



(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)



1-0036562

(51)<sup>2020.01</sup> G06T 7/00

(13) B

(21) 1-2020-04454

(22) 31/07/2020

(45) 25/08/2023 425

(43) 26/10/2020 391A1

(73) TẬP ĐOÀN VINGROUP – CÔNG TY CP (VN)

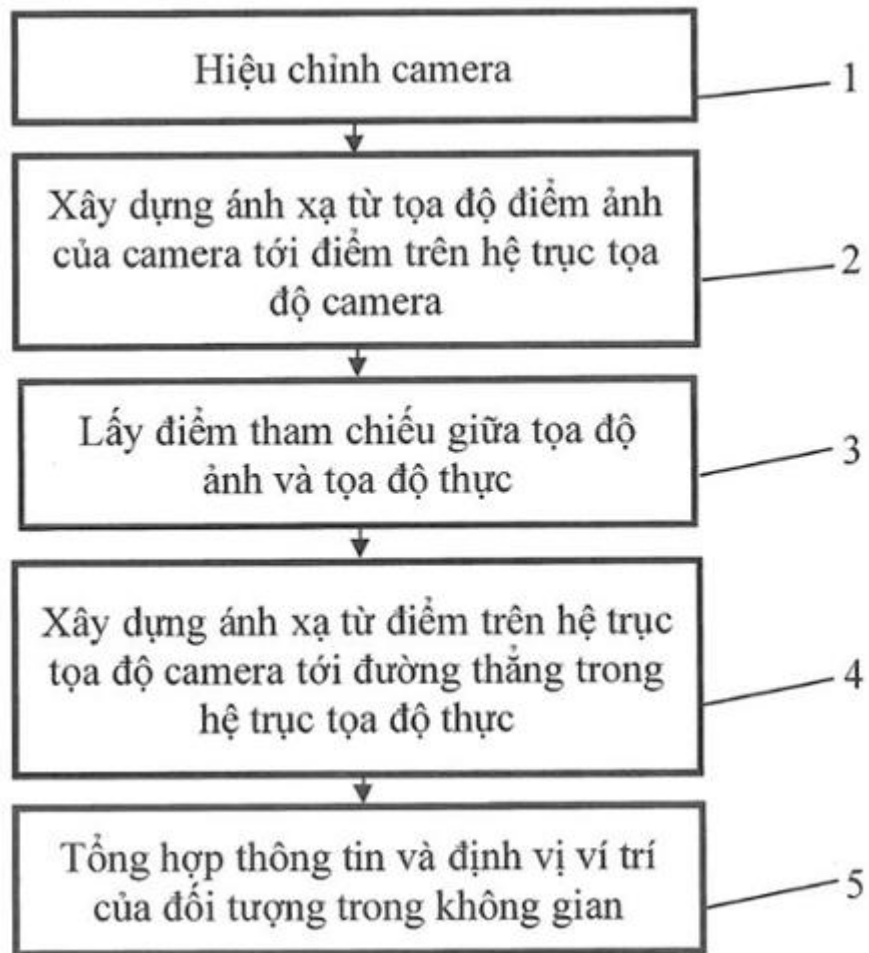
Số 7, đường Bằng Lăng 1, Khu đô thị sinh thái Vinhomes Riverside, phường Việt Hưng, quận Long Biên, thành phố Hà Nội

(72) Lê Bùi Phúc (VN); Đinh Đỗ Thủy (VN); Phan Tất Hải Triều (VN); Tạ Nguyên Long (VN); Trần Trung Tuyển (VN); Cao Sỹ Dũng (VN).

(74) Công ty Luật TNHH quốc tế BMVN (BMVN INTERNATIONAL LLC)

(54) PHƯƠNG PHÁP ĐỊNH VỊ ĐỐI TƯỢNG BẰNG HỆ THỐNG CAMERA

(57) Sáng chế đề xuất phương pháp định vị đối tượng bằng hệ thống camera (camera network - CAMNET), bao gồm: i) hiệu chỉnh camera; ii) xây dựng ánh xạ từ tọa độ điểm ảnh của camera tới điểm trên hệ trục tọa độ camera; iii) lấy điểm tham chiếu giữa tọa độ ảnh và tọa độ thực; iv) xây dựng ánh xạ từ điểm trên hệ trục tọa độ camera tới đường thẳng trong hệ trục tọa độ thực; và v) tổng hợp thông tin và định vị vị trí của đối tượng trong không gian.



### Lĩnh vực kỹ thuật được đề cập

Sáng chế liên quan đến việc nhận dạng và định vị các đối tượng để kiểm soát an ninh bằng thiết bị camera, cụ thể là, sáng chế liên quan đến phương pháp định vị đối tượng bằng hệ thống camera (camera network - CAMNET).

### Tình trạng kỹ thuật của sáng chế

Việc phối hợp nhiều camera để kết nối thông tin có vai trò đặc biệt quan trọng trong việc giám sát các đối tượng, giúp hạn chế lặp lại việc nhận dạng đối tượng, theo dõi được hành trình của đối tượng theo thời gian thực.

Các mô-đun camera 3D hiện có trên thị trường đều có giá thành cao. Việc sử dụng cặp 2 camera tương hỗ đặt gần nhau hoặc sử dụng cảm biến chiều sâu thường có sai số lớn ở không gian rộng và phù hợp hơn với những trường hợp cần độ chính xác cao trong phạm vi không gian nhỏ.

Các phương pháp hiệu chỉnh camera, tính toán các thông số vị trí góc nhìn của camera thường (sử dụng mô hình camera lỗ kim (pinhole camera)) đòi hỏi việc hiệu chỉnh tính toán tham số nội camera phải rất chính xác, và thường gặp khó khăn khi phải xử lý đối với các camera có tiêu cự dài.

### Bản chất kỹ thuật của sáng chế

Để khắc phục các nhược điểm nêu trên, các tác giả sáng chế đã nghiên cứu các thuật toán để nội suy góc nhìn của từng camera, từ đó phối hợp tính toán vị trí của đối tượng cần quan sát trong không gian.

