



(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)



1-0026100

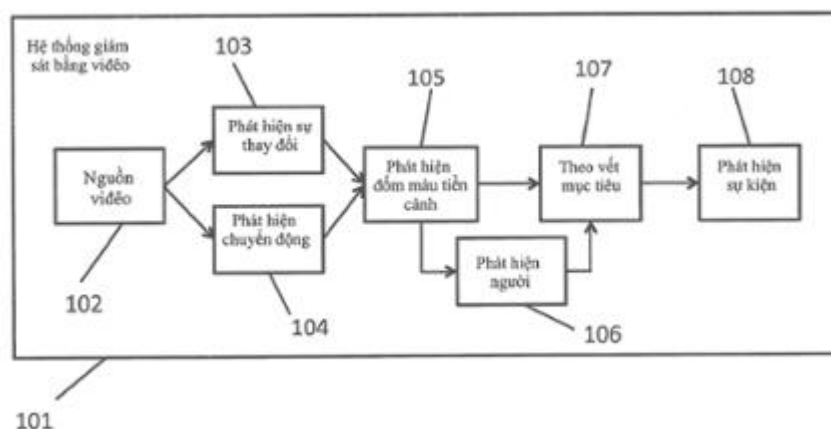
(51)⁷ G06K 9/00

(13) B

(21) 1-2015-00830 (22) 12/09/2013
(86) PCT/US2013/059471 12/09/2013 (87) WO 2014/043353 20/03/2014
(30) 61/700,033 12/09/2012 US; 13/838,511 15/03/2013 US
(45) 26/10/2020 391 (43) 25/05/2015 326A
(73) Avigilon Fortress Corporation (CA)
2900 - 550 Burrard Street, Vancouver, British Columbia, Canada, V6C 0A3
(72) Zhong ZHANG (CN); Weihong YIN (CN); Peter VENETIANER (HU).
(74) Công ty TNHH Tầm nhìn và Liên danh (VISION & ASSOCIATES CO.LTD.)

(54) PHƯƠNG PHÁP VÀ HỆ THỐNG PHÁT HIỆN ĐỐI TƯỢNG TRONG VIDEO

(57) Sáng chế đề cập đến phương pháp và hệ thống để thực hiện việc phân tích nội dung video để phát hiện người hoặc các đối tượng khác cần quan tâm trong hình ảnh video. Việc phát hiện người có thể được sử dụng để đếm số người, để phát hiện vị trí của mỗi người và/hoặc thực hiện việc phân tích đám đông của mỗi khu vực cần theo dõi.



Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập đến việc giám sát bằng video, như hệ thống và phương pháp giám sát bằng video và hệ thống và phương pháp xác nhận bằng video. Hệ thống, thiết bị và phương pháp giám sát bằng video được đề cập có thể phát hiện người. Hệ thống, thiết bị và phương pháp giám sát bằng video có thể đếm người và/hoặc theo dõi tình huống đông người trong các dòng video.

Tình trạng kỹ thuật của sáng chế

Hệ thống giám sát thông minh bằng video (Intelligent Video Surveillance - IVS) có thể được sử dụng để phát hiện các sự kiện quan trọng trong các nguồn video theo thời gian thực hoặc ngoại tuyến (chẳng hạn, bằng cách xem lại hình ảnh được ghi lại và được lưu trữ trước đó). Tác vụ này thường được hoàn thành bằng cách phát hiện và theo dõi các mục tiêu quan trọng. Điều này thường diễn ra thuận lợi khi hiện trường không đông đúc. Tuy nhiên, tính năng của hệ thống như vậy có thể giảm đáng kể khi hiện trường đông đúc. Trên thực tế, hiện trường đông đúc như vậy diễn ra thường xuyên, do đó khả năng để có thể phát hiện người trong đám đông là mối quan tâm lớn. Khả năng phát hiện người theo cách đó có thể được sử dụng để đếm và phân tích đám đông khác, như mật độ đám đông, sự hình thành đám đông và sự phân tán đám đông.

Việc phân tích đám đông trước đây giúp giải quyết một số tình huống tập trung rất đông người đặc biệt như các sự kiện thể thao hoặc tôn giáo nào đó. Tuy nhiên, cũng có nhu cầu tập trung vào các tình huống giám sát bình thường hơn ở nơi mà các đám đông lớn đôi khi có thể hình thành. Các nơi này bao gồm nơi công cộng như đường phố, trung tâm mua sắm, sân bay, bến xe buýt và nhà ga, v.v..

Gần đây, vấn đề ước lượng mật độ đám đông hoặc đếm người trong đám đông đang dành được sự quan tâm lớn trong cộng đồng nghiên cứu cũng như từ ngành công

nghiệp. Các giải pháp hiện tại chủ yếu bao gồm các giải pháp (gián tiếp) dựa trên bản ảnh xạ và/hoặc các giải pháp (trực tiếp) dựa trên sự phát hiện.

Giải pháp dựa trên bản ảnh xạ có thể cố gắng ánh xạ số mục tiêu con người với các dấu hiệu hình ảnh được trích xuất, như lượng điểm ảnh chuyển động, kích thước đám màu tiền cảnh, các mép nổi, nhóm gồm các góc nổi, và các đặc trưng của hình ảnh khác. Giải pháp dựa trên bản ảnh xạ thường đòi hỏi sự huấn luyện đối với các loại kịch bản video khác nhau. Nghiên cứu này chủ yếu tập trung vào việc tìm kiếm các đặc trưng tin cậy mà khá phù hợp với việc đếm người và tập trung vào cách làm thế nào giải quyết được một số vấn đề đặc biệt như các vùng tối và góc nhìn của máy quay. Trong nhiều tình huống, với đủ các video huấn luyện giải pháp dựa trên bản ảnh xạ có thể đưa ra các ước lượng đếm người khá chính xác. Tuy nhiên, hiệu suất thường phụ thuộc vào hiện trường, và có thể không thể có các vị trí thực tế của mỗi cá nhân được.

Giải pháp dựa trên sự phát hiện có thể đếm số người ở hiện trường bằng cách nhận dạng từng mục tiêu con người riêng rẽ. Nghiên cứu này đã tập trung vào việc phát hiện người, việc phát hiện các bộ phận của người và xem xét kết hợp việc phát hiện và theo dõi. Các giải pháp này có thể đưa ra kết quả phát hiện và đếm chính xác hơn trong các tình huống hơi đông đúc. Nếu có thể khiến cho có được vị trí của mỗi cá nhân thì, có thể tính mật độ đám đông cục bộ. Các thách thức chính của các giải pháp này là chi phí tính cao hơn, kỹ năng thực hiện phụ thuộc vào quan điểm và yêu cầu kích thước hình ảnh người tương đối lớn.

Các phương án được nêu trong bản mô tả này giải quyết được một số vấn đề nêu trên của các hệ thống hiện có.

Bản chất kỹ thuật của sáng chế

Sáng chế đề xuất phương pháp, thiết bị và hệ thống để phân tích thông minh các hình ảnh video để phát hiện các đối tượng, như các đối tượng con người.

Theo một số phương án nhất định, phương pháp phát hiện các đối tượng con người trong video bao gồm phát hiện các điểm ảnh nhất định của hình ảnh video là các điểm ảnh nổi, nhóm gồm các điểm ảnh nổi tạo thành tập hợp đốm màu tiền cảnh gồm một hoặc nhiều đốm màu tiền cảnh; đối với mỗi vị trí trong số N vị trí trong hình ảnh video, trong đó N là số nguyên, so sánh hình dạng định trước với tập hợp đốm màu tiền cảnh để thu được xác suất tương ứng của người tại vị trí này, nhờ đó thu được N xác suất tương ứng với N vị trí; và sử dụng N xác suất này, xác định X người được thể hiện bởi tập hợp đốm màu tiền cảnh, trong đó X là số nguyên.

Phương pháp phát hiện các đối tượng con người trong video, có thể bao gồm bước xác định các điểm ảnh của hình ảnh video của hiện trường thế giới thực là các điểm ảnh nổi, nhóm gồm các điểm ảnh nổi tạo thành tập hợp đốm màu tiền cảnh gồm một hoặc nhiều đốm màu tiền cảnh; và đối với từng vị trí trong số N vị trí trong hình ảnh video, trong đó N là số nguyên, so sánh hình dạng định trước với tập hợp đốm màu tiền cảnh để xác định X người được thể hiện bởi tập hợp đốm màu tiền cảnh, trong đó X là số nguyên.

Các phương pháp có thể bao gồm bước xác định vị trí của từng người trong số X người. Các vị trí của từng người trong số X người có thể định trước là vị trí trong mặt phẳng nằm ngang của thế giới thực, chẳng hạn như vị trí trên mặt phẳng đất vật lý của thế giới thực.

Việc phát hiện các đối tượng con người có thể được sử dụng để đếm người, để phân tích đám đông và để phát hiện sự kiện khác.

Hệ thống và thiết bị được đề cập có thể được tạo cấu hình để thực hiện các phương pháp nêu trên.

Vật ghi đọc được bởi máy tính chứa phần mềm mà có thể được sử dụng để tạo cấu hình máy tính nhằm thực hiện các hoạt động được mô tả trong bản mô tả và bao gồm các phương án khác của sáng chế.

Mô tả vắn tắt các hình vẽ

Các phương án làm ví dụ sẽ được hiểu rõ hơn từ phần mô tả chi tiết dưới đây có dựa vào các hình vẽ kèm theo. Các hình vẽ này thể hiện các phương án làm ví dụ, nhưng không chỉ giới hạn ở đó.

Fig.1 là hình vẽ minh họa ví dụ về hệ thống giám sát bằng video theo phương án làm ví dụ của sáng chế.

Fig.2 minh họa ví dụ về khung từ dòng video từ hệ thống giám sát bằng video theo phương án làm ví dụ của sáng chế.

Fig.3A minh họa ví dụ về lưu đồ để phát hiện và đếm mục tiêu theo phương án làm ví dụ của sáng chế.

Fig.3B minh họa ví dụ trong đó một vài mô hình người có mặt trong hình ảnh video hai chiều, mỗi mô hình tương ứng với một vị trí khác nhau trên hình ảnh video hai chiều.

Fig.3C minh họa một hàng tọa độ nhận dạng (x, y) 321, mỗi tọa độ liên quan đến mô hình người tương ứng 320.

Fig.3D minh họa phương pháp làm ví dụ để tính bản ánh xạ xác suất của người.

Fig.3E minh họa phương pháp làm ví dụ để thực hiện một lượt của bản ánh xạ xác suất như một phần trong việc tìm kiếm số lượng mô hình người lớn nhất trong hình ảnh video.

Fig.3F minh họa phương pháp thực hiện nhiều lượt của bản ánh xạ xác suất để tìm số lượng mô hình người lớn nhất trong hình ảnh video.

Fig.4 minh họa mô hình người tổng quát bao gồm mô hình hình trụ 3D và mô hình bao lồi 2D tương ứng của nó.

Fig.5 minh họa mô hình máy quay trái đất phẳng tổng quát mà có thể được hiệu chuẩn sử dụng vài mẫu hình ảnh người.

Các Fig.6A, Fig.6B và Fig.6C thể hiện các kết quả phát hiện làm ví dụ.

Các Fig.7A, Fig.7B và Fig.7C minh họa ví dụ về mật độ đám đông người dựa trên các kết quả phát hiện người.

Fig.8 minh họa các cách thực hiện làm ví dụ để phát hiện các sự kiện liên quan đến đám đông khác nhau.

Fig.9 minh họa phương pháp làm ví dụ về cách xác định và phát hiện khu vực đông người.

Fig.10 minh họa quy trình xử lý làm ví dụ đối với mỗi mục tiêu con người phát hiện được.

Fig.11 minh họa quy trình xử lý làm ví dụ đối với mỗi khu vực đám đông.

Fig.12 minh họa phương pháp có thể được sử dụng để xác định và phát hiện các sự kiện “tập trung” và “phân tán” đám đông.

Fig.13 minh họa một ví dụ về việc xác định điểm tập trung đông người.

Fig 14A và 14B là các Fig thể hiện ví dụ về điểm tập trung đông người.

Fig.15 minh họa phương pháp làm ví dụ để phát hiện các điểm tập trung đông người.

Fig.16 minh họa phương pháp làm ví dụ để cập nhật các điểm tập trung đông người và phát hiện các sự kiện “tập trung” và “phân tán” đám đông.

Fig.17 minh họa phương án thực hiện làm ví dụ sử dụng nhiều máy quay video.

Mô tả chi tiết sáng chế

Dưới đây, các phương án khác nhau làm ví dụ sẽ được mô tả chi tiết hơn có dựa vào các hình vẽ kèm theo, trong đó một số phương án làm ví dụ sẽ được thể hiện. Tuy nhiên, sáng chế có thể được thể hiện theo nhiều hình thức khác nhau và không được hiểu là chỉ giới hạn ở các phương án làm ví dụ đó. Các phương án làm ví dụ chỉ

đơn thuần là các ví dụ và nhiều phương án và nhiều thay đổi khác là có thể thực hiện được mà không cần nêu chi tiết ở đây. Cũng lưu ý rằng phần mô tả đưa ra các chi tiết về các ví dụ khác, nhưng việc liệt kê các ví dụ khác này là không toàn diện. Hơn thế nữa, không thể liệt kê mọi thay đổi có thể có đối với mỗi dấu hiệu được nêu trong bản mô tả. Ngôn ngữ của yêu cầu bảo hộ nên được xem xét khi xác định các yêu cầu của sáng chế. Trên các hình vẽ, các kích thước và các kích thước tương đối của các lớp và các vùng có thể được phóng to để cho rõ ràng. Các số chỉ dẫn giống nhau chỉ các chi tiết giống nhau trong toàn bộ bản mô tả.

Cần hiểu rằng, mặc dù các thuật ngữ thứ nhất, thứ hai, thứ ba v.v. có thể được sử dụng trong bản mô tả để mô tả các chi tiết khác nhau, nhưng các chi tiết này không chỉ giới hạn ở các thuật ngữ đó. Các thuật ngữ này được sử dụng để phân biệt chi tiết này với chi tiết khác. Do đó, chi tiết thứ nhất nêu dưới đây có thể được gọi là chi tiết thứ hai mà không nằm ngoài ý nghĩa của khái niệm trong sáng chế. Như được sử dụng ở đây, thuật ngữ "và/hoặc" bao gồm bất kỳ hoặc tất cả các kết hợp của một hoặc nhiều trong số các thuật ngữ liên quan được liệt kê.

Cần hiểu rằng khi chi tiết được coi như là "được nối" hoặc "được ghép" với chi tiết khác, nó có thể được nối hoặc được ghép một cách trực tiếp với chi tiết khác hoặc các chi tiết nằm xen giữa có thể có mặt. Trái lại, khi chi tiết được coi như là "được nối một cách trực tiếp" hoặc "được ghép một cách trực tiếp" với chi tiết khác, thì không có các chi tiết nằm xen giữa chúng. Các từ khác được dùng để mô tả mối quan hệ giữa các chi tiết sẽ được hiểu theo cách tương tự (chẳng hạn, "giữa" so với "ngay giữa," "liền kề" so với "liền ngay," v.v.).

Thuật ngữ được dùng ở đây chỉ nhằm mô tả chỉ các phương án làm ví dụ cụ thể và không nhằm giới hạn phạm vi của sáng chế. Như được sử dụng ở đây, các dạng số ít dự kiến bao gồm cả các dạng số nhiều, trừ khi hoàn cảnh có quy định rõ khác. Cũng cần hiểu rằng các thuật ngữ "bao gồm" và/hoặc "chứa," khi được sử dụng trong bản mô tả này, quy định sự có mặt của các đặc điểm, các tổng thể, các bước, các hoạt động,

các chi tiết, và/hoặc các bộ phận được nêu, nhưng không hạn chế sự có mặt hoặc bổ sung của một hoặc nhiều đặc điểm, các tổng thể, các bước, các hoạt động, các chi tiết, các bộ phận khác và/hoặc các tập hợp của chúng.

Trừ khi được định nghĩa khác, tất cả các thuật ngữ (bao gồm các thuật ngữ khoa học và kỹ thuật) được dùng trong bản mô tả có ý nghĩa giống như được hiểu bình thường bởi người có hiểu biết trung bình về lĩnh vực kỹ thuật liên quan mà sáng chế thuộc về. Cũng cần hiểu rằng các thuật ngữ, như các thuật ngữ được định nghĩa trong các từ điển thông dụng, sẽ được hiểu là có ý nghĩa phù hợp với ý nghĩa của chúng trong hoàn cảnh liên quan và sẽ không được hiểu theo hướng lý tưởng hóa hoặc quá trang trọng trừ khi được định nghĩa như trong bản mô tả này.

Các định nghĩa. Trong quá trình mô tả sáng chế, các định nghĩa sau có thể được dùng trong toàn bộ bản mô tả (bao gồm các định nghĩa nêu trên).

"Viđeo" có thể đề cập đến các hình ảnh động được thể hiện ở dạng tín hiệu tương tự (analog) và/hoặc dạng số. Các ví dụ về viđeo có thể bao gồm: các chương trình truyền hình; phim; chuỗi hình ảnh từ máy quay viđeo hoặc máy quan sát khác; chuỗi hình ảnh từ nguồn trực tiếp; chuỗi hình ảnh được tạo bởi máy tính; chuỗi hình ảnh từ công cụ đồ họa của máy tính; chuỗi hình ảnh từ thiết bị lưu trữ, như vật ghi đọc được bởi máy tính, đĩa viđeo kỹ thuật số (DVD), hoặc đĩa độ nét cao (high-definition disk - HDD); chuỗi hình ảnh từ giao diện dựa trên IEEE 1394; chuỗi hình ảnh từ bộ số hóa hình ảnh; hoặc chuỗi hình ảnh từ mạng.

"Dây viđeo" có thể liên quan đến một phần hoặc toàn bộ viđeo.

"Máy quay viđeo" có thể liên quan đến thiết bị để ghi hình trực quan. Các ví dụ về máy quay viđeo có thể bao gồm một hoặc nhiều thiết bị trong số các thiết bị sau: bộ tạo hình ảnh viđeo và cơ cấu thấu kính; máy quay viđeo; máy quay viđeo kỹ thuật số; máy quay màu; máy quay đen trắng; máy quay; máy quay viđeo xách tay có kèm theo bộ phận ghi hình; máy quay PC; webcam; máy quay viđeo hồng ngoại (infrared - IR); máy quay viđeo ánh sáng yếu; máy quay viđeo nhiệt; máy quay truyền hình mạch kín

(closed-circuit television - CCTV); máy quay quét ngang, quét dọc, thu phóng (pan, tilt, zoom - PTZ); và thiết bị thụ cảm video. Máy quay video có thể được bố trí để thực hiện việc giám sát khu vực quan tâm.

"Xử lý video" có thể liên quan đến việc xử lý và/hoặc phân tích video bất kỳ, bao gồm, ví dụ, nén, chỉnh sửa, giám sát, và/hoặc xác nhận.

"Khung" có thể liên quan đến hình ảnh cụ thể hoặc đơn vị rời rạc khác trong video.

"Máy tính" có thể liên quan đến một hoặc nhiều thiết bị và/hoặc một hoặc nhiều hệ thống mà có khả năng tiếp nhận đầu vào có cấu trúc, xử lý đầu vào có cấu trúc tuân theo các quy tắc nhất định, và tạo ra các kết quả xử lý như đầu ra. Các ví dụ về máy tính có thể bao gồm: máy tính; máy tính để bàn và/hoặc máy tính xách tay; máy tính có một bộ xử lý, nhiều bộ xử lý, hoặc bộ xử lý nhiều lõi, mà có thể hoạt động song song và/hoặc không song song; máy tính thông dụng; siêu máy tính; máy tính lớn; máy tính siêu nhỏ; máy tính bỏ túi; trạm làm việc; máy vi tính; máy chủ; máy khách; truyền hình tương tác; thiết bị web; thiết bị viễn thông có truy cập internet; tổ hợp ghép của máy tính và truyền hình tương tác; máy tính xách tay; máy tính cá nhân (PC); máy trợ giúp cá nhân kỹ thuật số (personal digital assistant - PDA); điện thoại di động; phần cứng chuyên dụng để mô phỏng máy tính và/hoặc phần mềm, ví dụ, bộ xử lý tín hiệu kỹ thuật số (digital signal processor - DSP), mảng cổng có thể lập trình được theo trường (field-programmable gate array - FPGA), mạch tích hợp chuyên dụng (application specific integrated circuit - ASIC), bộ xử lý tập lệnh chuyên dụng (application specific instruction-set processor-ASIP), chip, các chip, hoặc bộ vi mạch; hệ thống trên chip (system on chip - SoC), hoặc hệ thống nhiều bộ xử lý trên chip (multiprocessor system-on-chip - MPSoC); máy tính quang; máy tính lượng tử; máy tính sinh học; và thiết bị mà có thể tiếp nhận dữ liệu, có thể xử lý dữ liệu theo một hoặc nhiều chương trình phần mềm được lưu trữ, có thể tạo ra các kết quả, và thông

thường có thể bao gồm khối đầu vào, khối đầu ra, bộ nhớ, khối số học, khối logic, và khối điều khiển.

"Phần mềm" có thể liên quan đến các quy tắc quy định để vận hành máy tính. Các ví dụ về phần mềm có thể bao gồm: phần mềm; các đoạn mã; các lệnh; các applet; mã biên dịch trước; mã biên dịch; mã thông dịch; các chương trình máy tính; và logic được lập trình.

"Vật ghi đọc được bởi máy tính" có thể liên quan đến thiết bị lưu trữ bất kỳ được dùng để lưu trữ dữ liệu truy cập được bởi máy tính. Các ví dụ về vật ghi đọc được bởi máy tính có thể bao gồm: đĩa từ cứng; đĩa mềm; đĩa quang, như CD-ROM và DVD; băng từ; bộ nhớ chớp tháo được; chip nhớ; và/hoặc các loại vật ghi khác có thể lưu trữ các lệnh đọc được bằng máy trên đó.

"Hệ thống máy tính" có thể liên quan đến hệ thống có một hoặc nhiều máy tính, mỗi máy tính có thể có vật ghi đọc được bởi máy tính chứa phần mềm để vận hành máy tính. Các ví dụ về hệ thống máy tính có thể bao gồm: hệ thống máy tính phân tán để xử lý thông tin thông qua các hệ thống máy tính được liên kết với nhau bởi mạng lưới; hai hoặc nhiều hệ thống máy tính được nối với nhau thông qua mạng lưới để truyền và/hoặc tiếp nhận thông tin giữa các hệ thống máy tính; và một hoặc nhiều thiết bị và/hoặc một hoặc nhiều hệ thống mà có thể tiếp nhận dữ liệu, có thể xử lý dữ liệu theo một hoặc nhiều chương trình phần mềm lưu trữ, có thể tạo ra các kết quả, và thông thường có thể bao gồm khối đầu vào, khối đầu ra, khối lưu trữ, khối số học, khối logic, và khối điều khiển.

"Mạng" có thể liên quan đến nhiều máy tính và các thiết bị liên quan mà có thể được nối bởi các phương tiện truyền thông. Mạng lưới có thể bao gồm các kết nối lâu bền như cáp hoặc các kết nối tạm thời như các kết nối được thực hiện thông qua điện thoại hoặc các liên kết viễn thông khác. Mạng lưới có thể còn bao gồm các kết nối có dây vật lý (chẳng hạn, cáp đồng trục, cặp dây xoắn, sợi quang, các ống dẫn sóng, v.v.) và/hoặc kết nối không dây (chẳng hạn, các dạng sóng tần số vô tuyến, các

dạng sóng quang trong không gian tự do, các dạng sóng sóng âm, v.v.). Các ví dụ về mạng lưới có thể bao gồm: liên mạng, như là internet; intranet; mạng cục bộ (LAN); mạng diện rộng (WAN); và kết hợp các mạng lưới, như internet và intranet. Các mạng lưới làm ví dụ có thể hoạt động với giao thức bất kỳ, như giao thức mạng internet (Internet protocol - IP), phương thức truyền không đồng bộ (asynchronous transfer mode - ATM), và/hoặc mạng quang đồng bộ (synchronous optical network - SONET), giao thức gói dữ liệu người dùng (user datagram protocol - UDP), IEEE 802.x, v.v.

