đối tượng trong các khung ảnh riêng lẻ của chuỗi video, hệ thống tạo mặt nạ đối tượng cho mỗi khung trong chuỗi video được sử dụng tạo mô hình 3D và mỗi mặt nạ liên kết với một khung được sử dụng để xác định đối tượng. Theo một phương án khác, hệ thống có thể sử dụng xác định vị trí của mô hình với sự định hướng mà khớp với khung trong chuỗi video và sử dụng vị trí của đối tượng đã gán nhãn trong mô hình để xác định khu vực trong khung tương ứng với đối tượng.

Sau khi đối tượng được xác định trong mỗi khung của chuỗi video, hệ thống sẽ gán nhãn đối tượng trong mỗi khung ảnh riêng lẻ để tạo ra tập các khung ảnh chứa các trường hợp đã gán nhãn của đối tượng.

Hệ thống sử dụng dữ liệu là tập các ảnh đã gán nhãn để huấn luyện mô hình nhận dạng đối tượng. Sau quá trình huấn luyện, hệ thống có thể nhận dạng các đối tượng tương tự với các đối tượng đã được gán nhãn trong dữ liệu huấn luyện.

Mô hình hệ thống theo một phương án của giải pháp hữu ích được mô tả như Hình 3. Trong đó:

Đầu vào của mô hình:

Video từ camera: Các video được thu nhận từ camera giám sát, đóng vai trò là đầu vào của hệ thống.

Khối xử lý gồm:

Khối phát hiện đối tượng: Thông qua mô hình phát hiện đối tượng thu được đối tượng cần quan tâm.

Khối theo vết đối tượng: Thực hiện nhiệm vụ theo vết đối tượng khi đối tượng di chuyển giữa các khung hình liên tiếp nhau.

Tập hợp các hình ảnh của đối tượng trong video: Các hình ảnh được trích xuất.

Phát hiện thuộc tính đối tượng: Thông qua mô hình nhận diện thuộc tính để phát hiện các thuộc tính quan tâm.

Khối thuộc tính đối tượng: Được trích xuất thông qua khối phát hiện thuộc tính đối tượng.

Khối database (cơ sở dữ liệu): Khối cơ sở dữ liệu dùng để lưu các thuộc tính của đối tượng.

Khối xử lý background (khối xử lý tác vụ ngầm trong chương trình): Xử lý video từ camera và lưu vào cơ sở dữ liệu.

Đầu ra của mô hình:

Bản tóm tắt video: Bản tóm tắt video được tổng hợp từ dữ liệu trong cơ sở dữ liệu.

Quá trình MOT (model object tracking): Đối tượng được theo vết thực hiện trích xuất liên tục trên các khung hình và trong suốt quá trình hoạt động của phương pháp tóm tắt video. Việc kết xuất các hình ảnh trong cơ sở dữ liệu tạo ra kết quả là bản tóm tắt video như mong muốn.

Khởi tìm kiếm đối tượng: Cuối cùng dựa vào các trích xuất thuộc tính trong cơ sở dữ liệu mô hình trả về đối tượng cần tìm kiếm tức thì sau thời gian ngắn của quá trình trích xuất.

Hình 4 mô tả mô hình sau khi được huấn luyện

Mô hình nhận diện đối tượng

Hiện nay có rất nhiều mô hình nhận diện đối tượng, việc tự xây dựng hay tự phát triển mô hình mới là không thực sự cần thiết khi các mô hình này đã được cung cấp trước đó. Giải pháp hữu ích này dùng các mô hình học máy là một trong những mô hình phát hiện đối tượng tốt nhất trong thời điểm hiện tại.

Mô hình nhận diện thuộc tính của đối tượng

Tìm mô hình nhận diện thuộc tính phù hợp: Là các mô hình nhận diện thuộc tính phù hợp với mục tiêu cho trước của hệ thống.