Theo đó, sáng chế đề xuất phương pháp định vị đối tượng bằng hệ thống camera (camera network - CAMNET), còn được gọi là "Hệ thống Camnet", bao gồm các bước sau:

i) hiệu chỉnh từng camera trong hệ thống camera để thu thập các tham số nội của từng camera này, bao gồm: tiêu cự (focal length), điểm chính (principal point), trục xiên (axis skew) và hệ số méo (distortion coefficients);

ii) xây dựng ánh xạ từ điểm trên hệ trục tọa độ camera tới đường thẳng trong hệ trục tọa độ thực dựa vào mô hình camera lỗ kim (pinhole camera model):

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

trong đó:

- $(u, v)$  là tọa độ điểm ảnh (pixel) trên ảnh camera,

- $f_x, f_y$  là tiêu cự (focal lengths) theo trục  $x, y$  tương ứng tính theo đơn vị điểm ảnh,
- $(c_x, c_y)$  là tọa độ điểm chính thường ở trung tâm của ảnh, và
- $(x, y, z)$  là tọa độ điểm tương ứng trên hệ trục tọa độ gắn với camera,

iii) lấy điểm tham chiếu giữa tọa độ ảnh và tọa độ thực để sử dụng cho thuật toán nội suy bằng cách đo các điểm trong thực tế tương ứng với các điểm trên mặt phẳng ảnh lấy được từ camera, từ đó tìm ra các tham số của phép biến đổi Homography giữa mặt phẳng ảnh và mặt phẳng trong không gian thực;

iv) xây dựng ánh xạ từ điểm trên hệ trục tọa độ camera tới đường thẳng trong hệ trục tọa độ thực; và

v) tổng hợp thông tin và xác định vị trí của đối tượng trong không gian bằng cách sử dụng thông tin của đối tượng trong không gian thực (ánh xạ từ hình ảnh đối tượng trên camera) được gửi lên từ các camera thành phần, kết hợp thông tin của ít nhất 2 camera và kích thước ước lượng của đối tượng để có thể xác định được vị trí chính xác của đối tượng trong không gian thực.

Hệ thống CAMNET theo sáng chế chỉ sử dụng các camera đơn có tính năng tối thiểu của camera, đó là cung cấp luồng hình ảnh, các bước tiền xử lý đơn giản có tính khả thi cao khi triển khai công nghiệp.

Các camera trong hệ thống CAMNET theo sáng chế được bố trí ở các vị trí xa nhau với góc nhìn khác nhau giúp cho việc tính toán vị trí đối tượng ở không gian rộng có sai số đủ nhỏ.

Số lượng camera trong hệ thống CAMNET không bị giới hạn cụ thể, thường bao gồm từ 2 đến 10 camera.

Bước hiệu chỉnh camera để tính toán tham số nội camera, chỉ sử dụng để loại bỏ sai số do hiện tượng méo của ảnh camera trong tính toán từ tọa độ ảnh sang tọa độ camera chứ không sử dụng để tính toán các thông số vị trí góc nhìn của camera bằng mô hình camera lỗ kim (thay vào đó là xây dựng ánh xạ tọa độ camera tới tọa độ thực sử dụng phương pháp chuyển đổi Homography), camera có tiêu cự càng dài thì hiện tượng méo càng giảm cường độ, chính vì vậy, phương pháp theo sáng chế sẽ không gặp phải khó khăn tại bước hiệu chỉnh camera.

### **Mô tả vắn tắt các hình vẽ**

Hình 1 là hình vẽ minh họa lưu đồ của phương pháp định vị đối tượng bằng hệ thống camera (camera network - CAMNET) theo sáng chế.

Hình 2 là hình vẽ minh họa so sánh ảnh hưởng của sai số lệch góc camera tới tính toán vị trí đối tượng.

Hình 3 là hình vẽ minh họa việc hiệu chỉnh camera để hạn chế sai số do hiện tượng méo.

Hình 4 là hình vẽ mô tả các hệ trục tọa độ trong mô hình camera lỗ kim.

Hình 5 là hình vẽ mô tả cách dựng hệ trục tọa độ phẳng tương tượng trên mặt phẳng (P) chứa 4 điểm A, B, C, D.

Hình 6 là hình vẽ minh họa phép chiếu phối cảnh giữa mặt phẳng trong hệ tọa độ camera với mặt phẳng trong hệ tọa độ thực.

### **Mô tả chi tiết sáng chế**

Theo nguyên tắc cơ bản của camera, tập hợp tất cả các điểm trong không gian được quan sát thấy tại điểm  $(u,v)$  trên tọa độ ảnh camera là một đường thẳng đi qua tâm camera.

Hình 1 là hình vẽ minh họa lưu đồ của phương pháp định vị đối tượng bằng hệ thống camera (camera network - CAMNET) theo sáng chế. Như được thể hiện trên Hình 1, các bước 1, 2, 3, 4 sẽ được thực hiện cho từng camera trong hệ thống CAMNET với mục tiêu xây dựng một ánh xạ đi từ mỗi điểm  $(u,v)$  trên hệ trục tọa độ ảnh camera tới một đường thẳng đi qua điểm tâm camera và các điểm trong không gian thực được quan sát thấy tại điểm  $(u,v)$  của ảnh.

#### **1) Hiệu chỉnh camera (thực hiện cho từng camera trong hệ thống CAMNET)**

Mục đích của việc hiệu chỉnh từng camera trong hệ thống camera là để thu thập các tham số nội của từng camera này, bao gồm: tiêu cự (focal length), điểm chính (principal point), trục xiên (axis skew) và hệ số méo (distortion coefficients);

Với một camera thông thường do cấu tạo của ống kính (lens) để có thể bao quát một vùng rộng, các ảnh thu được từ camera thường bị méo qua đó các hình chiếu của vật thể trong không gian thực ở trong khung hình camera thường không chính xác và dẫn tới các tính toán sai lệch trên không gian ảnh. Hệ số méo được chia thành:

- Tham số méo xuyên tâm  $k_x$  (radial distortion coefficient) do sự hiện diện của hiệu ứng "mắt cá" (fish-eye) và "thùng" (barrel).
- Tham số méo tiếp tuyến  $p_x$  (tangential distortion coefficient) do ống kính chụp ảnh không song song hoàn toàn với mặt phẳng hình ảnh.