Trong một số phương án, phương pháp, hệ thống và thiết bị ước lượng mật độ đám đông có thể được dựa trên các phương pháp, các hệ thống và các thiết bị phân tích nội dung video hiện có. Ngoài yêu cầu độ chính xác ước lượng cơ bản, giải pháp có thể bao gồm một hoặc nhiều trong số sau:

- Tính độc lập cảnh quay máy quay có thể cho phép các phương án hoạt động được trong phạm vi các kịch bản ứng dụng rộng bất kể có sự thay đổi về vị trí máy quay, góc quan sát, số điểm ảnh trên mục tiêu, v.v..
- Chi phí tính toán tương đối thấp, nên có thể chạy theo thời gian thực. Các phương án này có thể được thực hiện trên hệ thống nhúng.
- Có thể giảm và/hoặc loại bỏ việc thiết lập và huấn luyện ban đầu phức tạp, cho phép sử dụng thuận tiện hơn và chi phí sở hữu thấp hơn.

Một số ví dụ được bộc lộ ở đây bao gồm giải pháp dựa trên sự phát hiện và có thể không cần đòi hỏi sự đào tạo. Các ví dụ này có thể được thực hiện với hệ thống IVS thông thường, mà đã thực hiện các tác vụ phát hiện và theo dõi cơ bản và tạo ra mặt nạ tiền cảnh tin cậy. Mô hình hình ảnh người dạng vùng lõi có thể được tính cho mọi điểm ảnh của hình ảnh, mà có thể được sử dụng để ước lượng số mục tiêu con người trong mỗi vùng tiền cảnh. Dữ liệu hiệu chỉnh máy quay có thể tạo ra bản ánh xạ từ mặt phẳng hình ảnh đến mặt phẳng đất trong thế giới thực, mà có thể được sử dụng để tạo ra các kết quả đo mật độ đám đông thực tế ở các khu vực trong cảnh quay máy quay. Bằng cách sử dụng (các) kết quả đo mật độ đám đông thực tế, các sự kiện quan

trọng khác có thể được phát hiện, ví dụ, "điểm nóng đông người", "sự tập trung đông người", "sự phân tán đám đông", v.v..

Fig.1 minh họa hệ thống giám sát bằng video 101 theo các phương án đưa ra làm ví dụ của sáng chế. Hệ thống giám sát bằng video này có thể được tạo cấu hình để phát hiện và giám sát các hoạt động đông người trong các dòng video. Hệ thống giám sát bằng video 101 có thể được sử dụng trong nhiều ứng dụng trong đó việc phát hiện người được quan tâm, như sử dụng để phân tích mật độ đám đông. Ví dụ, các phương án này có thể được sử dụng để phát hiện sự tập trung đông người khả nghi, thu thập thông kê tình hình giao thông cho người đi bộ, sự hình thành và/hoặc phân tán đám đông bất thường, v.v.. Hệ thống giám sát bằng video 101 có thể bao gồm nguồn video 102 (chẳng hạn, máy quay video hoặc bộ nhớ, chẳng hạn như ổ cứng, có lưu trữ video), môđun phát hiện sự thay đổi 103, môđun phát hiện chuyển động 104, môđun phát hiện đếm màu tiền cảnh 105, môđun phát hiện người 106, môđun theo vết mục tiêu 107 và môđun phát hiện sự kiện 108. Trong ví dụ này, nguồn video (chẳng hạn, máy quay video) là cố định. Tuy nhiên, chuyên gia trong lĩnh vực sẽ thừa nhận sáng chế cũng sử dụng các nguồn video di động. Trong ví dụ này, nguồn video tạo ra một dòng video. Tuy nhiên, sáng chế cũng dự tính sử dụng và xử lý nhiều dòng video.

Hệ thống giám sát bằng video có thể được thực hiện bằng hệ thống IVS sử dụng nền tảng tính điện hình. Chẳng hạn, xem patent US 7,868,912 được cấp cho Venetianer và đồng tác giả và patent US 7,932,923 được cấp cho Lipton và đồng tác giả, mà nội dung của cả hai patent US nêu trên được kết hợp vào đây bằng cách viện dẫn, để làm ví dụ mô tả chi tiết về hệ thống IVS mà có thể được sử dụng để thực hiện các phương án được nêu trong bản mô tả này. Patent US 7,868,912 và patent US 7,932,923 cũng được kết hợp vào đây bằng cách viện dẫn để làm ví dụ mô tả chi tiết về sự tạo thành video gốc (hoặc siêu dữ liệu) và sự xử lý sau đó (mà có thể là xử lý thời gian thực hoặc xử lý sau) để thu được thông tin từ video này, như phát hiện sự kiện, sử dụng các video gốc được tạo ra, mà có thể được sử dụng với các phương án

được mô tả trong bản mô tả. Mỗi môđun 103-108, cũng như các bộ phận riêng của chúng, riêng rẽ hoặc liên kết với các môđun/các bộ phận khác, có thể được thực hiện bởi phần cứng chuyên dụng (hệ mạch), phần mềm và/hoặc phần sụn. Ví dụ, máy tính thông dụng được lập trình bằng phần mềm có thể thực hiện tất cả các môđun này. Như được hiểu theo cách thông thường, vật ghi đọc được bởi máy tính chứa phần mềm mà có thể được sử dụng để tạo cấu hình máy tính để thực hiện các hoạt động được mô tả trong bản mô tả bao gồm các phương án khác của sáng chế. Về các ví dụ khác, để thực hiện các hệ thống, các thiết bị và các phương pháp được mô tả trong bản mô tả, các bộ phận quang và tính toán khác nhau có thể được sử dụng, chẳng hạn như một hoặc nhiều thiết bị trong số các thiết bị sau: máy tính thông dụng; siêu máy tính; máy tính lớn; máy tính siêu nhỏ; máy tính bỏ túi; trạm làm việc; máy vi tính; máy chủ; truyền hình tương tác; tổ hợp lai của máy tính và truyền hình tương tác; điện thoại thông minh; máy tính bảng; và phần cứng ứng dụng đặc biệt để mô phỏng máy tính và/hoặc phần mềm. Các thiết bị này có thể bao gồm một hoặc nhiều bộ xử lý, một trong số nhiều mảng cổng có thể lập trình được theo trường (field programmable gate array-FPGA), bộ nhớ máy tính, vật ghi đọc được bởi máy tính, ví dụ, thiết bị lưu trữ bất kỳ được dùng để lưu trữ dữ liệu truy cập được bởi máy tính (chẳng hạn, bộ xử lý có thể thực hiện các thuật toán khác nhau dựa trên dữ liệu được tiếp nhận từ thiết bị quay phim, và sau đó bộ nhớ máy tính có thể lưu trữ thông tin về các điểm ảnh khác nhau và có thể lưu trữ các kết quả của việc phát hiện đốm màu, phát hiện mục tiêu, và phát hiện sự kiện). Các ví dụ về vật ghi đọc được bởi máy tính bao gồm: đĩa từ cứng; đĩa mềm; đĩa quang, như CD-ROM và DVD; băng từ; chip nhớ; ổ nhớ thể rắn; và sóng mang được sử dụng để mang dữ liệu điện tử đọc được bởi máy tính, như các sóng mang được dùng trong truyền và nhận thư điện tử hoặc trong việc truy cập mạng lưới. Vật ghi đọc được bởi máy tính hữu hình bao gồm vật ghi đọc được bởi máy tính, như được nêu trên đây, mà là hữu hình. Ngoài ra, phần mềm có thể được sử dụng liên kết với các bộ phận tính toán và/hoặc quang học để thực hiện các phương pháp được mô tả trong bản mô tả. Phần mềm có thể bao gồm các quy tắc và/hoặc các thuật toán

để vận hành máy tính, và có thể bao gồm, ví dụ, các đoạn mã, các lệnh, các chương trình máy tính, và logic được lập trình. Nguồn video 102 và các môđun 103-108 có thể là trong cùng một hệ thống hoặc có thể được phân tán. Ví dụ, nguồn video 102 có thể bao gồm máy quay video tại khu vực cần được giám sát. Nguồn video 102 cấp dòng video đến vị trí giám sát (chẳng hạn, vị trí thứ hai tách biệt nằm ngoài vị trí cần giám sát) tại đó các môđun 103-107 được đặt. Môđun phát hiện sự kiện 108 có thể được bố trí ở vị trí thứ ba (chẳng hạn, nhà ga trung tâm) tách biệt khỏi vị trí theo dõi và vị trí thứ hai. Các môđun, các máy tính, các máy quay, và thiết bị hình ảnh khác được mô tả trong bản mô tả có thể được nối qua một mạng lưới, mà có thể bao gồm các kết nối lâu bền như cáp hoặc các kết nối tạm thời như các kết nối được thực hiện thông qua điện thoại hoặc các liên kết viễn thông khác, và cũng có thể bao gồm các liên kết truyền thông không dây. Các ví dụ về mạng lưới bao gồm: liên mạng, như internet; intranet; mạng cục bộ (LAN); mạng diện rộng (WAN); và kết hợp của các mạng lưới, như internet và intranet. Các ví dụ về phần mềm và phần cứng khác nhau được nêu trên đây cũng được mô tả chi tiết hơn trong các tài liệu sáng chế sẽ được kết hợp vào đây bằng cách viện dẫn.

Các điểm ảnh thay đổi có thể được phát hiện bởi môđun phát hiện sự thay đổi 103 khi các điểm ảnh của hình ảnh video được cấp bởi nguồn video 102 khác với hình ảnh nền thu được từ trước. Hình ảnh nền có thể là ảnh động. Mô hình hình ảnh nền động có thể được dựng và cập nhật liên tục từ các khung video vừa mới nhập vào. Do đó, các thay đổi về ánh sáng, thời tiết, v.v., mà làm thay đổi hình ảnh video có thể được tính đến trong hình ảnh nền. Trong môđun 104, tính toán sai khác khung có thể được sử dụng để phát hiện các điểm ảnh dịch chuyển. Trong 105, một hoặc cả hai điểm ảnh thay đổi từ môđun 103 và các điểm ảnh dịch chuyển từ môđun 104 được xem xét để xác định các điểm ảnh nổi mà được nhóm thành các đốm màu tiền cảnh về mặt không gian. Hình ảnh video có thể được xử lý bởi các hệ thống và các phương pháp phân tích nội dung video hiện có để trích xuất hình ảnh nổi, các đốm màu tiền cảnh và các đốm màu tiền cảnh quan tâm (như các đốm màu tiền cảnh của người),

như được mô tả trong patent US 7,825,954, được cấp cho Zhang và đồng tác giả, được công bố ngày 2/11/2010, mà toàn bộ nội dung của tài liệu này được kết hợp ở đây bằng cách viện dẫn. Thông tin cảm biến độ sâu có thể được sử dụng một cách tùy ý để ước lượng chiều cao hoặc kích thước thực của mỗi đối tượng được phát hiện là mục tiêu con người tiềm năng, và kết quả là, các điểm màu tương ứng với mục tiêu con người tiềm năng (trái ngược với các điểm màu không cần quan tâm) có thể định trước chính xác hơn. Thông tin cảm biến độ sâu có thể được sử dụng một cách tùy ý để loại bỏ các vùng tối, các vùng phản chiếu, các đối tượng được phát hiện nằm ngoài khu vực cần quan tâm, các đối tượng quá xa (chẳng hạn, có thể không đủ gần để cho phép phân tích chính xác), hoặc các yếu tố khác của hình ảnh video mà có thể làm tăng nguy cơ phân tích sai lệch hình ảnh video đó. Ví dụ về việc mô tả chi tiết cách sử dụng thông tin theo chiều sâu có thể được tìm thấy trong đơn yêu cầu cấp patent US số 13/744,254 của Zhang và đồng tác giả, mà toàn bộ nội dung của tài liệu này được kết hợp vào đây bằng cách viện dẫn. Các điểm màu được theo vết theo thời gian để tạo thành các mục tiêu không gian-thời gian trong môđun theo vết mục tiêu 107, và cuối cùng, môđun phát hiện sự kiện 108 phát hiện sự kiện quan tâm được định nghĩa bởi người sử dụng bằng cách sử dụng đầu ra của quá trình phát hiện và theo vết mục tiêu. Thay hoặc bổ sung cho việc nhóm các điểm ảnh nổi thành các điểm màu theo không gian một cách đơn giản, môđun phát hiện người 106 sử dụng thông tin hiệu chuẩn và mô hình người dạng vùng lỗi để phát hiện người ngay cả trong kịch bản đông người. Trong một số ví dụ, không cần hoặc chỉ cần sự đào tạo tối thiểu từ trước để phát hiện các đối tượng con người ở hiện trường. Và trong môđun phát hiện sự kiện 108, một số phương pháp phát hiện sự kiện mới có thể được thực hiện, mà có thể sử dụng các kết quả phát hiện người trong môđun phát hiện người 106.

Fig.2 thể hiện các hình ảnh video tương ứng với một số kịch bản ứng dụng thông thường cho hệ thống IVS 101 bao gồm quảng trường ngoài trời, đường phố, nơi thu hút khách du lịch, nhà ga xe lửa, trung tâm thương mại, bến tàu điện ngầm, v.v.. Như có thể thấy, phụ thuộc vào vị trí của máy quay so với hiện trường đang được ghi

hình, kích thước và hình dạng tương đối của người xuất hiện trong các hình ảnh video là khác nhau.

Fig.3A thể hiện sơ đồ khối đưa ra ví dụ mô tả chi tiết hơn về hệ thống giám sát bằng video 101. Môđun phát hiện đốm màu tiền cảnh 105 có thể giống như ở trên Fig.1. Các môđun 301, 302, 303, 304, 305 và 306 có thể là các chi tiết của môđun phát hiện người 106 trên Fig.1. Môđun phát hiện điểm ảnh thân người 301 phát hiện điểm ảnh thân người dựa trên điểm ảnh thay đổi thu được từ môđun phát hiện sự thay đổi 103. Các điểm ảnh này hoặc là khác đáng kể với mô hình hình ảnh nền (chẳng hạn, sự chênh lệch độ sáng và/hoặc sự chênh lệch màu sắc vượt quá ngưỡng cho phép tương ứng), hoặc được định vị giữa các điểm ảnh mép nổi tin cậy cao. Chúng được coi là phù hợp nhất với các điểm ảnh thân người trong hình ảnh. Chẳng hạn, xem 301a trên Fig.6A, là một ví dụ về các điểm ảnh thân người được phát hiện. Các điểm ảnh thay đổi khác có thể được loại trừ khỏi quá trình phát hiện người tiếp theo, do chúng có nhiều khả năng là các vùng tối hoặc các vùng phản xạ. Môđun phát hiện điểm ảnh biên người 302 phát hiện các điểm ảnh biên người ở nơi mà biên của các đốm màu tiền cảnh trùng khớp với các mép hình ảnh của khung video hiện tại. Chẳng hạn, xem 302a trên Fig.6A, là một ví dụ về các điểm ảnh biên người được phát hiện. Khi thực hiện việc phát hiện người, các phân tích khác có thể được thực hiện (bổ sung hoặc thay thế các phân tích được nêu trên đây) để hỗ trợ việc xác định thân người đã được phát hiện. Ví dụ, có thể yêu cầu mỗi đốm màu tiềm năng là người phải chứa số lượng nhất định các điểm ảnh ở mép nổi của biên. Về các ví dụ khác, quá trình xử lý khác có thể nhận biết (các) đốm màu có khả năng liên quan đến đối tượng không phải người (như xe cộ) và loại trừ (các) đốm màu này khỏi quá trình phát hiện người tiếp theo. Các đốm màu tiền cảnh khác không được coi là mục tiêu tiềm năng là người có thể được loại trừ khỏi tập hợp đốm màu tiền cảnh. Theo cách khác, đốm màu được phát hiện bất kỳ có thể là một phần của tập hợp đốm màu tiền cảnh.

Môđun mô hình người tổng quát 303 tạo ra mô hình người tổng quát 2D và 3D. Ví dụ, môđun mô hình người tổng quát 303 có thể chuyển đổi mô hình người 3D thành mô hình người 2D bằng cách ánh xạ hoặc chiếu mô hình người 3D trong thế giới thực lên mặt phẳng hình ảnh 2D của hình ảnh video. Fig.4 thể hiện ví dụ về mô hình 3D 303a được ánh xạ thành mô hình người 2D tương ứng 303b trên mặt phẳng hình ảnh 330. Mô hình người 3D 303a có thể là tập hợp của các hình dạng 3D đơn giản, như nhóm gồm các hình trụ (chẳng hạn, một hình trụ dùng cho chân, một hình trụ dùng cho thân và một hình trụ dùng cho đầu). Cùng một mô hình người 3D 303a (chẳng hạn, mô hình hình trụ) có thể được sử dụng với các vị trí khác nhau của máy quay video sao cho góc khác nhau của máy quay video này so với mặt đất (mặt phẳng đất của thế giới thực) có thể được sử dụng để thu được mô hình người 2D có hình dạng khác nhau 303b trong mặt phẳng hình ảnh của máy quay video. Ví dụ, lấy mô hình người 3D hình trụ làm ví dụ, góc máy quay tạo góc nhìn từ trên xuống của vị trí cụ thể có thể ánh xạ thành hình tròn trong mặt phẳng hình ảnh 2D, trong khi đó góc máy quay có góc nhìn xiên ở cùng vị trí có thể ánh xạ mô hình người 3D hình trụ thành hình dạng khác có dạng thiên dài. Trong ví dụ được thể hiện trên Fig.17, máy quay 1702 có thể có góc nhìn từ trên xuống mô hình người 3D 303a so với máy quay 1704, máy quay này có thể có góc nhìn ở bên cạnh của mô hình người 3D 303a nhiều hơn so với máy quay 1702. Nếu khoảng cách của các máy quay 1702 và 1704 từ mô hình người 3D 303a là giống nhau, mô hình người 2D tương ứng được ánh xạ lên mặt phẳng hình ảnh của máy quay 1702 là nhỏ gọn hơn (chẳng hạn, ngắn hơn) so với mô hình người 2D được ánh xạ lên mặt phẳng hình ảnh của máy quay 1704. Mô hình người 2D có thể có hình dạng lồi có thể thu được bằng cách nội suy các điểm thuộc các mép ngoài của phép chiếu mô hình người 3D lên mặt phẳng hình ảnh 2D.

Fig.4 minh họa mô hình người tổng quát bao gồm mô hình hình trụ 3D 303a và mô hình bao lồi 2D tương ứng của nó 303b được ánh xạ lên mặt phẳng hình ảnh 2D 330. Mô hình người 3D 303 bao gồm chân hình trụ, thân hình trụ và đầu hình trụ. Chiều dài và bán kính của mỗi hình trụ có thể tương ứng với dữ liệu thống kê thực

chất thể hiện các kích thước đặc trưng của người bình thường. Như được thể hiện trên Fig.4, ba hình trụ này có bốn mặt phẳng chính: mặt phẳng đầu, mặt phẳng vai, mặt phẳng hông và mặt phẳng chân. Để thu được mô hình người 2D tương ứng tại vị trí cụ thể, có thể lấy mẫu đồng đều quanh chu vi của bốn mặt phẳng chính này và chiếu mỗi điểm mẫu 3D lên mặt phẳng hình ảnh 2D sử dụng các thông số hiệu chuẩn máy quay để xác định kích thước và hướng thích hợp đối với vị trí cụ thể trong không gian hình ảnh 2D. Các điểm mẫu hình ảnh tương ứng này sau đó có thể được sử dụng để tạo ra thân lồi trên hình ảnh bằng phương pháp tạo hình lồi, mà có thể được sử dụng như mô hình hình ảnh người 2D.

Fig.5 minh họa mô hình máy quay trái đất phẳng tổng quát mà có thể được hiệu chuẩn bằng cách sử dụng vài mẫu hình ảnh người. Mô hình máy quay này có thể chỉ có ba thông số: độ cao máy quay so với mặt đất, góc lia lên của nó và tiêu cự của máy quay. Các thông số này có thể được ước lượng sử dụng ba hoặc nhiều hơn ba mẫu người từ các khung video như được mô tả trong "A Robust Human Detection and Tracking System Using a Human-Model-Based Camera Calibration" (The 8th International Workshop on Visual Surveillance, 2008, Z. Zhang, P. L. Venetianer and A. J. Lipton) và patent US 7,801,330, của Zhang và đồng tác giả, được công bố ngày 21/9/2010, mà nội dung của chúng được kết hợp vào đây bằng cách viện dẫn.

Theo cách khác, hoặc ngoài ra, mô đun mô hình người tổng quát 303 có thể có mô hình 2D định trước mà có thể được thay đổi (chẳng hạn, kéo, co, làm nghiêng so với trục thẳng đứng của mặt phẳng hình ảnh 2D, v.v.) tương ứng với góc máy quay của máy quay video đang quay hình ảnh video. Một vài mô hình người tổng quát có thể được tạo ra bằng mô đun mô hình người tổng quát 303. Các mô hình người cũng có thể bao gồm mô hình dùng cho các phụ kiện đặc trưng. Ví dụ, khi sử dụng hệ thống ngoài trời, mô hình người thứ nhất có thể được sử dụng cho thời tiết ấm áp, mô hình người thứ hai lớn hơn có thể được sử dụng cho thời tiết lạnh (lúc mà áo khoác ngoài có thể được mặc và có thể là phần được xem xét của mô hình người) và mô hình

người thứ ba có thể được sử dụng lúc trời mưa (lúc mà những chiếc ô có thể được sử dụng và có thể là phần được xem xét của mô hình người).

Môđun mô hình người tổng quát 303 cũng tạo ra giá trị ước lượng về kích thước của mô hình người 2D tại các vị trí tương ứng trong không gian hình ảnh. Không gian hình ảnh có thể tương ứng với không gian hai chiều của hình ảnh trong khung của video được tạo ra bởi nguồn video 102. Không gian hình ảnh có thể được đo theo các số gia điểm ảnh, sao cho các vị trí trong không gian hình ảnh định trước bởi các tọa độ điểm ảnh. Máy quay video có thể chụp hình ảnh video, bao gồm hình ảnh hai chiều của thế giới thực ba chiều. Khi người có mặt tại vị trí nhất định trong thế giới thực, người có thể cần phải chiếm lượng màu nổi nhất định tại vị trí nhất định trong hình ảnh video hai chiều. Nếu người cách xa máy quay video, kích cỡ hình ảnh của người có thể sẽ tương đối nhỏ so với kích cỡ hình ảnh của người gần với máy quay video. Đối với từng vị trí trong số nhiều vị trí trong không gian hình ảnh video hai chiều, môđun mô hình người tổng quát 303 có thể tạo ra mô hình người có kích thước tương ứng với vị trí trong không gian hình ảnh hai chiều. Đối với mỗi vị trí này, mô hình người 2D có thể có các kích thước và/hoặc kích cỡ tương ứng với vị trí tương ứng trong không gian hình ảnh của hình ảnh video hai chiều. Hướng của các mô hình người này cũng có thể tương ứng với vị trí trong không gian hình ảnh hai chiều. Ví dụ, một số ống kính máy quay (chẳng hạn, các thấu kính trường rộng) có thể biểu diễn hướng thẳng đứng trong thế giới thực là hướng thứ nhất ở một phía của khung hình ảnh video và hướng thứ hai khác ở phía thứ hai của khung hình ảnh video. Các mô hình người 2D có thể có các hướng khác nhau ở các phía khác nhau của khung hình ảnh video (và các vị trí khác) tương ứng với các mặt của hướng thẳng đứng trong thế giới thực.

