(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) **1-0022077**

(51)[7] **G10L 15/00**, 13/00, G06K 9/00

(13) **B**

(21) 1-2017-03914
(22) 04.10.2017
(45) 25.10.2019 379
(43) 25.01.2018 358

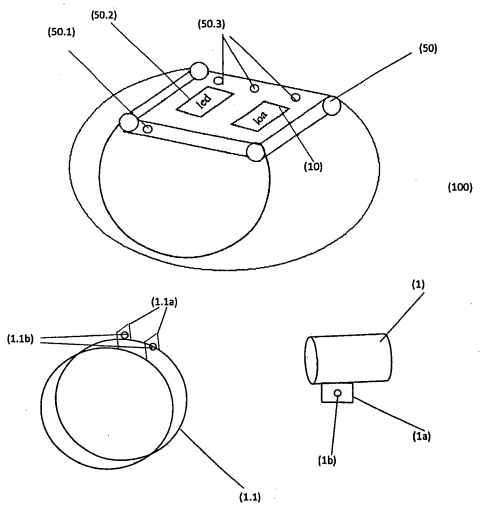
(73) VIỆN CÔNG NGHỆ THÔNG TIN (VN)
Nhà A3, 18 Hoàng Quốc Việt, phường Nghĩa Đô, quận Cầu Giấy, thành phố Hà Nội

(72) Nguyễn Trường Thắng (VN), Nguyễn Thế Hoàng Anh (VN), Phí Tùng Lâm (VN), Trần Mạnh Đông (VN), Mai Văn Thủy (VN)

(54) PHƯƠNG PHÁP, THIẾT BỊ VÀ VẬT GHI CÓ THỂ ĐỌC ĐƯỢC BẰNG MÁY TÍNH ĐỂ NHẬN DẠNG CHỮ VIẾT IN TIẾNG VIỆT VÀ TIẾNG ANH VÀ CHUYỂN THÀNH GIỌNG NÓI

(57) Sáng chế đề cập đến thiết bị đọc các chữ in tiếng Việt và phát thành giọng nói có thể mang theo người và có thể tháo lắp được, thiết bị này bao gồm: bộ thu video (1) để thu nhận dữ liệu hình ảnh/video, bộ thu video này được lắp trên gá đỡ (1a) có dạng vòng tròn như chiếc nhẫn, sao cho gá đỡ này có thể lồng vào ngón tay người, khi ngón tay di chuyển, bộ thu video sẽ thu hình ảnh/video phía trên ngón tay; và bộ xử lý trung tâm (50) để xử lý dữ liệu hình ảnh/video thu được từ bộ thu video (1) thông qua phương tiện kết nối để chuyển thành giọng nói, trong đó bộ xử lý trung tâm này bao gồm: bộ xử lý đa phương tiện (2) để xử lý dữ liệu hình ảnh/video thu được sao cho dữ liệu lối ra của bộ xử lý đa phương tiện này là ảnh gồm chuỗi các chữ in, có số lượng nằm trong khoảng từ 3 đến 5 chữ để phù hợp với tốc độ di chuyển ngón tay và góc mở của camera, để nhận dạng ký tự quang học; bộ nhận dạng ký tự quang học (3) để chuyển đổi hình ảnh các chữ thu được ở đầu ra của bộ xử lý ảnh thành chữ dưới dạng ký tự đọc được bằng máy tính; bộ tổng hợp tiếng nói (5) sẽ chuyển các chữ đã được xử lý trong bộ xử lý ký tự (4) thành giọng nói; và loa (10) để phát ra âm thanh.

Ngoài ra, sáng chế cũng đề cập đến phương pháp và vật ghi ghi chương trình máy tính để đọc các chữ in tiếng Việt và phát thành giọng nói tương ứng.



Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập tới phương pháp, thiết bị và vật ghi có thể đọc được bằng máy tính để đọc các chữ viết và phát thành giọng nói có thể mang theo người và có thể tháo lắp được.

Tình trạng kỹ thuật của sáng chế

Việt Nam có khoảng hơn 1 triệu người khiếm thị và trong số đó có hơn 300.000 người mù. Việc thu thập thông tin, nhất là những thông tin được thể hiện dưới dạng văn bản như sách vở, báo in đối với đối tượng này thường diễn ra rất khó khăn hoặc phải nhờ sự giúp đỡ của những người xung quanh. Với mục đích giúp đỡ những người có hoàn cảnh như vậy, chúng tôi đã kết nối các công nghệ phức tạp, chế tạo thành công một sản phẩm thân thiện và có khả năng giúp người sử dụng đọc bằng cách phát ra âm thanh các đoạn chữ viết in được trỏ tới.

Đã biết một số các thiết bị trên thị trường có thể đọc được văn bản và chuyển các ký tự nhận dạng được thành giọng nói cho các tiếng nước ngoài ví dụ Tiếng Anh.

Các thiết bị đều có bộ phận chụp ảnh để chụp các hình ảnh trong khu vực nhất định, sau đó, hình ảnh sẽ được chuyển đến bộ xử lý để định dạng lại hình ảnh, sau đó hình ảnh này sẽ được nhận dạng quang học. Kết quả nhận dạng quang học sẽ được gửi tới bộ tổng hợp tiếng nói, sau đó sẽ được phát ra loa.

Qua thực tế, thấy rằng các thiết bị này còn một số vấn đề cần giải quyết, như thiết bị sẽ không đọc được nếu hình ảnh chụp được không được rõ nét. Các thiết bị đã biết này chủ yếu sử dụng để đọc và phát âm thành từng từ, có nghĩa, không thể tiến hành đọc thành cả câu. Quan trọng hơn, các thiết bị này chưa có chức năng đọc

và phát âm chữ tiếng Việt cho người Việt hoặc người nước ngoài muốn đọc chữ tiếng Việt.

Ngoài ra, các thiết bị đọc văn bản này thường có vấn đề bỏ câu khi người đọc di chuyển ngón tay quá nhanh.

Sáng chế được đề xuất nhằm khắc phục các nhược điểm nêu trên.

Bản chất kỹ thuật của sáng chế

Mục đích của sáng chế đề xuất phương pháp, thiết bị và vật ghi để đọc các chữ viết in tiếng Việt và phát thành giọng nói có thể mang theo người và có thể tháo lắp được.

Để đạt được mục đích nêu trên, theo một khía cạnh, sáng chế đề xuất thiết bị đọc các chữ in tiếng Việt và phát thành giọng nói có thể mang theo người và có thể tháo lắp được, thiết bị này bao gồm:

bộ thu video (1) để thu nhận dữ liệu hình ảnh/video, bộ thu video này được lắp trên gá đỡ (1a) có dạng vòng tròn như chiếc nhẫn, sao cho gá đỡ này có thể lồng vào ngón tay người, khi ngón tay di chuyển, bộ thu video sẽ thu hình ảnh/video phía trên ngón tay; và

bộ xử lý trung tâm (50) để xử lý dữ liệu hình ảnh/video thu được từ bộ thu video (1) thông qua phương tiện kết nối để chuyển thành giọng nói, trong đó bộ xử lý trung tâm này bao gồm:

bộ xử lý đa phương tiện (2) để xử lý dữ liệu hình ảnh/video thu được sao cho dữ liệu lối ra của bộ xử lý đa phương tiện này là ảnh gồm chuỗi các chữ in, có số lượng nằm trong khoảng từ 3 đến 5 chữ để phù hợp với tốc độ di chuyển ngón tay và góc mở của camera, để nhận dạng ký tự quang học, trong đó bộ xử lý đa phương tiện này bao gồm:

bộ cảnh báo tốc độ để cảnh báo tốc độ khi người sử dụng di chuyển ngón tay quá nhanh,

bộ kiểm tra dữ liệu đầu vào (2.4) để thông báo cho người sử dụng biết khoảng trắng, hết dòng hoặc hình vẽ,

bộ xử lý ảnh (2.5) để xử lý dữ liệu đầu vào sao cho có kết quả tốt nhất cho bộ nhận dạng ký tự quang học, như chỉnh nghiêng dựa trên đo góc nghiêng của dòng chữ với biến đổi Hough hoặc biến đổi tương tự, khử nhiễu với bộ lọc trung vị (median filter), tăng cường sáng cho ảnh dựa trên thuật toán biến đổi Retinex hoặc thuật toán biến đổi tương tự;

bộ nhận dạng ký tự quang học (3) để chuyển đổi hình ảnh các chữ thu được ở đầu ra của bộ xử lý ảnh thành chữ dưới dạng ký tự đọc được bằng máy tính;

bộ xử lý ký tự (4) để xử lý ký tự nhận dạng thu được để ghép nhận dạng các ký tự tránh các trường hợp bị trùng, trong đó, bộ xử lý ký tự này bao gồm bộ nhớ lưu ký tự (4.1) và bộ trích ghép ký tự (4.2);

bộ tổng hợp tiếng nói (5) sẽ chuyển các chữ đã được xử lý trong bộ xử lý ký tự (4) thành giọng nói, trong đó, bộ tổng hợp tiếng nói bao gồm:

khối xử lý ngôn ngữ tự nhiên tiếng Việt (5.1) và khối xử lý tổng hợp tiếng nói tiếng Việt(5.2),

bộ khuếch đại âm thanh (7) để khuếch đại âm thanh từ bộ tổng hợp tiếng nói,

loa (10) để phát ra âm thanh;

bộ điều khiển điện (8) để điều khiển và cấp nguồn điện cho các bộ phận nêu trên; và

nguồn điện (9) để cấp điện cho bộ điều khiển điện.

Bên cạnh đó, thiết bị nêu trên không bị giới hạn ở chức năng đọc tiếng Việt mà thiết bị này còn được tạo cấu hình để có thể đọc các chữ in tiếng Anh và phát thành giọng nói. Cần hiểu rằng, chức năng đọc các ngôn ngữ khác cũng có thể được bổ sung cho thiết bị theo sáng chế dưới dạng phần mềm hoặc các môđun mở

rộng được thực hiện bởi phần mềm, hoặc tương tự, do đó cũng được coi là không nằm ngoài phạm vi của sáng chế.

Tốt hơn là, bộ xử lý ký tự (4) và bộ tổng hợp tiếng nói (5) có thể được tạo cấu hình để xử lý văn bản tiếng Việt và tiếng Anh.

Theo một phương án, bộ nhận dạng ký tự quang học (3), bộ xử lý ký tự (4), bộ tổng hợp tiếng nói (5) được tích hợp trên máy chủ từ xa hoặc tích hợp bằng công nghệ điện toán đám mây.

Tốt hơn là, thiết bị nêu trên còn bao gồm:

bộ nhận biết hết dòng khi bộ thu video di chuyển đến khoảng trắng, bộ nhận biết hết dòng này sẽ nhận thấy hết dòng;

bộ cảm biến độ nghiêng để nhận biết độ nghiêng của ngón tay người sử dụng khi bộ thu video được lắp tháo được trên ngón tay của người sử dụng;

bộ cảm biến tốc độ di chuyển ngón tay để xác định tốc độ ngón tay di chuyển, nếu tốc độ ngón tay di chuyển quá nhanh so với tốc độ cho trước.

Theo một khía cạnh khác, sáng chế đề xuất phương pháp xử lý để đọc các chữ in tiếng Việt và phát thành giọng, phương pháp này bao gồm:

thu dữ liệu hình ảnh/video phía trên ngón tay người sử dụng bởi bộ thu video (1), trong đó hình ảnh này được thu nhờ bộ thu video này được lắp trên gá đỡ (1a) có dạng vòng tròn như chiếc nhẫn, sao cho gá đỡ này có thể lồng vào ngón tay người, khi ngón tay di chuyển, bộ thu video sẽ thu hình ảnh phía trên ngón tay;

xử lý dữ liệu hình ảnh/video thu được bởi bộ xử lý đa phương tiện (2), trong đó dữ liệu lối ra của bộ xử lý đa phương tiện này là ảnh gồm chuỗi các chữ in, có số lượng nằm trong khoảng từ 3 đến 5 chữ để phù hợp với tốc độ di chuyển ngón tay và góc mở của camera, để nhận dạng ký tự quang học, trong đó bước xử lý đa phương tiện còn bao gồm:

cảnh báo tốc độ bởi bộ cảnh báo dữ liệu khi người sử dụng di chuyển ngón tay quá nhanh,

kiểm tra dữ liệu đầu vào bởi bộ kiểm tra dữ liệu đầu vào (2.4) để thông báo cho người sử dụng biết khoảng trắng, hết dòng hoặc hình vẽ,

xử lý ảnh bởi bộ xử lý ảnh (2.5) để xử lý dữ liệu đầu vào sao cho có kết quả tốt nhất cho bộ nhận dạng ký tự quang học, như chỉnh nghiêng dựa trên đo góc nghiêng của dòng chữ với biến đổi Hough hoặc biến đổi tương tự, khử nhiễu với bộ lọc trung vị (median filter), tăng cường sáng cho ảnh dựa trên thuật toán biến đổi Retinex hoặc thuật toán biến đổi tương tự;

nhận dạng ký tự quang học bởi bộ nhận dạng ký tự quang học (3) để chuyển đổi hình ảnh các chữ in tiếng Việt, thu được ở đầu ra của bộ xử lý ảnh thành chữ dưới dạng ký tự đọc được bằng máy tính;

xử lý ký tự bởi bộ xử lý ký tự (4) để xử lý ký tự nhận dạng thu được để ghép nhận dạng các ký tự tránh các trường hợp bị trùng;

tổng hợp tiếng nói bởi bộ tổng hợp tiếng nói (5) để chuyển các chữ đã được xử lý trong bộ xử lý ký tự (4) thành giọng nói tiếng Việt;

khuếch đại âm thanh bởi bộ khuếch đại âm thanh (7) để khuếch đại âm thanh từ bộ tổng hợp tiếng nói.