Hình 2 là hình vẽ so sánh ảnh hưởng của sai số lệch góc camera tới tính toán vị trí đối tượng trong 2 trường hợp:

- (a) 2 camera đặt cách nhau 10cm, cùng hướng nhìn, với khoảng cách tới camera là 300 cm, một camera bị lệch 0,5 độ có thể gây ra sai số 62 cm.
- (b) 2 camera đặt cách xa nhau, hướng nhìn vuông góc, với khoảng cách tới camera là 300 cm, một camera bị lệch 0,5 độ có thể gây ra sai số chỉ 3 cm.

Thông thường để hiệu chỉnh một camera ta cần một tập hợp các điểm trên ảnh thu được từ camera ứng với đối tượng thực tế. Trong trường hợp này sẽ là các điểm lưới trên một bàn cờ (xem hình 3). Bàn cờ được sử dụng phải đáp ứng một vài tiêu chí:

- Các ô vuông trên bàn cờ phải có cùng một kích cỡ.
- Các đoạn thẳng chia bàn cờ thành các ô bàn cờ phải thẳng.
- Các điểm lưới trên bàn cờ cần phải rõ nét.

Bàn cờ được đặt trước camera tại các vị trí khác nhau sao cho camera có thể nhìn thấy tất cả các lưới bàn cờ và các vị trí đặt bàn cờ phải thỏa mãn lưới bàn cờ thu được từ ảnh bao quát hầu hết tất cả vùng nhìn thấy trên ảnh. Từ đó, sử dụng thư viện mã nguồn mở OpenCV (Open Source Computer Vision) để lấy được các tham số nội của camera (tham khảo tại đường link:

[https://docs.opencv.org/2.4/doc/tutorials/calib3d/camera\\_calibration/camera\\_calibration.html](https://docs.opencv.org/2.4/doc/tutorials/calib3d/camera_calibration/camera_calibration.html)).

## 2) Xây dựng ánh xạ từ tọa độ điểm ảnh của camera tới điểm trên hệ trục tọa độ camera

Bước đầu tiên ta sẽ ánh xạ tọa độ ảnh của camera tới tọa độ trên hệ trục tọa độ gắn với camera dựa vào mô hình camera lỗ kim (pinhole camera model) (xem hình 4):

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

Với:

- $(u, v)$  là tọa độ điểm ảnh trên ảnh camera;
- $f_x, f_y$  là tiêu cự (focal lengths) theo trục  $x, y$  tương ứng tính theo đơn vị điểm ảnh'
- $(c_x, c_y)$  là tọa độ điểm chính thường ở trung tâm của ảnh; và
- $(x, y, z)$  là tọa độ điểm tương ứng trên hệ trục tọa độ gắn với camera.

Theo mô hình camera lỗ kim, ta có thể tính  $(u, v)$  từ  $x, y, z$  theo công thức:

$$u = f_x \frac{x}{z} + c_x$$

$$v = f_y \frac{y}{z} + c_y$$

Tuy nhiên, trong thực tế thì các camera thông thường sẽ bị ảnh hưởng bởi hiện tượng méo do ánh sáng bị đổi hướng khi đi qua thấu kính của camera (gọi là hiện tượng distortion). Do đó, cách tính  $(u, v)$  từ  $x, y, z$  sẽ thay đổi như sau:

$$x' = \frac{x}{z}$$

$$y' = \frac{y}{z}$$

$$x'' = x' \frac{1 + k_1 r^2 + k_2 r^4 + k_3 r^6}{1 + k_4 r^2 + k_5 r^4 + k_6 r^6} + 2p_1 x' y' + p_2 (r^2 + 2x'^2)$$

$$y'' = y' \frac{1 + k_1 r^2 + k_2 r^4 + k_3 r^6}{1 + k_4 r^2 + k_5 r^4 + k_6 r^6} + p_1 (r^2 + 2y'^2) + 2p_2 x' y'$$

$$u = f_x x'' + c_x$$

$$v = f_y y'' + c_y$$

Các hệ số  $k_1, k_2, k_3, k_4, k_5, k_6, p_1, p_2, f_x, f_y, c_x, c_y$  được tính toán qua bước hiệu chỉnh camera bằng cách sử dụng thư viện mã nguồn mở OpenCV.

Sử dụng API *projectPoints* (với  $rvec=tvec=(0,0,0)$ ) của Thư viện mã nguồn mở OpenCV (tham khảo tại đường link: [https://docs.opencv.org/2.4/modules/calib3d/doc/camera\\_calibration\\_and\\_3d\\_reconstruction.html](https://docs.opencv.org/2.4/modules/calib3d/doc/camera_calibration_and_3d_reconstruction.html)), với mỗi điểm  $(u, v)$  trên tọa độ điểm ảnh của ảnh sẽ tính được đầu ra là điểm  $(x', y', z')$  trên hệ trục tọa độ camera và sau cùng điểm  $(u, v)$  sẽ được ánh xạ thành điểm  $(x, y)$  với  $x = x'/z'$  và  $y = y'/z'$ , cần chú ý  $(x, y, 1)$  là giao điểm của đường thẳng đi qua 2 điểm  $(0, 0, 0)$  (là vị trí tâm của camera) và  $(x', y', z')$  với mặt phẳng  $z=1$ . Hệ trục tọa độ camera sẽ nói ở các bước sau là hệ trục tọa độ 2 chiều nằm trên mặt phẳng  $z=1$  và có trục tung, trục hoành cùng chiều với hệ trục tọa độ camera 3 chiều trong mô hình camera lỗ kim.