Các vị trí của mỗi trong số các mô hình người trong không gian hình ảnh video 2D có thể được kết hợp trong việc xác định các tọa độ trong không gian hình ảnh video 2D. Các tọa độ định vị có thể tương ứng với các vị trí điểm ảnh của không gian

hình ảnh video 2D. Ví dụ, vị trí tương ứng với hàng thứ 10, cột thứ 22 của mảng điểm ảnh có thể tương ứng với tọa độ định vị (10, 22). Đối với mỗi vị trí trong số các vị trí trong không gian hình ảnh video 2D, mô đun mô hình người tổng quát 303 có thể ánh xạ điểm cụ thể của mô hình người với tọa độ định vị có liên quan. Ví dụ, điểm cụ thể của mô hình người có thể là đỉnh của mô hình người tương ứng với đầu người, đáy của mô hình người tương ứng với bàn chân người, trọng tâm của hình dạng mô hình người tương ứng với giữa người. Phần còn lại của mô hình người có thể được ánh xạ vào không gian hình ảnh video 2D có liên quan đến tọa độ định vị được liên kết và kích thước của mô hình người dựa trên mối quan hệ cố định giữa điểm cụ thể của mô hình người và phần còn lại của mô hình người. Như một ví dụ, giả sử mô hình người là đường tròn. Đối với mỗi điểm ảnh trong không gian hình ảnh video 2D, tâm của đường tròn tương ứng được ánh xạ (chẳng hạn, được liên kết với các tọa độ (x, y) của không gian hình ảnh video 2D), trong đó phần còn lại của hình dạng đường tròn được ánh xạ lên không gian hình ảnh video 2D khi xem xét kích thước tương ứng của đường tròn (và mối quan hệ đã biết của đường tròn này với tâm của nó). Vị trí của phần cụ thể của người (như đỉnh đầu của người, lòng bàn chân người, phần giữa người) trong thế giới thực ba chiều có thể có sự tương ứng đơn nhất với vị trí của nó trong hình ảnh video hai chiều, và do đó, sự tồn tại của điểm cụ thể này của người trong hình ảnh video hai chiều có thể được sử dụng để xác định vị trí của người trong thế giới thực ba chiều.

Mô đun mô hình người tổng quát 303 có thể còn xác định kích thước của mô hình người đối với mỗi vị trí xác định trong không gian hình ảnh 2D. Kích thước của mô hình người có thể thu được từ sự hiệu chuẩn của hệ thống giám sát bằng video 101. Ví dụ, mô hình hiệu chuẩn có kích thước đã biết có thể dịch chuyển xung quanh khu vực cần được giám sát trong khi hệ thống giám sát bằng video 101 quay video phục vụ cho các mục đích hiệu chuẩn. Mô hình hiệu chuẩn có thể là người có chiều cao đã biết đi bộ quanh khu vực cần giám sát. Trong quá trình hiệu chuẩn, hệ thống này có thể xác định mô hình hiệu chuẩn trong video dưới dạng đốm màu tiền cảnh và nhận

biết (chẳng hạn, bằng cách truy cập thông tin hiệu chuẩn được cung cấp cho hệ thống giám sát bằng video 101 liên quan đến kích thước của mô hình hiệu chuẩn) đốm màu tiền cảnh này tương ứng với kích thước định trước (chẳng hạn, chiều cao định trước). Ở đây, khi mô hình hiệu chuẩn dịch chuyển qua khu vực cần giám sát trong quá trình hiệu chuẩn video, đối với các vị trí khác nhau trong hình ảnh video, hệ thống này có thể kết hợp chiều cao đã biết của mô hình hiệu chuẩn với kích thước trong hình ảnh video 2D. Ví dụ, khi phần giữa của mô hình hiệu chuẩn ở vị trí (x_1, y_1) , chiều cao của mô hình hiệu chuẩn có thể là 15 pixel (hoặc có thể được đo theo một vài cách đo khác). Khi phần giữa của mô hình hiệu chuẩn ở vị trí (x_2, y_2) , mô hình hiệu chuẩn có thể có chiều cao 27 pixel. Do đó, hệ thống giám sát bằng video 101 có thể kết hợp các kích thước của hình ảnh video 2D tại các vị trí cụ thể (chẳng hạn, các tọa độ (x, y)) trong hình ảnh video 2D với các kích cỡ (chẳng hạn, các chiều cao) trong thế giới thực bằng cách kết hợp kích thước của hình ảnh video 2D với kích thước đã biết (chẳng hạn, chiều cao) của mô hình hiệu chuẩn. Dựa trên sự kết hợp đã biết (thu được thông qua sự hiệu chuẩn này) giữa các kích thước của thế giới thực và các kích thước trong hình ảnh video 2D tại các vị trí cụ thể (chẳng hạn, các tọa độ (x, y)) trong hình ảnh 2D, kích thước 2D của mô hình người trong không gian hình ảnh video 2D có thể được tính cho mỗi vị trí trong số các vị trí khác nhau (các tọa độ (x, y)) trong hình ảnh video 2D để tương ứng với kích cỡ trung bình của người trong thế giới thực 3D.

Đối với các ví dụ về các quy trình hiệu chuẩn, xem patent US 7,932,923 được cấp cho Lipton và đồng tác giả và patent US 7,801,330, được cấp cho Zhang và đồng tác giả, mà nội dung của chúng được kết hợp vào đây bằng cách viện dẫn. Nói chung, bằng cách sử dụng các thông số được nhập hoặc thu được thông qua quy trình hiệu chuẩn, như chiều cao máy quay (H), góc trường nhìn ngang và dọc của máy quay (θ_H , θ_V), và góc lia của máy quay (α) và các thông tin khác, như các biên ngoài được phát hiện của đối tượng (chẳng hạn, điểm trên cùng và dưới cùng của người), hệ thống máy quay này nói chung có thể xác định kích thước và hình dạng ở thế giới thực của đối tượng phục vụ cho mục đích nhận dạng.

Mô hình hiệu chuẩn máy quay trên cơ sở người 304 có thể tiếp nhận và lưu trữ mô hình người có kích cỡ thích hợp từ mô đun mô hình người tổng quát 303 cùng với các vị trí tương ứng thích hợp trong không gian hình ảnh video. Các mô hình người này và các vị trí tương ứng có thể được lưu trữ trong bảng tìm kiếm. Ví dụ, mỗi trong số các tọa độ (x, y) trong và ngoài không gian hình ảnh video có thể được sử dụng để nhận dạng mô hình người tương ứng. Ví dụ, khi tọa độ định vị (x, y) tương ứng với trọng tâm của mô hình người, khi ước lượng sự tồn tại của đối tượng con người trong hình ảnh video có tâm ở vị trí (x_1, y_1) , bảng tìm kiếm của mô hình hiệu chuẩn máy quay trên cơ sở người 304 có thể tiếp nhận vị trí (x_1, y_1) làm đầu vào và kết xuất mô hình người tương ứng (bao gồm kích cỡ và vị trí của mô hình trong không gian hình ảnh 2D). Ví dụ, đầu ra này có thể bao gồm đường bao trong không gian hình ảnh 2D hoặc có thể bao gồm bộ điểm ảnh hoàn chỉnh (chẳng hạn, các tọa độ (x, y) của tất cả các điểm ảnh) trong không gian hình ảnh để mô tả mô hình người tương ứng.

Fig.3B minh họa ví dụ trong đó một vài mô hình người có mặt trong hình ảnh video hai chiều, mỗi người tương ứng với một vị trí khác nhau so với hình ảnh video hai chiều. Như được minh họa, bốn mô hình người 320a, 320b, 320c và 320d được liên kết với các tọa độ nhận dạng (x, y) khác nhau so với hình ảnh video hai chiều. Mô hình người 320a là nhỏ nhất, tương ứng với vị trí xa nhất từ nguồn video trong thế giới thực ba chiều. Các mô hình người 320b, 320c và 320d tương ứng với các vị trí trong thế giới thực ba chiều mà lần lượt gần với nguồn video hơn. Các mô hình người 320a, 320b, 320c và 320d có thể đều thu được từ cùng mô hình người đầy đủ. Tuy nhiên, có thể ước lượng rằng chỉ có một phần của mô hình người đầy đủ có thể xuất hiện trên hình ảnh video hai chiều tại các vị trí nhất định. Ở đây, mô hình người đầy đủ tương ứng với các hình người 320c và 320d được ước lượng chỉ chiếm một phần trong không gian hình ảnh video hai chiều 330; mô hình người 320c được ước lượng như sự kết hợp giữa thân và đầu của mô hình người đầy đủ, trong đó mô hình người 320d chỉ tương ứng với phần đầu của mô hình người đầy đủ.

Mỗi mô hình người 320a, 320b, 320c và 320d được liên kết với tọa độ định vị (x, y) đối với hình ảnh video hai chiều. Trong ví dụ này, các tọa độ định vị của các mô hình người 320a, 320b, 320c tương ứng với trọng tâm của mô hình người. Các tọa độ nhận dạng (x, y) liên kết với các hình dạng được ước lượng 320a, 320b và 320c lần lượt là 321a, 321b và 321c, và nằm trong các tọa độ (x, y) của hình ảnh video. Tọa độ định vị (x, y) liên kết với hình dạng được ước lượng 320d nằm ngoài các tọa độ (x, y) của hình ảnh video. Tức là, trong ví dụ này, trọng tâm của mô hình người liên kết với 320d nằm dưới hình ảnh video và do đó tọa độ định vị (x, y) của nó có giá trị âm trên trục y, mà trong ví dụ này, là nằm ngoài các tọa độ của hình ảnh video (và không được thể hiện trên Fig.3B). Để thuận tiện cho việc tính toán, các tọa độ nhận dạng (x, y) có thể tăng theo đơn vị pixel sao cho các tọa độ định vị 321a, 321b và 321c cũng xác định các điểm ảnh của hình ảnh video.

Fig.3B minh họa chỉ bốn mô hình người liên kết với bốn tọa độ định vị tương ứng, để thuận tiện cho việc giải thích. Tuy nhiên, mô hình hiệu chuẩn máy quay trên cơ sở người 304 có thể lưu trữ mô hình người với số lượng lớn các tọa độ nhận dạng (x, y) , một vài trong số này là các mô hình người có thể chồng chập lên nhau. Fig.3C minh họa một hàng tọa độ nhận dạng 321 (x, y) , mỗi tọa độ liên quan đến mô hình người tương ứng 320. Để thuận tiện cho việc giải thích, chỉ có một hàng được minh họa, nhưng các mô hình người có thể được tạo ra với nhiều hàng tọa độ nhận dạng (x, y) , mà có thể được phân bố đồng đều theo các chiều x và y trong không gian hình ảnh 330. Như đã được mô tả, kích cỡ của các hình dạng có thể khác nhau đối với các vị trí khác nhau (mặc dù chúng được thể hiện là có cùng kích cỡ trên Fig.3C). Ví dụ, mô hình hiệu chuẩn máy quay trên cơ sở người 304 có thể lưu trữ hình dạng người đối với mỗi điểm ảnh trong không gian hình ảnh 2D 330 dưới dạng các tọa độ nhận dạng (x, y) của không gian hình ảnh 2D 330 cũng như đối với các tọa độ (x, y) ngoài không gian hình ảnh 2D 330 được liên kết với mô hình người mà nằm ít nhất một phần trong không gian hình ảnh 2D 330. Ví dụ, đối với tất cả các tọa độ điểm ảnh (x, y) trong không gian hình ảnh video 330, mô hình hiệu chuẩn máy quay trên cơ sở người 304

có thể lưu trữ tọa độ định vị (x, y) và mô hình người liên kết (mà có thể bao gồm biên hoặc tập hợp điểm ảnh) của không gian con trong không gian hình ảnh video 330 dự kiến được chiếm bởi người khi trọng tâm của mô hình người được đặt ở tọa độ định vị (x, y) đó trong không gian hình ảnh video 330 của hình ảnh video. Các tọa độ nhận dạng (x, y) cũng có thể bao gồm tất cả các tọa độ nhận dạng (x, y) bên ngoài không gian hình ảnh video 330 có liên quan đến mô hình người trong không gian con trong không gian hình ảnh video 330 (tức là, một phần của mô hình người đầy đủ có thể được đặt trong không gian con của không gian hình ảnh video 330). Trong một số trường hợp, không gian con có thể bao gồm toàn bộ không gian hình ảnh video 330 (tương ứng với việc ước lượng khi người được định vị chiếm toàn bộ hình ảnh video). Mô hình hiệu chuẩn máy quay trên cơ sở người 304 có thể lưu trữ các tọa độ nhận dạng (x, y) và mô hình người liên kết dưới dạng bảng tìm kiếm. Mặc dù trọng tâm của mô hình dạng người đầy đủ tương ứng với các tọa độ nhận dạng (x, y) của mô hình người trong ví dụ này, thế nhưng các điểm nhận dạng khác của mô hình dạng người này có thể được sử dụng (chẳng hạn, mắt, mũi, giữa đầu, đỉnh đầu, ngón chân, bàn chân, v.v.).

Mô hình tính bản ánh xạ xác suất của người 305 sử dụng tập hợp đốm màu tiền cảnh của khung cụ thể của hình ảnh video được kết xuất bởi môđun phát hiện đốm màu tiền cảnh 105 và các mô hình người với các tọa độ định vị tương ứng của chúng được đưa ra từ mô hình hiệu chỉnh máy quay dựa trên người 304 để tính xác suất mục tiêu là người đối với mỗi vị trí trong số các vị trí trong hình ảnh video hai chiều, chẳng hạn như đối với mỗi vị trí điểm ảnh của hình ảnh. Các xác suất được tính này có thể được liên kết với các vị trí để tạo ra bản ánh xạ xác suất. Các vị trí này có thể là giống như các tọa độ nhận dạng (x, y) của các mô hình người.

Đối với mỗi tọa độ định vị (x, y) , việc tính toán được tiến hành để xác định xác suất tương ứng của việc tồn tại đối tượng con người trong hình ảnh video. Khi các tọa độ nhận dạng (x, y) có sự tương ứng một một với các điểm ảnh của hình ảnh video, thì

việc tính xác suất được tiến hành đối với mỗi trong số các điểm ảnh của hình ảnh video. Ví dụ, đối với mỗi điểm ảnh của hình ảnh, xác suất là người tương ứng có thể được tính bằng khả năng có mặt mục tiêu là người mà tâm hình ảnh của nó nằm ở điểm ảnh được xem xét. Bản ánh xạ xác suất có thể được tạo ra bằng cách ánh xạ mỗi trong số các kết quả toán xác suất với mỗi tọa độ định vị (x, y) . Bản ánh xạ xác suất có thể được lưu trữ trong bảng tìm kiếm, kết hợp mỗi tọa độ (x, y) (là đầu vào) với xác suất tính được tương ứng. Bảng tìm kiếm này có thể giống như bảng tìm kiếm của môđun mô hình hiệu chuẩn máy quay trên cơ sở người 304 (lưu trữ các mô hình người như đầu vào) hoặc có thể là bảng tìm kiếm thứ hai riêng rẽ.

Như nêu trên, các tọa độ định vị có thể nằm ngoài không gian hình ảnh video, và do đó việc tính toán có thể được tiến hành để xác định xác suất tương ứng của sự có mặt đối tượng con người trong hình ảnh video (đối với một phần của mô hình người đầy đủ 2D tương ứng nằm trong không gian hình ảnh (mô hình người) được liên kết với các tọa độ định vị này). Ví dụ, nếu trọng tâm của mô hình người đầy đủ 2D tương ứng với các tọa độ định vị, trọng tâm này có thể nằm bên ngoài không gian hình ảnh video, nhưng có thể tương ứng với mô hình người 2D trong không gian hình ảnh video là một phần của mô hình người đầy đủ. Ví dụ, vai và đầu của mô hình người đầy đủ có thể tạo ra mô hình người 2D (vai và đầu nằm trong không gian hình ảnh) mặc dù trọng tâm của mô hình người đầy đủ này (chẳng hạn, gần rốn của mô hình người đầy đủ) nằm ngoài không gian hình ảnh (trọng tâm tương ứng với các tọa độ định vị được dùng để xác định vai/đầu của mô hình người 2D tương ứng). Trong một số ví dụ, tỷ lệ % nhất định của mô hình người đầy đủ 2D phải nằm trong không gian hình ảnh để việc tính xác suất được tiến hành (được xem xét). Ví dụ, khi nhỏ hơn 10% hoặc nhỏ hơn 20% mô hình người đầy đủ 2D nằm trong không gian hình ảnh (hoặc, khi mô hình người nhỏ hơn 10% hoặc nhỏ hơn 20% mô hình người đầy đủ 2D), thì giá trị xác suất được liên kết với các tọa độ định vị có thể được thiết lập bằng 0 hoặc được bỏ qua. Trong một số ví dụ, khi nhỏ hơn 40% mô hình người đầy đủ 2D

nằm trong không gian hình ảnh, thì giá trị xác suất được liên kết với các tọa độ định vị có thể được thiết lập bằng 0.

Việc tính xác suất đối với mỗi tọa độ định vị (x, y) có thể là độ phủ mô hình người được liên kết với tọa độ định vị (x, y) tương ứng và tập hợp đốm màu tiền cảnh. Ví dụ, việc tính xác suất đối với mỗi tọa độ định vị (x, y) có thể là độ phủ các điểm ảnh thân người và các điểm ảnh biên người trong mô hình người được liên kết với tọa độ định vị (x, y) tương ứng. Mô hình người được liên kết với tọa độ định vị (x, y) tương ứng có thể được đưa ra từ mô đun mô hình hiệu chuẩn máy quay trên cơ sở người 304 (chẳng hạn, được lưu trữ trong bảng tìm kiếm của mô đun 304). Tập hợp đốm màu tiền cảnh có thể được đưa ra từ mô đun phát hiện đốm màu tiền cảnh 105. Việc gọi lại hình dạng đã được ước lượng với tập hợp đốm màu tiền cảnh có thể được tính bằng tỷ lệ diện tích mô hình người chồng chập với tập hợp đốm màu tiền cảnh với diện tích mô hình người. Các kết quả tính xác suất mà không vượt quá ngưỡng nhất định có thể được bỏ qua. Ví dụ, các xác suất tính được nhỏ hơn 0,4 (trên thang từ 0 đến 1) có thể chỉ ra rằng không có mục tiêu là người định tâm ở vị trí đó. Các bước lượng không phải bước lượng độ phủ có thể được tiến hành để xác định xác suất có mặt của đối tượng con người trong hình ảnh video tương ứng với mỗi trong số các hình dạng được ước lượng. Cần hiểu rằng các xác suất tính được là các giá trị ước lượng. Do đó, xác suất tính được bằng 1 (trên thang từ 0 đến 1) không biểu thị chắc chắn tuyệt đối sự có mặt của người tại vị trí liên quan tương ứng.

Fig.3D minh họa phương pháp tính bản ảnh xạ xác suất của người làm ví dụ, mà có thể được thực hiện bởi hệ thống trên Fig.3A. Ở bước S340, mô hình máy quay đã được hiệu chuẩn trong 304 có thể được sử dụng để ánh xạ mặt phẳng hình ảnh của không gian hình ảnh 2D lên trên mặt phẳng đất ở thế giới thực. Ở bước S342, mô hình người có thể thu được đối với N vị trí trong không gian hình ảnh 2D (N là số nguyên lớn hơn hoặc bằng 2). Mô hình máy quay đã được hiệu chuẩn 304 có thể được sử dụng để thu được mô hình người dạng thân lõi tương ứng làm mô hình người cho mỗi

vị trí điểm ảnh của hình ảnh trong không gian hình ảnh 2D. Mỗi trong số các mô hình người có thể được liên kết với tọa độ định vị trong không gian hình ảnh 2D. Ví dụ, điểm trọng tâm người của mô hình người có thể được sử dụng làm điểm mốc khi thực hiện việc ánh xạ đến tọa độ định vị. Giả sử tọa độ định vị của không gian hình ảnh 2D là trọng tâm của người trong không gian hình ảnh, thì vị trí vết chân thực tế tương ứng của nó trên mặt phẳng đất của thế giới thực có thể được tính thông qua mô hình máy quay đã được hiệu chuẩn (chẳng hạn, như được thể hiện trên Fig.5). Sau đó mô hình người tổng quát 3D (chẳng hạn, nhiều hình trụ) được đặt ở vị trí vết chân đó. Kích cỡ của mô hình 3D này có thể tương ứng với dữ liệu hiệu chuẩn thu được từ trước. Mô hình người tổng quát 3D có thể được chiếu hoặc được ánh xạ lên trên mặt phẳng hình ảnh 2D để thu được mô hình người trong không gian hình ảnh 2D. Ví dụ, việc chiếu mô hình người dạng nhiều hình trụ 3D có thể được sử dụng để tạo ra thân lõi dạng hình ảnh 2D tương ứng làm mô hình người dạng hình ảnh có trọng tâm ở tọa độ nhận dạng được liên kết (chẳng hạn, điểm hình ảnh được xem xét). Bằng cách này, mọi điểm ảnh của hình ảnh hợp lệ có thể có mô hình người dạng vùng lõi tương ứng (làm mô hình người) thể hiện kích thước và hình dạng tương tự với người tại vị trí hình ảnh đó. Để giảm chi phí tính, các mô hình người dạng vùng lõi có thể được tính trước ở lúc khởi tạo hệ thống, hộp giới hạn hình chữ nhật của mô hình người lõi có thể được sử dụng để thu được tỷ lệ độ phủ người gần đúng bằng cách sử dụng hình ảnh tích hợp. Ở bước S344, tập hợp đốm màu tiền cảnh có thể được trích xuất từ hình ảnh video. Tập hợp đốm màu tiền cảnh có thể bao gồm một hoặc nhiều đốm màu tiền cảnh được phát hiện sử dụng các điểm ảnh nổi của người được trích xuất bởi môđun 301 và/hoặc các điểm ảnh biên người được trích xuất bởi môđun 302. Ở bước S346, đối với mỗi vị trí trong số N vị trí, xác suất có mặt của người tại vị trí đó được tính để thu được bản ánh xạ xác suất. Số đo xác suất là người có thể được định nghĩa là tỷ lệ độ phủ đối với người được đưa ra có đầy đủ các điểm ảnh biên người trong mô hình người lõi dạng hình ảnh. Tỷ lệ độ phủ đối với người trong ví dụ này là số lượng điểm ảnh nổi của người được tính ở 301 trong mô hình người lõi dạng hình ảnh trên tổng diện tích của

mô hình lỗi của người này. Thứ tự các bước của quy trình trên Fig.3D có thể được thực hiện theo thứ tự không phải như được thể hiện. Ví dụ, bước 344 có thể được thực hiện trước một hoặc cả hai bước 340 và 342.

Liên quan đến Fig.3A, dựa trên bản ánh xạ xác suất là người được tính ở 305, môđun ước lượng mục tiêu là người 306 có thể tìm số lượng mô hình người lớn nhất (chẳng hạn, các đối tượng con người) trong hình ảnh video và các vị trí của chúng. Phương pháp tối ưu toàn cục có thể được sử dụng để tìm số lượng mô hình người lớn nhất và các vị trí của chúng. Nếu m ($m_1, \dots, m_M$) biểu thị tập hợp M của mô hình người từ tất cả các mô hình người có tiềm năng trong không gian hình ảnh, thì mục tiêu là tìm tập hợp tối ưu n^* sao cho hàm số tiêu chuẩn $f(n^*)$ đạt giá trị lớn nhất toàn cục. Tức là, mục tiêu là để tìm:

$$\operatorname{argmax}_{n \in m} f(n)$$

trong đó n là tập hợp cụ thể của các mô hình người trong không gian hình ảnh, và $f(n)$ là hàm được tính cho tập hợp các mô hình người này.

Như tiếp tục được mô tả dưới đây, hàm số $f(n)$ được tính cho mỗi trong số nhiều tập hợp mô hình người được lựa chọn, mỗi tập hợp lựa chọn m_i vị trí từ bản ánh xạ xác suất (m_i vị trí được lựa chọn cho mỗi lượt, trong đó số m_i có thể khác nhau đối với mỗi lượt trong số các lượt này). Mỗi tập hợp mô hình người có thể được lựa chọn với một lượt (hoặc quét) của bản ánh xạ xác suất, với tiêu chuẩn ràng buộc nhất định được dùng để lựa chọn các vị trí được thay đổi đối với mỗi lượt. Ở đây, hàm số $f(n)$ được định nghĩa là:

$$f(n) = w_R * R(n) + w_P * P(n) - w_O * O(n)$$

trong đó R là tỷ lệ độ phù đối với người, mà được định nghĩa là tỷ lệ % của diện tích nổi là người trên toàn bộ diện tích của nhóm gồm n mô hình người được lựa chọn; P là độ chính xác của người, là tỷ lệ % của diện tích nổi chồng chập với nhóm gồm n mô hình người được lựa chọn và O là tỷ lệ chồng chập người, là tỷ lệ diện tích

chồng chập của mô hình bất kỳ trong số n mô hình người được lựa chọn với nhau với diện tích bị chiếm bởi tất cả n mô hình người được lựa chọn, và w_R , w_P và w_O là các trọng số. Có thể có lợi nếu tìm ra sự trùng khớp nhất giữa vùng tiền cảnh (tập hợp điểm màu tiền cảnh) và sự kết hợp của các mô hình người (bộ gồm m mô hình người) mà không chồng chập quá nhiều người. Thực tế, cách để xác định ba trọng số nêu trên có thể tác động đáng kể đến các kết quả phát hiện, ví dụ, nếu yêu cầu để giảm tỷ lệ chồng chập người được đặt ra càng lớn, thì có thể dẫn đến kết quả đếm người càng thấp.

