

PHÁT HIỆN VÀ ƯỚC LƯỢNG KHOẢNG CÁCH VẬT CẢN, ỨNG DỤNG TRỢ GIÚP DẪN ĐƯỜNG CHO NGƯỜI KHIẾM THỊ

Nguyễn Quốc Hùng[†],
Trần Thị Thanh Hải*, Vũ Hải*, Hoàng Văn Nam*,
Nguyễn Quang Hoan[‡]

* Viện nghiên cứu quốc tế MICA, Trường ĐHBK HN - CNRS/UMI - 2954 - INP Grenoble

[†] Trường Cao đẳng Y tế Thái Nguyên

[‡] Trường Đại học sư phạm kỹ thuật Hưng Yên

Tóm tắt: Bài báo này trình bày một phương pháp phát hiện và ước lượng khoảng cách vật cản sử dụng camera duy nhất gắn trên robot ứng dụng trong trợ giúp dẫn đường cho người khiếm thị. Với vật cản tĩnh ít di chuyển vị trí trong môi trường, chúng tôi lưu trữ thông tin về loại vật cản, vị trí và hình ảnh của vật cản vào CSDL. Trong quá trình di chuyển, chúng tôi thực hiện đối sánh nhanh quan sát hiện tại với quan sát tương ứng được xác định bởi giải thuật định vị [1]. Sau đó sự có mặt của vật cản tại quan sát trong CSDL sẽ được kiểm tra và xác định vị trí trong quan sát hiện tại. Với vật cản động, cụ thể là người di chuyển, chúng tôi sử dụng giải thuật HOG-SVM là một bộ phát hiện người hiệu quả đề xuất bởi Dalal và các cộng sự [2]. Việc ước lượng khoảng cách từ camera tới vật cản chỉ sử dụng một camera RGB là một bài toán không đơn giản. Trong bài báo này, chúng tôi đề xuất giải pháp xây dựng bản đồ chênh lệch từ quan sát hiện tại và quan sát trước đó để ước lượng khoảng cách tương đối từ vật cản tới robot. Các kết quả thực nghiệm được tiến hành khi camera di chuyển trên hành lang có chiều dài 60 m trong các điều kiện chiếu sáng khác nhau cho thấy phương pháp phát hiện và ước lượng khoảng cách vật cản đề xuất là phù hợp, giúp cho người khiếm thị có thể nhận biết và tránh được các vật cản nguy hiểm trong khi di chuyển.

Từ khóa: Phát hiện vật cản; Ước lượng khoảng cách vật cản; Robot dẫn đường.

Tác giả liên hệ: Nguyễn Quốc Hùng,
email: Quoc-Hung.Nguyen@mica.edu.vn;
mobile: (+84) 912 251 253

Đến tòa soạn: 12/2/2016, chỉnh sửa: 12/4/2016, chấp nhận
đăng: 12/5/2016.

Một phần kết quả của bài báo này đã được trình bày tại quốc
gia ECIT'2015.

I. GIỚI THIỆU

Phát hiện và ước lượng khoảng cách vật cản là một chủ đề thu hút sự quan tâm của các nhà khoa học trong thời gian dài bởi ý nghĩa và tính ứng dụng của nó trong các bài toán dẫn đường tránh vật cản cho robot, xe tự hành. Đã có rất nhiều phương pháp đề xuất sử dụng công nghệ khác nhau như GPS, LIDAR, RFID, Camera nhằm tăng độ chính xác phát hiện, giảm độ sai số ước lượng và thời gian tính toán.

Mục tiêu của chúng tôi là nghiên cứu và phát triển hệ thống robot thông minh di động, có khả năng trợ giúp dẫn đường cho người khiếm thị (NKT) trong môi trường trong nhà. Các nghiên cứu liên quan đến việc biểu diễn môi trường, định vị, dẫn hướng đã được trình bày trong các bài báo trước của chúng tôi [1]. Trong bài báo này, chúng tôi trình bày một phương pháp phát hiện và ước lượng khoảng cách vật cản nhằm hoàn thiện hệ thống cuối cùng là dẫn đường và cảnh báo vật cản. Phương pháp mà chúng tôi đề xuất chỉ sử dụng một camera RGB duy nhất gắn trên robot. Hình ảnh thu nhận được từ camera sẽ được đối sánh để xác định vị trí của robot trên bản đồ, sau đó xác định sự có mặt của vật cản. Để ước lượng khoảng cách từ vật cản đến đối tượng, chúng tôi đề xuất giải pháp sử dụng hai khung nhìn của camera ở hai thời điểm khác nhau để xây dựng bản đồ chênh lệch, từ đó ước lượng độ sâu của vật cản.

II. CÁC NGHIÊN CỨU LIÊN QUAN

Trong phần này, chúng tôi trình bày một số nghiên cứu liên quan đến phát hiện và ước lượng

khoảng cách vật cản trong ứng dụng dẫn đường cho robot. Các hướng tiếp cận được chia thành ba nhóm chính: (i) sử dụng 01 camera ; (ii) sử dụng camera kép (camera-stereo); (iii) sử dụng cảm biến ảnh và độ sâu (RGB-D).

1) *Hướng tiếp cận sử dụng 01 camera*: Hướng tiếp cận sử dụng 01 camera khá phù hợp với bài toán phát hiện đối tượng động và tĩnh. Tuy nhiên việc sử dụng chỉ một camera gặp khó khăn trong việc dự đoán khoảng cách vật cản. Jeongdae Kim 2012 [3] sử dụng 01 camera xây dựng bản đồ chênh lệch nhằm phát hiện người di chuyển trong môi trường bằng cách dự đoán chuyển động của các vùng phát hiện được (Block-Based Motion Estimation). Taylor 2004 [4] đề xuất phương pháp ROP (Radial Obstacle Profile) xây dựng bản đồ vật cản sử dụng 01 camera nhằm xác định phạm vi vật cản gần nhất trong bất kỳ hướng nào khi robot di chuyển. Erik Einhorn 2009 [5] trình bày phương pháp sử dụng các đặc trưng SIFT, SURF bất biến với các phép biến đổi kết hợp với bộ lọc kalman mở rộng (EKF) xử lý một chuỗi các hình ảnh chụp bằng máy ảnh duy nhất được gắn ở phía trước của một robot di động nhằm tái tạo lại môi trường phục vụ cho bài toán phát hiện đối tượng.