### 3) Lấy điểm tham chiếu giữa tọa độ ảnh và tọa độ thực

Lấy các điểm tham chiếu để sử dụng cho thuật toán nội suy bằng cách đo các điểm trong thực tế tương ứng với các điểm trên mặt phẳng ảnh lấy được từ camera, từ đó tìm ra các tham số của phép biến đổi Homography giữa mặt phẳng ảnh và mặt phẳng trong không gian thực.

Các camera trong hệ thống CAMNET khi lấy điểm tham chiếu cần lấy hệ trục tọa độ giống nhau. Với mỗi hệ camnet ta cần chọn một hệ trục tọa độ 3 chiều OXYZ trong vùng hoạt động của hệ. Gốc tọa độ thường được chọn trên mặt đất để tiện cho việc đo đạc. Trục OZ là trục thẳng đứng hướng lên trên, trục OX và OY vuông góc với nhau và nằm trên mặt đất. Cần lưu ý rằng hệ trục tọa độ này là duy nhất cho toàn bộ hệ thống CAMNET chứ không phải là cho riêng từng camera trong hệ thống CAMNET này.

Một điểm tham chiếu là một điểm được chọn trong không gian thực có tọa độ  $(X, Y, Z)$  trong hệ trục tọa độ không gian thực đã được chọn trước, điểm này cần nằm trong vùng quan sát của camera và có tọa độ trên ảnh của camera là  $(u, v)$  (xem hình 4). Mỗi điểm tham chiếu  $(X, Y, Z)$  được lấy sẽ ghi lại đồng thời điểm  $(u, v)$  tương ứng.

Lấy 2 nhóm điểm tham chiếu, mỗi nhóm 4 điểm (lấy tổng cộng 8 điểm) sao cho các điểm trong cùng 1 nhóm đồng phẳng và tạo thành một tứ giác lồi.

### 4) Xây dựng ánh xạ từ điểm trên hệ trục tọa độ camera tới đường thẳng trong hệ trục tọa độ thực

Dựa vào mỗi nhóm điểm tham chiếu ở bước 3), ta sẽ dựng lên một ánh xạ từ điểm trên hệ trục tọa độ camera tới 1 điểm trên mặt phẳng tạo bởi các điểm tham chiếu của nhóm (tương ứng với 2 nhóm sẽ tạo được ra 2 ánh xạ). Như vậy, ta sẽ có ánh xạ đi từ một điểm trên hệ trục tọa độ camera tới cặp 2 điểm trong không gian thực, lưu ý rằng cặp 2 điểm biểu thị cho 1 đường thẳng (ngược lại 1 đường thẳng cũng được biểu diễn bởi 2 điểm). Đây chính là ánh xạ từ điểm tới đường thẳng cần xây dựng ở bước này.

Phương pháp xây dựng ánh xạ từ điểm trên hệ trục tọa độ camera tới 1 điểm trên mặt phẳng tạo bởi các điểm của 1 nhóm 4 điểm tham chiếu được mô tả như sau.

Phần mô tả sau đây sẽ sử dụng các chữ cái in nét thường biểu thị các hằng số, các chữ cái in nét đậm để biểu thị véc-tơ (hay điểm); sử dụng các phép tính toán cơ bản của véc-tơ trong hình học giải tích không gian là phép “.” (tích vô hướng), phép “ $\times$ ” (tích có hướng), phép “ $\|$ ”, (độ dài ơ-clit), phép “+” (cộng), phép “-” (trừ), phép “ $*$ ” (nhân với hằng số), phép “/” (chia cho hằng số).

Giả sử có 4 điểm tham chiếu  $A=(X1,Y1,Z1)$ ,  $B=(X2,Y2,Z2)$ ,  $C=(X3,Y3,Z3)$ ,  $D=(X4,Y4,Z4)$  nằm trên mặt phẳng (P) có tọa độ trên ảnh của camera lần lượt là  $(u1,v1)$ ,  $(u2,v2)$ ,  $(u3,v3)$ ,  $(u4,v4)$

Trước hết các điểm  $(u1,v1)$ ,  $(u2,v2)$ ,  $(u3,v3)$ ,  $(u4,v4)$  sẽ được thực hiện ánh xạ như ở bước 2 và sẽ tương ứng với các điểm  $(x1,y1)$ ,  $(x2,y2)$ ,  $(x3,y3)$ ,  $(x4,y4)$

Một hệ trục tọa độ phẳng tưởng tượng nằm trên mặt phẳng (P) được xây dựng với tâm là  $O' = (A+B+C+D)/4$ , véc-tơ đơn vị theo trục hoành là  $i=(A-O')/|A-O'|$ , véc-tơ đơn vị theo trục tung là  $j=y/|y|$  với  $y=i \times (B-O') \times i$  (xem hình 5)

Tọa độ các điểm A, B, C, D trên hệ trục tọa độ tâm  $O'$  lần lượt là  $A'=(A.i,A.j)=(X1',Y1')$ ,  $B'=(B.i,B.j)=(X2',Y2')$ ,  $C'=(C.i,C.j)=(X3',Y3')$ ,  $D'=(D.i,D.j)=(X4',Y4')$