Mỗi trong số m_i mô hình người được lựa chọn có thể được lựa chọn bằng cách dựa vào bản ánh xạ xác suất được đưa ra bởi mô hình tính bản ánh xạ xác suất của người 305. Một vài lượt có thể được tiến hành để thực hiện việc tính $f(n)$, mỗi lượt lựa chọn tập hợp con gồm m_i mô hình người từ các mô hình người 2D được tạo ra bởi mô đun mô hình người tổng quát 303 và được liên kết với tọa độ định vị (x, y) trong mô hình hiệu chuẩn máy quay trên cơ sở người 304 (chẳng hạn, trong bảng tìm kiếm). Như nêu trên, giá trị m_i có thể khác nhau đối với mỗi lượt trong số các lượt này. Tiêu chuẩn lựa chọn của các mô hình người có thể khác nhau đối với mỗi lượt sao cho các mô hình người khác nhau được chọn đối với các lượt khác nhau (và có thể, số lượng m_i khác nhau của các mô hình người được chọn đối với các lượt khác nhau). Tiêu chuẩn lựa chọn có thể bao gồm việc yêu cầu mô hình người được lựa chọn được liên kết với ngưỡng xác suất P_{th} như được đưa ra bởi bản ánh xạ xác suất. Tiêu chuẩn lựa chọn cũng có thể bao gồm mô hình người 2D được lựa chọn tiếp theo để có khoảng cách nhỏ nhất D_{min} tính từ các mô hình người 2D được lựa chọn trước đó bất kỳ. Khoảng cách nhỏ nhất D_{min} có thể là khoảng cách trên mặt phẳng đất của thế giới thực. Ví dụ, các trọng tâm của các mô hình người 2D có thể được ánh xạ hoặc được dịch tới các vị trí trong thế giới thực 3D và các khoảng cách giữa chúng có thể được tính. Khoảng cách nhỏ nhất D_{min} có thể được tính trong mặt phẳng hình ảnh 2D, nhưng các khoảng cách trong mặt phẳng hình ảnh 2D có thể phản ánh các vị trí 3D tương ứng, sao cho đối với các mô hình người gần nguồn hình ảnh video, có thể cần sự tách biệt

lớn hơn đến trong mặt phẳng hình ảnh 2D so với sự tách biệt đối với các mô hình người xa hơn.

Trong một số phương án làm ví dụ, một hoặc nhiều bước quét một lượt nhanh của bản ảnh xạ xác suất được dùng để xác định số người được đếm và các vị trí tương ứng. Fig.3E minh họa phương pháp thực hiện một lượt của bản ảnh xạ xác suất như một phần trong việc tìm kiếm số lượng mô hình người lớn nhất trong hình ảnh video. Phương pháp trên Fig.3E có thể được thực hiện bởi mô đun ước lượng mục tiêu là người 306. Ở bước S350, bản ảnh xạ xác suất được quét để tìm giá trị cực đại địa phương (mà có thể được yêu cầu bởi tiêu chuẩn lựa chọn nhất định). Bản ảnh xạ xác suất có thể được quét để định vị giá trị cực đại địa phương hợp lệ không được lựa chọn mà tương ứng với vị trí trong thế giới thực gần nhất với nguồn video. Đáy của bản ảnh xạ xác suất có thể tương ứng với đáy của hình ảnh video. Trong nhiều phương án thực hiện, máy quay video thực hiện chức năng giám sát có thể được gắn ở vị trí cao hơn so với mức đầu người trong khu vực cần giám sát. Do đó, đáy của hình ảnh video có thể tương ứng với vị trí gần nhất với nguồn video. Việc quét bản ảnh xạ xác suất từ dưới lên trên trong ví dụ này cho phép lựa chọn các mô hình người ít có khả năng tương ứng với đối tượng bị giữ trong hình ảnh video.

Bản ảnh xạ xác suất có thể được quét từ dưới lên trên để tìm điểm cực đại cục bộ, thể hiện giá trị cực đại địa phương của các xác suất tính được trước đó (được lưu trong bản ảnh xạ xác suất) đối với mỗi vị trí trong số các vị trí trong không gian hình ảnh. Giá trị cực đại địa phương có thể là tọa độ định vị (x, y) (chẳng hạn, điểm ảnh) có giá trị xác suất cao hơn các giá trị xác suất của mỗi trong số các tọa độ nhận dạng (x, y) ngay cạnh nhau (chẳng hạn, các điểm ảnh ngay cạnh nhau). Khi giá trị cực đại địa phương được tìm thấy, mô hình người liên kết với điểm lớn nhất cục bộ này khi các tọa độ định vị của mô hình này được lựa chọn làm một trong số tập hợp của m_i mô hình người ở bước S352. Ở bước S354, tất cả các điểm ảnh trong vùng trong của mô hình được lựa chọn này (chẳng hạn, nằm trong biên của mô hình người 2D) và các

điểm ảnh tương ứng với khoảng cách tối thiểu $D_{\min}$ cách xa mô hình được lựa chọn này (chẳng hạn, các điểm ảnh trong hình ảnh video thể hiện khoảng cách tối thiểu trên mặt phẳng đất của thế giới thực) được loại trừ ra khỏi sự xem xét tiếp ở lượt này (và có thể được loại bỏ tạm thời khỏi bản ánh xạ xác suất đối với lượt này). Lưu ý rằng trong ví dụ này, các điểm ảnh tương ứng với các tọa độ định vị của các mô hình người và phần mô tả này cũng có thể áp dụng được cho các tọa độ định vị mà không phải các vị trí điểm ảnh. Trong một số ví dụ, bản thân hình ảnh video không cần được phân tích tiếp ở giai đoạn này, và các điểm ảnh này có thể được loại trừ khỏi sự xem xét tiếp đơn giản bằng cách loại bỏ tạm thời chúng ra khỏi bản ánh xạ xác suất. Bản ánh xạ xác suất được quét lại để lựa chọn điểm lớn nhất cục bộ khác của các xác suất của bản ánh xạ xác suất là người được liên kết với các điểm ảnh mà lớn hơn so với ngưỡng xác suất P_{th} và đã không được loại trừ. Ở bước S356, các điểm ảnh hợp lệ bất kỳ định trước đã được xem xét hay chưa. Tức là, xác suất có thể được xem xét lại đối với các giá trị không được loại trừ bởi tiêu chuẩn lựa chọn và cũng không được loại trừ bởi sự lựa chọn các mô hình người khác trong quá trình quét bản ánh xạ xác suất này. Quá trình quét bản ánh xạ xác suất được tiếp tục cho đến khi tất cả các điểm ảnh hợp lệ được xem xét và được loại bỏ khỏi bản ánh xạ này. Do đó, m_i mô hình người có thể được lựa chọn nhờ quá trình quét bản ánh xạ xác suất này. Đối với lượt này, hàm số $f(m_i)$ được tính đối tập hợp gồm m_i mô hình người.

Các bước quét bổ sung bản ánh xạ xác suất có thể được thực hiện, mỗi lần quét có tập hợp tiêu chuẩn lựa chọn khác nhau. Fig.3F minh họa phương pháp thực hiện nhiều lượt của bản ánh xạ xác suất để tìm số lượng mô hình người lớn nhất trong hình ảnh video. Phương pháp trên Fig.3F có thể được thực hiện bởi môđun ước lượng mục tiêu là người 306. Ở đây, giá trị của ít nhất một trong số $D_{\min}$ (khoảng cách tối thiểu) và P_{th} (ngưỡng xác suất) có thể là khác nhau đối với mỗi lần quét. Ở bước S360, tiêu chuẩn lựa chọn là tập hợp dùng cho việc quét một lượt cụ thể. Số phương án khác nhau của tiêu chuẩn lựa chọn là bao nhiêu (và do đó số lần quét là bao nhiêu) có thể định trước trên cơ sở từng trường hợp cụ thể, có xét đến độ chính xác mong muốn và

tổng chi phí tính toán. Ở bước S362, quá trình quét bản ánh xạ xác suất được tiến hành để lựa chọn tập hợp gồm m mô hình người theo tiêu chuẩn lựa chọn. Giá trị m là số nguyên lớn hơn hoặc bằng 0 và có thể khác nhau đối với mỗi lựa chọn (chẳng hạn, đối với mỗi vòng lặp trên Fig.3F thực hiện bước S362). Bước S362 có thể tương ứng với phương pháp trên Fig.3E. Ở bước S364, hàm số tiêu chuẩn được tính đối với m_i mô hình người được lựa chọn, chẳng hạn, $f(m_i)$ tương ứng được tính đối với m_i mô hình người được lựa chọn ở bước quét này. Các bước quét bổ sung có thể được thực hiện theo tiêu chuẩn lựa chọn mới (S366). Khi tất cả các bước quét bản ánh xạ xác suất được hoàn thành, giá trị lớn nhất của $f(n)$, $ne\{m_i, \dots, m_M\}$ của nhóm gồm các lần quét định trước. Tập hợp các mô hình người tương ứng với giá trị lớn nhất này định trước để tương ứng với các đối tượng con người trong hình ảnh video (S368). Bằng cách sử dụng các tọa độ định vị (x, y) (chẳng hạn, các vị trí điểm ảnh) của các mô hình người định trước để thể hiện các đối tượng con người trong hình ảnh video, vị trí thể giới thực trên mặt phẳng đất có thể định trước.

Theo phương án khác, nếu m biểu thị bộ gồm các mô hình người từ tất cả các mô hình người có tiềm năng trong không gian hình ảnh, mục tiêu có thể là để tìm tập hợp tối ưu m^* sao cho hàm số tiêu chuẩn $g(m^*)$ đạt giá trị lớn nhất toàn cục. Tức là, mục tiêu là để tìm giá trị lớn nhất của:

$$g(m) = \sum_{n=1}^m f(n)$$

trong đó n là một mô hình cụ thể trong số các mô hình người trong không gian hình ảnh, m là số mô hình người được lựa chọn (mà có thể thay đổi đối với các phép tính tổng khác nhau), và $f(n)$ là hàm số được tính đối với mỗi mô hình trong số m mô hình người, chứ không phải nhóm các mô hình.

Ở đây, hàm số $f(n)$ được định nghĩa là:

$$f(n) = w_R * R(n) + w_P * P(n) - w_O * O(n)$$

trong đó R là tỷ lệ độ phủ đối với người, mà được định nghĩa là tỷ lệ % của diện tích nổi là người trên toàn bộ diện tích của mô hình người được lựa chọn; P là độ chính xác của người, là tỷ lệ % của diện tích nổi là người chồng chập với mô hình người được lựa chọn và O là tỷ lệ chồng chập người, là phần chồng chập của mô hình người thứ n được lựa chọn có các diện tích bị chiếm bởi các mô hình người từ thứ nhất đến người thứ $n-1$ [các diện tích bị chiếm bởi các mô hình người được lựa chọn từ trước trong lượt hiện tại khi tính $\sum f(n)$], và w_R , w_P và w_O là các trọng số. Mỗi trong số các lượt quét bản ảnh xạ xác suất được thảo luận trên đây có thể được liên kết với việc tính $\sum f(n)$, với các ràng buộc khác nhau đối với tiêu chuẩn lựa chọn khi lựa chọn các điểm lớn nhất cục bộ của bản ảnh xạ xác suất đối với mỗi lượt. Các hàm số $f(n)$ khác có thể được sử dụng ngoài các hàm số được mô tả trong bản mô tả.

Các Fig.6A, Fig.6B và Fig.6C thể hiện các kết quả phát hiện của hệ thống giám sát bằng video 101 theo một ví dụ. Đối với một khung đầu vào, Fig.6A là đầu ra của mô đun phát hiện thân người 301 và mô đun phát hiện điểm ảnh biên người 302, trong đó các điểm ảnh 301a chỉ các điểm ảnh thân người được phát hiện và các điểm ảnh 302a thể hiện các điểm ảnh biên người. Tập hợp đốm màu tiền cảnh được thể hiện trên Fig.6A dưới dạng kết hợp của các điểm ảnh thân người được phát hiện 301a và các điểm ảnh biên người 302a. Các điểm ảnh thân người và các điểm ảnh biên người được phát hiện xếp chồng lên toàn bộ khung hình ảnh video gốc xác định không gian hình ảnh video 330. Trong ví dụ này, phần còn lại của hình ảnh video trong khung hình ảnh video này (không phải tập hợp đốm màu tiền cảnh) là một phần của hình ảnh nền.

Fig.6B minh họa bản ảnh xạ xác suất của người được tính từ Fig.6A. Trong ví dụ này, bản ảnh xạ xác suất của người thể hiện các xác suất tính được trên thang màu xám, với màu đen tương ứng với xác suất bằng không (0) và màu trắng tương ứng với xác suất bằng một (1). Mỗi xác suất tính được được thể hiện ở vị trí trong không gian

hình ảnh 330 tương ứng với điểm ảnh tương ứng với các tọa độ định vị của mô hình người tương ứng.

Fig.6C thể hiện kết quả phát hiện người, minh họa các mô hình người 320 (đường bao dạng lõi màu hồng) là người tương ứng được phát hiện. Mỗi mô hình người này có thể được kết hợp bởi tọa độ định vị (như trọng tâm) mà có thể xác định vị trí của người được phát hiện trong thế giới thực 3D và được ánh xạ lên mặt phẳng đất của thế giới thực (không được thể hiện).

Các Fig.7A, Fig.7B và Fig.7C minh họa ví dụ về việc đo mật độ đám đông người dựa trên các kết quả phát hiện người. Fig.7A minh họa các kết quả làm ví dụ về các kết quả phát hiện của hệ thống giám sát bằng video 101, thể hiện các mô hình người 2D 320 (các bao lõi màu hồng), mỗi kết quả là người tương ứng được phát hiện, nằm chồng lên hình ảnh video gốc. Fig.7B minh họa việc ánh xạ những người được phát hiện lên mặt phẳng đất vật lý của thế giới thực, thể hiện hình biểu diễn nhìn từ trên xuống của hình ảnh video trên Fig.7A với mỗi đường tròn thể hiện mô hình người 320 khi được ánh xạ lên mặt phẳng đất vật lý của thế giới thực và nhờ đó xác định vị trí của người được phát hiện trong thế giới thực. Các mục tiêu là người được phát hiện có thể được ánh xạ lên trên mặt phẳng đất vật lý do sự hiệu chuẩn đã tạo ra sự tương quan giữa kích cỡ đã biết của mô hình hiệu chuẩn, vị trí trong hình ảnh 2D, và kích cỡ tương ứng trong các không gian hình ảnh. Với các vị trí đã biết, các tính toán có thể được tiến hành để đếm số người trong khu vực định trước nhất định (chẳng hạn, được lựa chọn bởi người sử dụng) hoặc trong toàn bộ hiện trường. Các tính toán cũng có thể được tiến hành để xác định số người trên một diện tích. Các kết quả đo mật độ đám đông thực ở mỗi vị trí mặt đất cũng có thể được tính một cách trực tiếp. Định nghĩa thực tế của kết quả đo mật độ đám đông có thể phụ thuộc vào ứng dụng thực tế, cụ thể, phụ thuộc vào kích cỡ của đám đông cần được theo dõi. Ví dụ, đối với trường hợp được thể hiện trên các Fig 6A, Fig.6B và Fig.6C, có thể sử dụng số người có trong bán kính 2 m làm kết quả đo mật độ đám đông. Trong khi đối với trường hợp trên các

Fig.7A, Fig.7B và Fig.7C, mật độ đám đông của vị trí có thể định trước là số người có trong bán kính 6 m. Fig.7C minh họa bản ánh xạ mật độ đám đông sử dụng bán kính 6 m với màu hồng có cường độ lớn hơn thể hiện mật độ đám đông cao hơn.

Dựa trên kết quả đo mật độ đám đông đối với mỗi khung video, có thể phát hiện nhiều sự kiện liên quan đến đám đông như được thể hiện trên Fig.8, bao gồm sự phát hiện đám đông, sự tập trung đông người và sự phân tán đám đông, mà có thể được phát hiện bởi các môđun 801, 802 và 803 tương ứng của môđun phát hiện sự kiện 108 trên Fig.1. Fig.9 minh họa phương pháp làm ví dụ về cách xác định và phát hiện khu vực đông người. Khối 901 minh họa cách để xác định sự kiện liên quan đến khu vực đông người. Đầu tiên người sử dụng có thể lựa chọn khu vực quan tâm trên hình ảnh (chẳng hạn, trong không gian hình ảnh). Tiếp theo, một vài ngưỡng mật độ đám đông có thể được sử dụng để xác định mức độ đám đông được quan tâm. Các ngưỡng này có thể là số người trong diện tích có bán kính nhất định. Các ngưỡng trễ có thể được sử dụng để có hiệu quả mạnh hơn. Ví dụ, nếu xác định mật độ đám đông là số người trong khu vực bán kính 3 m, có thể thiết lập hai ngưỡng mật độ đám đông: $T_{high} = 10$ và $T_{low} = 8$. Một khu vực có thể được coi là khu vực đám đông chỉ khi mật độ đám đông tương ứng lớn hơn hoặc bằng T_{high} . Khu vực đám đông trở nên không đông chỉ khi mật độ đám đông tương ứng nhỏ hơn hoặc bằng T_{low} . Khu vực đám đông có thể định trước bởi đám đông định trước và có thể thay đổi vị trí và/hoặc hình dạng từ khung này đến khung khác. Tâm của khu vực đám đông có thể được sử dụng để mô tả vị trí đám đông. Ngưỡng thời lượng tối thiểu có thể định trước là thời lượng tối thiểu khu vực đám đông được duy trì là đám đông trước khi kích hoạt việc phát hiện sự kiện. Đối với đầu vào khung video mới, khối 902 xem xét kỹ tất cả các mục tiêu là người được phát hiện để xem liệu nó thuộc khu vực đám đông hay không, sau đó khối 903 kiểm tra tất cả các khu vực đám đông để cập nhật tình trạng của chúng. Khi được phát hiện, các đám đông và các vị trí của chúng có thể được theo dõi từng khung hình ảnh video một. Ví dụ, miễn là đám đông đã được phát hiện và tiếp tục thỏa mãn ngưỡng tối thiểu T_{low} , các mô hình người liên kết với khu vực đám đông có thể xác định đám đông

trong các khung tiếp theo của hình ảnh video miễn là chúng còn tồn tại trong diện tích đáp ứng mật độ đám đông tối thiểu. Các mô hình người bổ sung có thể được bổ sung vào đám đông đã được phát hiện khi chúng dịch chuyển vào khu vực đám đông đã được phát hiện.

Fig.10 minh họa quy trình xử lý làm ví dụ đối với mỗi mục tiêu con người phát hiện được. Khối 1001 kiểm tra liệu mục tiêu hiện tại có bên trong hay gần khu vực đám đông hiện có hay không. Nếu "có", khối 1001 cập nhật số người được đếm đối với khu vực đó. Nếu "không", khối 1002 tính mật độ đám đông ở vị trí của mục tiêu hiện tại, sau đó khối 1004 kiểm tra liệu kết quả đo mật độ đám đông có lớn hơn hoặc bằng ngưỡng T_{high} hay không. Nếu "có", khu vực đám đông mới được tạo có tâm ở mục tiêu hiện tại. Nếu "không", tiếp tục xử lý mục tiêu là người tiếp theo.

Fig.11 minh họa quy trình xử lý làm ví dụ đối với mỗi khu vực đám đông. Khối 1101 cập nhật diện tích khu vực và đếm đám đông dựa trên các kết quả xử lý mục tiêu; Khối 1102 kiểm tra liệu mật độ đếm được vẫn lớn hơn ngưỡng định trước bởi người sử dụng hay không; nếu "không", khu vực đám đông này được loại bỏ khỏi danh sách theo dõi. Khối 1104 tiếp tục kiểm tra liệu thời lượng duy trì đám đông của khu vực đám đông đang được xử lý có cao hơn hoặc bằng ngưỡng định trước bởi người sử dụng hay không. Nếu "có", khối 1105 tiếp tục kiểm tra liệu sự kiện đông người tương ứng có được báo cáo hay không, nếu không, khối 1106 sẽ thực hiện hoạt động, chẳng hạn như báo cáo sự kiện đám đông và đánh dấu vùng đám đông này là "được báo cáo".

Fig.12 minh họa phương pháp có thể được sử dụng để xác định và phát hiện các sự kiện "tập trung" và "phân tán" đám đông. Ở đây "tập trung" và "phân tán" liên quan đến hai quá trình hình thành và kết thúc điểm tập trung đông người. Trong ví dụ này, điểm tập trung đông người liên quan đến khu vực có mật độ đám đông tĩnh cục bộ cao và khác với đám đông đang di chuyển như khi diễu hành. Tuy nhiên, sáng chế không chỉ giới hạn ở đó và phương pháp này cũng có thể được áp dụng để phát hiện

sự dịch chuyển của các điểm tập trung đông người. Khối 1201 minh họa cách điểm tập trung đông người có thể định trước. Đầu tiên người sử dụng có thể lựa chọn khu vực quan tâm trên hình ảnh. Tiếp theo, một số ngưỡng mật độ đám đông có thể được sử dụng để xác định có bao nhiêu đám đông là quan trọng. Ngưỡng thời lượng tối thiểu có thể xác định thời lượng tối thiểu khu vực đám đông phải duy trì đông đúc để được coi là điểm tập trung hợp lệ. Khối 1202 phát hiện các điểm tập trung đông người. Khối 1203 cập nhật và theo dõi các điểm tập trung đông người được phát hiện và phát hiện các sự kiện “tập trung” và “phân tán” đám đông.

Fig.13 minh họa một ví dụ về việc xác định điểm tập trung đông người. Bao gồm khu vực bên trong được biểu thị bởi số chỉ dẫn 1301 và khu vực bên ngoài được biểu thị bởi số chỉ dẫn 1302. Hai khu vực này có thể định trước bởi điểm tâm O , bán kính ngắn r và bán kính dài R . Trong ví dụ này, điểm tập trung đông người có thể thỏa mãn hai tiêu chuẩn sau:

- Mật độ đám đông của khu vực bên trong phải lớn hơn hoặc bằng ngưỡng định trước;
- Người được đếm trong khu vực bên ngoài phải nhỏ hơn (chẳng hạn, nhỏ hơn 2 lần, 4 lần, 10 lần, v.v.) so với người được đếm trong khu vực bên trong. Theo cách khác, mật độ đám đông ở khu vực bên ngoài phải nhỏ hơn (chẳng hạn, nhỏ hơn 2 lần, 4 lần, 10 lần, v.v.) so với mật độ đám đông ở khu vực bên trong.

Hai tiêu chuẩn nêu trên có thể chỉ ra rằng khu vực bên trong là điểm tập trung đông người, không chỉ khu vực bên trong đám đông lớn.

Các Fig 14A và 14B thể hiện ví dụ về điểm tập trung đông người. Fig.14A và Fig.14B đều thể hiện khung video và các mục tiêu là người được phát hiện được ánh xạ lên trên mặt phẳng đất vật lý của thế giới thực. Mặc dù Fig.14A có các mục tiêu là người nhiều hơn, nhưng chỉ Fig.14B chứa điểm tập trung đông người như được định nghĩa nêu trên.