Tốt hơn là, bước xử lý dữ liệu hình ảnh/video và phát ra tiếng nói nêu trên còn được tạo cấu hình để có thể đọc các chữ in tiếng Anh và phát thành giọng nói.

Tốt hơn nữa là, bộ xử lý ký tự (4) và bộ tổng hợp tiếng nói (5) có thể được tạo cấu hình để xử lý văn bản tiếng Việt và tiếng Anh.

Theo một khía cạnh khác nữa, sáng chế đề xuất vật ghi ghi chương trình máy tính có thể đọc được bằng máy tính để khi được chạy trên máy tính, chương trình máy tính này có thể làm cho máy tính thực hiện phương pháp nêu trên.

Mô tả vắn tắt các hình vẽ

Hình 1 là hình vẽ sơ lược thể hiện thiết bị đọc mang theo người và có thể tháo lắp được theo một phương án của sáng chế;

Hình 2 là hình vẽ sơ lược thể hiện các bộ phận của thiết bị đọc mang theo người và có thể tháo lắp được theo một phương án của sáng chế;

Hình 3 là hình vẽ sơ lược thể hiện các bộ phận của thiết bị mang theo người và có thể tháo lắp được theo một phương án khác của sáng chế;

Hình 4 là lưu đồ thể hiện các bước thực hiện của bộ cảnh báo tốc độ theo một phương án của sáng chế;

Hình 5 là lưu đồ thể hiện bước kiểm tra dữ liệu theo một phương án của sáng chế;

Hình 6 là lưu đồ thể hiện các bước thực hiện của bộ xử lý ảnh chữ viết theo một phương án của sáng chế;

Hình 7 là lưu đồ thể hiện các bước thực hiện nhận dạng ký tự theo một phương án của sáng chế;

Hình 8 là lưu đồ thể hiện các bước thực hiện của bộ trích ghép ký tự theo một phương án của sáng chế;

Hình 9 là lưu đồ thể hiện các bước thực hiện tổng hợp tiếng nói theo một phương án của sáng chế.

Mô tả chi tiết các phương án ưu tiên thực hiện sáng chế

Dưới đây, sáng chế sẽ được bộc lộ trong phần mô tả chi tiết dưới đây và có dựa vào các hình vẽ kèm theo.

Hình 1 là hình vẽ sơ lược thể hiện thiết bị đọc mang theo người và có thể tháo lắp được theo một phương án của sáng chế.

Như được thể hiện trên hình vẽ, thiết bị đọc các chữ viết và chuyển thành giọng nói 100 có thể mang theo người và có thể tháo lắp được, bao gồm Camera 1, và bộ xử lý trung tâm 50.

Camera 1 để thu nhận video, theo một phương án của sáng chế, Camera 1 này là camera có độ phân giải tối thiểu 480 x 640.

Camera 1 này được lắp tháo lắp và điều chỉnh được trên gá đỡ 1.1 có dạng vòng tròn như chiếc nhẫn, sao cho gá đỡ 1.1 này có thể lồng vào ngón tay người sử dụng được.

Theo một phương án của sáng chế, camera 1 được liên kết với gá đỡ 1.1 sao cho có thể tháo lắp được và camera 1 có thể điều chỉnh góc so với trục xuyên tâm của gá đỡ 1a. Gá đỡ 1.1 được đeo vào ngón tay (thường là ngón tay trỏ) của người sử dụng sao cho có thể chỉnh được vị trí của bộ đỡ 1.1 để camera có khoảng cách và góc nghiêng phù hợp để thiết bị thực hiện chức năng.

Như cũng được thể hiện trên hình vẽ, bên dưới camera 1 có phần nhô ra 1a và phần hở 1b ở giữa phần nhô ra 1a này, tốt hơn là phần hở 1b có dạng tròn. Trên gá đỡ 1.1 cũng có hai phần nhô 1.1a và trên hai phần nhô 1.1a này có các phần hở 1.1b, tốt hơn là phần hở 1.1b này có dạng hình tròn. Các phần hở 1.1b này được tạo ra sao cho khi phần nhô ra 1a trên camera 1 lắp vào bộ đỡ 1.1 thì các phần hở 1.1b và phần hở 1b sẽ đồng tâm với nhau. Camera 1 và gá đỡ 1.2 sẽ được liên kết với nhau bằng phương tiên liên kết có thể tháo lắp được (không được thể hiện trên hình vẽ). Theo một phương án của sáng chế, phương tiên liên kết có thể tháo lắp được bu-lông và đai ốc.

Như được thể hiện trên hình vẽ, thiết bị 100 bao gồm bộ xử lý 50 để chuyển đổi hình ảnh thu được từ camera thành giọng nói. Theo một phương án của sáng chế, bộ xử lý 50 được tích hợp vòng để có thể đeo vào tay người sử dụng.

Theo một phương án của sáng chế, camera 1 được nối tín hiệu với bộ xử lý 50 bằng giao thức truyền dẫn không dây (ví dụ, Bluetooth,..) hoặc theo một phương án khác của sáng chế là truyền dẫn có dây (ví dụ cáp FPC…).

Bộ xử lý 50 được tích hợp sẵn lối ra âm thanh là loa 10 và lỗ cắm tai nghe 50.1, hệ thống màn hình Led 50.2 để hiển thị các chức năng của thiết bị, và các phím bấm chức năng 50.3.

Hình 2 thể hiện hình vẽ sơ lược thể hiện các bộ phận của thiết bị đọc mang theo người và có thể tháo lắp được theo một phương án của sáng chế bao gồm camera 1 và các bộ phận có trong bộ xử lý 50.

Như được thể hiện trên hình vẽ thiết bị đọc các chữ viết và phát thành giọng nói có thể mang theo người và có thể tháo lắp được, bao gồm camera 1 để thu nhận video. Khi ngón tay di chuyển, camera sẽ thu video phía trên ngón tay với nội dung bao gồm các dòng chữ viết in. Hình ảnh thu được từ camera 1 sẽ được gửi tới bộ xử lý đa phương tiện 2 để xử lý video thành hình ảnh có thể nhận dạng được để gửi tới bộ nhận dạng ký tự quang học 3.

Bộ xử lý đa phương tiện 2 sẽ xử lý các dữ liệu thu được từ camera thông qua các bộ phận bao gồm Bộ lữu trữ video 2.2, bộ cảnh báo tốc độ 2.3, bộ kiểm tra dữ liệu đầu vào 2.4, bộ xử lý ảnh 2.5.

Bộ lưu dữ liệu video 2.2 là bộ nhớ để lưu dữ liệu video thu được.

Bộ cảnh báo tốc độ 2.3 sẽ kiểm tra tốc độ di chuyển ngón tay của người sử dụng có phù hợp hay không, nếu không thu được hình ảnh tốt thì sẽ phát tín hiệu cảnh báo qua loa 10.

Nếu tốc độ di chuyển ngón tay phù hợp, dữ liệu sẽ được chuyển tới bộ kiểm tra dữ liệu đầu vào 2.4. Bộ kiểm tra dữ liệu đầu vào 2.4 này sẽ phát hiện ra các tình trạng "Khoảng trống", "hết dòng" và "hình vẽ" để phát ra cảnh báo cho người sử dụng.

Nếu bộ kiểm tra dữ liệu đầu vào 2.4 không phát hiện các trường hợp nêu trên, dữ liệu sẽ được chuyển tới bộ xử lý ảnh 2.5.

Bộ xử lý ảnh 2.5 sẽ xử lý dữ liệu là 15 giây dữ liệu hình ảnh, mỗi giây sẽ chọn ra 5 khung hình để chọn ra ảnh nét nhất. Sau đó thực hiện các bước xử lý ảnh như chỉnh các ảnh nghiêng thành thẳng, tăng cường độ sáng tối của ảnh, xử lý nhiễu trên ảnh và phân đoạn ảnh để lấy ảnh có chuỗi các chữ, thường là 3-5 chữ, ngay trong dòng phía trên của ngón tay của người sử dụng.

Hình ảnh thu được từ bộ xử lý ảnh sẽ được gửi tới bộ nhận dạng ký tự quang học 3 để chuyển đổi hình ảnh các chữ thu được ở đầu ra của bộ xử lý ảnh thành chữ dưới dạng ký tự. Theo một phương án của sáng chế, dữ liệu thu được sau khi xử lý của bộ nhận dạng ký tự quang học sẽ được lưu vào tệp tin chứa các ký tự có thể được chỉnh sửa bằng các chương trình soạn thảo văn bản.

Bộ nhận dạng ký tự quang học 3 thực hiện quá trình chuyển đổi ảnh văn bản sang dạng văn bản có thể chỉnh sửa trong máy tính. Đầu vào của quá trình này là tập tin hình ảnh và đầu ra sẽ là các tập tin văn bản chứa nội dung là các chữ viết, ký hiệu có trong hình ảnh đó.

Dữ liệu thu được từ bộ nhận dạng quang học 3 sẽ được chuyển tới bộ xử lý ký tự 4. Bộ xử lý ký tự 4 bao gồm bộ nhớ lưu ký tự 4.1 để lưu dữ liệu được chuyển từ bộ nhận dạng ký tự quang học 3 và bộ trích ghép ký tự 4.2 vốn sẽ thực hiện việc so sánh chuỗi ký tự, kết hợp/ghép các chữ viết in và loại bỏ các ký tự không có nghĩa, thừa, không cần thiết.

Dữ liệu thu được từ bộ xử lý ký tự 4 sẽ được gửi tới bộ tổng hợp tiếng nói 5.

Bộ tổng hợp tiếng nói 5 bao gồm khối xử lý tổng hợp tiếng nói 5.1 và khối xử lý ngôn ngữ tự nhiên 5.2.

Khối xử lý tổng hợp tiếng nói 5.1 sẽ xử lý dữ liệu nhận được từ bộ xử lý ký tự 4 để tạo ra tiếng nói của con người một cách nhân tạo. Việc tổng hợp tiếng nói từ văn bản (Text-To-Speech, viết tắt là TTS) là quá trình chuyển đổi tự động một văn bản có nội dung bất kỳ thành lời nói.

Đã biết, các bộ tổng hợp tiếng nói được sử dụng cho mục đích này còn gọi là hệ thống tổng hợp tiếng nói và còn có thể cài đặt bằng phần mềm hoặc trong sản phẩm phần cứng.

Cấu trúc một hệ thống tổng hợp tiếng nói, nếu đầu vào của một hệ thống tổng hợp tiếng nói là văn bản, thì hệ thống này được gọi là tổng hợp tiếng nói từ văn bản (TTS). Trong trường hợp các hệ thống tổng hợp tiếng nói với bộ từ vựng hạn chế, chẳng hạn như các máy trò chơi, các hệ thống trả lời tự động với các mẫu

âm thanh thu âm trước, đôi khi có thể coi đó là một hệ thống TTS hạn chế cho một bài toán cụ thể, có giới hạn đầu vào.

Về cơ bản, một hệ thống tổng hợp tiếng nói về cơ bản bao gồm hai khối chức năng: (i) khối phân tích xử lý ngôn ngữ tự nhiên NLP) hay còn gọi là khối tổng hợp mức cao; và (ii) khối xử lý tổng hợp tiếng nói (SSP) có nhiệm vụ tổng hợp tiếng nói hay còn gọi là khối tổng hợp mức thấp.

Tổng hợp mức cao có nhiệm vụ chuyển đổi chuỗi các ký tự văn bản đầu vào thành một dạng chuỗi các nhãn ngữ âm đã được thiết kế trước của hệ thống TTS. Nghĩa là, chuyển đổi chuỗi văn bản đầu vào thành dạng biểu diễn ngữ âm, xác định cách đọc nội dung văn bản. Quá trình này cũng đòi hỏi khả năng dự đoán ngôn điệu từ văn bản đầu vào với thông tin ngữ âm và ngữ điệu tương ứng. Từ các thông tin ngôn điệu và ngữ âm là chuỗi các nhãn phụ thuộc ngữ cảnh mức âm vị của văn bản đầu vào, khối tổng hợp mức thấp sẽ chọn ra các tham số thích hợp từ tập các giá trị tần số cơ bản, phổ tín hiệu, trường độ âm thanh (bao gồm âm vị, âm tiết). Sau đó, tiếng nói ở dạng sóng tín hiệu sẽ được tạo ra bằng một kỹ thuật tổng hợp.

Dữ liệu thu được, sau khi đi qua khối xử lý ngôn ngữ tự nhiên 5.1, sẽ được chuyển tới khối tổng hợp tiếng nói 5.2.