2) *Hướng tiếp cận sử dụng camera kép*: Camera kép (camera-stereo) là thiết bị chuyên dụng cho các bài toán liên quan đến việc ước lượng khoảng cách. Điểm mạnh của loại thiết bị này là khả năng tái tạo chính xác không gian 3D trên bản đồ chênh lệch các điểm ảnh. Tuy nhiên đây là thiết bị giá thành cao, việc hiệu chỉnh tương đối phức tạp.

Lazaros Nalpantidis 2009 [6] trình bày thuật toán ra quyết định (Decision Making) tránh vật cản dựa vào thông tin hình ảnh thu nhận từ camera-stereo. Ming Bai 2010 [7] trình bày phương pháp phát hiện vật cản cho phép robot tìm đường an toàn trong các tình huống phức tạp sử dụng thông tin hình ảnh được thu thập từ camera-stereo. Rostam Affendi Hamzah 2011 [8] sử dụng phương pháp xây dựng bản đồ chênh lệch từ hai quan sát nhằm ước lượng khoảng cách vật cản phía trước giúp robot tránh được va chạm khi di chuyển. Lagisetty 2013 [9] đề xuất phương pháp phát hiện và tránh vật cản sử dụng camera-stereo gắn trên robot di động trong môi trường có cấu trúc nhằm giải quyết 02 bài toán cơ bản là xác định vị trí, hướng của robot và xác định kích thước, hình dạng, khoảng cách phạm vi của vật

cản có trong môi trường.

3) *Hướng tiếp cận sử dụng cảm biến hình ảnh độ sâu (RGB-D)*: Các cảm biến cung cấp hình ảnh và độ sâu như Microsoft Kinect có giá thành rẻ đang được sử dụng rất rộng rãi trong các ứng dụng giải trí và nghiên cứu. Các cảm biến này thường được áp dụng cho môi trường trong nhà. Khi ở ngoài trời có ánh sáng tự nhiên (hàng lang) thì thiết bị này không thích hợp.

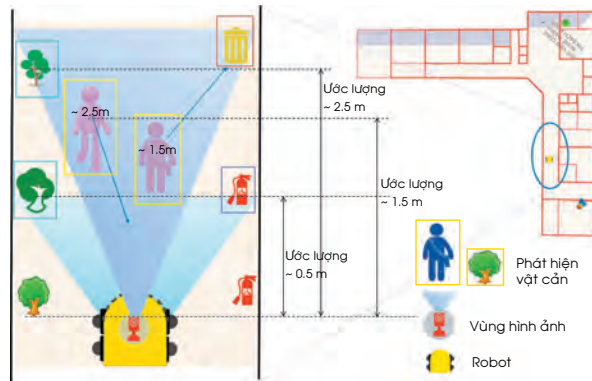
Diogo Santos 2012 [10] đề xuất phương pháp nhận dạng các cấu trúc khác nhau của môi trường trong nhà (con đường phía trước, bên phải, bên trái) sử dụng kỹ thuật mạng neuron nhân tạo trên dữ liệu hình ảnh và độ sâu thu được từ cảm biến Kinect. Sharon Nissimov 2015 [11] đề xuất mô hình xe gắn cảm biến Kinect để phát hiện vật cản phía trước sử dụng đồng bộ thông tin màu (RGB) và độ sâu (Depth). Việc quyết định vùng chứa vật cản được thực hiện bằng cách sử dụng thông tin cường độ điểm ảnh nằm trong vùng độ dốc xác định so với các điểm ảnh lân cận. Brian Peasley 2013 [12] trình bày phương pháp phát hiện vật cản sử dụng cảm biến Kinect bằng cách chiếu các điểm ảnh 3D lên mặt phẳng nhằm xây dựng một bản đồ 2D cho phép xác định xem có tồn tại vật cản trong môi trường. Sau đó vận tốc tịnh tiến và quay của robot được hiệu chỉnh để robot có thể tránh được vật cản. Các thử nghiệm với nhiều kịch bản trong nhà bao gồm các vật cản cố định và di chuyển với độ cao khác nhau, đặc biệt hệ thống không phụ thuộc nhiều vào điều kiện môi trường như ánh sáng và hoạt động trong thời gian thực.

Căn cứ vào các phân tích đánh giá phía trên, trong ngữ cảnh trợ giúp NKT trong môi trường trong nhà, chúng tôi lựa chọn đi theo hướng tiếp cận sử dụng 01 camera với mục đích thu nhận được hình ảnh có góc nhìn tốt nhất và thời gian tính toán nhanh cho cả hai bài toán phát hiện và ước lượng khoảng cách. Phần tiếp theo sẽ trình bày chi tiết của phương pháp đề xuất.

III. PHƯƠNG PHÁP ĐỀ XUẤT

Việc phát hiện và ước lượng vật cản được minh họa như trong Hình 1. Trong mô hình này, robot gắn camera RGB thông thường di chuyển với tốc độ nào đó. Trong quá trình di chuyển, robot có thể gặp các vật cản cố định trong môi trường (chậu hoa, bình cứu hỏa, thùng rác) hoặc các vật cản động xuất hiện bất ngờ (người). Giả thiết của

bài toán là robot di chuyển trên 01 mặt sàn bằng phẳng. Chuyển động của robot theo một lộ trình đã được xác định.



Hình 1. Mô hình phát hiện và định vị vật cản ước lượng khoảng cách

A. Khung làm việc tổng quát

Tại thời điểm k , camera trên robot thu nhận hình ảnh I_k . Với hình ảnh này, vị trí của robot trong môi trường đã được xác định bởi mô đun định vị (xem chi tiết trong bài báo [13]). Vị trí đó là một điểm $P_{(x,y,z=0)}$ trong hệ quy chiếu đã được định nghĩa từ trước; $z = 0$ vì giả thiết robot chuyển động trên một mặt phẳng. Bài toán phát hiện và ước lượng khoảng cách vật cản được định nghĩa như sau:

- + **Đầu vào:** Ảnh I_k , vị trí của robot $P_{(x,y,z=0)}$.
- + **Đầu ra:** Tập các vật cản và vị trí của nó trong hệ quy chiếu đã định nghĩa từ trước:
 $O_k = \{O_k(x, y), k \in [1, n]\}$.