Theo định lý Homography (thảo khảo tại đường link: <https://en.wikipedia.org/wiki/Homography>), thực hiện phép chiếu phối cảnh từ mặt phẳng (P) tới mặt phẳng  $z=1$  trong hệ trục tọa độ camera (xem hình 6), mỗi điểm  $(X',Y')$  trên hệ trục tọa độ tâm  $O'$  tương ứng với điểm  $(x,y)$  trên hệ trục tọa độ camera theo công thức:

$$X' = \frac{ax + by + c}{gx + hy + 1}, Y' = \frac{dx + ey + f}{gx + hy + 1}$$

Với a, b, c, d, e, f, g, h là các hằng số. Do đã biết các điểm  $(X1',Y1')$ ,  $(X2',Y2')$ ,  $(X3',Y3')$ ,  $(X4',Y4')$  tương ứng với các điểm  $(x1,y1)$ ,  $(x2,y2)$ ,  $(x3,y3)$ ,  $(x4,y4)$  nên các hằng số a, b, c, d, e, f, g, h là nghiệm của hệ phương trình:

$$\begin{bmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -x_1X'_1 & -y_1X'_1 \\ x_2 & y_2 & 1 & 0 & 0 & 0 & -x_2X'_2 & -y_2X'_2 \\ x_3 & y_3 & 1 & 0 & 0 & 0 & -x_3X'_3 & -y_3X'_3 \\ x_4 & y_4 & 1 & 0 & 0 & 0 & -x_4X'_4 & -y_4X'_4 \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -x_1Y'_1 & -y_1Y'_1 \\ 0 & 0 & 0 & x_2 & y_2 & 1 & -x_2Y'_2 & -y_2Y'_2 \\ 0 & 0 & 0 & x_3 & y_3 & 1 & -x_3Y'_3 & -y_3Y'_3 \\ 0 & 0 & 0 & x_4 & y_4 & 1 & -x_4Y'_4 & -y_4Y'_4 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \\ e \\ f \\ g \\ h \end{bmatrix} = \begin{bmatrix} X'_1 \\ X'_2 \\ X'_3 \\ X'_4 \\ Y'_1 \\ Y'_2 \\ Y'_3 \\ Y'_4 \end{bmatrix}$$

Từ đây tính được:

$$\begin{bmatrix} a \\ b \\ c \\ d \\ e \\ f \\ g \\ h \end{bmatrix} = \begin{bmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -x_1X'_1 & -y_1X'_1 \\ x_2 & y_2 & 1 & 0 & 0 & 0 & -x_2X'_2 & -y_2X'_2 \\ x_3 & y_3 & 1 & 0 & 0 & 0 & -x_3X'_3 & -y_3X'_3 \\ x_4 & y_4 & 1 & 0 & 0 & 0 & -x_4X'_4 & -y_4X'_4 \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -x_1Y'_1 & -y_1Y'_1 \\ 0 & 0 & 0 & x_2 & y_2 & 1 & -x_2Y'_2 & -y_2Y'_2 \\ 0 & 0 & 0 & x_3 & y_3 & 1 & -x_3Y'_3 & -y_3Y'_3 \\ 0 & 0 & 0 & x_4 & y_4 & 1 & -x_4Y'_4 & -y_4Y'_4 \end{bmatrix}^{-1} \begin{bmatrix} X'_1 \\ X'_2 \\ X'_3 \\ X'_4 \\ Y'_1 \\ Y'_2 \\ Y'_3 \\ Y'_4 \end{bmatrix}$$

Cuối cùng, ánh xạ từ điểm (x,y) trên hệ trục tọa độ camera tới một điểm (X,Y,Z) trên mặt phẳng tạo bởi các điểm của 1 nhóm 4 điểm tham chiếu được tính toán theo công thức sau:

$$X' = \frac{ax + by + c}{gx + hy + 1}$$

$$Y' = \frac{dx + ey + f}{gx + hy + 1}$$

$$(X,Y,Z) = i * X' + j * Y'$$

Với hai nhóm 4 điểm tham chiếu, tương tự xây dựng được ánh xạ từ điểm (x,y) trên hệ trục tọa độ camera tới đường thẳng đi qua điểm (X1,Y1,Z1), (X2,Y2,Z2) trong hệ trục tọa độ thực

5) Tổng hợp thông tin từ các camera để xác định vị trí của đối tượng trong không gian.

Với một camera khi phát hiện được đối tượng có tâm ở vị trí (u,v) trên tọa độ ảnh, sử dụng ánh xạ đã xây dựng được ở bước 2) và bước 4) thì có thể xác định được đối tượng này sẽ nằm trên đường thẳng đi qua 2 điểm (X1,Y1,Z1), (X2,Y2,Z2) trong hệ trục tọa độ thực. Các camera trong hệ thống CAMNET sẽ gửi thông tin các đường thẳng đi qua các đối tượng mà các camera này quan sát thấy lên đơn vị tổng hợp thông tin.

Trong trường hợp có đối tượng được quát sát thấy bởi hai camera. Đơn vị tổng hợp thông tin sẽ nhận được từ hai camera hai đường thẳng (d1) và (d2) trên lý thuyết ta sẽ có (d1) và (d2) cắt nhau tại chính vị trí tâm của đối tượng, nhưng trên thực tế (d1) và (d2) thường sẽ không cắt nhau tuyệt đối. Vì vậy, đơn vị tổng hợp thông tin sẽ tính khoảng cách

giữa hai đường thẳng này nếu nhỏ hơn một ngưỡng đặt trước (thường được chọn bằng khoảng một nửa kích thước trung bình của đối tượng camera đang quan sát) sẽ coi đây là hai đường thẳng cắt nhau và giao điểm sẽ tính là trung điểm của đoạn vuông góc chung của hai đường thẳng.

Ngoài đường thẳng đi qua tâm đối tượng, camera sẽ gửi thêm thông tin đường thẳng đi qua các mép trái phải của đối tượng để đơn vị tổng hợp có thể tính toán kích thước của đối tượng khi đã xác định được vị trí.

Dựa vào thông tin về đường thẳng đi qua tâm và các mép của đối tượng được gửi từ các camera, đơn vị tổng hợp thông tin sẽ chọn lọc các đường thẳng có khả năng cùng đi qua 1 đối tượng, tính toán vị trí của đối tượng, đánh giá tính khả thi của vị trí đối tượng và kích thước của đối tượng để đưa ra kết luận về kết quả định vị các đối tượng trong vùng kiểm soát của hệ thống CAMNET

Các lợi ích đạt được của sáng chế

Phương pháp theo sáng chế là công nghệ lõi giúp cho các nhà phát triển ứng dụng trong các bài toán cần nhận dạng chủ thể và hướng chuyển động để ứng dụng trong các bài toán thực tế đặt ra như nhận diện để chấm công nhân viên sử dụng khuôn mặt, nhận diện khách hàng VIP để hành xử riêng, v.v.