Khối xử lý tổng hợp tiếng nói 5.2 sẽ chuyển đổi dữ liệu nhận được để chuyển thành âm thanh. Âm thanh này sau khi được khuếch đại có thể phát ra được bởi loa.

Theo một phương án của sáng chế, khối xử lý ngôn ngữ tự nhiên 5.1 sẽ tạo ra các thông tin về ngữ âm và ngữ điệu cho việc đọc văn bản đầu vào.

Đã biết, thông tin ngữ âm cho biết những âm nào sẽ được phát ra, trong ngữ cảnh cụ thể nào. Thông tin ngữ điệu mô tả điệu tính của các âm được phát.

Việc xử lý ngôn ngữ tự nhiên bao gồm: chuẩn hóa văn bản, phân tích cú pháp, phân tích ngữ cảnh và ngữ nghĩa, chuyển đổi hình vị sang âm vị, dự đoán và phát sinh thông tin ngữ âm và ngữ điệu.

Khối xử lý ngôn ngữ tự nhiên được chia thành ba phần chính gồm thành phần phân tích văn bản, thành phần chuyển đổi hình vị sang âm vị và thành phần dự đoán và sinh ngôn điệu cho văn bản.

Khối xử lý tổng hợp tín hiệu tiếng nói 5.2 đảm nhiệm việc thực hiện việc tạo ra tín hiệu tiếng nói từ các thông tin ngữ âm và ngữ điệu do khối phân tích xử lý ngôn ngữ tự nhiên 5.1 cung cấp.

Đã biết, chất lượng tiếng nói tổng hợp được đánh giá thông qua hai khía cạnh: mức độ dễ hiểu nội dung và mức độ tự nhiên. Mức độ dễ hiểu đề cập đến nội dung của tiếng nói tổng hợp có thể hiểu được dễ dàng không. Mức độ tự nhiên của tiếng nói tổng hợp là sự so sánh độ giống nhau giữa giọng nói tổng hợp và giọng nói tự nhiên của con người. Một hệ thống tổng hợp tiếng nói lý tưởng cần phải vừa dễ hiểu vừa tự nhiên, và mục tiêu xây dựng hệ thống tổng hợp tiếng nói là cải thiện đến mức tối đa hai tính chất này. Có nhiều phương pháp tổng hợp tiếng nói khác nhau được áp dụng, một số thiên về mức độ dễ hiểu hơn hoặc mức độ tự nhiên hơn, tùy thuộc vào mục đích mà các phương pháp tổng hợp được lựa chọn. Nhưng mục đích cơ bản của bất kỳ phương pháp tổng hợp là tạo ra tiếng nói với chất lượng dễ hiểu nội dung. Hiện nay, có ba phương pháp chính thường được dùng là tổng hợp mô hình hoá hệ thống phát âm, tổng hợp cộng hưởng tần số và tổng hợp ghép nối, ngoài ra cũng có các phương pháp khác phát triển từ ba phương pháp trên.

Theo một phương án của sáng chế, bộ tổng hợp tiếng nói 5 sẽ được thiết kế để "học" và phát âm tiếng Việt Nam.

Khối xử lý ngôn ngữ tự nhiên 5.1 bao gồm các khối con chuẩn hóa văn bản, phân tích cú pháp, phân tích ngữ cảnh, phân tích ngôn điệu và chuyển đổi hình vị - âm vị của tiếng Việt nam.

Khối xử lý tổng hợp tiếng nói 5.2 được thiết kế bao gồm các mô hình toán học dựa trên Mô hình Makov ẩn, các thuật toán và phương pháp tính toán.

Khối xử lý ngôn ngữ tự nhiên 5.2 sẽ phân tích âm tố, âm vị để đưa ra tiếng nói. Đối với âm tố, khi phát âm các âm tiết tan và lan, chúng ta nhận thấy giữa chúng có sự khác nhau. Sự khác nhau ở đây rõ ràng là do "t" và "l" gây ra. Như vậy có thể phân tích âm tiết thành những yếu tố nhỏ hơn, "tan" do 3 âm "t", "a", "n" phối hợp thành, và "lan" do 3 âm "l", "a", "n" phối hợp thành. Người ta gọi các yếu tố vừa tách ra khỏi 2 âm tiết trên là ***âm tố***. Âm tố được ghi vào giữa hai kí hiệu [], ví dụ: âm tố [a], [b], [c], v.v…

Âm tố là đơn vị ngữ âm nhỏ nhất trong lời nói. Một âm tố "a" ở ba người nói sẽ có ba cách phát âm khác nhau. Thậm chí, một người khi phát âm "a" ở ba thời điểm phát âm khác nhau, thì âm "a" khi phát ra cũng không hoàn toàn giống nhau. Đứng về mặt phát âm, chúng ta có vô số âm tố khác nhau. Có 3 loại âm tố là ***nguyên âm***, ***phụ âm***, ***bán âm*** (bán nguyên âm hay bán phụ âm).

Nguyên âm có đặc điểm là khi phát âm không bị luồng hơi cản lại, ví dụ âm a, u, i, e, o,…

Phụ âm có đặc điểm là khi phát âm thì luồng hơi bị cản lại, ví dụ âm p, b, t, m, n,….

Bán âm có đặc điểm giống nguyên âm về mặt cấu tạo, và giống phụ âm về mặt chức năng (nên còn được gọi là bán nguyên âm hay bán phụ âm), ví dụ /u/ (ngắn), /i/ (ngắn).

Đối với âm vị, như đã nói ở phần âm tố, cách phát âm một âm "a" của mỗi người và ngay ở một người, trong những thời điểm khác nhau, cũng không hoàn toàn như nhau. Và do đó, ta có vô số âm cụ thể của "a". Dựa vào những nét chung nhất, người ta quy nó về một đơn vị khu biệt, có chức năng phân biệt nghĩa, gọi là ***âm vị***.

Âm vị trong tiếng Việt là đơn vị ngữ âm nhỏ nhất có chức năng khu biệt nghĩa. Nếu số lượng âm tố là vô số, thì số lượng âm vị là có hạn, khoảng vài chục đơn vị trong một ngôn ngữ. Để khu biệt với âm tố, người ta ghi âm vị ở giữa hai kí hiệu //, ví dụ: âm vị /a/, /u/, /o/, v.v…

Đối với tiếng nói, khi người Việt phát âm các âm tiết để tạo nên chuỗi lời nói trong một hoàn cảnh giao tiếp cụ thể, đơn vị được dùng trong chuỗi lời nói là "tiếng". ***Tiếng*** trong tiếng Việt thường được hiểu là *âm tiết*, về mặt là đơn vị có nghĩa, dùng trong chuỗi lời nói. Khi phát âm, mỗi tiếng bao giờ cũng phát ra một hơi, có mang một thanh điệu nhất định. Tuy nhiên, trong lời nói hàng ngày, thường người ta nói đến *tiếng* nhiều hơn là *âm tiết*. Ví dụ trong phát ngôn *Cháu nó mới nói được hai tiếng "bà" và "mẹ"*, thì chúng ta có thể nói phát ngôn đó gồm 10 âm tiết hoặc 10 tiếng, nhưng không ai nói *Cháu nó mới nói được hai âm tiết "bà" và "mẹ"*, mặc dù "*bà*" và "*mẹ*" khi phân tích về mặt phát âm là 2 âm tiết.

Trên chữ viết, mỗi tiếng được ghi thành một ***chữ***. Tiếng có thể trực tiếp hay gián tiếp gắn liền với một ý nghĩa nhất định và không thể chia ra thành những đơn vị có nghĩa nhỏ hơn nữa. Vì vậy có thể hiểu *tiếng* trùng với *hình vị* và từ: *ăn*, *nói*, *đi*, *đứng*, *và*, *sẽ*,... là những tiếng trong tiếng Việt.

Cần chú ý: Trong phát ngôn "***tiếng*** *trong tiếng Việt*" thì "*tiếng*" được hiểu như ở trên, còn "*tiếng*" được hiểu với nghĩa là chỉ một ngôn ngữ cụ thể nào đó, ví dụ: *tiếng Anh, tiếng Nga, biết nhiều thứ tiếng*,...

Đối với hình vị, khi phân tích một phát ngôn (một đoạn của lời nói) người ta có thể phân xuất ra những đơn vị có ý nghĩa nhỏ nhất, đơn vị đó là hình vị.

Ví dụ trong phát ngôn "*Ngày mai tôi nghỉ học*" sẽ có 5 hình vị có ý nghĩa là "*ngày / mai / tôi / nghỉ / học*".

Hình vị thường có hình thức cấu tạo một âm tiết, tức là mỗi *hình vị* trùng với *âm tiết*, trên chữ viết mỗi hình vị được viết thành một *chữ*. Hình vị trong tiếng Việt có thể một mình đóng vai trò như một từ cũng có thể làm thành tố cấu tạo từ, nhưng nó chỉ được phân xuất ra nhờ phân tích bản thân các từ.

Tóm lại, khi phân tích chuỗi âm thanh của lời nói, người ta nhận thấy có những đơn vị ngữ âm được phát ra với một luồng hơi liên tục, không bị cắt đoạn ra trong dòng ngữ lưu, đơn vị đó gọi là ***âm tiết***. Trong tiếng Việt, một âm tiết thường mang một

thanh điệu và được ghi lại thành một ***chữ***. Khi phân tích một phát ngôn (một đoạn của lời nói), người ta phân xuất ra được những đơn vị nhỏ nhất trùng với *âm tiết*, đó là ***tiếng***. Tiếng thường trực tiếp hoặc gián tiếp gắn với với một ý nghĩa nhất định cho nên trùng với ***hình vị*** và ***từ***. Hình vị có hình thức cấu tạo một âm tiết, tức là mỗi hình vị trùng với âm tiết, nó có vai trò như từ nhưng nó không phải là từ vì từ bao gồm nó. *Âm tiết*, *hình vị* và *từ* là đơn vị của ngôn ngữ, còn *tiếng* là đơn vị của lời nói.

Theo một phương án của sáng chế, thiết bị 100 còn bao gồm bộ khuếch đại âm thanh 7 để khuếch đại âm thanh từ bộ tổng hợp tiếng nói 5 đưa ra.

Tín hiệu từ bộ tổng hợp tiếng nói 5 sẽ được chuyển qua bộ khuếch đại âm thanh 7 để khuếch đại âm thanh và cuối cùng, được chuyển tới Loa 10.

Theo một phương án của sáng chế, thiết bị 100 bao gồm bộ điều khiển điện 8 để điều khiển điện được cấp từ nguồn điện 9 và cấp nguồn điện cho các bộ phận nêu trên.

Nguồn điện 10 để cấp điện cho bộ điều khiển điện. Theo một phương án của sáng chế, nguồn điện có thể là pin có thể sạc được hoặc nguồn điện lưới đã qua bộ điều chỉnh với điện áp vào là 5v.

Hình 3 thể hiện hình vẽ sơ lược thể hiện các bộ phận của thiết bị đọc mang theo người và có thể tháo lắp được theo một phương án khác của sáng chế.

Theo đó, thiết bị theo sáng chế sẽ được chia thành hai phần, phần A và phần B. Phần A bao gồm Camera 1, bộ xử lý đa phương tiện 2, bộ phát tín hiệu 11a để phát tín hiệu là hình ảnh đã qua qua xử lý của bộ xử lý đa phương tiện 2. Các bộ phận thuộc phần A của thiết bị 100 này là thiết bị có thể tháo lắp được và mang theo người sử dụng.

Chi tiết bộ xử lý đa phương tiện 2 như đã được mô tả trên Hình 2.

Theo một phương án khác của sáng chế, bộ phát tín hiệu 11a có thể là bộ phát sử dụng công nghệ wifi 11.1 hoặc 3G 11.2.

Theo một phương án khác của sáng chế thiết bị còn có thêm bộ mã hóa hình ảnh, (không được thể hiện trên hình vẽ) để mã hóa hình ảnh.

Đã biết có nhiều phương pháp để mã hóa hình ảnh khác nhau trên thị trường.

Phần B bao gồm bộ thu tín hiệu hình ảnh 11b, bộ nhận dạng ký tự quang học 3, bộ xử lý ký tự 4 và bộ tổng hợp tiếng nói 5 và bộ phát tín hiệu 12b.

Chi tiết mô tả bộ nhận dạng ký tự quang học 3, bộ xử lý ký tự 4 và bộ tổng hợp tiếng nói 5 như đã được mô tả trên Hình 2.

Bộ thu tín hiệu hình ảnh 11b để thu tín hiệu từ bộ phát tín hiệu 11a.

Theo một phương án khác của sáng chế, bộ thu tín hiệu 11b có thể là bộ thu sử dụng công nghệ wifi 11.3 hoặc 3G 11.4 tương ứng với bộ phát tín hiệu 11a.

Theo một phương án khác của sáng chế thiết bị còn có thêm bộ giải mã hình ảnh, (không được thể hiện trên hình vẽ) để giải mã hình ảnh khi phần A có sử dụng bộ mã hóa.