Fig.15 minh họa phương pháp làm ví dụ để phát hiện các điểm tập trung đông người. Đối với mỗi mục tiêu con người phát hiện được, các khối 1501 kiểm tra mục tiêu này liệu có thuộc điểm tập trung đông người hiện có hay không. Nếu “có”, mục tiêu này được sử dụng để cập nhật tình trạng hiện tại của điểm tập trung đông người tương ứng trong khối 1502. Nếu “không”, khối 1503 tiếp tục kiểm tra liệu mục tiêu hiện tại có ở chính giữa của điểm tập trung đông người mới hay không. Nếu “có”, khối 1504 khởi tạo điểm tập trung đông người mới để tiếp tục theo dõi. Nếu “không”, mô đun này tiếp tục kiểm tra việc phát hiện người tiếp theo.

Fig.16 minh họa phương pháp làm ví dụ để cập nhật các điểm tập trung đông người và phát hiện các sự kiện “tập trung” và “phân tán” đám đông. Khối 1601 cập nhật vị trí và diện tích của điểm tập trung đông người sử dụng các kết quả phát hiện người mới trên khung video đang được xem xét. Khối 1602 kiểm tra liệu sự kiện “tập trung” đông người có được phát hiện từ điểm tập trung đông người hiện tại hay không. Nếu “không”, khối 1603 tiếp tục phát hiện sự kiện “tập trung” đông người bằng cách kiểm tra liệu điểm tập trung đông người có được cập nhật thành công trong thời lượng nhất định hay không. Ngưỡng thời lượng này có thể được thiết lập bởi người sử dụng tại thời điểm xác định quy tắc. Khi điểm tập trung đông người đã tạo ra sự kiện “tập trung”, khối 1604 tiếp tục theo dõi điểm tập trung để phát hiện sự kiện “phân tán”. Ở đây, sự kiện “phân tán” đám đông được định nghĩa là điểm tập trung đông người trở thành vết trống hoặc vết có mật độ thấp (chẳng hạn, dưới ngưỡng mật độ đám đông tối thiểu T_{low}) trong thời lượng ngắn. Khối 1604 phát hiện hai thời lượng cụ thể rất ngắn của điểm tập trung đông người: thời điểm để điểm trở nên không đông đúc và thời điểm tụ điểm trở nên trống hoặc ít. Nếu thời gian giữa hai thời điểm này ngắn hơn so với ngưỡng được định nghĩa bởi người sử dụng, thì hiện tượng “phân tán” đám đông đã được phát hiện.

Fig.17 minh họa ví dụ về hệ thống nhiều máy quay mà sáng chế có thể được ứng dụng ở đó. Trong ví dụ này, hai máy quay 1702 và 1704 chụp riêng rẽ các hình

ảnh video của hiện trường cần quan tâm từ các góc nhìn khác nhau. Đối với mỗi máy quay 1702 và 1704 hệ thống giám sát bằng video 101 và các phương pháp được mô tả trong bản mô tả có thể là giống như được mô tả trong bản mô tả của môđun phát hiện sự thay đổi 103, môđun phát hiện chuyển động 104, môđun phát hiện đếm màu tiền cảnh 105, môđun mô hình người tổng quát 303, mô hình hiệu chuẩn máy quay trên cơ sở người 304 và mô hình tính bản ánh xạ xác suất của người 305 - tức là, bản thân mỗi máy quay có thể có môđun hoặc chức năng môđun của chính chúng (nếu hệ mạch được dùng chung) của các môđun này.

Các mô hình người 2D với không gian hình ảnh tương ứng được tạo ra bởi mô hình hiệu chuẩn máy quay trên cơ sở người 304 của mỗi máy quay video 1702, 1704, cũng có thể được liên kết với tọa độ của mặt phẳng đất vật lý của thế giới thực. Ví dụ, đối với môđun mô hình hiệu chuẩn máy quay trên cơ sở người 304 của mỗi máy quay, đầu vào bổ sung có thể được tạo ra đối với tọa độ mặt phẳng đất vật lý tương ứng để nhờ đó liên kết mỗi mô hình trong số N mô hình người với nhau. Khi tính bản ánh xạ xác suất của người đối với mỗi máy quay 1702, 1704, xác suất của mỗi bản ánh xạ xác suất có thể được ánh xạ lên mặt phẳng đất vật lý không phải không gian hình ảnh 2D.

Trong một ví dụ, môđun ước lượng mục tiêu là người 306, phát hiện số người tốt nhất có thể thực hiện các bước quét bản ánh xạ xác suất thứ nhất của một máy quay theo cách được nêu trên đây, tức là, với các ràng buộc của tiêu chuẩn tìm kiếm, tìm kiếm giá trị cực đại địa phương của bản ánh xạ xác suất thứ nhất. Khi tính hàm số tiêu chuẩn để xác định giá trị lớn nhất đối với M tập hợp mô hình người $m(m_1, \dots, m_M)$, mục tiêu là để tìm:

$$\operatorname{argmax}_{n \in m} f_1(n) + f_2(n)$$

trong đó n là bộ cụ thể của các mô hình người 3D, mà có thể có các tọa độ định vị trong mặt phẳng đất vật lý mà các xác suất của mỗi trong hai bản ánh xạ xác suất của người được ánh xạ lên. Tức là, dựa vào việc lựa chọn điểm trong thế giới thực

được liên kết với mô hình người đối với tập hợp mô hình, các mô hình người trong không gian hình ảnh 2D được liên kết với điểm này được nhận biết đối với mỗi hệ thống máy quay, với một mô hình người được dùng để tính $f_1(n)$ và mô hình còn lại để tính $f_2(n)$. $f_1(n)$ và $f_2(n)$ có thể là giống như các hàm số được mô tả trong bản mô tả (tương ứng với tập hợp đốm màu tiền cảnh của người hoặc diện tích nổi là người được trích xuất từ hình ảnh video thích hợp):

$$f(n) = w_R * R(n) + w_P * P(n) - w_O * O(n)$$

trong đó (đối với n mô hình người 2D được lựa chọn tương ứng được liên kết với hình ảnh video và diện tích nổi là người của hình ảnh video đó) R là tỷ lệ độ phủ đối với người, mà được định nghĩa là tỷ lệ % của diện tích nổi là người trên toàn bộ diện tích của nhóm gồm n mô hình người được lựa chọn; P là độ chính xác của người, là tỷ lệ % của diện tích nổi là người mà đang chồng chập với nhóm gồm n mô hình người được lựa chọn và O là tỷ lệ chồng chập của người, là tỷ lệ của diện tích phần chồng chập của mô hình bất kỳ trong số n mô hình người được lựa chọn với nhau với diện tích bị chiếm bởi tất cả n mô hình người được lựa chọn và mô hình người thứ n được lựa chọn với các diện tích được chiếm bởi các mô hình người từ người thứ nhất đến người thứ $n-1$ [các diện tích được chiếm bởi các mô hình người được lựa chọn từ trước trong lượt hiện tại khi tính $f(n)$], và w_R , w_P và w_O là các trọng số. Lưu ý rằng các trọng số có thể khác nhau giữa các hàm số $f_1(n)$ và $f_2(n)$. Việc loại trừ các điểm ảnh để xem xét tiếp theo khi lựa chọn giá trị cực đại địa phương tiếp theo có thể chiếu mô hình người 3D được liên kết với tọa độ mặt phẳng đất của mô hình người được lựa chọn từ trước về mỗi trong hai bản ảnh xạ xác suất trong mặt phẳng hình ảnh tương ứng.

Trong một phương án khác, một bản ảnh xạ xác suất có thể được sử dụng cho nhiều máy quay. Trong ví dụ trên Fig.17, các bước lượng xác suất có thể được tiến hành đối với mỗi hình ảnh video 2D, như được mô tả trong bản mô tả và tạo ra hai bản ảnh xạ xác suất trên mặt phẳng hình ảnh, mỗi bản ảnh xạ tương ứng với mặt

phẳng hình ảnh 2D tương ứng. Các xác suất của bản ánh xạ xác suất trên mặt phẳng hình ảnh có thể được thiết lập bằng 0 nếu chúng không cao quá ngưỡng nhất định (mà có thể là giống nhau hoặc khác nhau đối với mỗi bản ánh xạ xác suất trên mặt phẳng hình ảnh). Các tọa độ định vị trong mỗi bản ánh xạ xác suất trên mặt phẳng hình ảnh có thể được dịch đến tọa độ trên mặt phẳng đất trong thế giới thực đối với mỗi bản ánh xạ xác suất trên mặt phẳng hình ảnh, tạo ra bản ánh xạ xác suất trên mặt phẳng đất đối với mỗi hình ảnh video. Hai bản ánh xạ xác suất trên mặt phẳng đất có thể được hợp nhất bằng cách nhân các xác suất mà dùng chung cùng các tọa độ trên mặt phẳng đất để tạo ra bản ánh xạ xác suất hợp nhất. Bản ánh xạ xác suất trên mặt phẳng đất hợp nhất có thể được quét để tìm các giá trị cực đại địa phương. Mỗi giá trị cực đại địa phương tìm được có thể xác định các mô hình người riêng rẽ đối với mỗi hình ảnh video trong không gian hình ảnh tương ứng của chúng mà sau đó có thể được dùng để tính $f_1(n)$ hoặc $f_2(n)$ (được nêu trên đây) nếu thích hợp. Việc thực hiện nhiều bước quét bản ánh xạ xác suất trên mặt phẳng đất hợp nhất đối với các giá trị cực đại địa phương có thể được thực hiện để tìm các mô hình người tiếp theo (mô hình dùng cho mỗi hình ảnh video) và để tính

$$f_1(n) + f_2(n)$$

Các ràng buộc khi lựa chọn (như ngưỡng xác suất tối thiểu và khoảng cách tối thiểu trong thế giới thực 3D) có thể được thay đổi và lượt quét mới được tiến hành để tìm tập hợp tối ưu của m mô hình người 3D (trong ví dụ này tương ứng với 2 mô hình người 2D).

Trong một ví dụ khác, mô đun ước lượng mục tiêu là người 306, việc phát hiện số người tốt nhất có thể thực hiện các bước quét bản ánh xạ xác suất thứ nhất của một máy quay theo cách được nêu trên đây, tức là, trong các ràng buộc của tiêu chuẩn tìm kiếm, tìm kiếm giá trị cực đại địa phương của bản ánh xạ xác suất thứ nhất. Khi tính hàm số tiêu chuẩn để xác định giá trị lớn nhất của các tập hợp gồm m mô hình người, mục tiêu là để tìm giá trị lớn nhất của:

$$\sum_{n=1}^m f_1(n) + f_2(n)$$

trong đó n là tọa độ định vị trong mặt phẳng đất vật lý mà các xác suất của mỗi trong số hai bản ảnh xạ xác suất của người được ánh xạ. Tức là, dựa trên việc lựa chọn điểm trong thế giới thực, các mô hình người trong không gian hình ảnh 2D được liên kết với điểm này định trước đối với mỗi hệ thống máy quay, với một mô hình người được dùng để tính $f_1(n)$ và mô hình còn lại để tính $f_2(n)$. $f_1(n)$ và $f_2(n)$ có thể là giống như hàm số được nêu trên đây (tương ứng với tập hợp đốm màu tiền cảnh của người hoặc diện tích màu nổi của người được trích xuất từ hình ảnh video thích hợp):

$$f(n) = w_R * R(n) + w_P * P(n) - w_O * O(n)$$

trong đó R là tỷ lệ độ phủ đối với người, mà được định nghĩa là tỷ lệ % của diện tích nổi là người trên toàn bộ diện tích của mô hình người được lựa chọn; P là độ chính xác của người, là tỷ lệ % của diện tích nổi mà đang chồng chập với mô hình người được lựa chọn và O là tỷ lệ chồng chập của người, mà là chồng chập của mô hình người thứ n được lựa chọn với các diện tích được chiếm bởi các mô hình người từ thứ nhất đến người thứ $n-1$ [các diện tích được chiếm bởi các mô hình người được lựa chọn từ trước ở lượt hiện tại khi tính $\sum f(n)$], và w_R , w_P và w_O là các trọng số.

Lưu ý rằng các trọng số này có thể khác nhau giữa các hàm số $f_1(n)$ và $f_2(n)$. Ngoại trừ các điểm ảnh để tiếp tục xem xét khi lựa chọn giá trị cực đại địa phương tiếp theo có thể chiếu mô hình người 3D được liên kết với tọa độ mặt phẳng đất của mô hình người được lựa chọn từ trước lên mỗi trong hai bản ảnh xạ xác suất trong mặt phẳng hình ảnh tương ứng.

Trong một phương án khác, một bản ảnh xạ xác suất có thể được sử dụng cho nhiều máy quay. Trong ví dụ trên Fig.17, các bước lượng xác suất có thể được tiến hành đối với mỗi hình ảnh video 2D, như được mô tả trong bản mô tả và tạo ra hai bản ảnh xạ xác suất trên mặt phẳng hình ảnh, mỗi bản ảnh xạ này tương ứng với mặt phẳng hình ảnh 2D tương ứng. Các xác suất của bản ảnh xạ xác suất trên mặt phẳng hình ảnh có thể được thiết lập bằng 0 nếu chúng không cao quá ngưỡng nhất định (mà

có thể là giống nhau hoặc khác nhau đối với mỗi bản ảnh xạ xác suất trên mặt phẳng hình ảnh). Các tọa độ định vị trong mỗi bản ảnh xạ xác suất trên mặt phẳng hình ảnh có thể được dịch đến tọa độ trên mặt phẳng đất trong thế giới thực đối với mỗi bản ảnh xạ xác suất trên mặt phẳng hình ảnh, tạo ra bản ảnh xạ xác suất trên mặt phẳng đất đối với mỗi hình ảnh video. Hai bản ảnh xạ xác suất trên mặt phẳng đất có thể được hợp nhất bằng cách nhân các xác suất mà dùng chung cùng các tọa độ trên mặt phẳng đất để tạo ra bản ảnh xạ xác suất hợp nhất. Bản ảnh xạ xác suất trên mặt phẳng đất hợp nhất có thể được quét để tìm giá trị cực đại địa phương. Mỗi giá trị cực đại địa phương tìm được có thể xác định các mô hình người riêng rẽ đối với mỗi hình ảnh video trong không gian hình ảnh tương ứng của chúng mà có thể sau đó được dùng để tính $f_1(n)$ hoặc $f_2(n)$ (được nêu trên đây) nếu thích hợp. Thực hiện nhiều bước quét bản ảnh xạ xác suất trên mặt phẳng đất hợp nhất để các giá trị cực đại địa phương có thể được thực hiện để tìm các mô hình người tiếp theo (một mô hình dùng cho mỗi hình ảnh video) và để tính:

$$\sum_{n=1}^m f_1(n) + f_2(n)$$

Các ràng buộc khi lựa chọn (như ngưỡng xác suất tối thiểu và khoảng cách tối thiểu trong thế giới thực 3D) có thể được thay đổi và lượt quét mới được tiến hành để tìm tập hợp tối ưu của m mô hình người 3D (tương ứng trong ví dụ này là 2 m mô hình người 2D).

Phần mô tả trên đây minh họa các phương án làm ví dụ và không được hiểu là chỉ giới hạn ở đó. Mặc dù một vài phương án làm ví dụ đã được mô tả, những người có hiểu biết trung bình về lĩnh vực kỹ thuật liên quan có thể thấy rằng có thể thực hiện được nhiều cải biến từ các phương án làm ví dụ mà không nằm ngoài bản chất của sáng chế. Ví dụ, mặc dù phần mô tả đã bộc lộ việc phát hiện các đối tượng con người trong hình ảnh video, nhưng sáng chế không được coi là chỉ giới hạn ở đó và các đối tượng khác cần quan tâm cũng có thể phát hiện được.

YÊU CẦU BẢO HỘ

1. Phương pháp phát hiện các đối tượng con người trong video bao gồm các bước:

phát hiện các điểm ảnh dịch chuyển trong hình ảnh video sử dụng tính toán sai khác khung;

xây dựng và cập nhật mô hình nền động sử dụng hình ảnh video;

xác định các điểm ảnh thay đổi của hình ảnh video là các điểm ảnh khác với nền;

xác định các đốm màu tiền cảnh bằng cách nhóm các điểm ảnh dịch chuyển được phát hiện và các điểm ảnh thay đổi định trước về mặt không gian;

theo dõi các đốm màu tiền cảnh theo thời gian để tạo thành các mục tiêu không gian-thời gian;

đối với mỗi vị trí trong số N vị trí trong hình ảnh video, trong đó N là số điểm ảnh của hình ảnh video, so sánh hình dạng người định trước tương ứng với các đốm màu tiền cảnh để thu được xác suất là người tương ứng ở vị trí này, nhờ đó thu được N xác suất tương ứng với N vị trí, trong đó hình dạng người được định trước tương ứng là hình dạng định trước có kích thước đối với mỗi vị trí trong số N vị trí định trước đáp ứng việc hiệu chuẩn hệ thống video được sử dụng để thu được hình ảnh video, việc hiệu chuẩn được thực hiện bằng cách sử dụng một số mẫu hình ảnh người;

sử dụng N xác suất này, xác định X người được thể hiện bởi tập hợp đốm màu tiền cảnh, trong đó X là số nguyên;

sử dụng việc xác định X người được thể hiện, xác định mật độ đám đông trong vùng thứ nhất của thế giới thực;

so sánh mật độ đám đông với ngưỡng thứ nhất; và

cung cấp việc phát hiện sự kiện bằng cách sử dụng các mục tiêu không gian-thời gian và xác định X người được thể hiện và báo cáo về sự kiện được phát hiện khi mật

độ đám đông vượt quá ngưỡng thứ nhất và thời lượng đám đông tập trung vượt quá ngưỡng thứ hai.

2. Phương pháp theo điểm 1, còn bao gồm các bước:

tạo ra bản ánh xạ xác suất có N xác suất và xác định các giá trị cực đại địa phương trong số các xác suất của bản ánh xạ xác suất; và

sử dụng bản ánh xạ xác suất để xác định vị trí của mỗi người trong số X người theo cách tùy chọn.

3. Phương pháp theo điểm 1, còn bao gồm ít nhất một trong số các bước:

tính toán tỷ lệ độ phủ đối với người cho mỗi vị trí trong số N vị trí; và

sử dụng N xác suất để xác định vị trí của mỗi người trong số X người.

4. Phương pháp theo điểm 3, trong đó đối với mỗi vị trí trong số N vị trí, tính toán tỷ lệ độ phủ đối với người bao gồm việc chia số lượng điểm ảnh nổi của người trong mô hình người lỗi dạng hình ảnh trên tổng diện tích của mô hình lỗi của người này.

5. Phương pháp theo điểm 1, trong đó hình dạng người định trước đối với ít nhất một vài vị trí trong số N vị trí có kích cỡ khác nhau.

6. Phương pháp theo điểm 1,

trong đó việc hiệu chuẩn hệ thống video bao gồm việc xác định kích cỡ hình ảnh của một phần hình ảnh video tương ứng với kích cỡ trung bình của người tại mỗi vị trí trong số N vị trí, và

trong đó kích cỡ của hình dạng người định trước đối với mỗi vị trí trong số N vị trí định trước đáp ứng kích cỡ hình ảnh tương ứng.

7. Phương pháp theo điểm 1,

trong đó, đối với mỗi vị trí trong số N vị trí, hình dạng người được định trước tương ứng bao gồm giá trị ước lượng của phần hình ảnh nổi sẽ bị chiếm trong hình

ảnh video khi có người tại vị trí tương ứng dựa trên việc chiếu mô hình người trong thế giới thực lên mặt phẳng hình ảnh của hình ảnh video.

8. Phương pháp theo điểm 1,

trong đó hình ảnh video bao gồm nhiều khung hình ảnh, mỗi khung hình ảnh bao gồm hình ảnh hai chiều có N vị trí, mỗi vị trí trong số N vị trí định trước bởi cặp tọa độ x, y tương ứng trong hình ảnh hai chiều.

9. Phương pháp theo điểm 1, trong đó việc so sánh hình dạng người định trước bao gồm, đối với mỗi hình dạng trong số N hình dạng định trước tại các vị trí tương ứng của N vị trí định trước, tính toán lượng diện tích chồng chập của hình dạng định trước tương ứng với các đốm màu tiền cảnh.

10. Hệ thống giám sát bằng video (101), bao gồm:

nguồn video (102) được cấu hình để cung cấp hình ảnh video của hiện trường của thế giới thực;

môđun phát hiện sự thay đổi (103) được cấu hình để xây dựng và cập nhật mô hình nền động sử dụng hình ảnh video, và để xác định các điểm ảnh thay đổi của hình ảnh video là các điểm ảnh khác với nền;

môđun phát hiện chuyển động (104) được cấu hình để phát hiện các điểm ảnh dịch chuyển trong hình ảnh video sử dụng tính toán sai khác khung;

môđun phát hiện đốm màu tiền cảnh (105) được cấu hình để xác định các đốm màu tiền cảnh bằng cách nhóm các điểm ảnh dịch chuyển được phát hiện và các điểm ảnh thay đổi định trước về mặt không gian;

môđun phát hiện người (106), trong đó đối với mỗi N vị trí trong hình ảnh video, trong N là số điểm ảnh của hình ảnh video, môđun phát hiện người (106) được cấu hình để so sánh hình dạng người được định trước tương ứng với các đốm màu tiền cảnh để thu xác suất tương ứng của người tại vị trí này, do đó thu được N xác suất tương ứng với N vị trí, trong đó hình dạng người được định trước tương ứng là hình

dạng định trước, trong đó môđun phát hiện người (106) được tạo cấu hình để xác định kích thước của hình dạng định trước đối với mỗi vị trí trong số N vị trí đáp ứng lại việc hiệu chuẩn nguồn video, việc hiệu chuẩn được thực hiện bằng cách sử dụng một số mẫu hình ảnh người, và trong đó môđun phát hiện người (106) được cấu hình để sử dụng N xác suất để xác định X người được thể hiện bằng các đốm màu tiền cảnh, trong đó X là một số nguyên;

môđun theo vết mục tiêu (107) được cấu hình để theo dõi các đốm màu tiền cảnh theo thời gian để tạo thành các mục tiêu không gian-thời gian; và

môđun phát hiện sự kiện (108) được tạo cấu hình để xác định mật độ đám đông sử dụng các mục tiêu không gian-thời gian và xác định X người được thể hiện, để so sánh mật độ đám đông với ngưỡng thứ nhất, và để báo cáo sự kiện khi mật độ đám đông vượt quá ngưỡng thứ nhất và thời lượng đám đông tập trung vượt quá ngưỡng thứ hai.