## YÊU CẦU BẢO HỘ

1. Phương pháp định vị đối tượng bằng hệ thống camera (camera network - CAMNET), bao gồm các bước sau:

i) hiệu chỉnh từng camera trong hệ thống camera để thu thập các tham số nội của từng camera này bao gồm: tiêu cự (focal length), điểm chính (principal point), trục xiên (Axis skew) và hệ số méo (distortion coefficients);

ii) xây dựng ánh xạ từ điểm trên hệ trục tọa độ camera tới đường thẳng trong hệ trục tọa độ thực dựa vào mô hình camera lỗ kim (pinhole camera model):

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

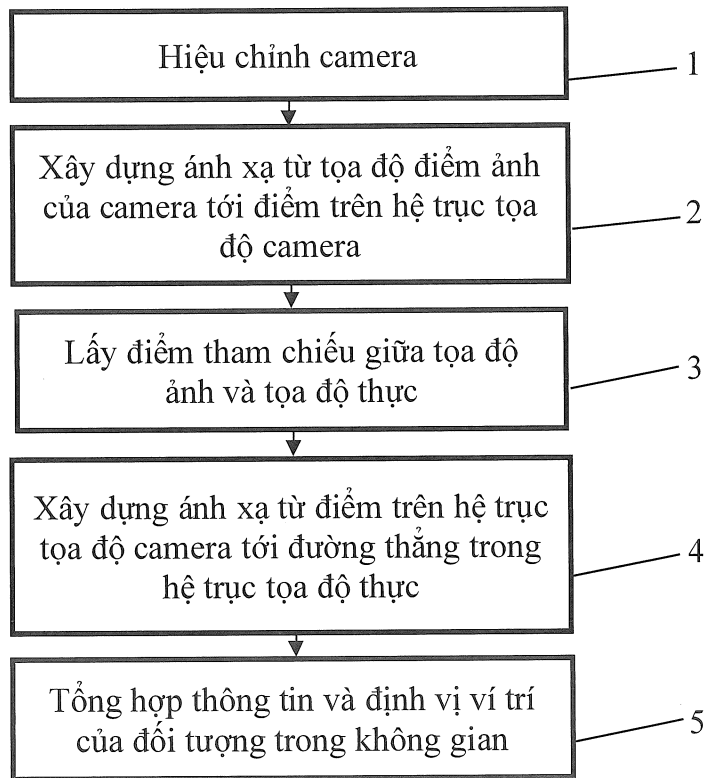
trong đó:

- (u, v) là tọa độ điểm ảnh trên ảnh camera,
- $f_x, f_y$  là tiêu cự (focal lengths) theo trục x, y tương ứng tính theo đơn vị điểm ảnh,
- (cx, cy) là tọa độ điểm chính thường ở trung tâm của ảnh, và
- (x,y,z) là tọa độ điểm tương ứng trên hệ trục tọa độ gắn với camera,

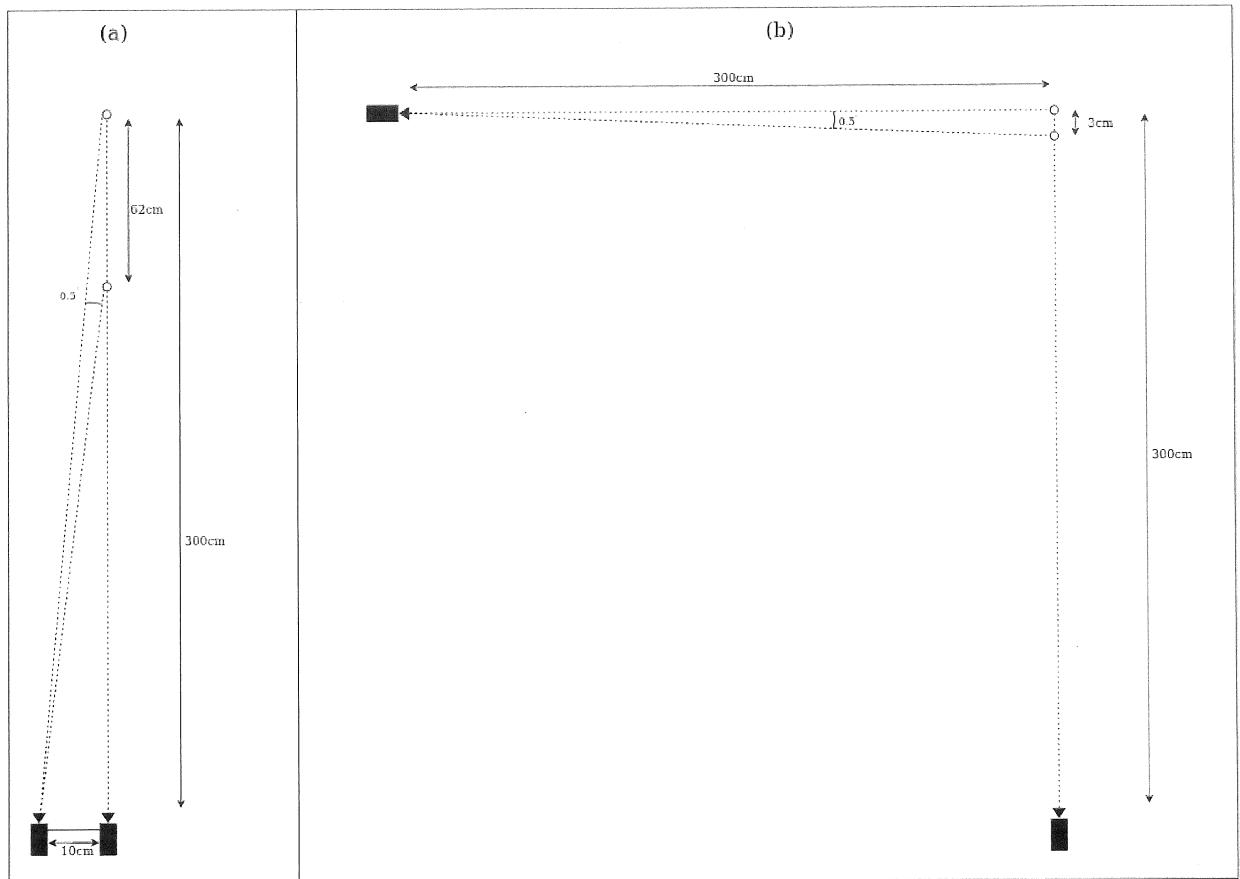
iii) lấy điểm tham chiếu giữa tọa độ ảnh và tọa độ thực để sử dụng cho thuật toán nội suy bằng cách đo các điểm trong thực tế tương ứng với các điểm trên mặt phẳng ảnh lấy được từ camera, từ đó tìm ra các tham số của phép biến đổi Homography giữa mặt phẳng ảnh và mặt phẳng trong không gian thực;