Mô hình phát hiện vật cản đề xuất gồm hai pha như minh họa trong Hình 2 gồm:

- **Phát hiện vật cản:** Chúng tôi phân các vật cản thành hai nhóm: vật cản tĩnh và vật cản động. Vật cản tĩnh là các đối tượng trong môi trường như chậu hoa, bình cứu hỏa, thùng rác trong khi vật cản động là các đối tượng người di chuyển trong môi trường.
- **Ước lượng khoảng cách vật cản:** Chúng tôi lấy ý tưởng dự đoán khoảng cách từ hệ thống camera kép mô phỏng như đôi mắt của người. Tuy nhiên trong ngữ cảnh của bài toán chúng tôi chỉ sử dụng duy nhất 01 camera duy nhất gắn trên robot chuyển động, quan sát hình ảnh tại hai thời điểm khác nhau.

Phần dưới đây, chúng tôi trình bày chi tiết kỹ thuật phát hiện và ước lượng khoảng cách từ vật cản tới robot.

B. Phát hiện vật cản

1) **Phát hiện vật cản cố định:** Mục tiêu là phát hiện các đồ vật chính xác và nhanh nhất có thể. Ý tưởng cơ bản của chúng tôi là học trước các vật cản cũng như vị trí của chúng trong hệ quy chiếu đã định nghĩa, các thông tin này sẽ được lưu lại trong CSDL biểu diễn môi trường. Với ảnh đầu vào, sau khi đã xác định một cách tương đối vị trí của robot trên bản đồ bằng giải thuật định vị trình bày [1], tương ứng với nó là các đối tượng trong môi trường. Pha phát hiện vật cản tĩnh chỉ kiểm tra và định vị lại cho chính xác hơn.

Trong hình 2, bản đồ môi trường được biểu diễn là một tập các điểm quan trọng trong môi trường $\mathcal{L}^N = \{L_1, L_2, \dots, L_i, \dots, L_N\}$. Với mỗi điểm L_i , tương ứng là ảnh I_i , đặc tả bởi quan sát Z^i và tập tất cả các đối tượng có thể quan sát được tại thời vị trí L_i : $\{O_{i_1}, O_{i_2}, \dots, O_{i_{n_i}}\}$. Tại thời điểm k , camera thu nhận ảnh I_k , nhờ giải thuật định vị đã được trình bày trong [1] [13], vị trí của robot được xác định tương ứng L_k^* . Tại vị trí L_k^* này, tương ứng với quan sát của robot I_k^* , tập các vật cản trong môi trường cũng đã được xác định và lưu sẵn: $\{O_{k_1}, O_{k_2}, \dots, O_{k_{n_k}}\}$.

Các bước thực hiện để phát hiện sự có mặt của các vật cản tại thời điểm k như sau:

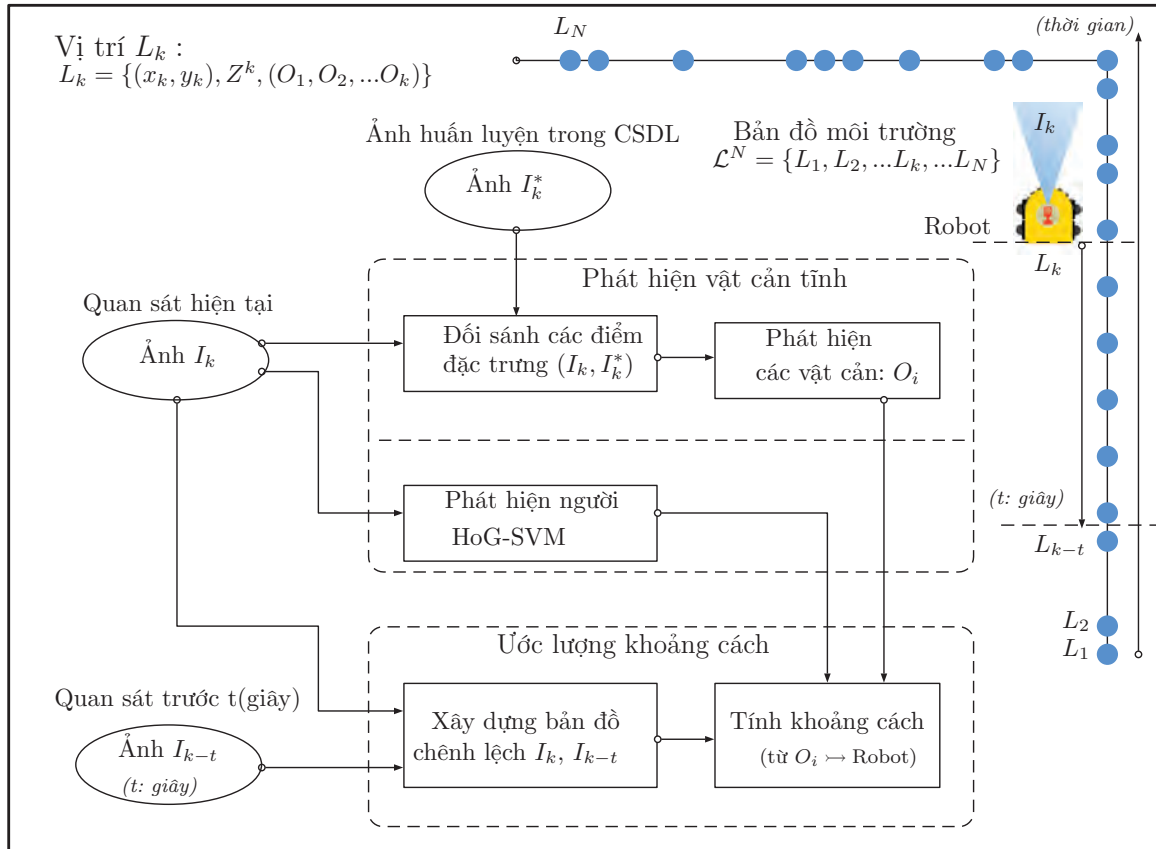
- 1) Trích chọn đặc trưng trên hai ảnh I_k và I_k^* và đối sánh điểm đặc trưng tương ứng trên hai ảnh này
- 2) Xác định vật cản trên ảnh I_k từ kết quả đối sánh
- 3) Xác định vùng chứa đối tượng

Phần dưới đây, chúng tôi trình bày chi tiết kỹ thuật các bước thực hiện.