Fig.1

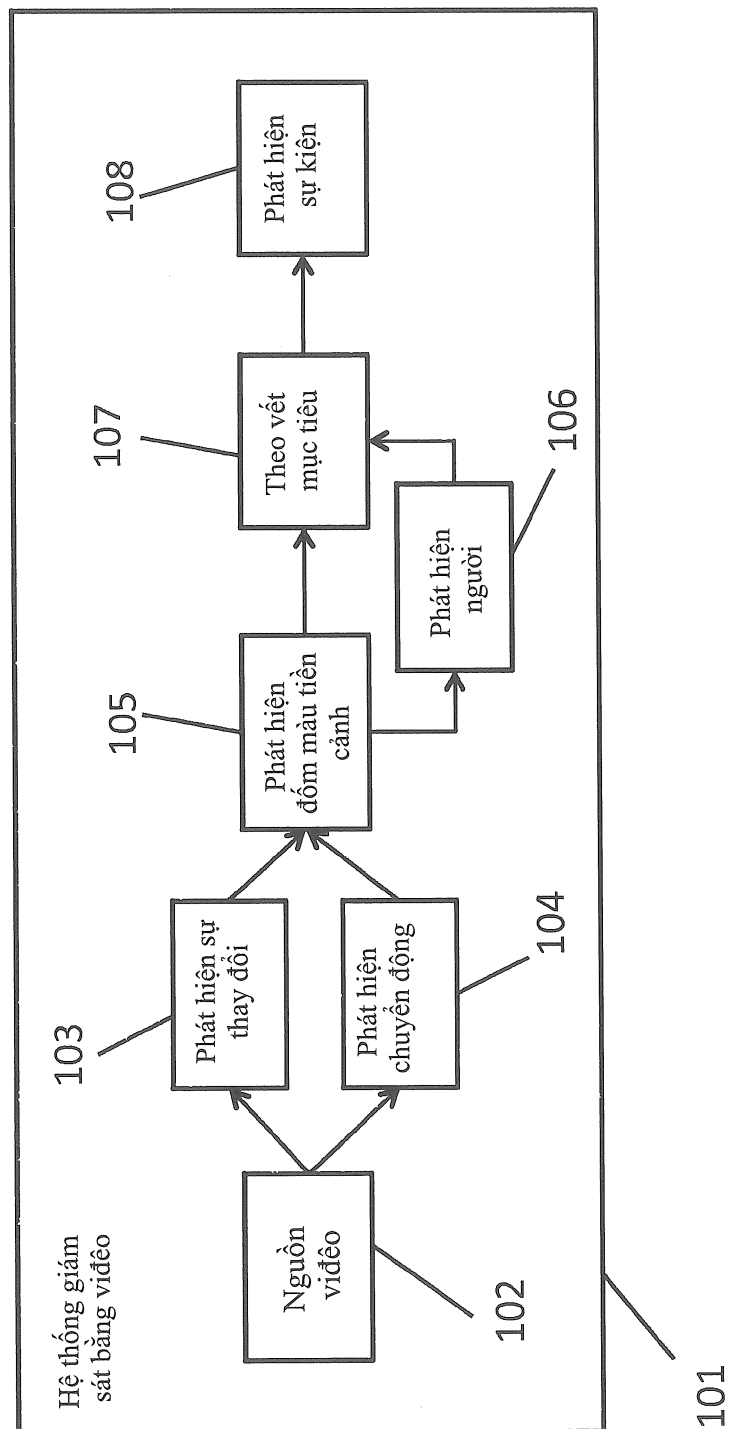


Fig.2

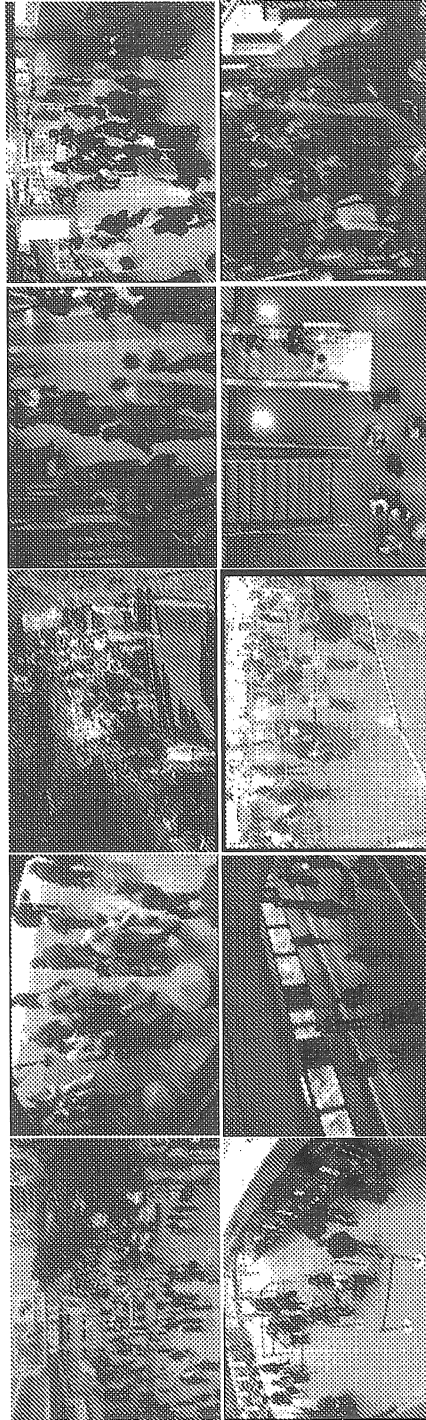


Fig.3A

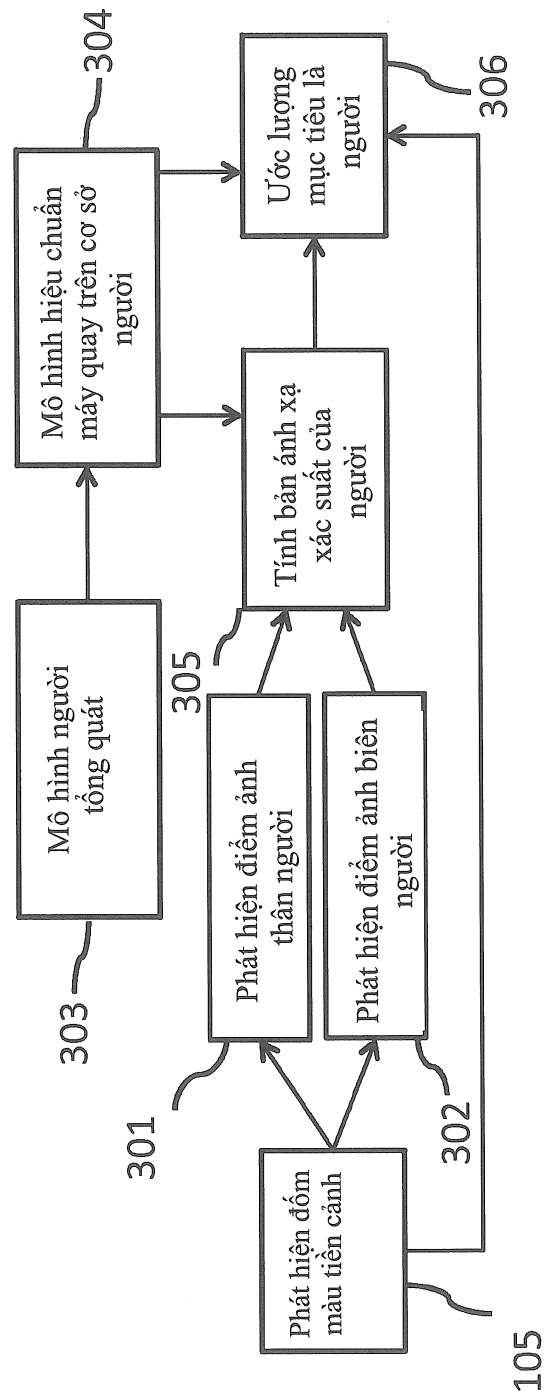


Fig.3B

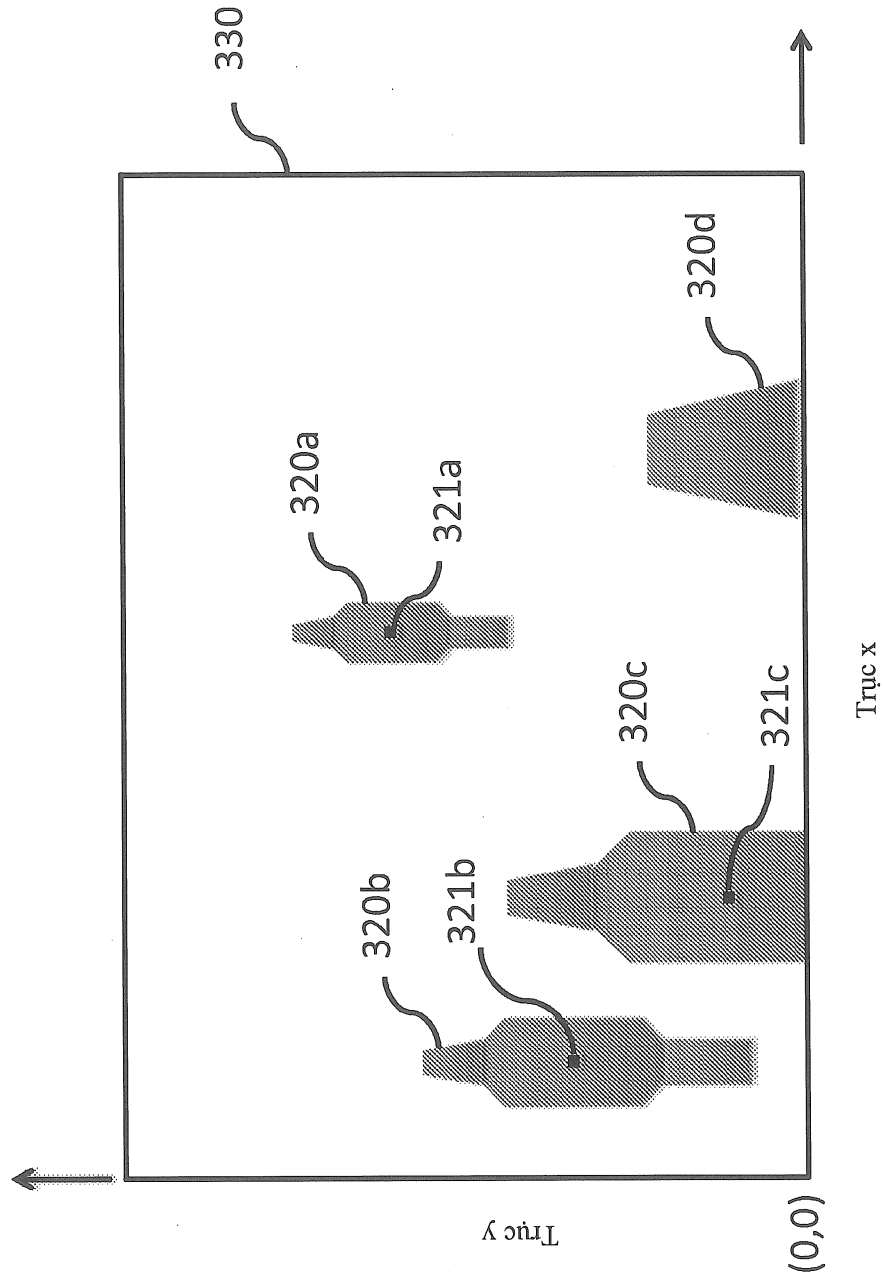


Fig.3C

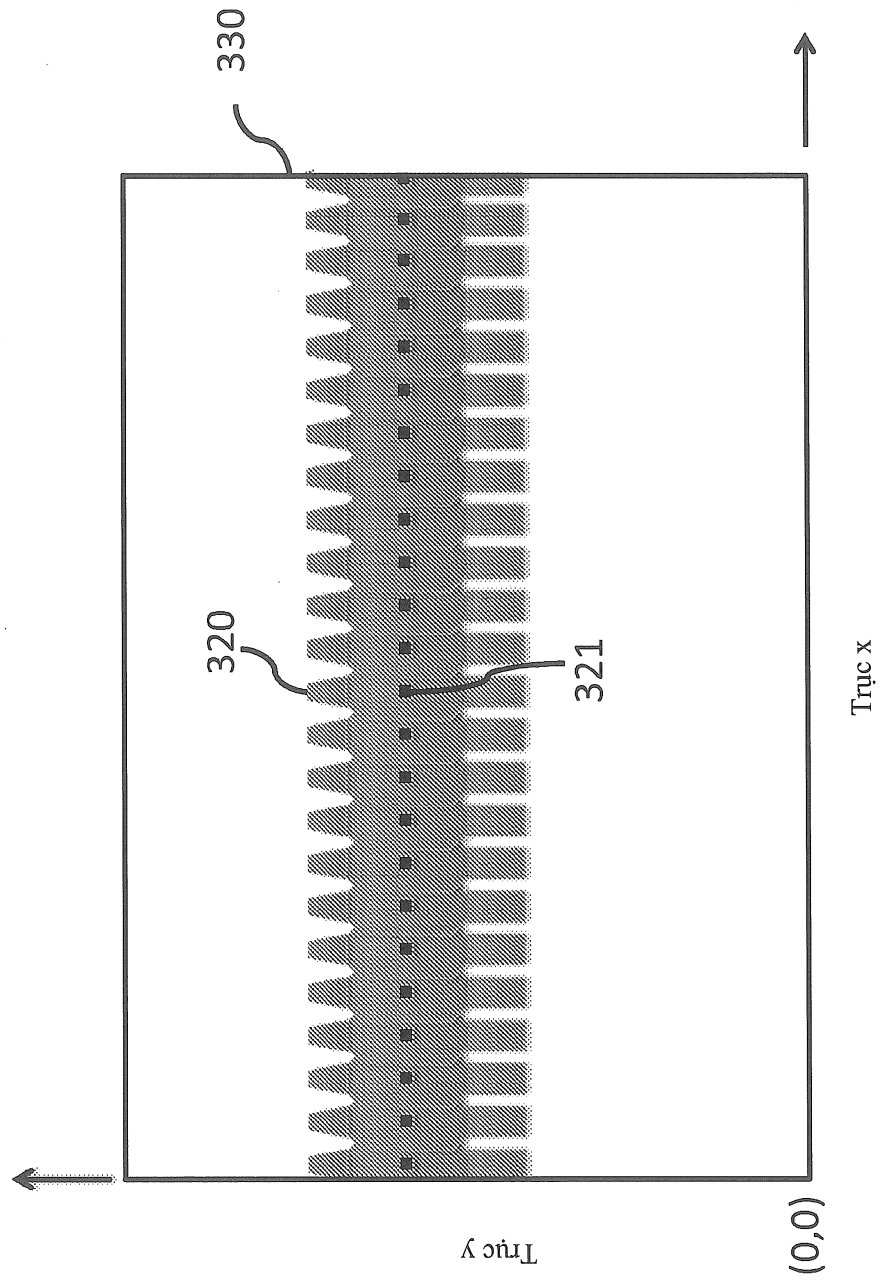


Fig.3D

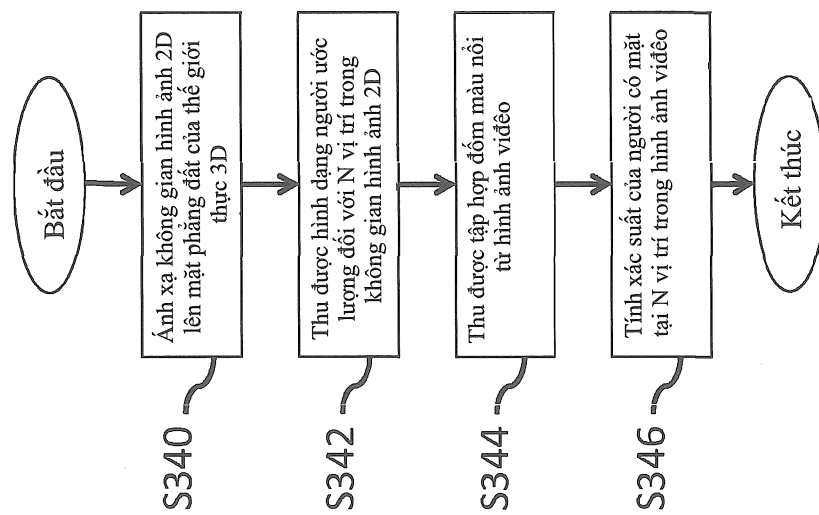


Fig.3E

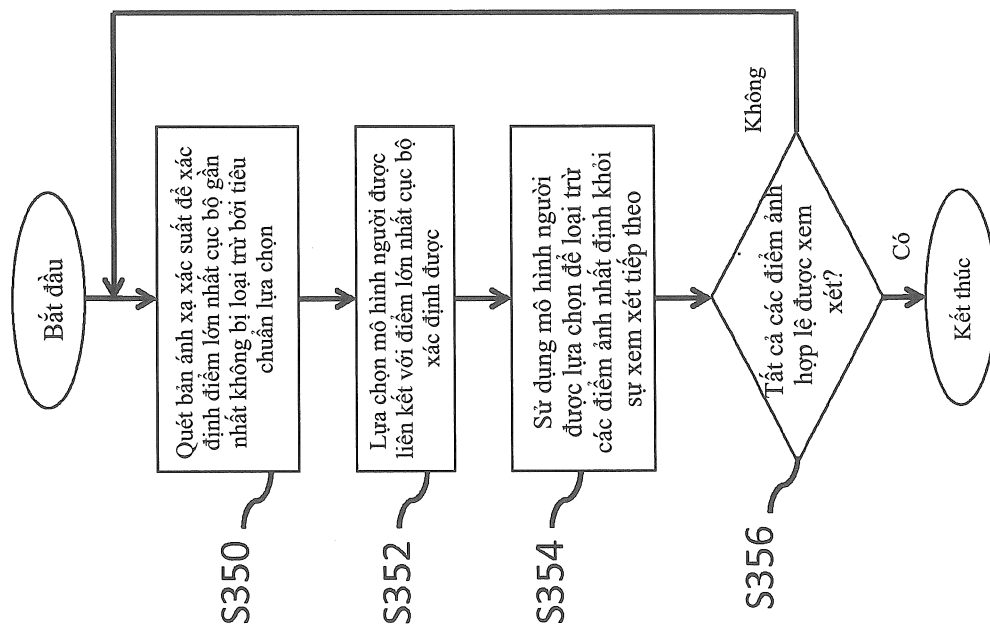