Dữ liệu thu được từ bộ thu tín hiệu 11b sẽ được chuyển tới bộ nhận dạng ký tự quang học 3, rồi tới bộ xử lý ký tự 4 và tới bộ tổng hợp tiếng nói 5. Dữ sau khi xử lý qua bộ tổng hợp tiếng nói 5 vốn là dữ liệu âm thanh sẽ dược chuyển tới bộ phát tín hiệu 12b.

Bộ phát tín hiệu 12b để phát tín hiệu âm thanh nhận được từ bộ tổng hợp tiếng nói 5.

Theo một phương án khác của sáng chế, bộ phát tín hiệu 112b có thể là bộ phát sử dụng công nghệ wifi 12.3 hoặc 3G 12.4 tương ứng.

Theo một phương án khác của sáng chế thiết bị còn có thêm bộ mã hóa âm thanh, để mã hóa âm thanh.

Theo một phương án của sáng chế, dữ liệu thu được từ bộ thu 11b sẽ được gửi tới máy chủ để thực hiện các công việc của bộ nhận dạng ký tự quang học 3, bộ xử lý ký tự 4 và bộ tổng hợp tiếng nói 5.

Theo một phương án của sáng chế, toàn bộ các bước thực hiện của bộ nhận dạng ký tự quang học 3, bộ xử lý ký tự 4 và bộ tổng hợp tiếng nói 5 sẽ được xử lý bằng điện toán đám mây.

Ngoài ra, phần A còn bao gồm bộ thu tín hiệu âm thanh 12a có thể thu qua công nghệ wifi 12.1 hoặc 3G 12.1, tương ứng

Theo một phương án khác của sáng chế thiết bị còn có thêm bộ giải mã âm thanh (không được thể hiện trên hình vẽ) tương ứng, để giải mã âm thanh thu được. Phần A còn có Bộ khuếch đại âm thanh 7; Loa 10; Bộ điều khiển điện 9; và Nguồn điện 10 như được mô tả trên Hình 2.

Theo một phương án khác của sáng chế, các bước thực hiện của bộ xử lý đa phương tiện 2, bộ nhận dạng ký tự quang học 3, bộ xử lý ký tự 4 và bộ tổng hợp tiếng nói 5, tốt hơn là được thực hiện bởi chương trình máy tính.

Hình 4 thể hiện lưu đồ xử lý cảnh báo tốc độ.

Như được thể hiện trên Hình 4, quá trình xử lý bộ cảnh báo tốc độ 2.3 sẽ được thực hiện theo các bước tuần tự.

Bộ cảnh báo tốc độ có chức năng cảnh báo cho người sử dụng biết nếu tốc độ di chuyển của ngón tay quá nhanh so với văn bản thì thiết bị sẽ phát ra âm thanh tương ứng. Bộ cảnh báo tốc độ đóng vai trò như một bước tiền kiểm tra để tránh trường hợp chất lượng hình ảnh văn bản không tốt gây ra do di chuyển camera quá nhanh.

Dữ liệu thu được sau khi camera 1 thực hiện chức năng ghi hình là một đoạn video được lưu trong bộ nhớ đệm lưu video 2.12.

Đoạn video được lưu này sau đó sẽ được kiểm tra đã được so sánh tốc độ 2.13 hay chưa.

Nếu đã thực hiện việc so sánh tốc độ, dữ liệu sẽ được chuyển cho khối kiểm tra dữ liệu đầu vào 2.12; và nếu xác định chưa thực hiện việc so sánh tốc độ, thì sẽ thực hiện hiện việc so sánh tốc độ.

Nếu xác định là chưa so sánh tốc độ, dữ liệu thu được sẽ được xử lý tiếp bằng cách trích xuất 2s đầu tiên 2.14 của đoạn video.

Camera được sử dụng sẽ thu video gồm 24 hoặc 30 khung hình trên một giây. Do đó mỗi 2 giây dữ liệu chứa 48 hoặc 60 khung hình.

Mỗi 2 khung hình đầu tiên số thứ tự 1 và 25 hoặc 31 sẽ được trích xuất ảnh đầu tiên của mỗi giây 2.15.

Sau đó mỗi khung hình được trích xuất này sẽ được thay đổi kích thước của ảnh thu được 2.16. Thông thường kích thước khung hình được thay đổi về dạng 3x30x100 điểm ảnh.

Hai khung hình này sau đó sẽ được chuyển đổi ảnh về mức xám 2.17, tức mỗi điểm ảnh có giá trị là số nguyên nằm trong dải từ 0 đến 255.

Các bước 2.16 và 2.17 đóng vai trò quan trọng để chuẩn hóa dữ liệu và giảm thời gian tính toán cho trong việc tính tỷ số tương quan chéo (alpha) 2.18.

Chỉ số tương quan chéo là đại lượng thể hiện sự giống nhau giữa hai chuỗi giá trị, ở đây là chuỗi giá trị các điểm ảnh của hai khung hình.

Sau khi tính tỷ số tương quan chéo (alpha) 2.18, giá trị chỉ số tương quan chéo sẽ được chuẩn hóa về khoảng 0 tới 100. Giá trị 0 đạt được khi hai khung hình hoàn toàn khác nhau và giá trị 100 đạt được khi hai khung hình giống hệt nhau. Giá trị thực nghiệm cho thấy nếu giá trị tương quan chéo bé hơn 50 thì tốc độ di chuyển camera quá nhanh và chất lượng hình ảnh không đảm bảo để thực hiện các công đoạn tiếp theo.

Sau đó, bước kiểm tra Alpha bé hơn 50 2.19 sẽ được thực hiện đối với chỉ số tương quan chéo được tạo ra ở bước 2.18. Nếu giá trị bé hơn 50, bộ xử lý cảnh báo tốc độ sẽ thực hiện chức năng cảnh báo "tốc độ cao quá" 2.10 bằng cách phát ra âm thanh "di chuyển nhanh quá".

Khi người sử dụng nghe thấy âm thanh này sẽ có đưa ra điều chỉnh tốc độ một cách phù hợp cho tới khi thiết bị không phát ra âm thanh cảnh báo.

Nếu giá trị lớn hơn 50, dữ liệu sẽ được chuyển về bước đã so sánh tốc độ 2.13 và chuyển sang xử lý tiếp ở bước kiểm tra dữ liệu đầu vào 2.2.

Hình 5 thể hiện lưu đồ kiểm tra dữ liệu.

Như được thể hiện trên Hình 5, lưu đồ thực hiện của khối kiểm tra dữ liệu đầu vào 2.2 được thể hiện. khối kiểm tra dữ liệu đầu vào 2.2 có nhằm phát hiện đoạn video tương ứng với ảnh trắng, chỉ vào vùng hết dòng hay hình vẽ và phát ra các âm thanh tương ứng là "khoảng trắng", "hết dòng" và "hình vẽ". Khi người sử dụng nghe thấy các âm thanh này sẽ thay đổi vị trí ngón tay phù hợp tới các vùng khác để đọc nội dung văn bản.

Trước tiên, khối kiểm tra dữ liệu đầu vào 2.2 sẽ trích xuất 15 giây video từ bộ lưu trữ video.

Khối kiểm tra dữ liệu đầu vào 2.2 sẽ kiểm tra các chuỗi khung hình có trắng, hết dòng hay là hình vẽ hay không. Bước kiểm tra dữ liệu 2.20 sẽ xác định các chức năng này đã được thực hiện hay chưa. Nếu được thực hiện rồi, dữ liệu sẽ được chuyển cho Khối xử lý ảnh 2.5. Nếu chưa được thực hiện, các bước sau sẽ được thực thi.

Bước kiểm tra chuỗi khung hình trắng 2.21a này được thực hiện bằng cách tính biểu đồ phân phối giá trị các điểm ảnh trên một chuỗi các khung hình. Sau đó lấy giá trị trung bình của các phân phối này. Giá trị này nằm trong khoảng từ 0 đến 100 tương ứng với việc ảnh không chứa nội dung gì (ảnh trắng hoàn toàn) hoặc cả khung hình là hình ảnh với tất cả các điểm ảnh đều nhận một giá trị khác 0.

Nếu giá trị trung bình nhỏ hơn ngưỡng giá trị là 5 thì bộ kiểm tra dữ liệu sẽ xác định được đây là đoạn video trắng và sẽ xử lý chuyển sang bước phát ra âm thanh "Khoảng trắng" 2.21b.

Nếu giá trị trung bình lớn hơn 5, các khung hình sẽ được chuyển tới bước kiểm tra hết dòng 2.22a để kiểm tra vị trí ngón tay di chuyển hết dòng chưa.

Việc kiểm tra hết dòng 2.22a được thực hiện bằng cách chia một khung hình thành hai khung hình bé hơn tương đương với việc kẻ một đường thẳng đi qua chính giữa khung hình đó. Sau đó tính giá trị tương quan chéo giữa hai khung hình. Giá trị tương quan chéo thuộc khoảng từ 0 tới 1, tương ứng với việc hai khung hình hoàn khác tới hoàn toàn giống nhau. Nếu giá trị tương quan chéo giữa hai khung hình này nhỏ hơn 0.5 thì bộ kiểm tra dữ liệu sẽ xác định được đây là hết dòng và sẽ xử lý chuyển sang bước phát ra âm thanh "hết dòng" 2.22b.

Nếu giá trị tương quan chéo giữa hai khung hình lớn hơn 0.5 thì các khung hình sẽ được chuyển tới bước xác định hình vẽ 2.23a.

Bước xác định hình vẽ 2.23a để kiểm tra xem khu vực được camera chỉ tới có phải là hình vẽ hay không. Bước xác định hình vẽ 2.23a được thực hiện thông qua việc xác định dòng cơ sở theo chiều dọc của khung hình.

Nếu giá trị hai phần ba số dòng theo chiều dọc của khung hình có giá trị lớn hơn 5 thì bộ kiểm tra dữ liệu sẽ xác định được đây là hình vẽ và sẽ xử lý chuyển sang bước phát ra âm thanh "Hình vẽ" 2.23b.

Nếu giá trị hai phần ba số dòng theo chiều dọc của khung hình có giá trị nhỏ hơn 5 thì bộ kiểm tra dữ liệu sẽ lệnh cho bộ xử lý ảnh xử lý dữ liệu video tiếp theo.

Hình 6 thể hiện lưu đồ thực hiện của bộ xử lý ảnh chữ viết.

Như được thể hiện trên Hình 6, sau khi được kiểm tra dữ liệu đầu vào tại khối 2.2, dữ liệu video được chuyển tới bước nhận một đoạn video 2.31.

Tiếp theo, 15 giây dữ liệu sẽ được trích xuất 2.32.

Tiếp theo, 15s video sẽ được trích xuất mỗi giây chọn 5 khung hình 2.33. Việc trích xuất này nhằm để giảm bớt khối lượng công việc cần tính toán, qua đó tiết kiệm thời gian xử lý của hệ thống. Theo một phương án của sáng chế, mỗi giây có 30 khung hình, ta chọn các khung hình thứ 3, 9, 15, 21, 27.

Sau đó, các khung hình thu được sẽ được xử lý bằng cách chọn ảnh nét nhất 2.34 của 5 khung hình trả lại bởi bước 2.33.

Để tính độ nét của mỗi khung hình/bức ảnh, bộ lọc Sobel được áp dụng trên toàn bộ bức ảnh. Kết quả trả về là mức độ biến thiên của mỗi giá trị điểm ảnh. Sau khi tính tổng các giá trị điểm ảnh này sẽ cho ra giá trị độ nét của mỗi bức ảnh. Giá trị tính được càng cao, bức ảnh càng nét và ngược lại.

Thực tế, trong quá trình di chuyển camera 1 để thu nhận thông tin văn bản, một vấn đề thường gặp là bức ảnh văn bản thu được thường bị nghiêng. Việc bức ảnh bị nghiêng sẽ ảnh hưởng đến kết quả của bộ nhận dạng ký tự, dẫn đến việc nhận dạng không chính xác. Vấn đề đặt ra là cần tính được góc nghiêng và cân chỉnh lại góc nghiêng để có được một bức ảnh có chữ thẳng hàng. Sáng chế đã áp dụng thuật toán Biến đổi Hough vốn là kỹ thuật được sử dụng để tính góc nghiêng của dòng chữ so với biên của ảnh văn bản.

Ảnh nét nhất sẽ được xác định nghiêng 2.35 hay không so với biên.

Nếu xác định dòng chữ bị nghiên, ảnh nét nhất sẽ được chuyển tiếp tới bước xử lý nắn thẳng 2.36 để nắn thẳng dựa trên góc nghiêng đã xác định trước đó tại 2.35.

Nếu chữ trong bức ảnh văn bản được xác định không bị nghiêng 2.35, bức ảnh văn bản được chuyển tới bước xác định ảnh tối 2.37.

Bước xác định ảnh tối 2.37 sẽ xác định độ sáng của văn bản.

Nếu độ sáng nhỏ hơn 50 (trong trường hợp ảnh mức xám, mỗi giá trị điểm ảnh thuộc dải từ 0 tới 255 bức ảnh bị xác định là tối thì bộ xử lý ảnh sẽ chuyển ảnh tới bước tăng sáng 2.38 để tăng cường mức sáng cho ảnh.