iv) xây dựng ánh xạ từ điểm trên hệ trục tọa độ camera tới đường thẳng trong hệ trục tọa độ thực; và

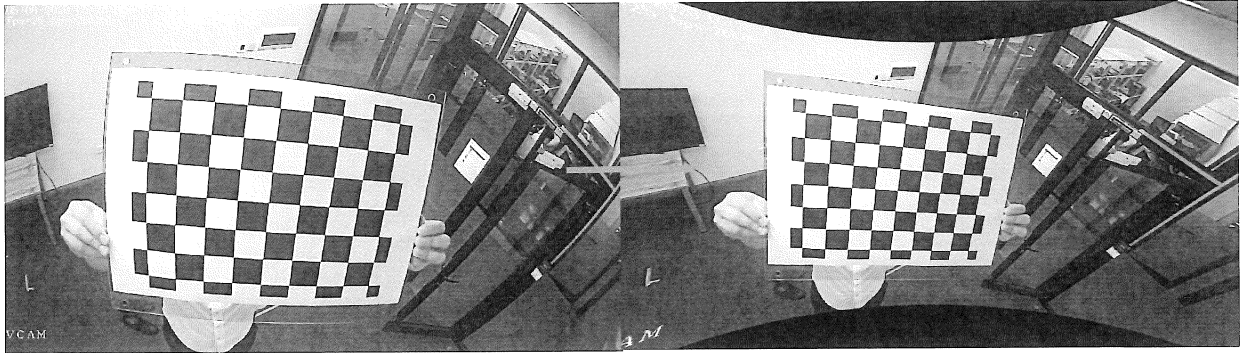
v) tổng hợp thông tin và xác định vị trí của đối tượng trong không gian bằng cách sử dụng thông tin của đối tượng trong không gian thực (ánh xạ từ hình ảnh đối tượng trên camera) được gửi lên từ các camera thành phần, kết hợp thông tin của ít nhất 2 camera và kích thước ước lượng của đối tượng để có thể xác định được vị trí chính xác của đối tượng trong không gian thực.