Tạo cơ sở dữ liệu: Tạo bộ dữ liệu để đưa vào quá trình huấn luyện mô hình.

Tinh chỉnh cơ sở dữ liệu: Bộ cơ sở dữ liệu đưa vào mô hình được tinh chỉnh sao cho phù hợp với mô hình.

Khởi huấn luyện mô hình: Dữ liệu được đưa vào để huấn luyện mô hình.

Đánh giá mô hình: Mô hình sau khi được huấn luyện và được đưa vào đánh giá để phân định độ chính xác của mô hình.

Mô hình nhận diện thuộc tính: Sau quá trình huấn luyện thu được mô hình phục vụ mục đích nhận diện thuộc tính.

Mô hình phát hiện đối tượng

Tìm mô hình phát hiện đối tượng: Các mô hình nhận diện đối tượng phù hợp với mục tiêu của hệ thống.

Tạo cơ sở dữ liệu: Tạo bộ dữ liệu để đưa vào huấn luyện mô hình.

Tinh chỉnh cơ sở dữ liệu: Dữ liệu đưa vào phải được tinh chỉnh sao cho phù hợp với mô hình.

Khởi huấn luyện mô hình: Dữ liệu được đưa vào để huấn luyện mô hình.

Đánh giá mô hình: Mô hình sau khi được huấn luyện và được đánh giá qua tập thử nghiệm để đưa ra độ chính xác của mô hình.

Mô hình phát hiện đối tượng: Sau quá trình huấn luyện mô hình thu được mô hình với độ chính xác tốt nhất phục vụ cho bài toán phát hiện đối tượng.

Ví dụ minh họa

Giải pháp đã thử nghiệm phương pháp tóm tắt video trên một số luồng video được thu nhận tại camera giao thông. Hình 6 là minh họa hình ảnh được thu thập từ camera đặt tại nước ngoài, trong khi Hình 5 và Hình 7 là từ hình ảnh được thu nhận bằng camera được đặt tại Việt Nam. Trong hầu hết các ví dụ, "đối tượng" hay "con người" là số lượng pixel chuyển động trong đó.

YÊU CẦU BẢO HỘ

1. Phương pháp tạo ra bản tóm tắt video sử dụng thiết bị tính toán, phương pháp bao gồm các bước:

thu được các video được ghi lại từ nhiều camera khác nhau, mỗi video bao gồm các khung hình chứa các đối tượng;

tóm tắt mỗi video từ đoạn video dài để tạo ra đoạn video ngắn hơn;

xác định đối tượng để tìm ra các đối tượng cụ thể xuất hiện trong đoạn video ngắn hơn;

tạo mô hình ba chiều (3D) của đối tượng đã gắn nhãn;

sử dụng mô hình 3D của đối tượng để xác định đối tượng trong các khung ảnh riêng lẻ của chuỗi video;

gán nhãn đối tượng trong mỗi khung ảnh riêng lẻ;

huấn luyện mô hình nhận dạng đối tượng sử dụng dữ liệu là tập các ảnh đã gắn nhãn;

kết xuất mô hình 3D của đối tượng để tạo ra bản tóm tắt video chứa ảnh đối tượng được gắn nhãn và sử dụng mô hình nhận dạng đối tượng để xác định đối tượng trong bản tóm tắt video;

trong đó,

bước tóm tắt mỗi video từ đoạn video dài để tạo ra đoạn video ngắn hơn bao gồm các bước:

phân tích video: đoạn video sẽ được phân tích để xác định các thông tin quan trọng, bao gồm các cảnh, sự kiện và đối tượng trong video;

chọn thông tin quan trọng: Sau khi phân tích video, các thông tin quan trọng sẽ được chọn để tạo ra phiên bản tóm tắt; việc này có thể bao gồm chọn các cảnh, đối tượng hoặc hành động quan trọng nhất trong video;

tạo đoạn video ngắn hơn: các thông tin được chọn sẽ được sắp xếp và tạo thành đoạn video ngắn hơn;