Theo một phương án của sáng chế, một số phương pháp có thể được sử dụng để tăng cường sáng như cân bằng độ phân bố, biến đổi gama hay thuật toán retinex. Nếu độ sáng lớn hơn 50, bộ xử lý ảnh 2.5 sẽ xác định bức ảnh văn bản không bị tối và chuyển dữ liệu ảnh này tới bước xác định nhiễu 2.39 hay không.

Độ nhiễu của bức ảnh được xác định trên tỉ số tín hiệu trên nhiễu SNR (Signal to Noise Ratio).

Nếu bức ảnh được xác định là nhiễu, bộ xử lý ảnh 2.5 sẽ chuyển bước ảnh tới bước loại nhiễu 2.310.

Theo một phương án của sáng chế, thuật toán bộ lọc trung bình (Median filter) được sử dụng để loại bỏ nhiễu.

Nếu bức ảnh được xác định không nhiều, bộ xử ảnh 2.5 sẽ chuyển bức ảnh tới bước phân đoạn ảnh 2.311 để khoanh vùng đoạn chữ cần đọc.

Thực tế, khi đặt camera lên gá đỡ đeo bởi ngón tay, khi camera tiến hành ghi hình, thông thường đầu ngón tay sẽ lọt vào khung hình. Người sử dụng được hướng dẫn để di chuyển camera sao cho vị trí của đầu ngón tay sẽ nằm tại khoảng trắng giữa hai dòng chữ và dưới đoạn chữ cần đọc. Do vị trí đặt giá đỡ, khung hình thu được có thể chứa một hoặc (thường là) nhiều dòng chữ và hình ảnh của đầu ngón tay. Vị trí của đầu ngón tay trong khung hình chính là điểm mốc để xác định dòng chữ và đoạn chữ cần được trích xuất trong công đoạn phân đoạn cụm chữ viết. Phân đoạn ảnh tiếng Việt cần phải tính đến các dấu và ký tự đặc biệt của tiếng Việt ví dụ các từ "để", "đối", "đắng"… Do khoảng cách của dấu thanh điệu sắc, huyền, hỏi, ngã của các chữ ở dòng dưới thường nằm rất gần với các chữ ở dòng trên, việc phân đoạn chữ in tiếng Việt thường khó khăn hơn việc phân đoạn chữ in tiếng Anh. Sau khi phân đoạn ảnh 2.311, ảnh thu thường chứa cụm chữ viết in sẽ được chuyển tới bộ nhận dạng ký tự quang học 3.

Hình 7 thể hiện lưu đồ khối các bước thực hiện của bộ nhận dạng ký tự quang học cho tiếng Việt và tiếng Anh.

Như được thể hiện trên Hình 7, các bước thực hiện của bộ nhận dạng ký tự quang học sẽ được bộc lộ. Ảnh văn bản đã được xử lý ở bộ xử lý ảnh 2.5 sẽ được thực hiện việc dò tìm chuỗi ký tự 3.1 để nhận dạng.

Chuỗi ký tự này bao gồm các từ đơn riêng lẻ được phân cách bởi khoảng trắng. Một số ký tự không đầy đủ lọt vào khung hình sẽ được loại bỏ. Bộ ký tự tiếng Việt bao gồm tập ký tự không dấu {A, B, C, D, Đ, E, G, H, I, K, L, M, N, O, Q, R, S, T, U, V, X, Y} và các ký tự có dấu {Ă, Â, À, Ả, Ã, Á, Ạ, Ằ, Ắ, Ẳ, Ẵ, Ặ,

Ầ, Ấ, Ẩ, Ẫ, Ậ, Ê, È, Ẻ, É, Ẹ, Ẽ, Ề, Ế, Ể, Ễ, Ệ, Ì, Ỉ, Í, Ĩ, Ị, Ô, Ơ, Ò, Ỏ, Õ, Ó, Ọ, Ồ, Ổ, Ố, Ộ, Ỗ, Ờ, Ở, Ớ, Ỡ, Ợ, Ư, Ừ, Ử, Ứ, Ữ, Ự, U, Ù, Ủ, Ú, Ũ, Ụ, Ỳ, Ỷ, Ỹ, Ý, Ỵ}.

Sau khi xác định được chuỗi ký tự tại bước dò tìm chuỗi ký tự 3.1, bộ nhận dạng ký tự quang học sẽ tiến hành việc phân vùng ký tự 3.2 để phân tách các ký tự trong từng từ đơn dựa vào khoảng trắng bé hơn nằm giữa mỗi hai từ đơn.

Tiếp theo, các đặc trưng về chiều cao, độ rộng, số nét chữ, số vùng liên thông của mỗi ký tự được trích xuất ở bước trích chọn đặc trưng 3.3. Dựa vào số thành phần liên thông, tiếng Việt được tách thành 3 nhóm có 1 vùng liên thông {A, B, C…}, 2 vùng liên thông {Ă, Â, Ơ, Ô…} và 3 vùng liên thông {Ẳ, Ấ, Ố, Ở…}.

Các đặc trưng này sau đó được chuyển tới bước phân lớp bằng mô hình học máy 3.4. Hiện nay có các mô hình phân loại bằng học máy (ví dụ Máy vect –tơ hỗ trợ hoặc Mạng Nơ-ron học sâu) để phân loại thành các ký tự có thể sửa đổi bằng phần mềm xử lý văn bản trên máy tính.

Kết quả của quá trình nhận dạng ký tự là các chuỗi ký tự tiếng Việt hoặc tiếng Anh được chuyển tới bộ xử lý ký tự 4.

Hình 8 thể hiện lưu đồ các bước thực hiện của bộ trích ghép ký tự 4.2

Như được thể hiện trên Hình 8 các bước thực hiện của bộ xử lý ký tự sẽ được bộc lộ.

Dữ liệu thu được của bộ nhận dạng ký tự quang học 3 sẽ được lưu vào bộ lưu ký tự 4.1 dưới dạng tệp tin dữ liệu có thể chỉnh sửa được bằng các chương trình soạn thảo văn bản vốn là các chuỗi ký tự có thể sửa được bởi các chương trình soạn thảo văn bản.

Bộ xử lý ký tự 4 sẽ thực hiện loại bỏ ký tự không có nghiã 5.1a, thường là các ký tự ở vị trí đầu hoặc cuối đoạn dữ liệu. Các ký tự này sinh ra khi hình ảnh văn bản đầu vào được camera thu nhận không chứa hết mà chỉ chứa một phần hình ảnh của các từ trong văn bản.

Tiếp theo, bộ xử lý ký tự sẽ so sánh hai đoạn ký tự 5.1, việc so sánh mỗi hai đoạn ký tự này thường dài khoảng 3 đến 5 chữ nếu áp dụng cho ngôn ngữ tiếng Việt Nam.

Nếu so sánh phát hiện có chữ trùng 5.2, bộ xử lý ký tự sẽ chuyển tới bước đối chiếu vị trí vật lý 5.3 của đoạn chữ thông qua việc xem xét lại các ảnh chứa các đoạn chữ này.

Việc đối chiếu vật lý 5.3 cho biết vị trí của các đoạn chữ và góp thêm căn cứ để bước loại bỏ phần giống 5.4 thực hiện vốn để loại bỏ phần chữ giống nhau giữa các đoạn chữ.

Nếu bước phát hiện trùng có kết quả sai, hoặc bước loại bỏ phần giống hoàn thành, thì bộ xử lý ký tự sẽ thực hiện bước ghép các đoạn ký tự 5.5 để ghép nối các đoạn ký tự sao cho thu được một đoạn ký tự duy nhất.

Sau đó, kết quả thu được tại bước ghép các đoạn ký tự 5.5 sẽ được chuyển tới bước bước xác định hết chữ trùng 5.6 kiểm tra xem đã hết chữ trùng nhau.

Nếu xác định hết chữ trùng thì chuyển sang bước xuất chữ 5.7 cho các khối tiếp theo.

Nếu xác định sai thì quay lại bước 5.1.

Hình 9 thể hiện lưu đồ thực hiện bộ tổng hợp tiếng nói.

Hình 9 thể hiện lưu đồ thực hiện bộ tổng hợp tiếng nói với đầu vào là chuỗi từ được trả bởi Bộ xử lý ký tự 4.

Bước 6.1 thực hiện việc xác định từ là số, từ viết tắt hay tên riêng. Việc xác định từ là số hay từ viết tắt có thể được thực hiện bằng cách đối chiếu từ cần xác định với bộ từ vựng được lưu trữ. Việc xác định tên riêng (Name entity recognition) được xác định dựa trên các thông tin như ký tự viết hoa hay viết thường, đứng đầu câu hay giữa câu bằng các ký thuật học máy như Chuỗi Markov ẩn (Hidden Markov Model).

Xác định từ loại 6.2 là việc phân chia các từ xuất hiện trong văn bản thành danh từ, động từ, tính từ, trạng từ… 6.2 có thể được thực hiện bằng nhiều phương

pháp trong đó có thể kể đến phương pháp Trường ngẫu nhiên có điều kiện (Conditional Random Field) hay Tối đa hóa biến thiên (Maximum Entropy).

Trong phương pháp này, mô hình phân loại từ được huấn luyện dựa trên tập các tập dữ liệu huấn luyện với mục đích phân chia được các loại từ dựa trên ngữ pháp của câu. Tính chất của tập dữ liệu huấn luyện quyết định chất lượng của việc phân loại từ. Theo đó, tập dữ liệu càng lớn và càng đa dạng về từ loại, ngữ nghĩa thì mô hình phân loại càng đưa ra kết quả chính xác.

Bước 6.3 thực hiện việc xem xét ngữ cảnh, tức là bối cảnh ngôn ngữ trong đó từ được sử dụng. Vai trò của ngữ cảnh là cơ sở cho việc lựa chọn nội dung cách thức ngôn ngữ được sử dụng từ đó chương trình sẽ đưa ra phiên âm phù hợp ngữ cảnh 6.4. Xác định đúng ngữ điệu, nhấn mạnh, độ kéo dài phần phát âm và phần nghỉ từ văn bản là một vấn đề quan trọng trong các hệ thống tổng hợp tiếng nói. Những đặc tính này gọi chung là ngôn điệu, tức là cách diễn đạt hay các đặc tính siêu đoạn và có thể được xem như giai điệu, nhịp điệu và sự nhấn mạnh của tiếng nói. Việc xác định ngữ cảnh 6.4 đối với tiếng Việt được tính tới cả sự thay đổi cao độ và trường độ các âm vị trong một âm tiết, thường là diễn tả một âm tiết khác.

Bước 6.5 thực hiện việc chuyển phiên âm thành tiếng nói tổng hợp dựa trên từ điển phát âm hoặc quy luật ngôn ngữ. Khác với tiếng anh, đặc điểm của tiếng Việt là ngôn ngữ đơn âm tiết, không có sự luyến âm, nuốt âm khi cấu âm. Một âm tiếng Việt có thể được chia thành 3 thành phần có mối liên kết gồm phụ âm đầu, vần và dấu thanh. Số lượng âm vị trong tiếng Việt là 39 và cấu thành nên khoảng 1600 bán âm vị và ngữ cảnh. Số lượng âm tiết khoảng 7000 là các âm tiết hay dùng nhất. Số lượng phụ âm gồm phụ âm vần có/không có dấu là khoảng 1000.

Có hai đặc tính cơ bản đánh giá chất lượng của giọng tổng hợp là tính tự nhiên và dễ hiểu. Hai đặc tính này được cải thiện bằng việc xác định tần số cơ bản, cường độ và trường độ 6.6 cũng như xác định vị trí âm vị so với âm vị khác 6.7.

Sau khi đã xác định được các đặc trưng ở các bước trên, đã biết nhiều phương pháp để tổng hợp tiếng nói 6.8 tiếng Việt dựa trên ghép nối, tổng hợp Formant hoặc phương pháp chuỗi xích.

Đầu ra của 6.8 là tín hiệu âm thanh thể hiện nội dung văn bản và được gửi tới bộ khuếch đại âm thanh 7.