- **Đối sánh các điểm đặc trưng:** Mục đích của công việc này là xác định các cặp điểm đặc trưng tương đồng giữa hai ảnh I_k và I_k^* thông qua một thủ tục đối sánh FLANN [14]. Sau đó sử dụng ngưỡng để loại bỏ các cặp điểm đối sánh sai nhằm tìm ra được tập các cặp điểm tương đồng. Quá trình này gồm các bước như sau:

- + **Trích chọn đặc trưng và bộ mô tả:** trong bài báo này chúng tôi sử dụng đặc trưng cục bộ SIFT [15]. SIFT là một loại đặc

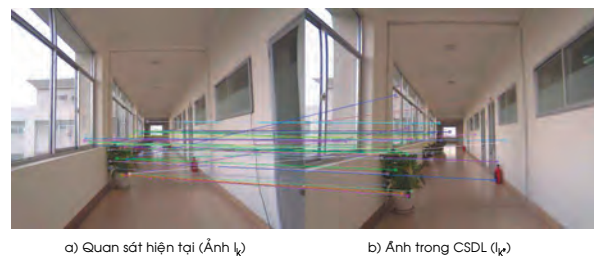
SƠ ĐỒ CÁC BƯỚC TÍNH TOÁN



Hình 2. Các bước phát hiện và ước lượng khoảng cách vật cản

- trưng đã được chứng minh là bất biến với sự thay đổi về tỉ lệ, chiều sáng, góc nhìn của đối tượng trong ảnh. Chi tiết kỹ thuật trích chọn đặc trưng SIFT có trong [15].
- + **Đối sánh các điểm đặc trưng:** Ý tưởng giải thuật FLANN [14] là tìm tập đặc trưng tương ứng ở hai ảnh I_k và I_k^* . Giả sử $F_k = \{p_{k_1}, p_{k_2}, \dots, p_{k_i}\}$ và $F_k^* = \{p_{k_1^*}, p_{k_2^*}, \dots, p_{k_j^*}\}$ là hai tập điểm đặc trưng trích chọn từ hai ảnh tương ứng I_k và I_k^* . Trong đó i, j là số điểm đặc trưng phát hiện từ mỗi ảnh. Khoảng cách Euclid trong không gian đặc trưng giữa hai điểm p_{k_m} và $p_{k_n^*}$, quy ước là $D(p_{k_m}, p_{k_n^*})$. Theo [14] hai điểm đặc trưng p_{k_m} và $p_{k_n^*}$ được coi là giống nhau nếu như $D(p_{k_m}, p_{k_n^*})$ là nhỏ nhất và tỷ số giữa khoảng cách nhỏ nhất và khoảng cách nhỏ hơn một ngưỡng cho trước.
 - + **Loại bỏ cặp điểm sai sử dụng ngưỡng:**

Mỗi cặp điểm được gọi là matching yếu nếu như khoảng cách Euclid giữa chúng nhỏ hơn hai lần khoảng cách nhỏ nhất trong số tất cả các cặp điểm hoặc lớn hơn một ngưỡng cố định $T_{dist} = 0.2$. Hình 3 minh họa kết quả đạt được sau khi loại bỏ các cặp matching yếu.



Hình 3. Kết quả khi loại bỏ một số cặp điểm đối sánh yếu

- **Phát hiện vật cản từ kết quả đối sánh**
 - + **Xác định cặp điểm tương ứng của vật cản trên hai ảnh liên tiếp:** Ảnh I_k^* đã có thông tin về vật cản trong ảnh (được

xác định bởi bao đóng chữ nhật). Chúng tôi thực hiện khoanh vùng các điểm đặc trưng nằm trong vùng chứa vật cản trong ảnh I_k^* . Sau đó tìm điểm tương ứng của chúng trên ảnh I_k .

- + **Tính ma trận chuyển tọa độ H :** Mục đích của việc này nhằm tính toán tọa độ đối tượng trong quan sát hiện thời I_k^* dựa vào các cặp điểm đặc trưng. Cụ thể là với các cặp điểm tương đồng $(p_{k_m}^*, p_{k_n})$, thực hiện ước lượng ma trận chuyển đổi H dựa vào ràng buộc:

$$p_{k_n} = H \cdot p_{k_m}^* \quad (1)$$

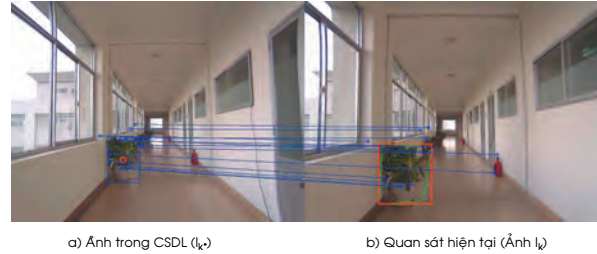
$$\text{Trong đó } p_{k_n} = \begin{bmatrix} x_{k_n} \\ y_{k_n} \\ 1 \end{bmatrix}; p_{k_m}^* = \begin{bmatrix} x_{k_m}^* \\ y_{k_m}^* \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} x_{k_n} \cdot w \\ y_{k_n} \cdot w \\ w \end{bmatrix}; H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}$$

h_{ij} là hệ số của ma trận H , (x_{k_m}, y_{k_m}) và $(x_{k_n}^*, y_{k_n}^*)$ là tọa độ cặp điểm tương đồng trong không gian ảnh, $w \neq 0$ là thành phần thứ 3 trong hệ tọa độ đồng nhất của p_{k_n} . Do ma trận H có 8 bậc tự do, mỗi cặp điểm cho ta 2 phương trình ràng buộc nên để giải được ma trận H ta cần ít nhất 4 cặp điểm tương đồng [16]. Tuy nhiên, trong thực tế nếu chọn chính xác 4 cặp điểm để xây dựng ma trận H thì có thể gây ra sai số rất lớn nếu như có một cặp đối sánh sai. Điều này rất hay xảy ra do trong môi trường tòa nhà có khá nhiều các vị trí mà tại đó các điểm đặc trưng có độ tương đồng lớn. Vì vậy, một kỹ thuật phổ biến và thường hay được sử dụng để khắc phục trường hợp này khi xây dựng ma trận H đó là kỹ thuật RANSAC [17] nhằm tìm ra 4 cặp điểm tương ứng xây dựng nên một đa giác, bằng phương pháp hình học chuẩn hóa về dạng hình chữ nhật bao lấy vùng chứa đối tượng.