bước xác định đối tượng để tìm ra các đối tượng cụ thể xuất hiện trong đoạn video ngắn hơn bao gồm:

chuẩn bị dữ liệu: các đoạn video được chia thành các khung hình (frames) và xác định các thông tin như độ phân giải, tốc độ khung hình, và định dạng video;

xác định đối tượng trong khung hình: bước này sẽ tìm kiếm các đối tượng trong từng khung hình và trích xuất các đặc trưng của đối tượng;

xác định đối tượng qua các khung hình: bước này sẽ theo dõi các đối tượng qua các khung hình khác nhau để tìm ra các đối tượng tương ứng với nhau trong các khung hình khác nhau;

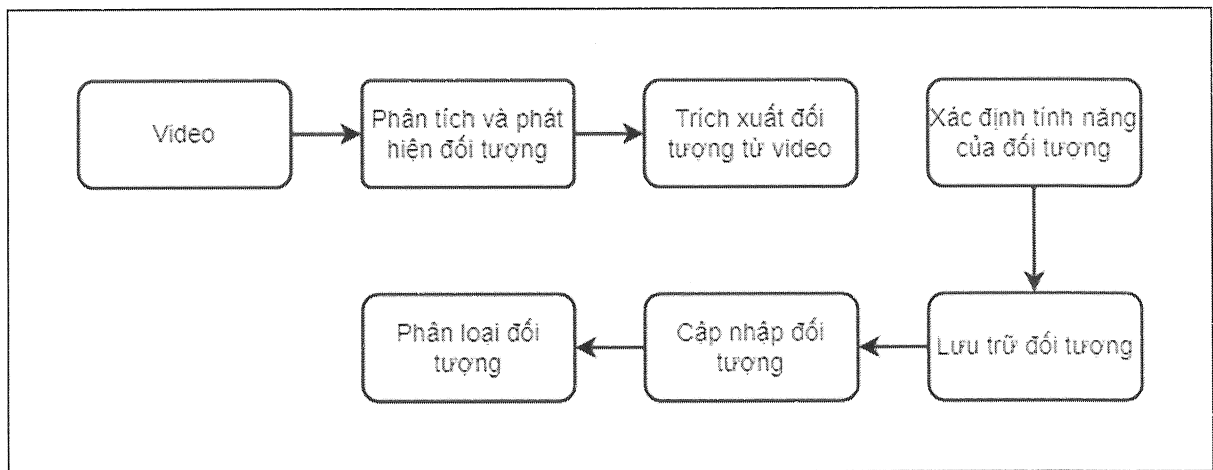
phân loại và gắn nhãn các đối tượng tương ứng;

bước tạo mô hình ba chiều (3D) của đối tượng với nhãn tương ứng bao gồm:

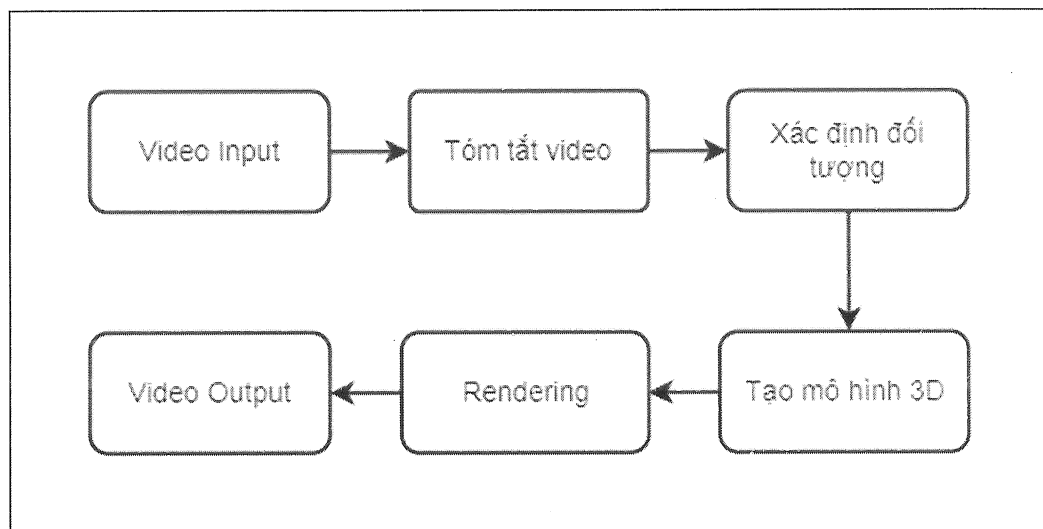
xây dựng mô hình 3D: các đối tượng được tái tạo lại dưới dạng các mô hình 3D, bao gồm các thông tin về hình dạng, kích thước, màu sắc và vị trí của chúng;