YÊU CẦU BẢO HỘ

1. Thiết bị đọc các chữ in tiếng Việt và phát thành giọng nói có thể mang theo người và có thể tháo lắp được, thiết bị này bao gồm:

bộ thu video (1) để thu nhận dữ liệu hình ảnh/video, bộ thu video này được lắp trên gá đỡ (1a) có dạng vòng tròn như chiếc nhẫn, sao cho gá đỡ này có thể lồng vào ngón tay người, khi ngón tay di chuyển, bộ thu video sẽ thu hình ảnh/video phía trên ngón tay; và

bộ xử lý trung tâm (50) để xử lý dữ liệu hình ảnh/video thu được từ bộ thu video (1) thông qua phương tiện kết nối để chuyển thành giọng nói, trong đó bộ xử lý trung tâm này bao gồm:

bộ xử lý đa phương tiện (2) để xử lý dữ liệu hình ảnh/video thu được sao cho dữ liệu lối ra của bộ xử lý đa phương tiện này là ảnh gồm chuỗi các chữ in, có số lượng nằm trong khoảng từ 3 đến 5 chữ để phù hợp với tốc độ di chuyển ngón tay và góc mở của camera, để nhận dạng ký tự quang học, trong đó bộ xử lý đa phương tiện này bao gồm:

bộ cảnh báo tốc độ để cảnh báo tốc độ khi người sử dụng di chuyển ngón tay quá nhanh,

bộ kiểm tra dữ liệu đầu vào (2.4) để thông báo cho người sử dụng biết khoảng trắng, hết dòng hoặc hình vẽ,

bộ xử lý ảnh (2.5) để xử lý dữ liệu đầu vào sao cho có kết quả tốt nhất cho bộ nhận dạng ký tự quang học, như chỉnh nghiêng dựa trên đo góc nghiêng của dòng chữ với biến đổi Hough hoặc biến đổi tương tự, khử nhiễu với bộ lọc trung vị (median filter), tăng cường sáng cho ảnh dựa trên thuật toán biến đổi Retinex hoặc thuật toán biến đổi tương tự;

bộ nhận dạng ký tự quang học (3) để chuyển đổi hình ảnh các chữ thu được ở đầu ra của bộ xử lý ảnh thành chữ dưới dạng ký tự đọc được bằng máy tính;

bộ xử lý ký tự (4) để xử lý ký tự nhận dạng thu được để ghép nhận dạng các ký tự tránh các trường hợp bị trùng, trong đó, bộ xử lý ký tự này bao gồm bộ nhớ lưu ký tự (4.1) và bộ trích ghép ký tự (4.2);

bộ tổng hợp tiếng nói (5) sẽ chuyển các chữ đã được xử lý trong bộ xử lý ký tự (4) thành giọng nói, trong đó, bộ tổng hợp tiếng nói bao gồm:

khối xử lý ngôn ngữ tự nhiên tiếng Việt (5.1) và khối xử lý tổng hợp tiếng nói tiếng Việt(5.2),

bộ khuếch đại âm thanh (7) để khuếch đại âm thanh từ bộ tổng hợp tiếng nói,

loa (10) để phát ra âm thanh;

bộ điều khiển điện (8) để điều khiển và cấp nguồn điện cho các bộ phận nêu trên; và

nguồn điện (9) để cấp điện cho bộ điều khiển điện.

2. Thiết bị theo điểm 1, trong đó thiết bị này còn được tạo cấu hình để có thể đọc các chữ in tiếng Anh và phát thành giọng nói.

3. Thiết bị theo điểm bất kỳ trong số các điểm nêu trên, trong đó bộ xử lý ký tự (4) và bộ tổng hợp tiếng nói (5) có thể được tạo cấu hình để xử lý văn bản tiếng Việt và tiếng Anh.

4. Thiết bị theo điểm bất kỳ trong số các điểm nêu trên, trong đó bộ nhận dạng ký tự quang học (3), bộ xử lý ký tự (4), bộ tổng hợp tiếng nói (5) được tích hợp trên máy chủ từ xa hoặc tích hợp bằng công nghệ điện toán đám mây.

5. Thiết bị theo điểm 1, trong đó thiết bị này còn bao gồm:

bộ nhận biết hết dòng khi bộ thu video di chuyển đến khoảng trắng, bộ nhận biết hết dòng này sẽ nhận thấy hết dòng;

bộ cảm biến độ nghiêng để nhận biết độ nghiêng của ngón tay người sử dụng khi bộ thu video được lắp tháo được trên ngón tay của người sử dụng;

bộ cảm biến tốc độ di chuyển ngón tay để xác định tốc độ ngón tay di chuyển, nếu tốc độ ngón tay di chuyển quá nhanh so với tốc độ cho trước.

6. Phương pháp xử lý để đọc các chữ in tiếng Việt và phát thành giọng nói, phương pháp này bao gồm:

thu dữ liệu hình ảnh/video phía trên ngón tay người sử dụng bởi bộ thu video (1), trong đó hình ảnh này được thu nhờ bộ thu video này được lắp trên gá đỡ (1a) có dạng vòng tròn như chiếc nhẫn, sao cho gá đỡ này có thể lồng vào ngón tay người, khi ngón tay di chuyển, bộ thu video sẽ thu hình ảnh phía trên ngón tay;

xử lý dữ liệu hình ảnh/video thu được bởi bộ xử lý đa phương tiện (2), trong đó dữ liệu lối ra của bộ xử lý đa phương tiện này là ảnh gồm chuỗi các chữ in, có số lượng nằm trong khoảng từ 3 đến 5 chữ để phù hợp với tốc độ di chuyển ngón tay và góc mở của camera, để nhận dạng ký tự quang học, trong đó bước xử lý đa phương tiện còn bao gồm:

cảnh báo tốc độ bởi bộ cảnh báo dữ liệu khi người sử dụng di chuyển ngón tay quá nhanh,

kiểm tra dữ liệu đầu vào bởi bộ kiểm tra dữ liệu đầu vào (2.4) để thông báo cho người sử dụng biết khoảng trắng, hết dòng hoặc hình vẽ,

xử lý ảnh bởi bộ xử lý ảnh (2.5) để xử lý dữ liệu đầu vào sao cho có kết quả tốt nhất cho bộ nhận dạng ký tự quang học, như chỉnh nghiêng dựa trên đo góc nghiêng của dòng chữ với biến đổi Hough hoặc biến đổi tương tự, khử nhiễu với bộ lọc trung vị (median filter), tăng cường sáng cho ảnh dựa trên thuật toán biến đổi Retinex hoặc thuật toán biến đổi tương tự;

nhận dạng ký tự quang học bởi bộ nhận dạng ký tự quang học (3) để chuyển đổi hình ảnh các chữ in tiếng Việt, thu được ở đầu ra của bộ xử lý ảnh thành chữ dưới dạng ký tự đọc được bằng máy tính;

xử lý ký tự bởi bộ xử lý ký tự (4) để xử lý ký tự nhận dạng thu được để ghép nhận dạng các ký tự tránh các trường hợp bị trùng;

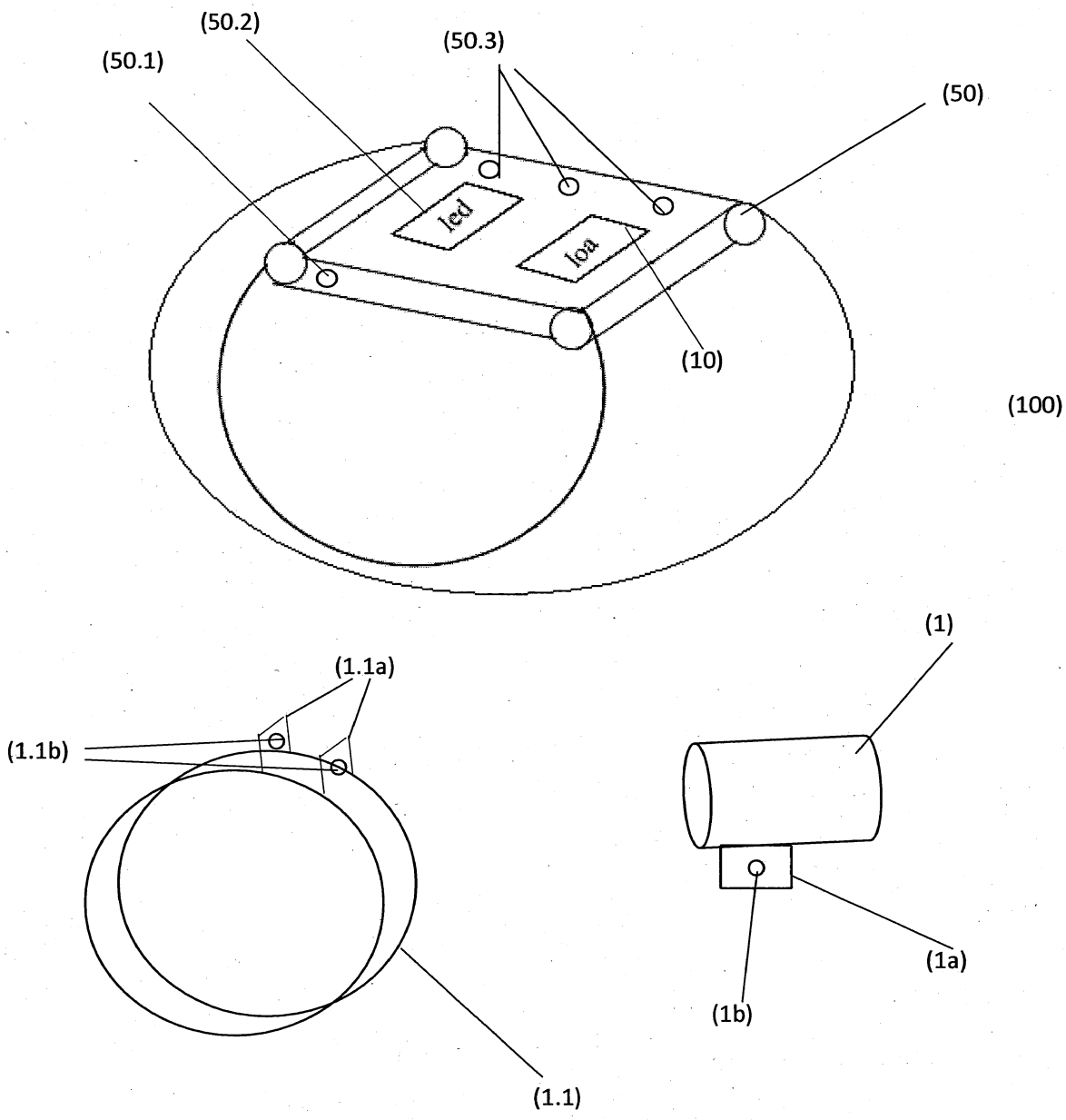
tổng hợp tiếng nói bởi bộ tổng hợp tiếng nói (5) để chuyển các chữ đã được xử lý trong bộ xử lý ký tự (4) thành giọng nói tiếng Việt;

khuếch đại âm thanh bởi bộ khuếch đại âm thanh (7) để khuếch đại âm thanh từ bộ tổng hợp tiếng nói.

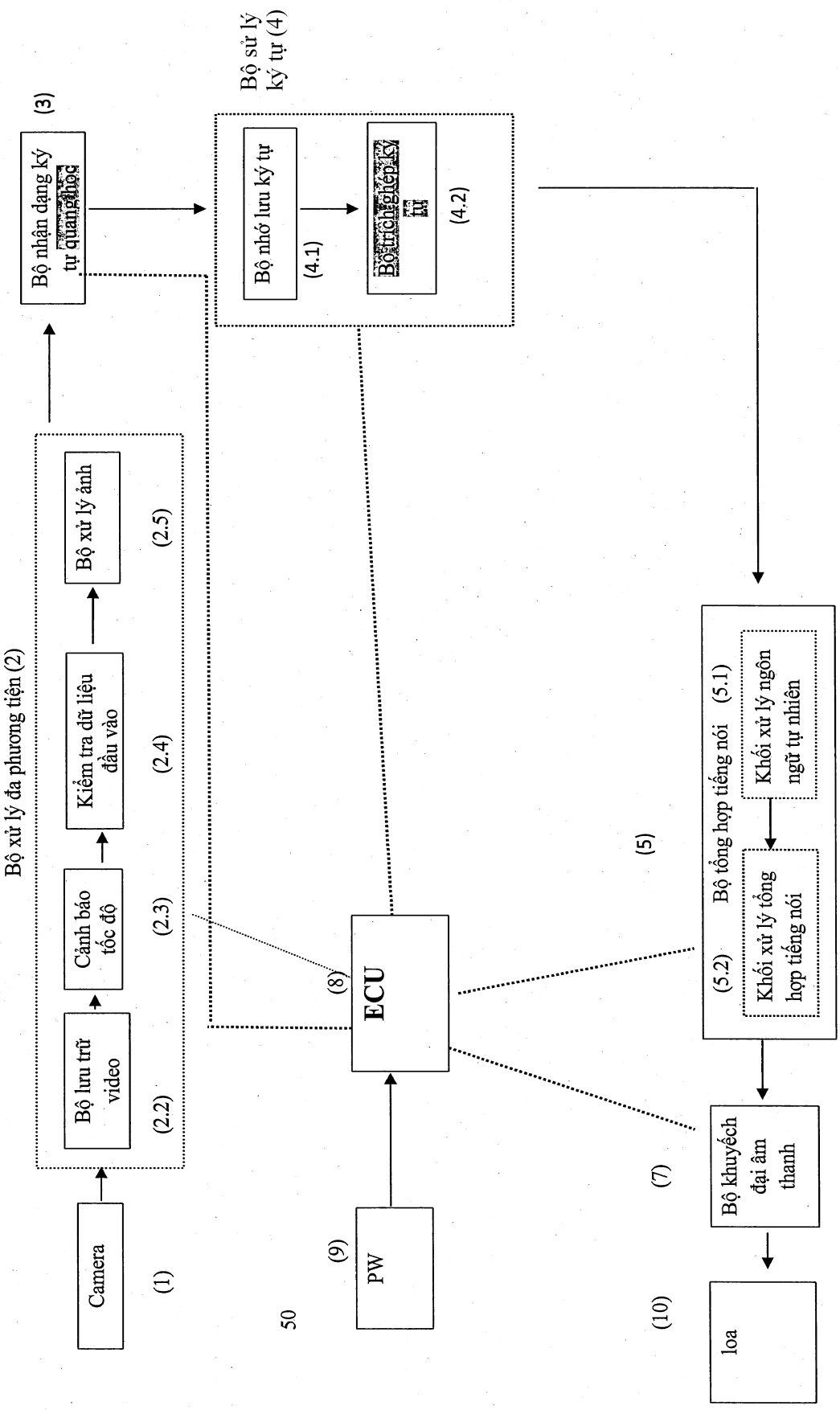
7. Phương pháp theo điểm 6, trong đó bước xử lý dữ liệu hình ảnh/video và phát ra tiếng nói còn được tạo cấu hình để có thể đọc các chữ in tiếng Anh và phát thành giọng nói.