- **Xác định vùng chứa đối tượng** Mục đích của việc này là xác định vị trí của đối tượng trên ảnh I_k của quan sát hiện tại. Từ 4 góc của vật cản trên ảnh I_k , xác định 4 góc vật cản trên ảnh I_k^* thông qua ma trận H vừa tính ở trên. Kết quả xác định vật cản cố định được minh họa ở hình 4, trong đó tâm của đối

tượng được xác định ở sử dụng kỹ thuật đối sánh ảnh FLANN [14] xác định đối tượng trong ảnh hiện thời.



Hình 4. Minh họa kết quả xác định vùng chứa đối tượng

Hình 4(a) Minh họa ảnh I_k^* chứa đối tượng với các điểm đặc trưng trích chọn và vùng chữ nhật khoanh đối tượng từ trước. Hình 4(b) là ảnh I_k với các điểm đặc trưng tương ứng. Kết quả hình chữ nhật màu đỏ khoanh vùng đối tượng chuẩn hóa bao lấy đối tượng và sử dụng kỹ thuật biến đổi hình học để chuẩn hóa về đa giác màu xanh.

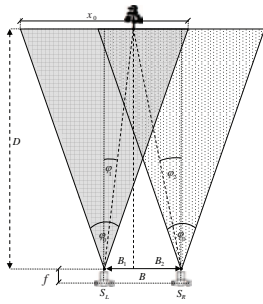
2) **Phát hiện vật cản động:** Đối với vật cản động, chúng tôi đề xuất phát hiện người (là người di chuyển với tốc độ trung bình với vận tốc $v=1.4m/s$ [18]) là đối tượng hay gặp trong các tình huống thử nghiệm tại các môi trường thực tế. Đây là chủ đề thu hút được nhiều nghiên cứu trong thời gian gần đây [19] nhằm nâng cao hiệu năng nhận dạng.

C. Ước lượng khoảng cách từ camera tới vật cản

1) **Nguyên lý ước lượng khoảng cách từ hai camera:** Mục đích của việc dự đoán khoảng cách là tái tạo lại không gian 3 chiều (3D), mô phỏng lại hệ thống thị giác của con người thông qua việc lấy đồng thời ảnh từ hai camera cùng quan sát một khung cảnh từ các góc nhìn khác nhau. Bằng phép biến đổi hình học tính toán được khoảng chênh lệch giữa hai quan sát trên ảnh để từ đó ước lượng khoảng cách trên thực địa như minh họa trong hình 5.

Trong đó:

- + S_L và S_R hai camera được đặt đồng trục trên cùng một mặt phẳng.
- + B khoảng cách nối tâm hai camera; B_1 khoảng cách từ tâm chiếu đối tượng tới camera thứ nhất, B_2 khoảng cách đến camera thứ hai.



Hình 5. Mô hình ước lượng khoảng cách từ hai quan sát

- + φ_0 góc quan sát đối tượng từ 02 camera, φ_1 và φ_2 là góc giữa trục quang học của camera và các đối tượng quan sát.
- + f tiêu cự ống kính hai camera; x_0 khoảng cách vùng quan sát của camera.

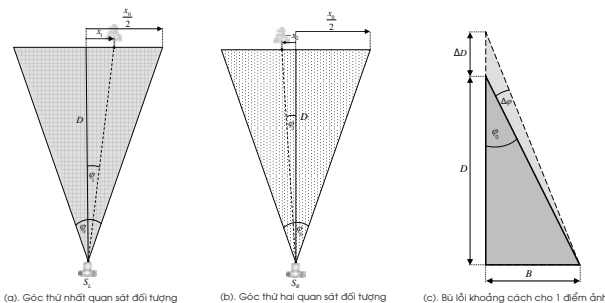
Xuất phát từ khoảng cách đường nối tâm camera B được xác định từ hai thành phần B_1 và B_2 ta có:

$$B = B_1 + B_2 = D \tan(\varphi_1) + D \tan(\varphi_2) \quad (2)$$

Do vậy khoảng cách ước lượng từ camera đến đối tượng tính bằng công thức sau:

$$D = \frac{B}{\tan(\varphi_1) + \tan(\varphi_2)} \quad (3)$$

Để xác định khoảng cách này, chúng ta sẽ phân tích hình ảnh của đối tượng được quan sát được ở hoành độ trên ảnh x_1 từ camera thứ nhất và hoành độ trên ảnh x_2 từ camera thứ hai.



Hình 6. Hình ảnh của đối tượng quan sát từ hai góc thu nhận

Áp dụng nguyên lý đồng dạng trong hình học như mô tả trong hình 6(a-b) ta có:

$$\frac{x_1}{(\frac{x_0}{2})} = \frac{\tan(\varphi_1)}{\tan(\frac{\varphi_0}{2})} \quad (4)$$

$$\frac{-x_2}{(\frac{x_0}{2})} = \frac{\tan(\varphi_2)}{\tan(\frac{\varphi_0}{2})} \quad (5)$$

Thay giá trị $\tan(\varphi_1)$, $\tan(\varphi_2)$ trong công thức 4 và công thức 5 vào công thức 3, khoảng cách D được tính toán như sau:

$$D = \frac{Bx_0}{2 \tan(\frac{\varphi_0}{2})(x_1 - x_2)} \quad (6)$$

Với x_0 là chiều rộng của ảnh, $(x_1 - x_2)$ là chênh lệch (Disparity) về vị trí của đối tượng quan sát trên camera thứ nhất và thứ hai cùng tính theo từng điểm ảnh.