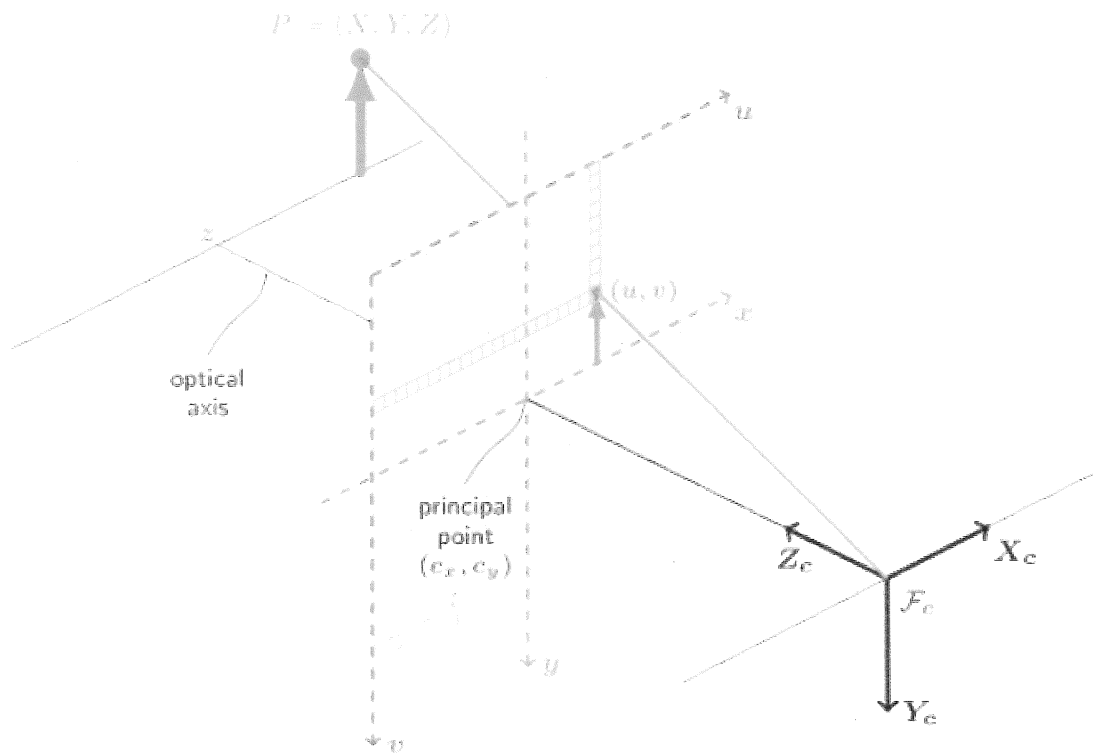
Hình 1



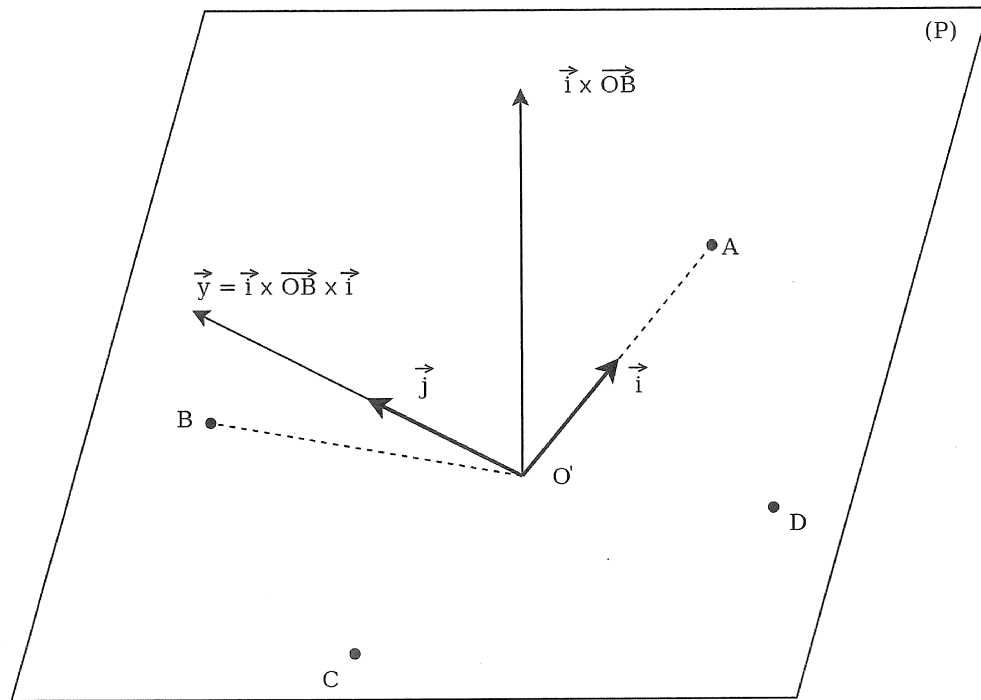
Hình 2



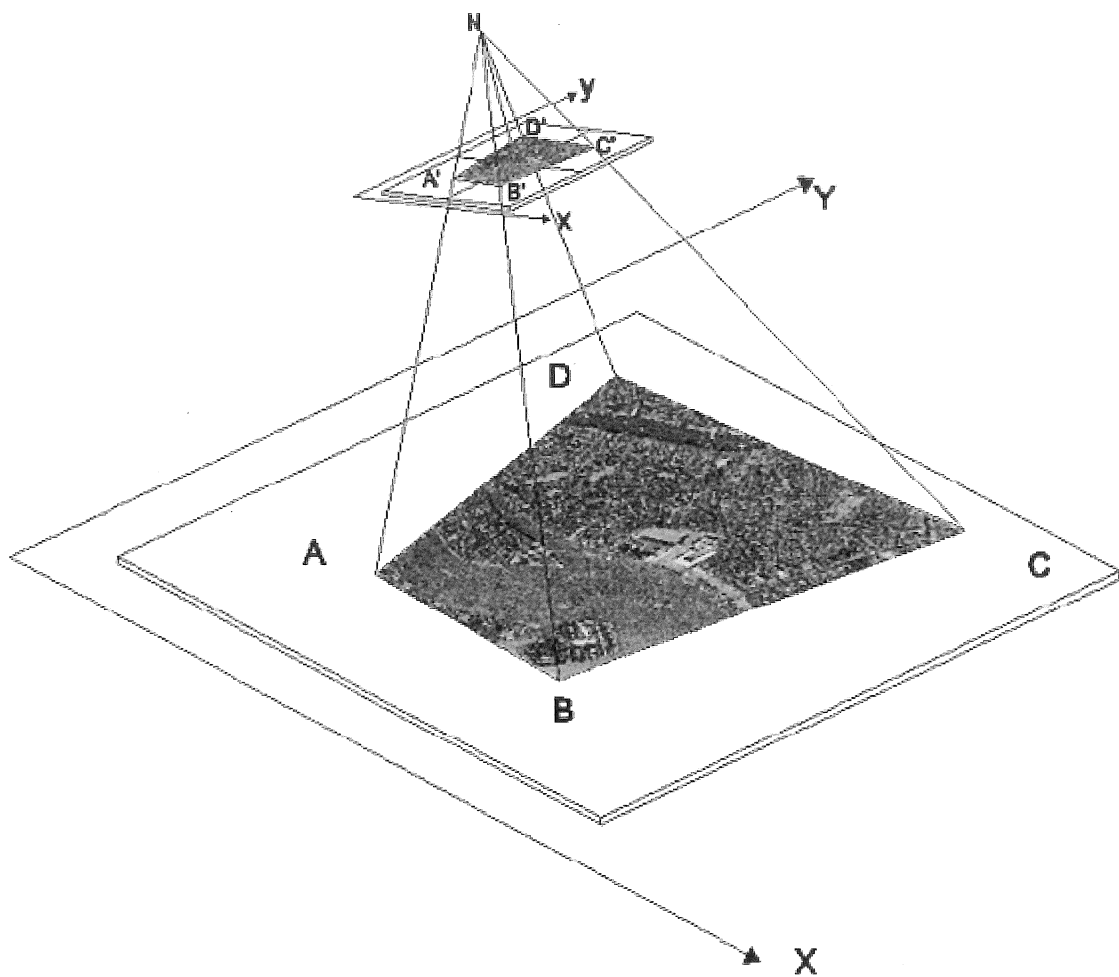
Hình 3



Hình 4



Hình 5



Hình 6