8. Phương pháp theo điểm 6, trong đó bộ xử lý ký tự (4) và bộ tổng hợp tiếng nói (5) có thể được tạo cấu hình để xử lý văn bản tiếng Việt và tiếng Anh.

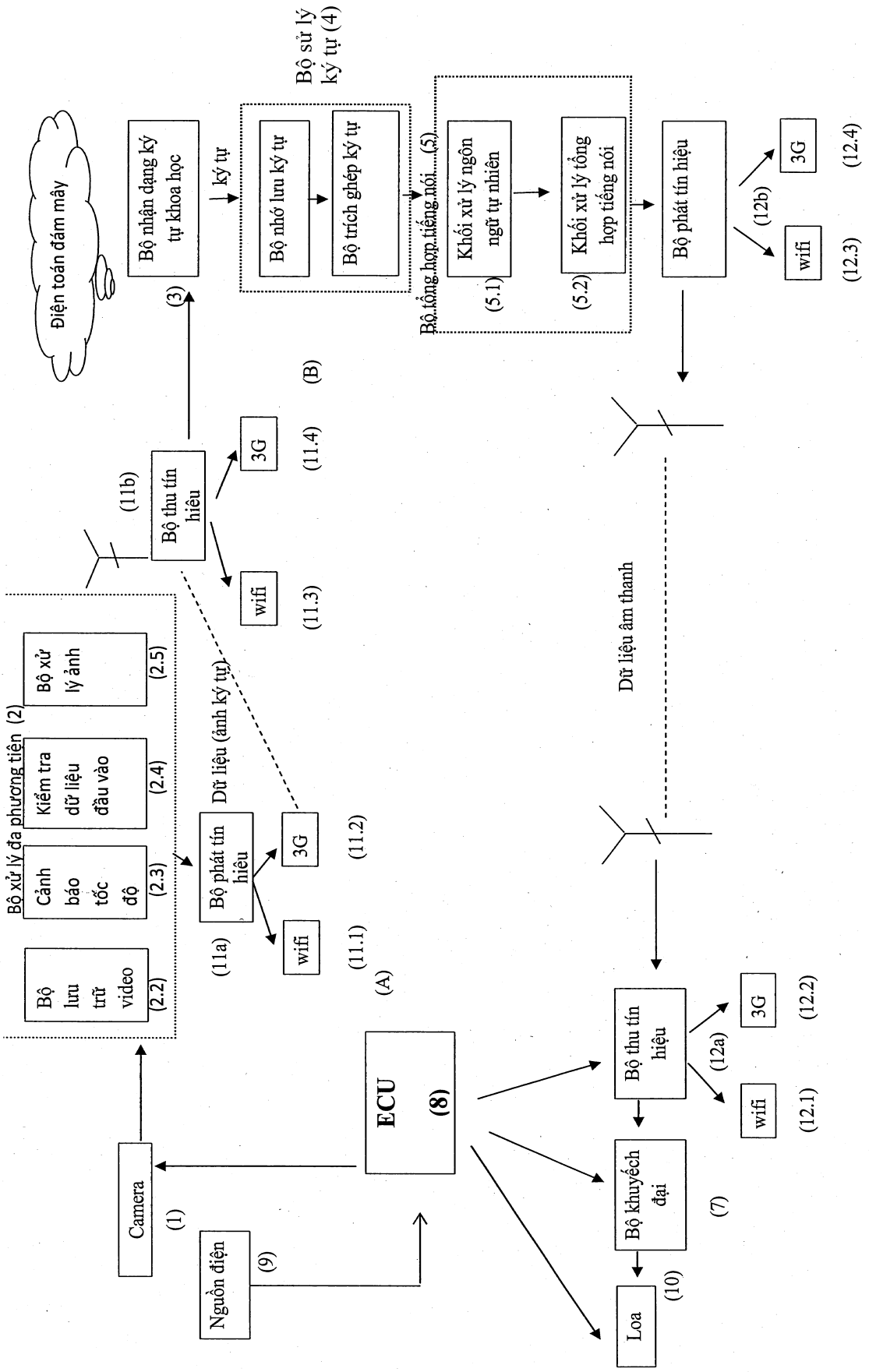
9. Vật ghi ghi chương trình máy tính có thể đọc được bằng máy tính để khi được chạy trên máy tính, chương trình máy tính này có thể làm cho máy tính thực hiện phương pháp theo điểm bất kỳ trong số các điểm từ 6 đến 8.



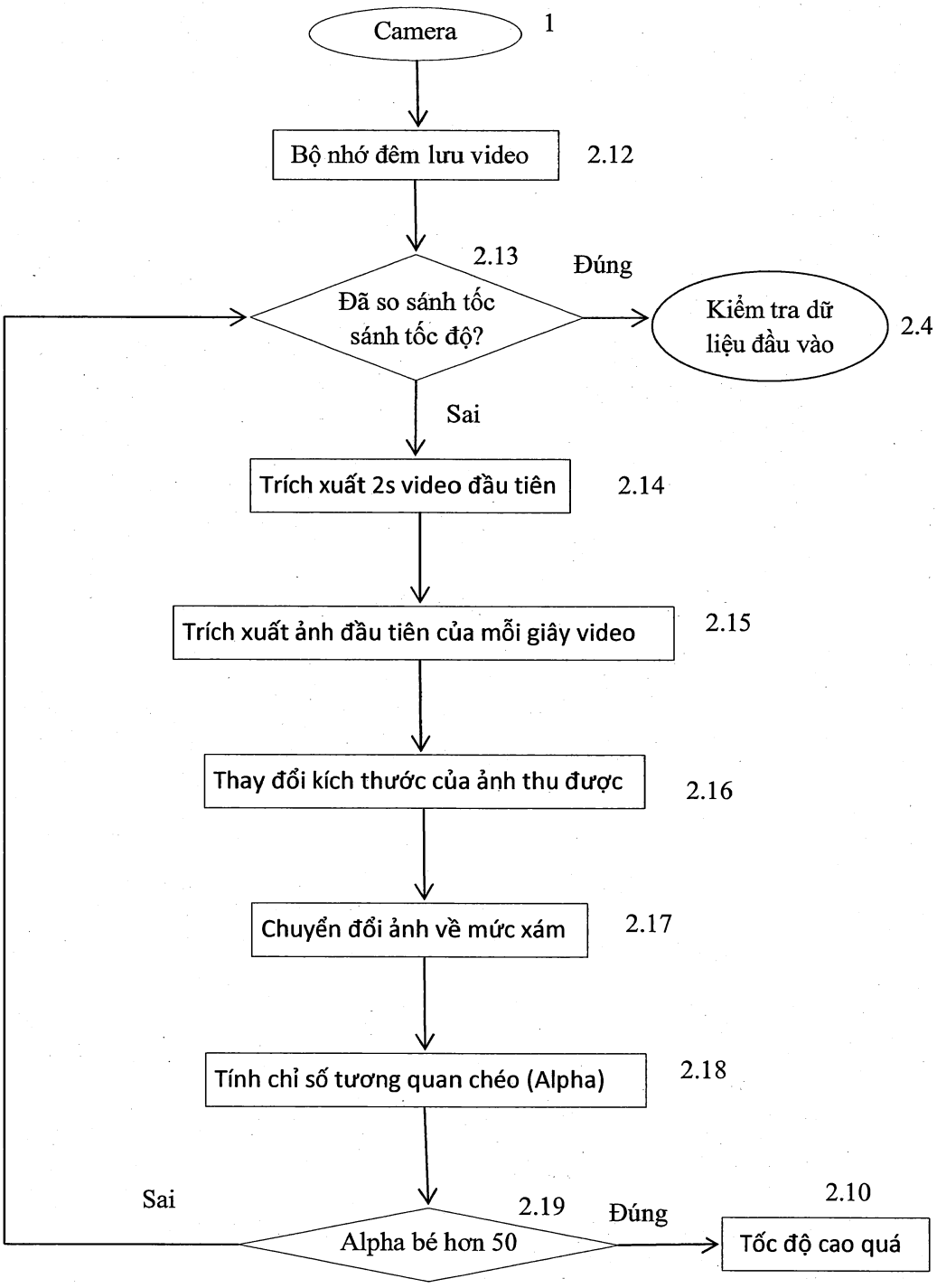
HÌNH 1


Bộ xử lý đa phương tiện (2)
Camera
(1)
Bộ lưu trữ video
(2.2)
Cảnh báo tốc độ
(2.3)
Kiểm tra dữ liệu đầu vào
(2.4)
Bộ xử lý ảnh
(2.5)
Bộ nhận dạng ký tự quang học
(3)
Bộ sử lý ký tự (4)
Bộ nhớ lưu ký tự
(4.1)
Bộ trích chọn ký tự
(4.2)
50
ECU
(8)
PW
(9)
(5)
Bộ tổng hợp tiếng nói (5.1)
Khối xử lý ngôn ngữ tự nhiên
(5.2)
Khối xử lý tổng hợp tiếng nói
(7)
Bộ khuyếch đại âm thanh
(10)
loa

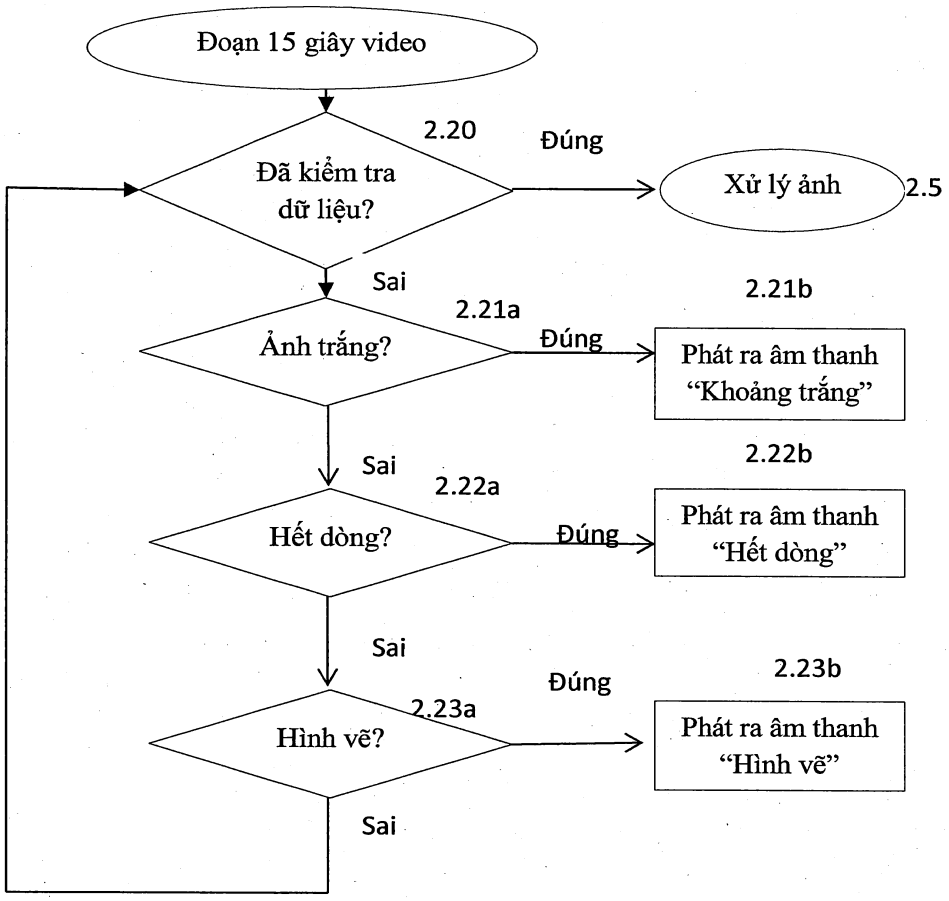
HÌNH 2


Điện toán đám mây
Bộ nhận dạng ký tự khoa học
(3)
ký tự
Bộ sử lý ký tự (4)
Bộ nhớ lưu ký tự
Bộ trích ghép ký tự
Bộ tổng hợp tiếng nói (5)
Khối xử lý ngôn ngữ tự nhiên
(5.1)
Khối xử lý tổng hợp tiếng nói
(5.2)
Bộ phát tín hiệu
(12b)
3G
wifi
(12.3)
(12.4)
(11b)
Bộ thu tín hiệu
3G
(11.4)
(B)
wifi
(11.3)
Dữ liệu (ảnh ký tự)
Dữ liệu âm thanh
Bộ xử lý đa phương tiện (2)
Bộ xử lý ảnh
(2.5)
Kiểm tra dữ liệu đầu vào
(2.4)
Cảnh báo tốc độ
(2.3)
Bộ lưu trữ video
(2.2)
(11a)
Bộ phát tín hiệu
3G
(11.2)
wifi
(11.1)
(A)
Camera
(1)
ECU
(8)
Bộ thu tín hiệu
(12a)
3G
wifi
(12.1)
(12.2)
Bộ khuyếch đại
(7)
Loa
(10)
Nguồn điện
(9)