Tuy nhiên, theo [20], [21], [22] đã chứng minh được khoảng cách D sẽ tỷ lệ nghịch với hiệu $(x_1 - x_2)$ vì vậy để bù lỗi cho góc quan sát φ_0 thì φ_0 tương ứng với mỗi điểm ảnh sẽ được cộng thêm một đại lượng $\Delta\varphi$. Tiếp tục áp dụng nguyên lý đồng dạng trong hình học như trong hình 6(c) ta có:

$$\frac{\tan(\varphi_0)}{\tan(\varphi_0 - \Delta\varphi)} = \frac{\Delta D + D}{D} \quad (7)$$

Sử dụng tính đồng nhất của lượng giác cơ bản khoảng cách lỗi được tính như sau:

$$\Delta D = \frac{D^2}{B} \tan(\Delta\varphi) \quad (8)$$

Như vậy, khoảng cách dự đoán D trong công thức 6 biến đổi thành:

$$D = \frac{Bx_0}{2 \tan(\frac{\varphi_0}{2} + \Delta\varphi)(x_1 - x_2)} \quad (9)$$

Từ đây công thức 9 đưa về biểu diễn về dạng hàm mũ như sau:

$$D = k * x^d \quad (10)$$

Trong đó:

- + k là hằng số được tính như sau:

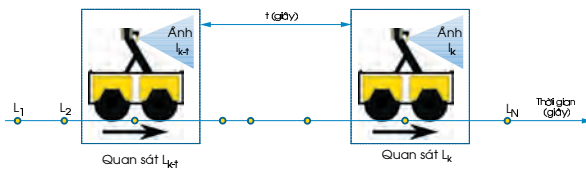
$$k = \frac{Bx_0}{2 \tan(\frac{\varphi_0}{2} + \Delta\varphi)} \quad (11)$$

- + $x = (x_1 - x_2)$, d là hằng số xác định giá trị độ chênh lệch (Disparity) giữa các điểm ảnh từ hai quan sát và được tính toán trên bản đồ chênh lệch của từng điểm ảnh khi hai camera quan sát đối tượng tại các góc khác nhau.

2) *Xây dựng bản đồ chênh lệch*: Như vậy để xác định độ sâu của đối tượng (khoảng cách từ đối tượng tới camera), bản đồ chênh lệch giữa các điểm ảnh trên hai quan sát phải được tính toán. Nhiều thuật toán đã được đề xuất để giải quyết vấn đề này như trong [23], [24] nhằm cải thiện độ chính xác dự đoán khoảng cách.

Trong hệ thống đề xuất, chúng tôi đi theo hướng tiếp cận xác định bản đồ chênh lệch sử dụng 01 camera duy nhất với các bước cơ bản như sau:

- 1) *Thu thập dữ liệu*: thu thập hình ảnh ở các thời điểm khác nhau, chúng tôi định nghĩa hai quan sát mô tả trong hình 7. Trong đó:



Hình 7. Minh họa thu thập dữ liệu khi camera chuyển động

L_k là vị trí quan sát hiện tại; $L_{k-\delta_T}$ là vị trí quan sát trước đó, với δ_T là một khoảng thời gian xác định trước đủ để phân biệt hai ảnh I_k và $I_{k-\delta_T}$.

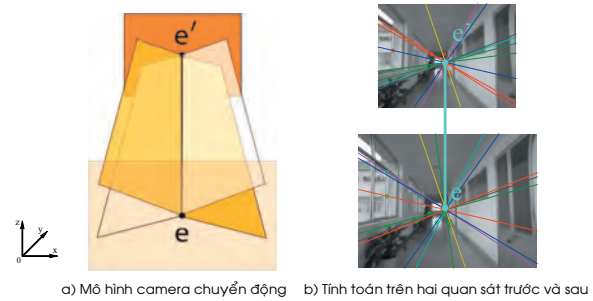
- 2) *Hiệu chỉnh ảnh*: Việc hiệu chỉnh ảnh là rất cần thiết để giảm độ phức tạp tính toán điểm ảnh tương ứng ở hai quan sát. Quá trình hiệu chỉnh gồm gồm có hai bước: (i) tính toán các tham số trong và ngoài của camera; (ii) hiệu chỉnh hình ảnh thu nhận sử dụng các biến đổi tuyến tính xoay, dịch và nghiêng hình ảnh sao cho đường epipolar của hình ảnh liên kết theo chiều ngang.

- **Tìm đường epipolar trên từng ảnh**: các tham số của camera được định nghĩa như sau:

- + Gọi O_F và O_T là tâm chiếu của hai camera, Π_F và Π_T là cặp mặt phẳng ảnh tương ứng.
- + Điểm P trong thế giới thực có một phép chiếu mặt phẳng ảnh Π_F là điểm P_F và mặt phẳng ảnh Π_T là điểm P_T .
- + Điểm e_T gọi là điểm epipole được định nghĩa là ảnh của tâm chiếu O_F lên mặt phẳng ảnh Π_T ; e_F là ảnh của tâm chiếu O_T lên mặt phẳng ảnh Π_F .

- + Đường epipolar l_F và l_T là đường nối giữa hai điểm $e_T P_T$ và $e_F P_F$ nằm trong hai mặt phẳng ảnh.

Chúng tôi thực hiện tìm các đường epipolar trên hai quan sát khi camera chuyển động, kết quả cho thấy các đường epipolar tìm được cắt nhau qua điểm epipole e và e' khi chiếu lên mặt phẳng nằm ngang thì $e \equiv e'$ (hình 8).



Hình 8. Kết quả tìm đường epipolar khi camera chuyển động

- **Hiệu chỉnh liên kết ngang của các epipolar**: là quá trình chiếu hình ảnh trên cùng một mặt phẳng sao cho các đường epipolar của hai điểm song song theo chiều ngang nhằm so sánh giữa hai cặp hình ảnh.

- + *Tính toán ma trận E (Essential matrix)*: xác định mối quan hệ giữa điểm P và hai điểm P_F và P_T từ phép chiếu lên hai mặt phẳng ảnh Π_F và Π_T xác định:

$$P_F = R(P_T - T) \quad (12)$$

Trong đó mặt phẳng ảnh Π_F chứa các vector P_T và T , do đó nếu chọn một vector $(P_T \times T)$ vuông góc với cả hai thì một phương trình cho tất cả các điểm P_T đi qua T và chứa cả hai vector được xác định như sau:

$$(P_T - T)^T (T \times P_T) = 0 \quad (13)$$

Thay $(P_T - T) = R^{-1} P_F$ và $R^T = R^{-1}$ vào công thức 13 ta có

$$(R^T P_F)^T (T \times P_T) = 0 \quad (14)$$

Khi thực hiện phép nhân ma trận thì luôn tồn tại một đường chéo của ma trận kết quả S nhận giá trị 0:

$$T \times P_T = S P_T \quad (15)$$

$$S = \begin{bmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{bmatrix} \quad (16)$$

Các công thức trên được viết lại:

$$P_F^T E P_T = 0 \quad (17)$$

Khi đó ma trận E được tính toán như sau: $E = RS$.