tích hợp dữ liệu của đối tượng: dữ liệu được tích hợp bao gồm các thông tin về vị trí, hướng, tốc độ và hành động của các đối tượng trong đoạn video;

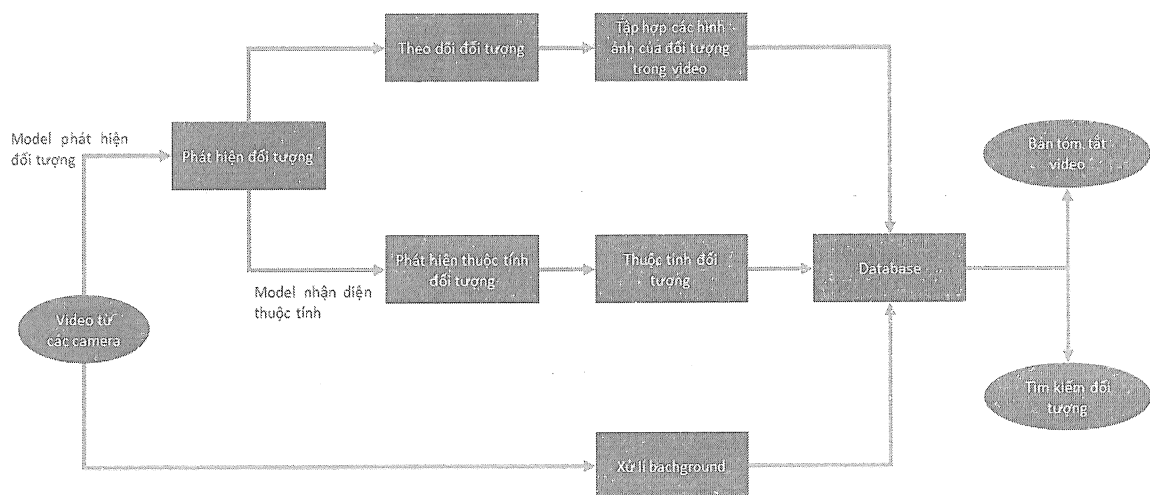
lắp ráp mô hình 3D để tạo thành một mô hình 3D toàn cảnh của đoạn video;
tối ưu hóa mô hình 3D;
bước kết xuất mô hình 3D của đối tượng để tạo ra bản tóm tắt video bao gồm:
thiết lập các thông số của mô hình: thiết lập các thông số cho mô hình 3D, bao gồm ánh sáng, vật liệu, cảnh quan và camera;
tạo ra các khung hình ảnh của mô hình 3D của đối tượng đã được gán nhãn với các góc khác nhau;
tạo ra các bản tóm tắt video các đối tượng trong mô hình 3D với nhãn được gán.