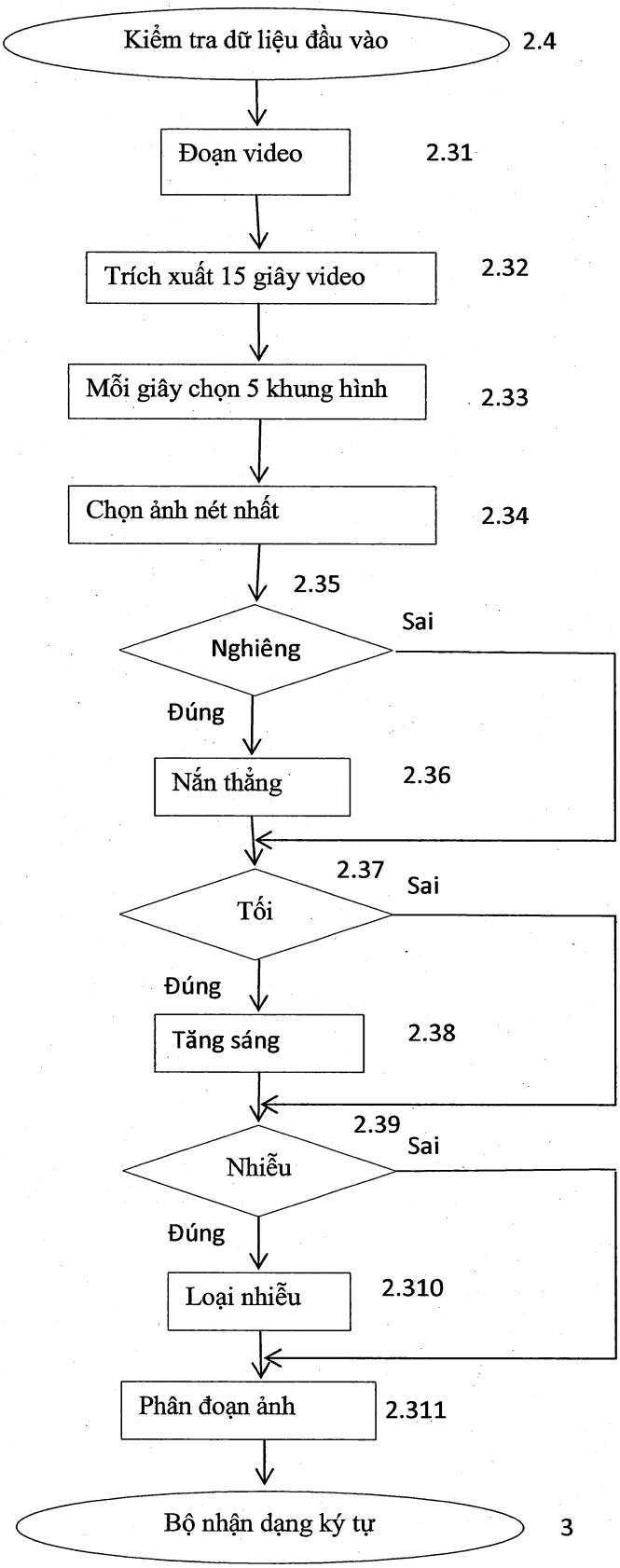
HÌNH 3



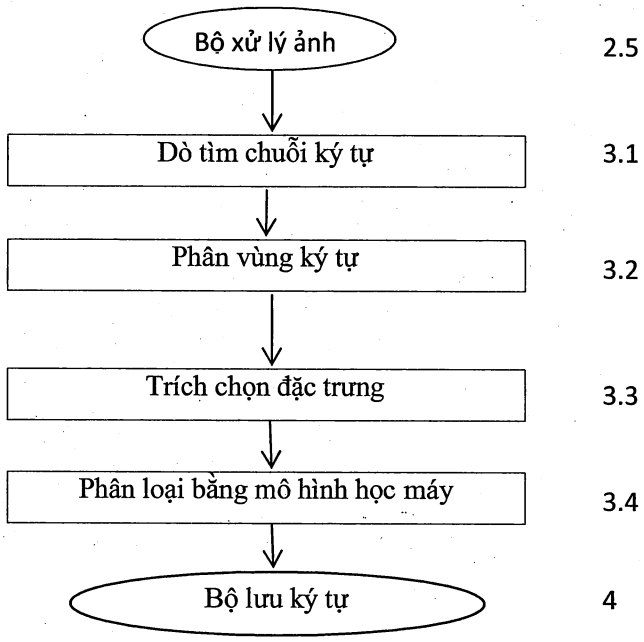
Hình 4. Lưu đồ xử lý cảnh báo tốc độ



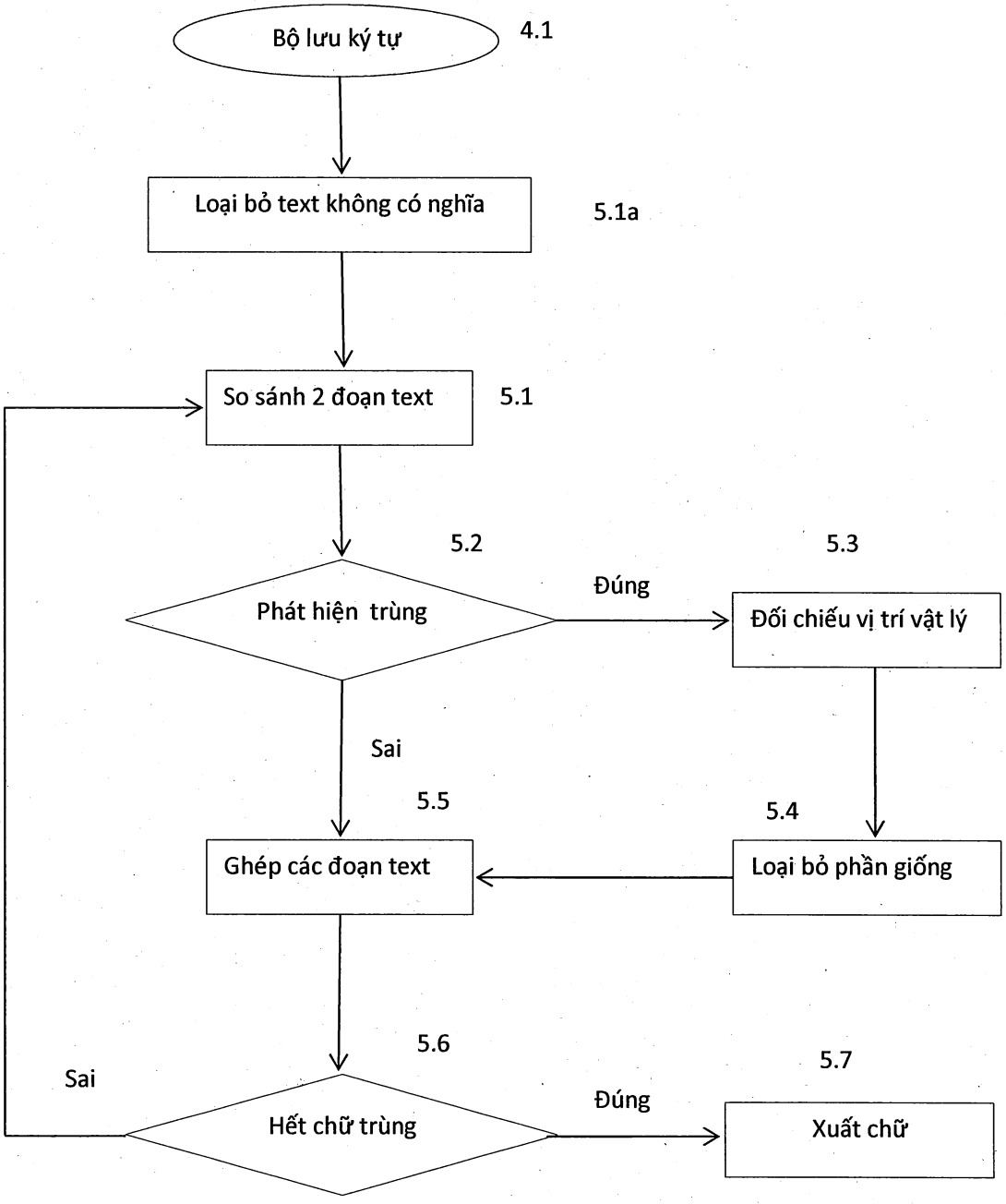
Hình 5. Lưu đồ kiểm tra dữ liệu



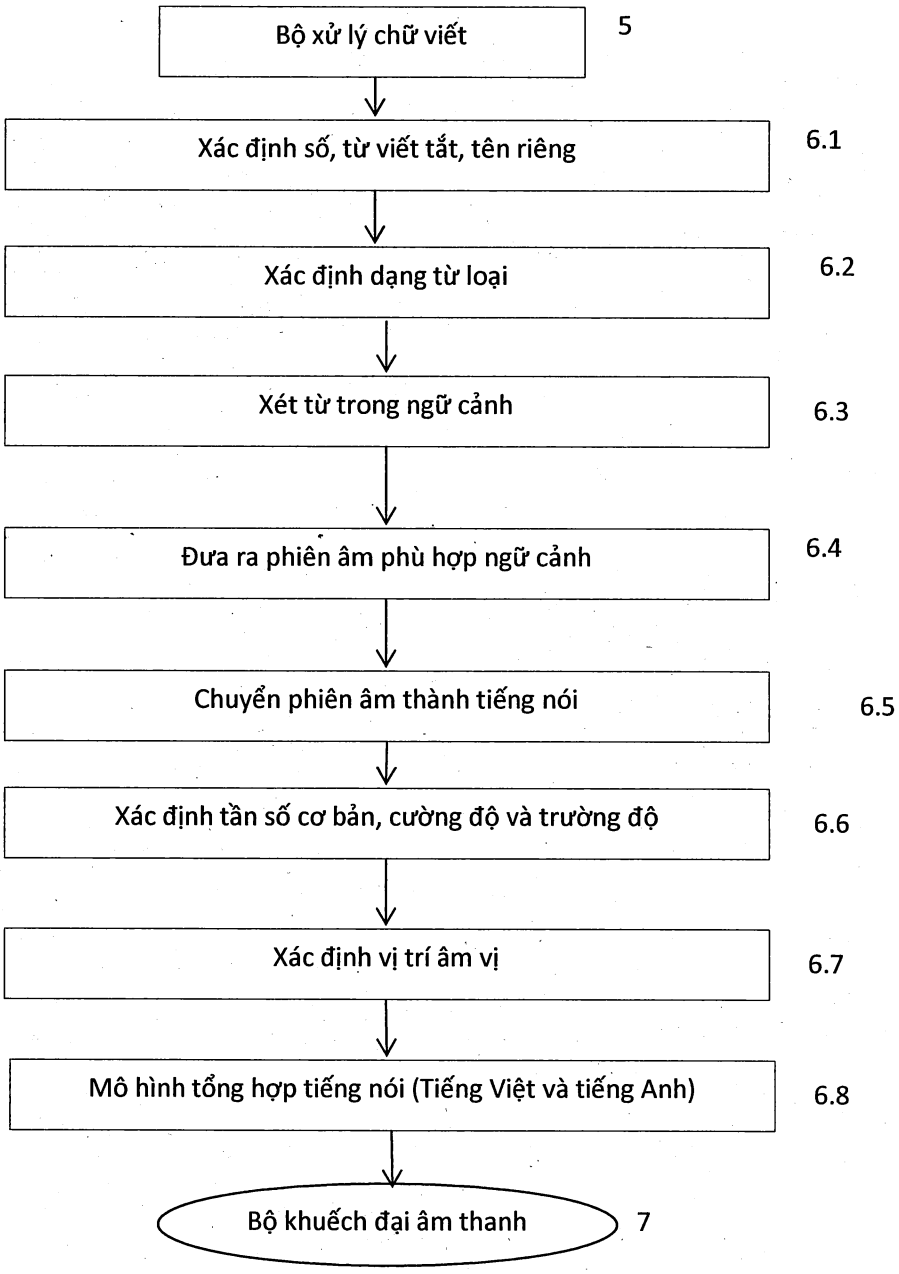
Hình 6. Bộ xử lý ảnh chữ viết



Hình 7.Lưu đồ khối nhận dạng ký tự



Hình 8. Lưu đồ Bộ trích ghép ký tự



Hình 9. Lưu đồ thực hiện tổng hợp tiếng nói