+ *Tính toán ma trận cơ bản F (Fundamental matrix):* Gọi M_T và M_F là tham số trong của camera O_T và O_F , $\overline{P_T}$ và $\overline{P_F}$ là tọa độ của P_T và P_F .

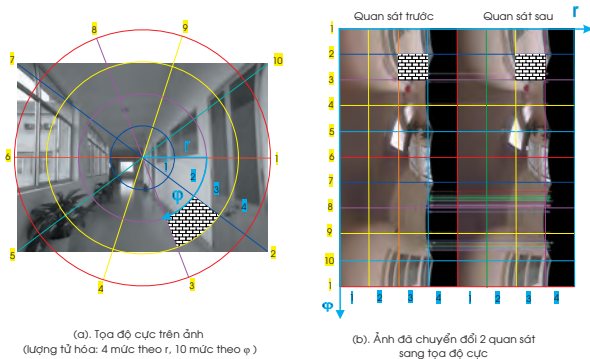
$$\begin{aligned} \overline{P_T} &= M_T P_T \\ \overline{P_F} &= M_F P_F \end{aligned} \quad (18)$$

Áp dụng công thức 17 để triển khai với ma trận F, ta có:

$$\overline{P_F}^T F \overline{P_T} = 0 \quad (19)$$

Khi đó $F = (M_F^{-1})^T E M_T^{-1} = (M_F^{-1})^T R S M_T^{-1}$ là ma trận cơ bản.

Chúng tôi áp dụng kỹ thuật của Pollefeys [25] để hiệu chỉnh hai ảnh thu nhận từ hai quan sát trước và sau khi camera chuyển động. Hình 9(a) minh họa phương pháp chuyển đổi hình ảnh từ tọa độ đề các thông thường về tọa độ cầu sao cho hai điểm epipolar e và e' trùng nhau. Hình 9(b) xoay hai hình ảnh về tọa độ cực sao cho các đường epipolar song song với nhau.



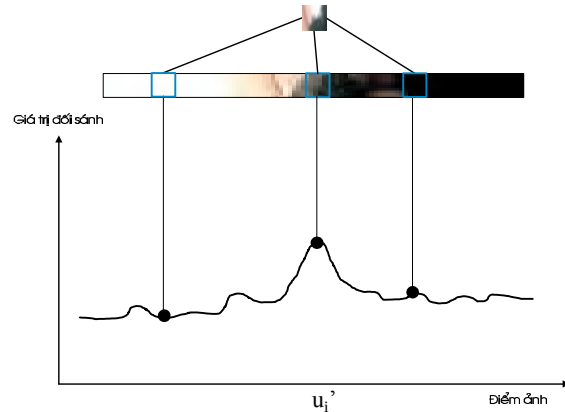
Hình 9. Kết quả hiệu chỉnh hình ảnh

3) **Đối sánh hình ảnh:** Mục đích là tính toán giá trị chênh lệch của một điểm vật lý trên hai ảnh I_T và I_F chúng tôi sử dụng phương

pháp tổng sự khác biệt tuyệt đối SAD [26]:

$$SAD(x, y, d) = \sum_{x, y \in W} |I_T(x, y) - I_F(x, y - d)| \quad (20)$$

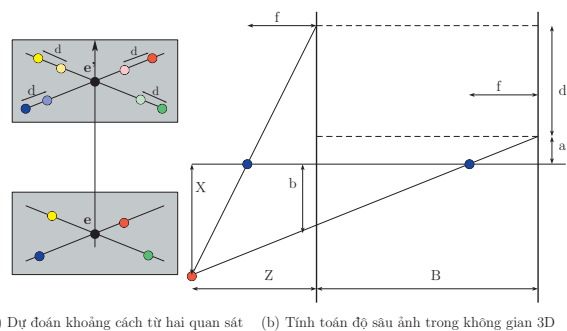
Trong đó: I_T và I_F là hai ảnh đưa vào tính toán; (x, y) tọa độ điểm ảnh; W là cửa sổ quét có kích thước (3×3) , (5×5) , (7×7) ; phạm vi chênh lệch $d < 120$.



Hình 10. Kết quả đối sánh ảnh sử dụng giải thuật SAD

Kết quả của hàm SAD cho biết tổng sự khác biệt của các khối dữ liệu dò tìm trên ảnh thứ hai khi đưa vào tính toán. Hình 10 mô tả quá trình tính toán trượt cửa sổ để tìm ra khối dữ liệu phù hợp, giá trị lớn nhất trong biểu đồ quyết định vị trí đối sánh chính xác.

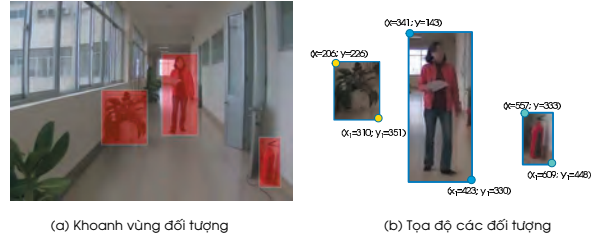
4) **Tính toán độ sâu:** Mục đích của việc này là tìm ra độ sâu của các điểm ảnh trên bản đồ chênh lệch (Disparity map) dựa vào phép đổi hình học để tính toán khoảng cách giữa các điểm ảnh tương ứng trên đường epipoline.



Hình 11. Minh họa phương pháp tính bản đồ chênh lệch

Hình 11(a) mô tả việc dự đoán khoảng cách

từ quan sát e tịnh tiến đến quan sát e' . Các điểm nằm trên đường epipole sẽ dự đoán chính xác khoảng cách d trong tọa độ thế giới thực. Hình 11(b) mô tả chi tiết cách tính toán khoảng cách từ hai quan sát đến vật thể, xuất phát từ cặp tam giác đồng dạng [25].



Hình 13. Minh họa chuẩn bị dữ liệu đánh giá phát hiện

IV. ĐÁNH GIÁ THỬ NGHIỆM

A. Môi trường thử nghiệm và thu thập dữ liệu đánh giá

Môi trường thử nghiệm được tiến hành tại hành lang tầng 10 – Viện MICA – Trường Đại học Bách khoa Hà Nội, tổng chiều dài của hành lang là 60m được mô tả trong hình 12.