Fig.3F

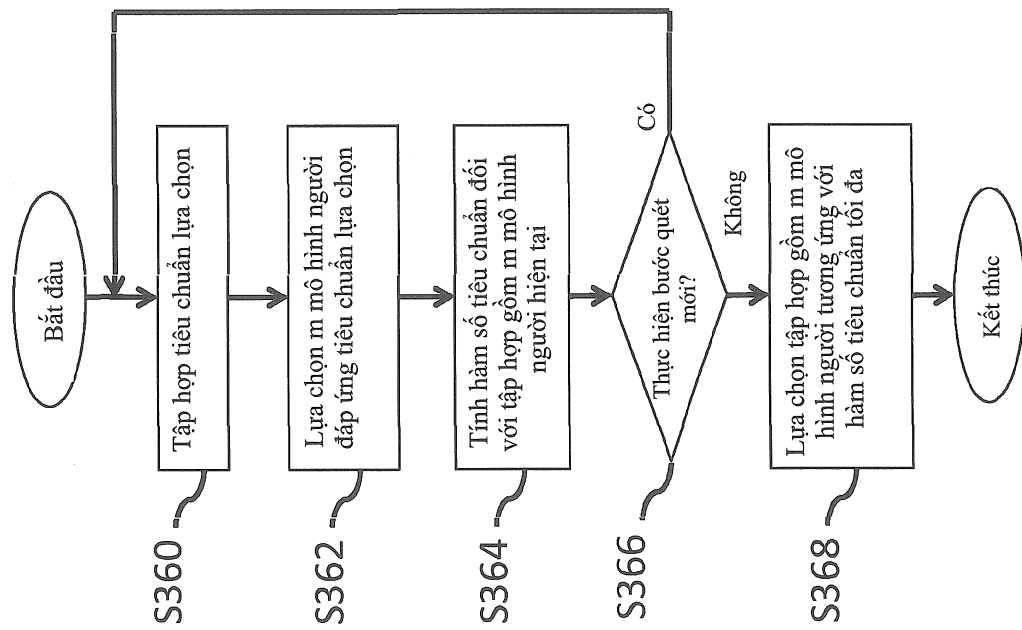


Fig.4

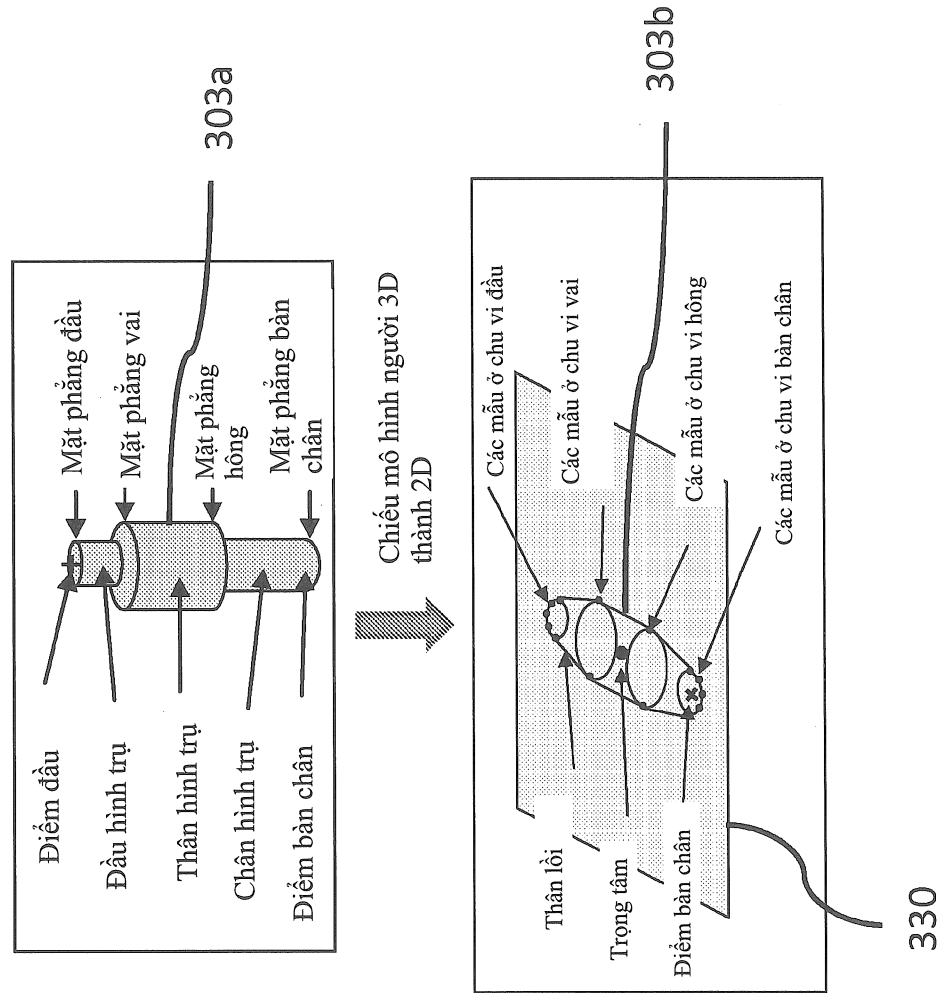


Fig.5

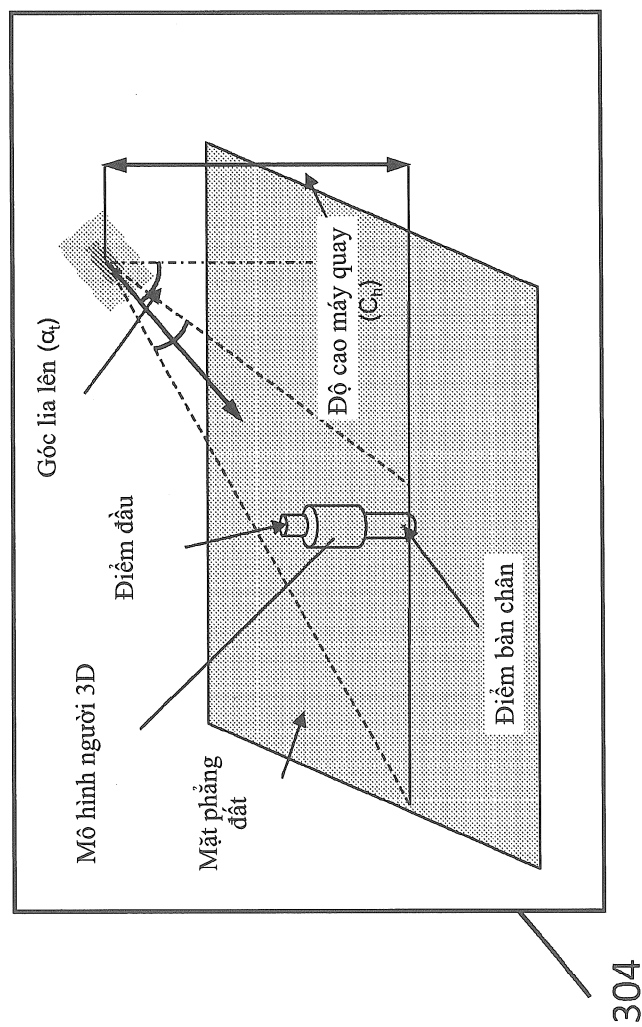


Fig.6A

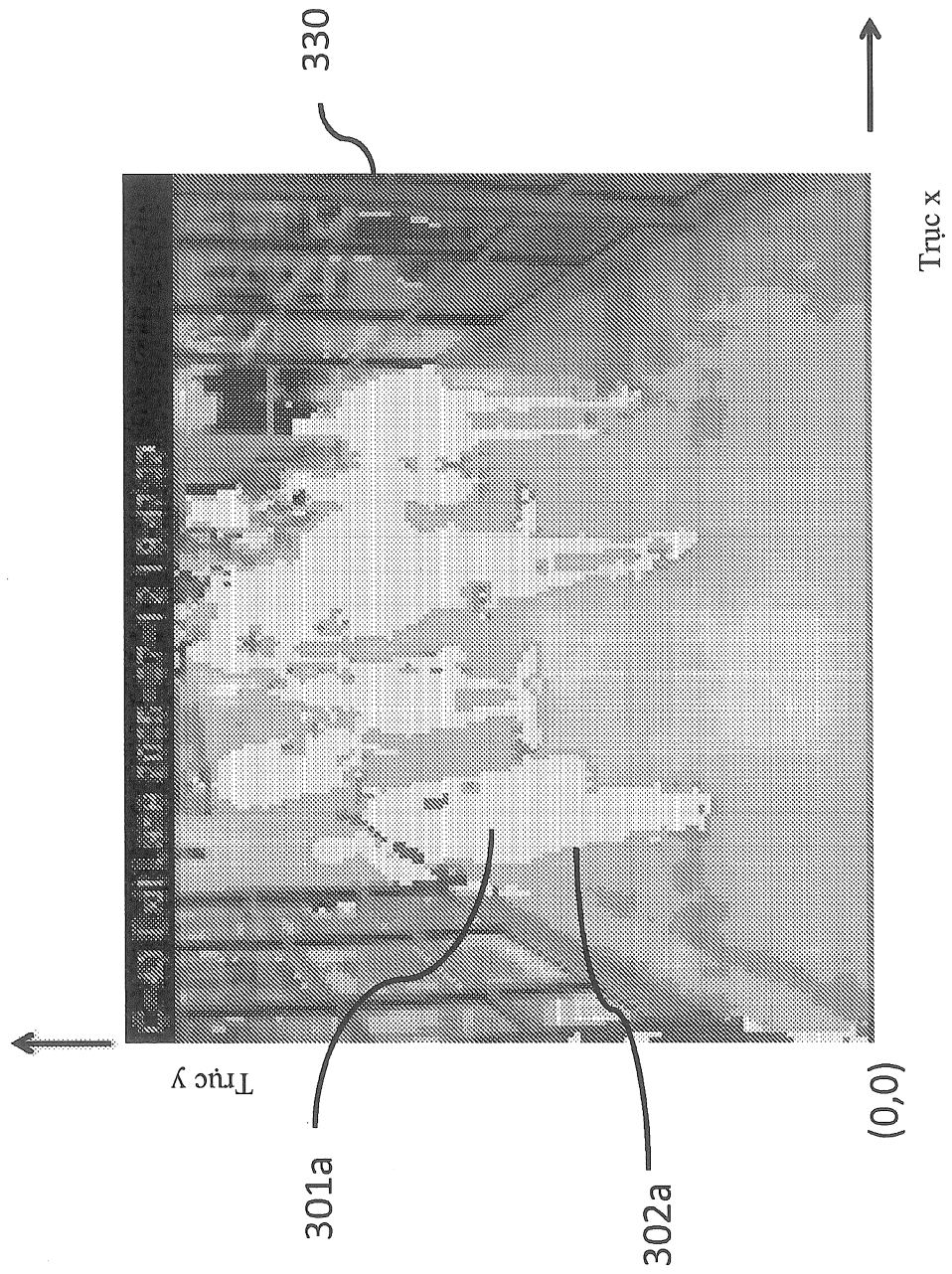


Fig.6B

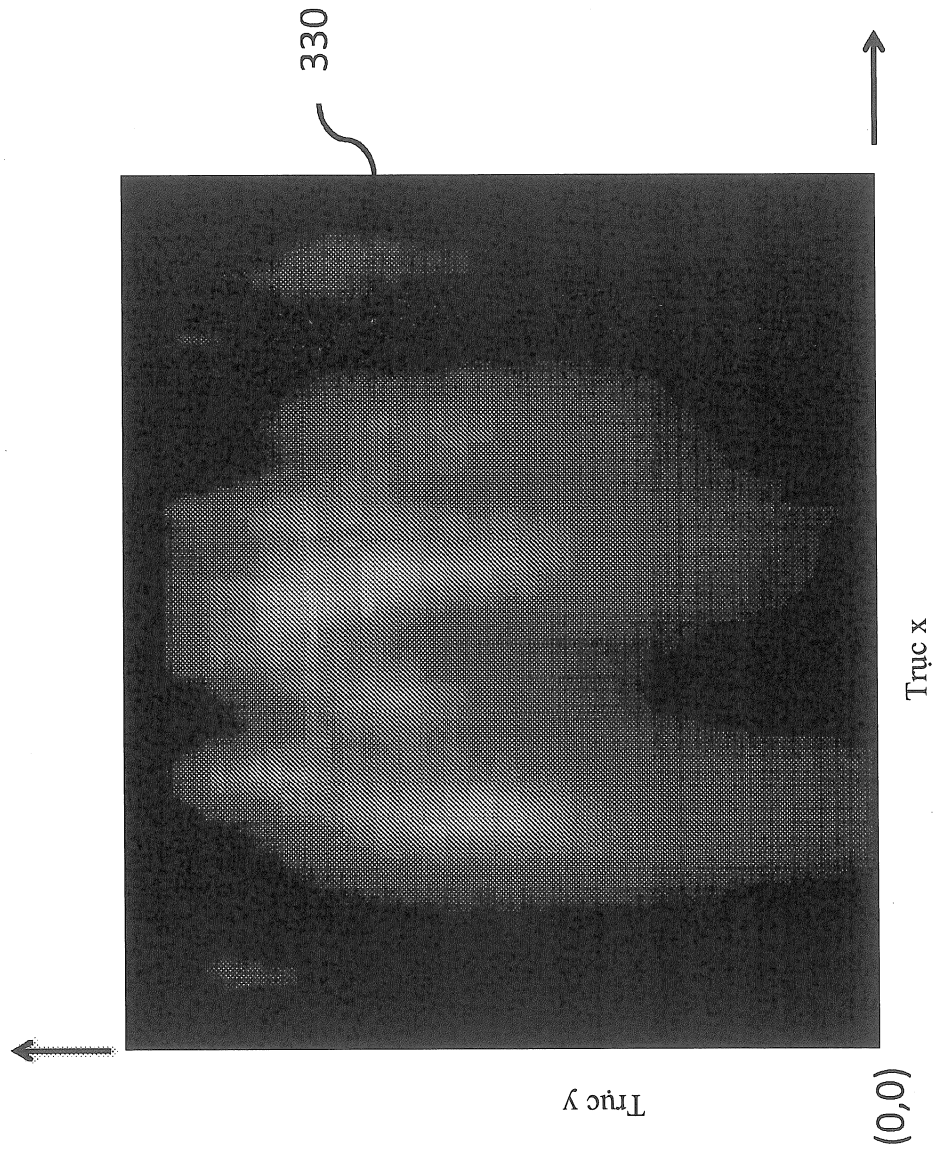


Fig.6C

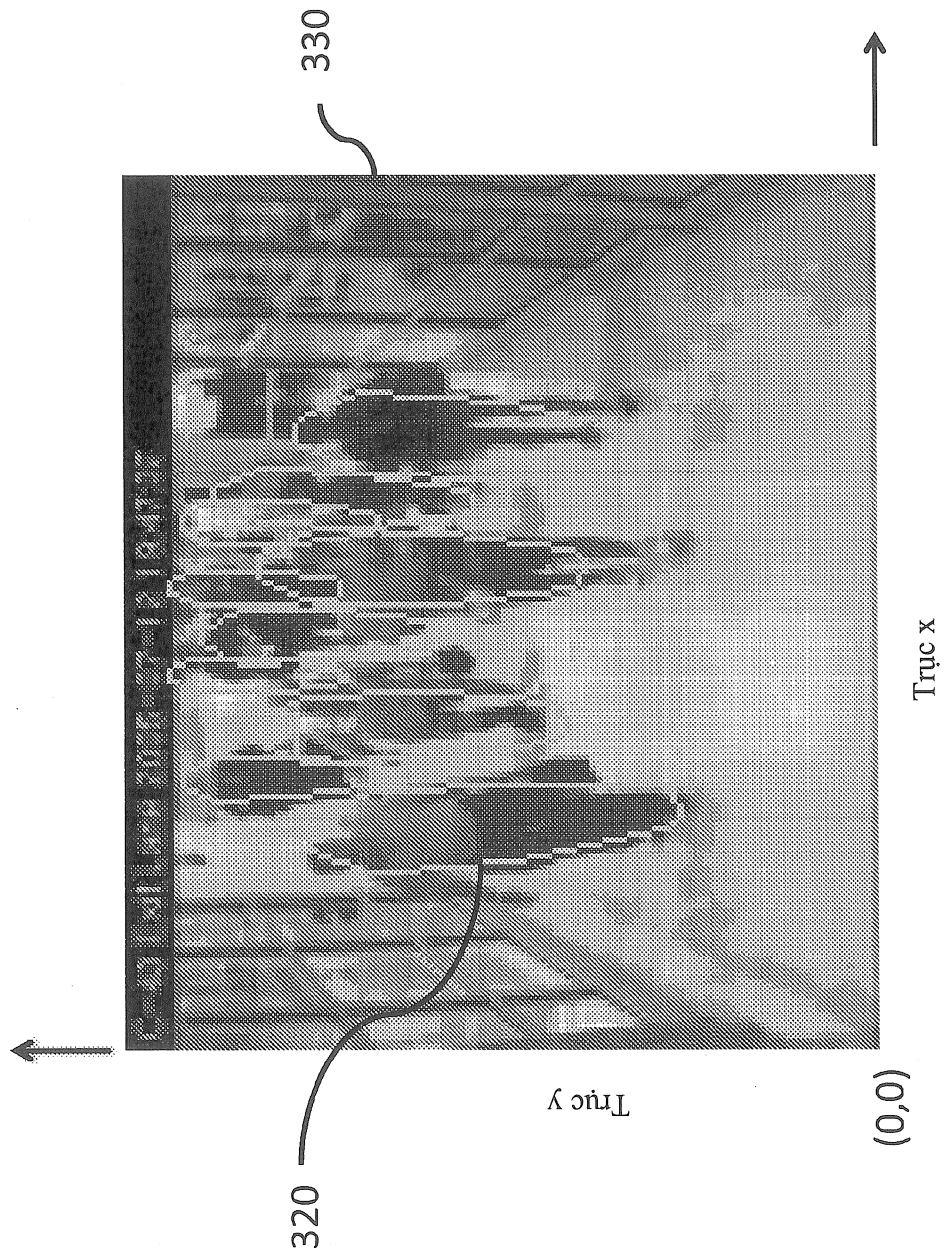
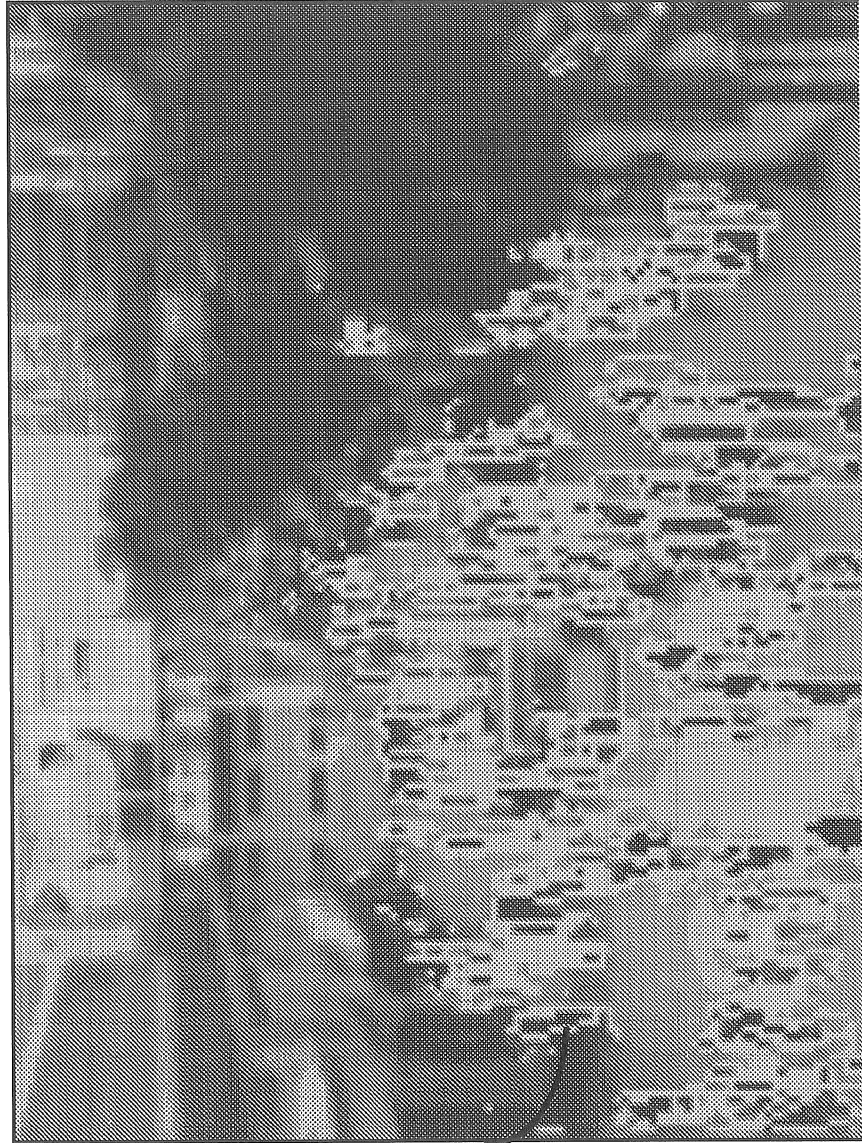