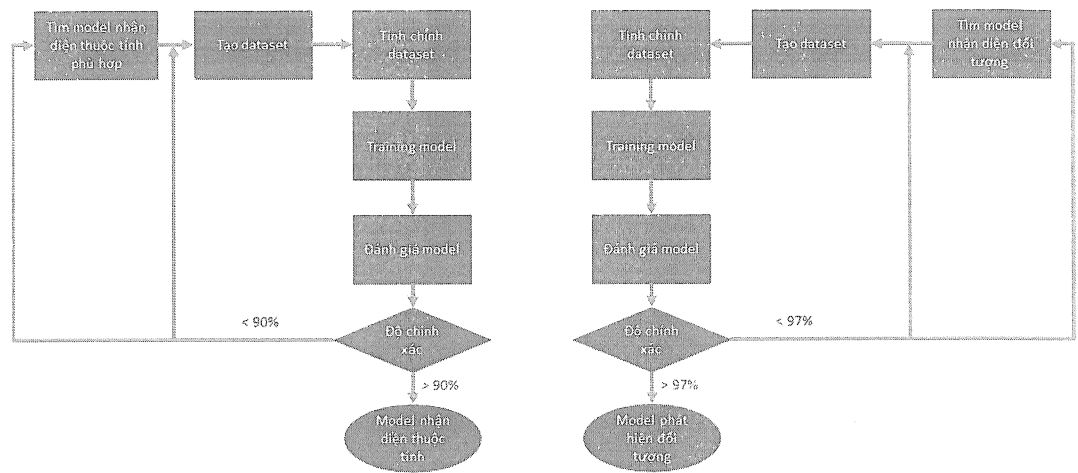
Hình 1



Hình 2



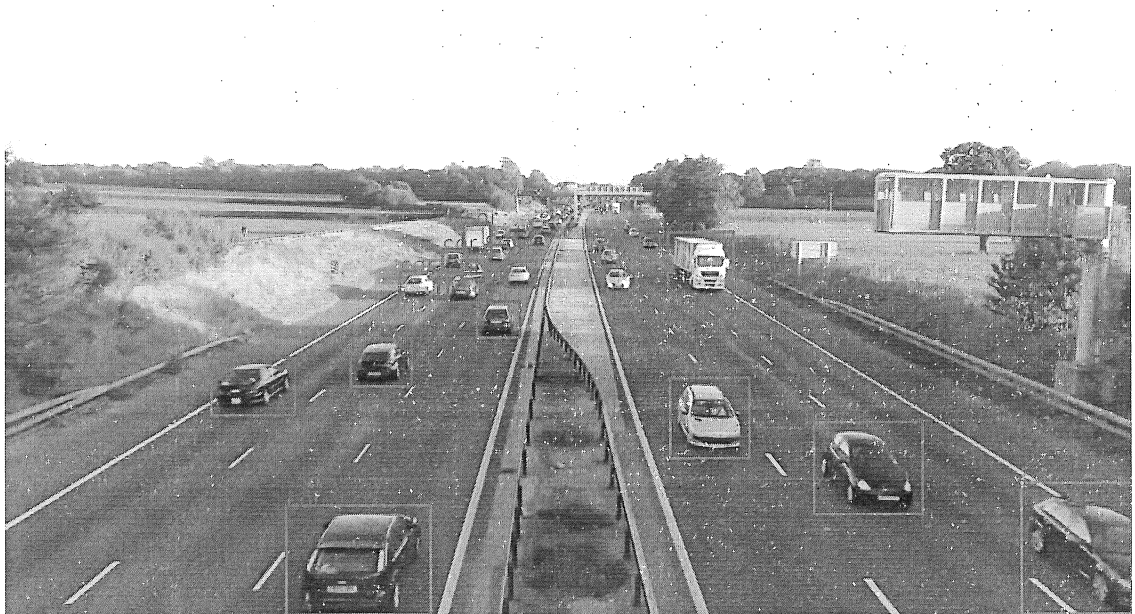
Hình 3



Hình 4



Hình 5



Hình 6



Hình 7



Hình 8