Fig.7A



320

Fig.7B

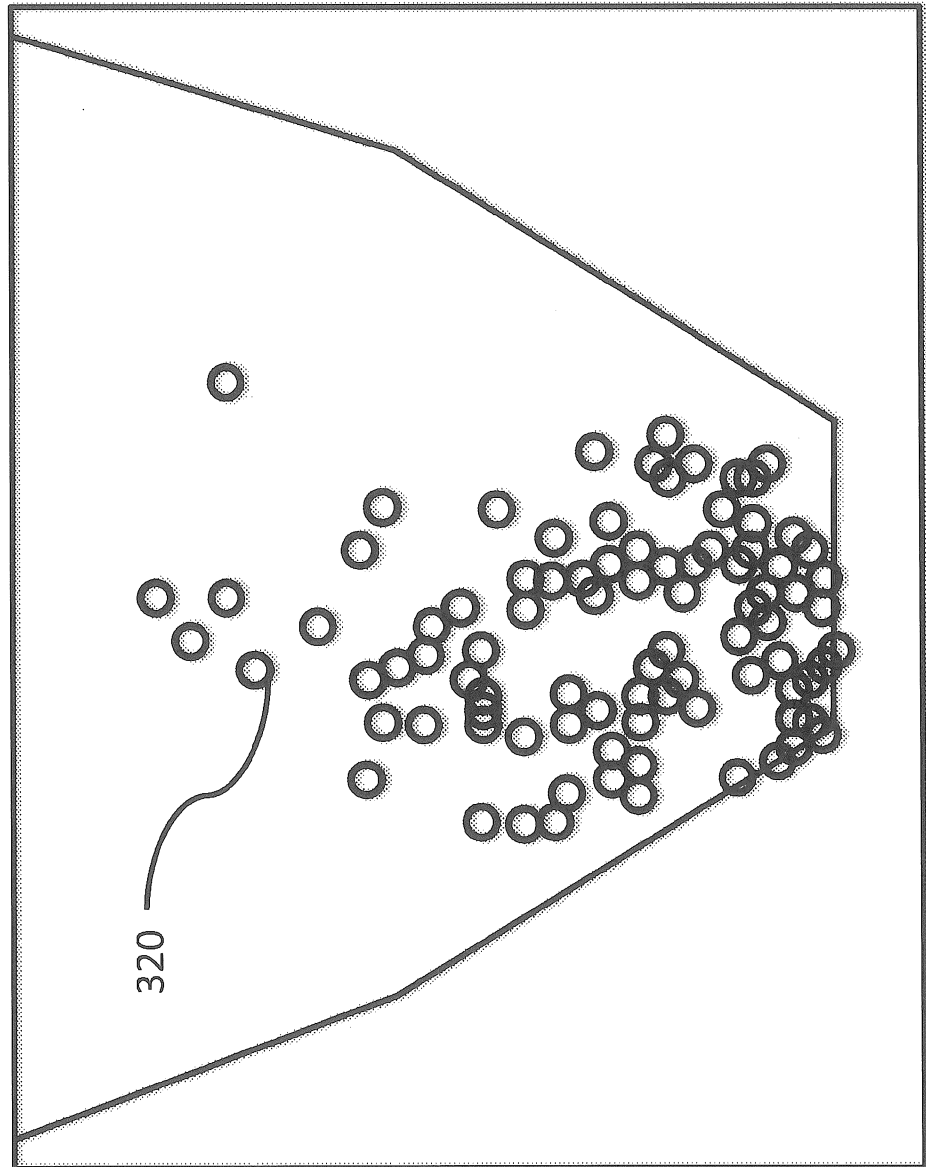


Fig.7C

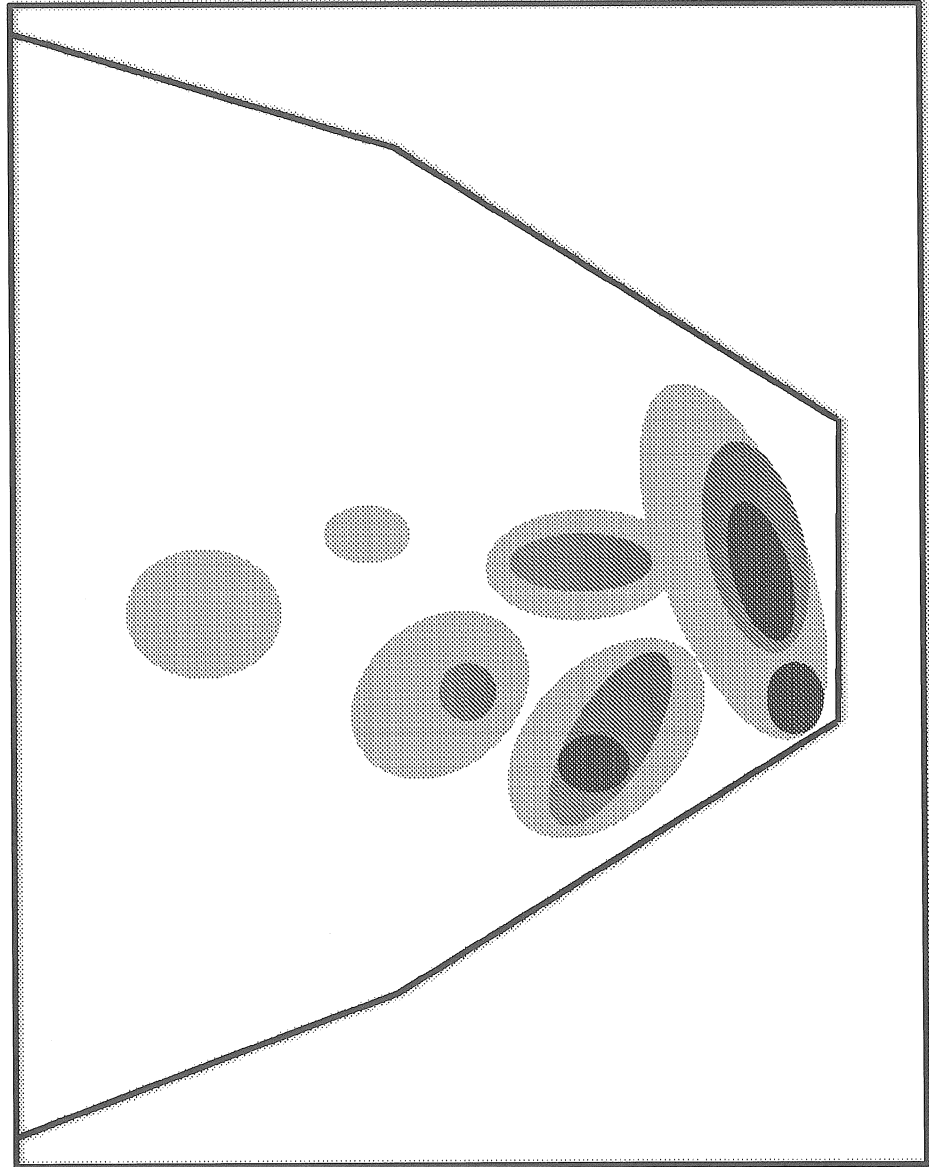


Fig.8

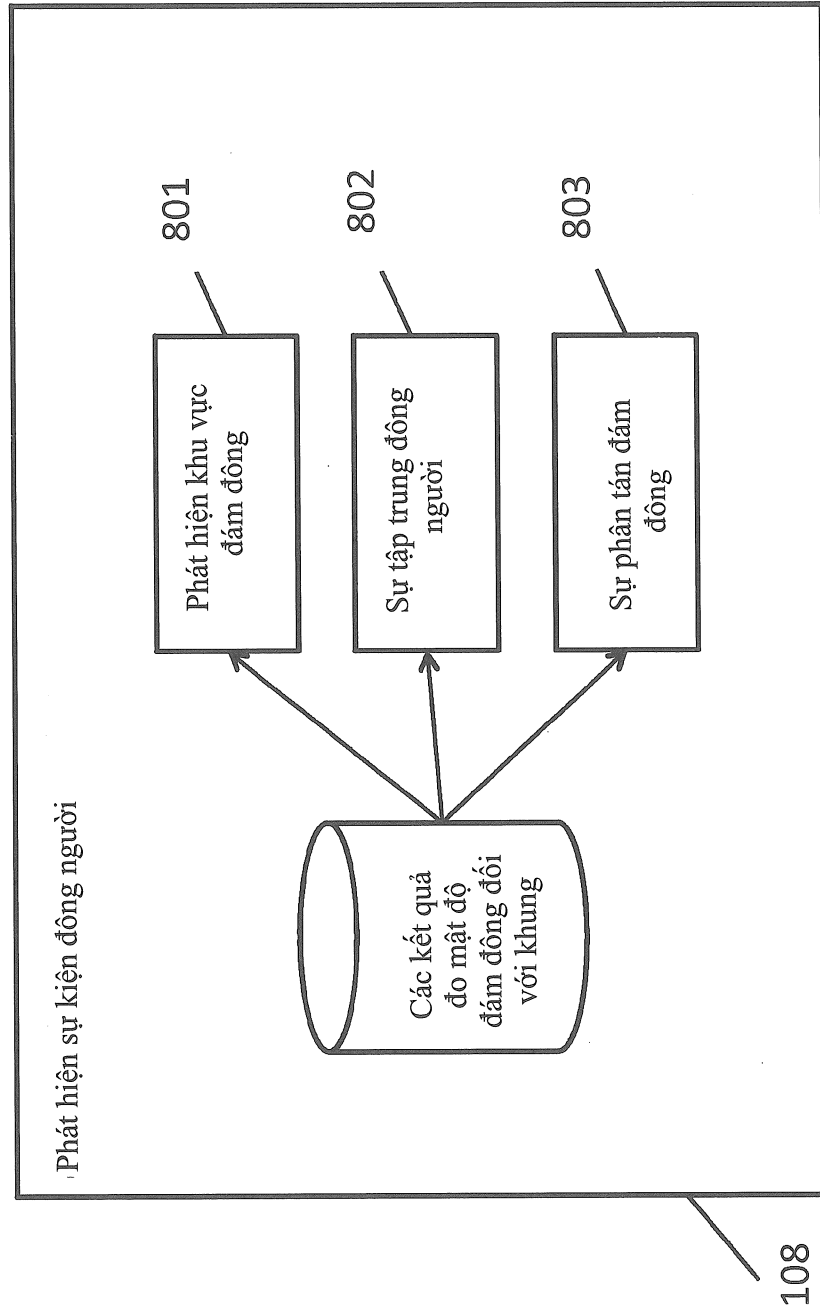


Fig.9

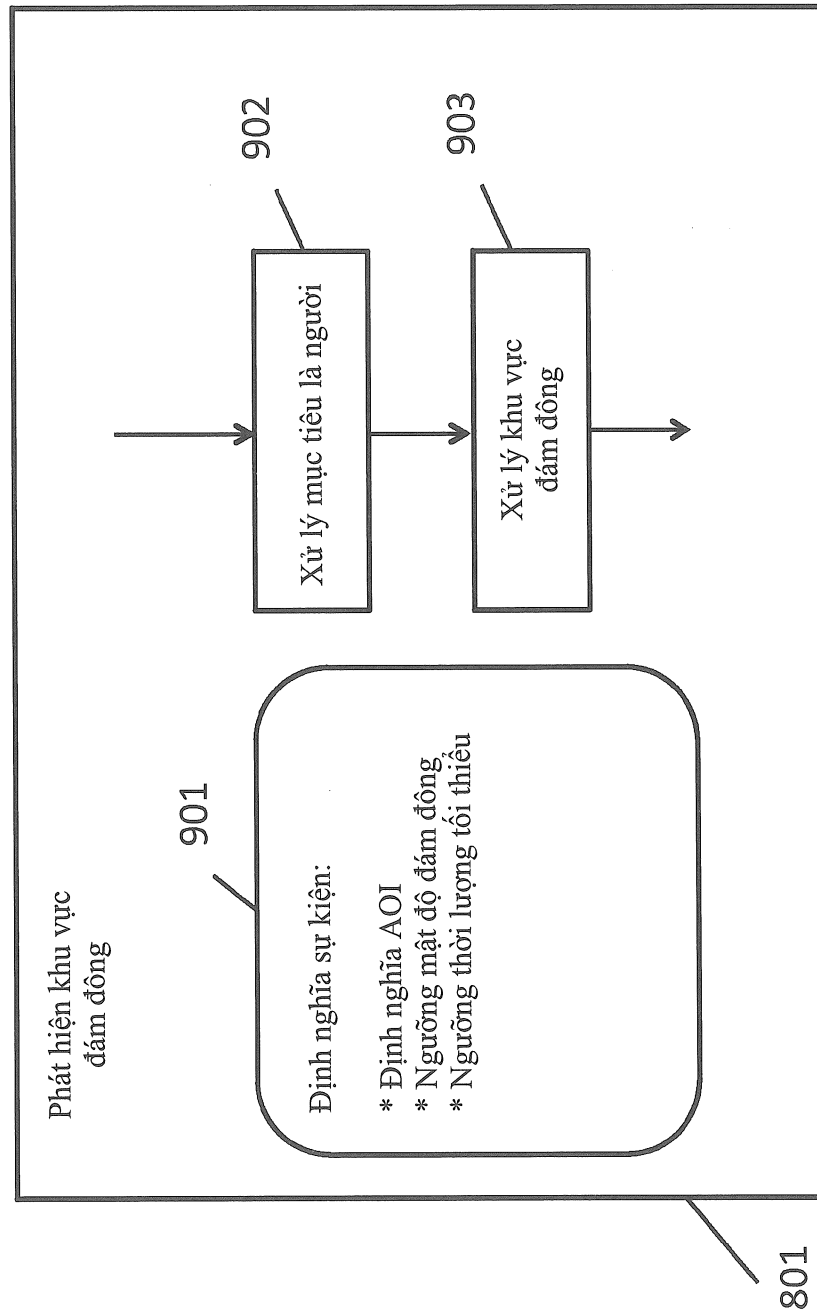


Fig.10

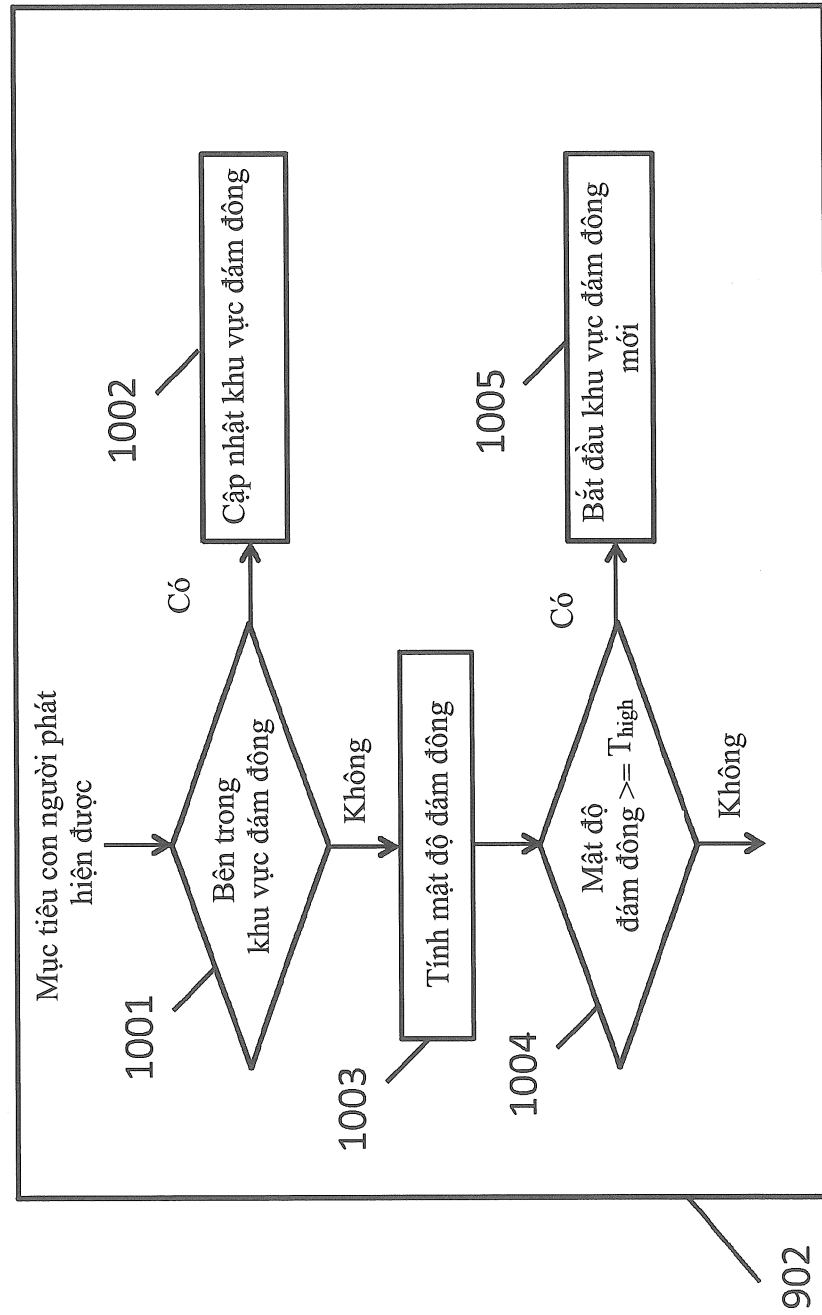


Fig.11

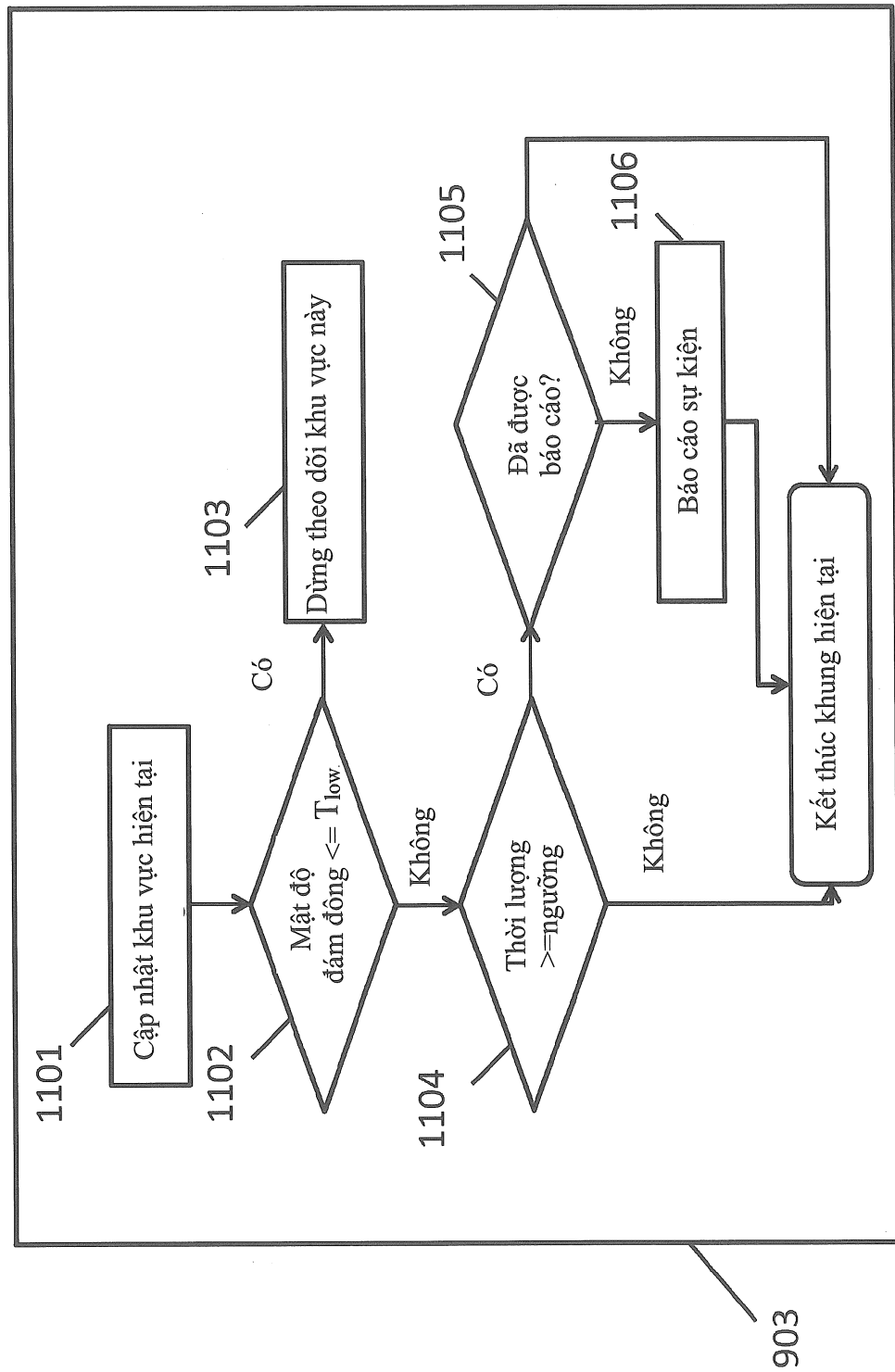


Fig.12

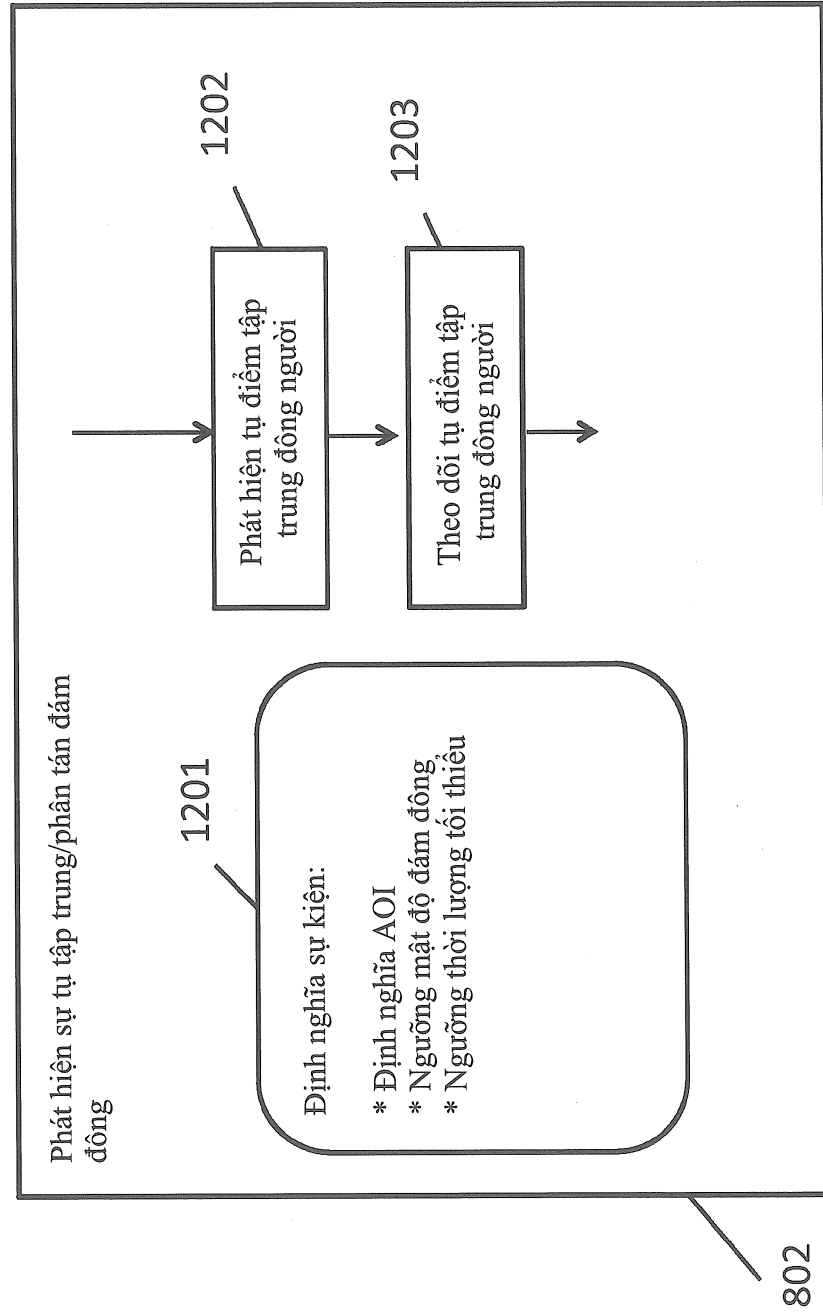
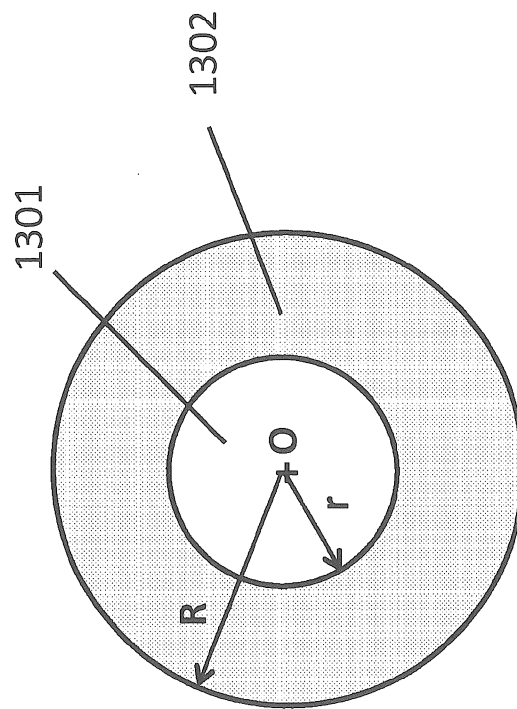


Fig.13



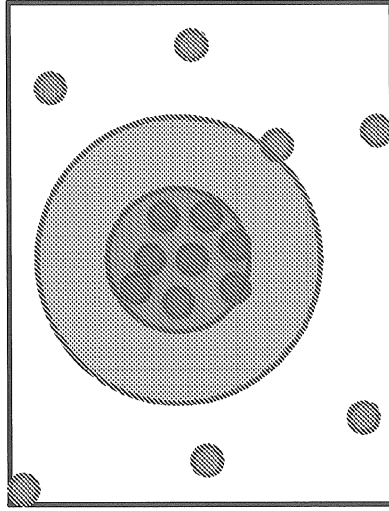


Fig. 14B

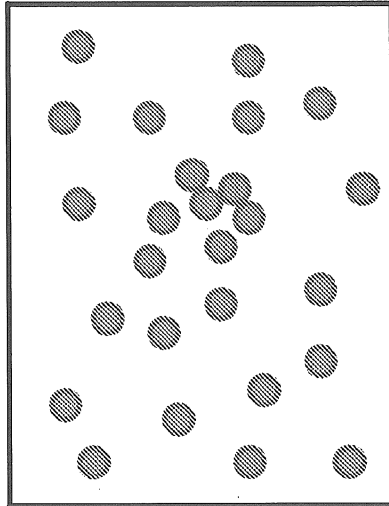


Fig. 14A

Fig.15

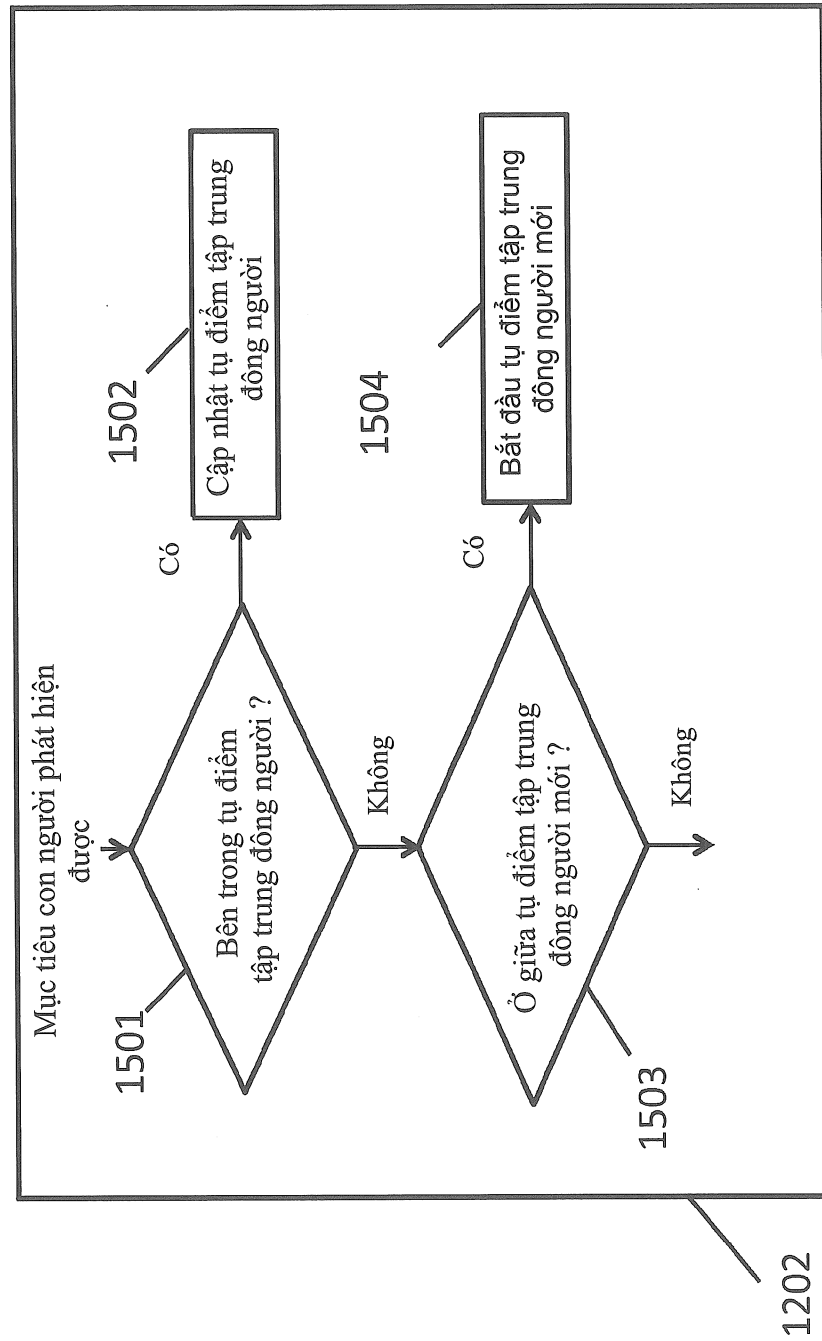


Fig.16

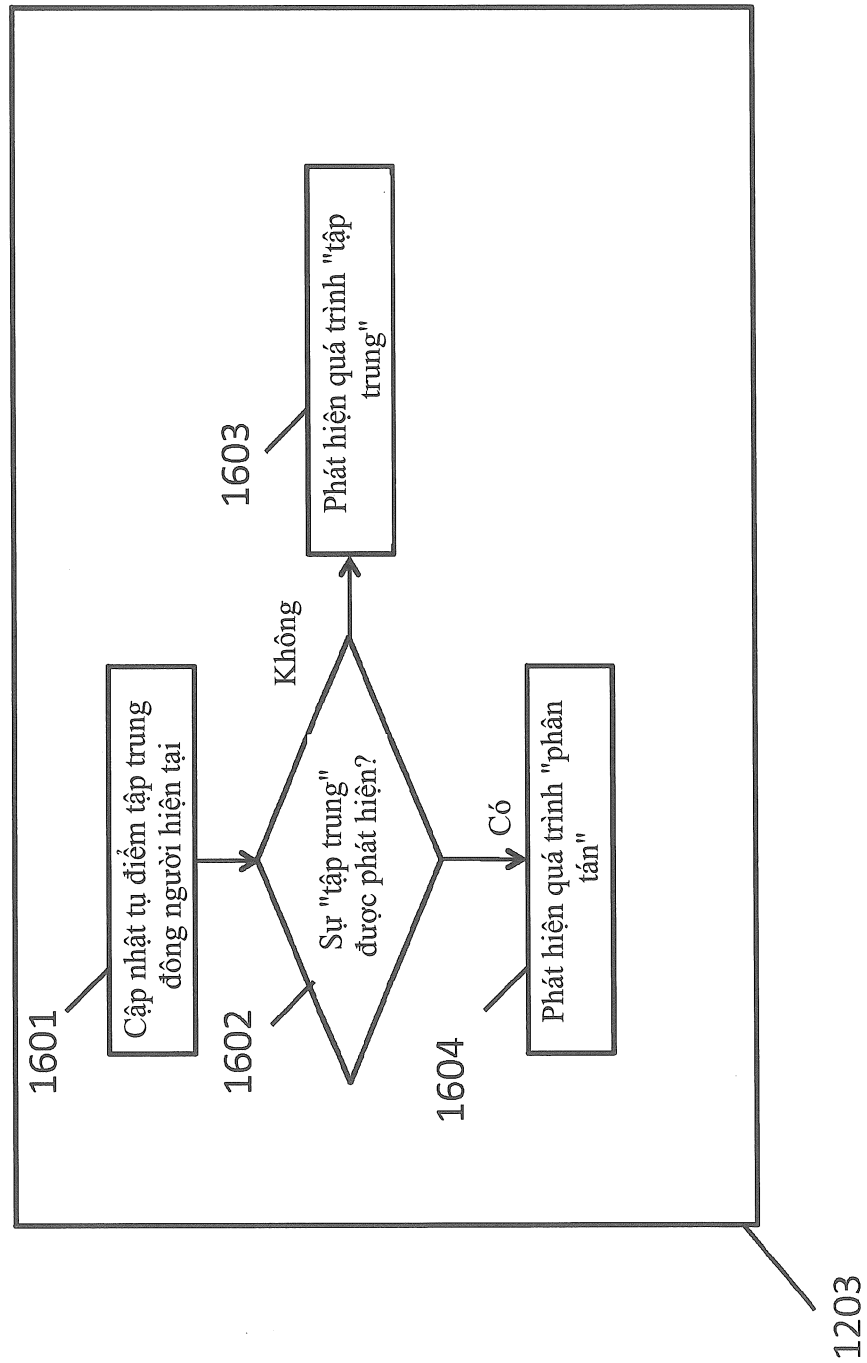


Fig. 17