Hình 12. Môi trường thử nghiệm robot dẫn đường

Chúng tôi gắn camera trên robot và cho robot di chuyển ở ba vận tốc khác nhau: $v_1 = 100\text{mm/s}$, $v_2 = 200\text{mm/s}$, $v_3 = 300\text{mm/s}$. Lộ trình di chuyển của robot là đi từ A đến D trong hình 12. Tổng số ảnh thu được trong ba lần di chuyển là 2597 khung hình.

1) *Đo khoảng cách từ camera tới vật cản*: Để đánh giá độ sai số ước lượng khoảng cách, khoảng cách thật từ camera tới vật cản phải được đo bằng tay sau này sẽ dùng để đối sánh với kết quả đo tự động. Với dữ liệu thu thập được, chúng tôi tiến hành đo và đánh dấu khoảng cách vị trí vật cản có trong môi trường. Khoảng cách của đối tượng so với gốc tọa độ được xác định.

2) *Chuẩn bị dữ liệu phát hiện đối tượng*: Trên luồng dữ liệu khung cảnh, chúng tôi khoanh vùng các vật cản đã định nghĩa ở trên và lưu vào CSDL biểu diễn môi trường. Quá trình này được thực hiện bằng tay như minh họa trong hình 13.

B. Kết quả đánh giá

1) Độ đo đánh giá:

+ *Phát hiện vật cản*: sử dụng độ triệu hồi (Recall) và độ chính xác (Precision) được định nghĩa trong công thức 21 và 22 để đánh giá hiệu năng phát hiện vật cản.

$$\text{Chính xác (Precision)} = \frac{tp}{tp + fp} \quad (21)$$

$$\text{Triệu hồi (Recall)} = \frac{tp}{tp + fn} \quad (22)$$

tp gọi là một phát hiện được coi là đúng nếu hệ số Jaccard Index [27] $JI \geq 0.5$, hệ số này được tính bởi tỷ lệ giữa vùng giao trên vùng hợp của hình chữ nhật phát hiện được bằng giải thuật B_p và vùng chữ nhật chứa đối tượng được xác định bằng tay B_{gt} .

$$JI = \frac{\text{area}(B_p \cap B_{gt})}{\text{area}(B_p \cup B_{gt})} \quad (23)$$

Ngược lại fp là một phát hiện sai nếu như $JI < 0.5$ và fn không phát hiện được đối tượng.

Thuật toán phát hiện đối tượng được cài đặt lên Robot PCBOT914 cấu hình (CHIP Intel(R) Core(TM)2 T7200@ 2.00 GHz x 2, RAM 8GB), kích thước trung bình của ảnh 640×480 điểm ảnh, tốc độ lấy mẫu 1Hz.

+ *Ước lượng khoảng cách vật cản*: sử dụng độ đo sai số tiêu chuẩn (RMSE).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\theta_i - \hat{\theta})^2} \quad (24)$$

Trong đó $\hat{\theta}$ là khoảng cách đo thực địa tới vật cản; θ là khoảng cách dự đoán trên bản đồ chênh lệch.

2) *Kết quả đánh giá phát hiện vật cản:* Bảng I trình bày chi tiết đánh giá phát hiện vật cản của phương pháp đề xuất. Trên thực tế, kết quả phát hiện vật cản tĩnh phụ thuộc vào kết quả đối sánh ảnh trong CSDL với ảnh hiện tại. Nếu như điều kiện (thời điểm, chiều sáng) thu thập dữ liệu để biểu diễn môi trường gần với điều kiện thử nghiệm thì kết quả sẽ tốt nhất. Trong thí nghiệm, CSDL xây dựng để biểu diễn môi trường là vào buổi sáng, vì vậy giải thuật đạt được độ triệu hồi và độ chính xác cao nhất.

Bảng I
KẾT QUẢ PHÁT HIỆN ĐỐI TƯỢNG PHƯƠNG PHÁP ĐỀ XUẤT

Tên lớp	Recall(%)	Precision(%)	Times(s)
Chậu hoa	98.30	90.23	0.47
Bình cứu hỏa	94.59	89.42	
Thùng rác	85.71	92.31	
Người	92.72	89.74	

Với việc phát hiện vật cản động, hiện tại mô đun sử dụng dữ liệu huấn luyện được cung cấp bởi tác giả của thuật toán gốc trên OpenCV, vì vậy kết quả phát hiện người cũng có bị ảnh hưởng. Sau này để cải thiện hiệu năng của giải thuật phát hiện vật cản tĩnh, cần tính toán đến các yếu tố ảnh hưởng này và có thể thực hiện huấn luyện lại bộ phát hiện người với dữ liệu thu thập được trong thời gian tới.

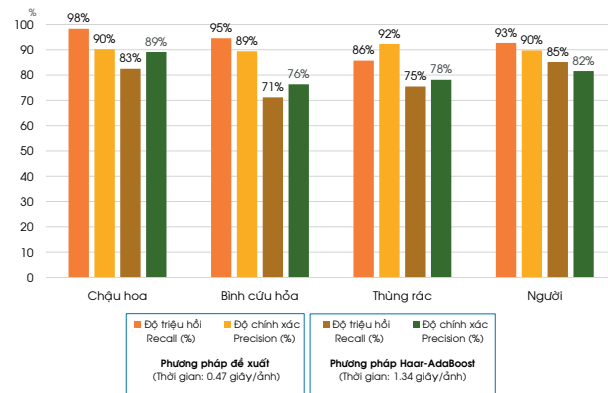
Tiếp theo, chúng tôi đánh giá khả năng phát hiện với phương pháp Haarlike-Adaboost [28]

Bảng II
KẾT QUẢ PHÁT HIỆN ĐỐI TƯỢNG PHƯƠNG PHÁP HAAR-ADABOOST

Tên lớp	Recall (%)	Precision (%)	Times(s)
Chậu hoa	82.52	89.15	1.34
Bình cứu hỏa	71.22	76.37	
Thùng rác	75.49	78.15	
Người	85.16	81.61	

Bảng II trình bày kết quả đánh giá phát hiện các lớp đối tượng. Độ chính xác đạt $Precision = 78.60\%$, độ triệu hồi đạt $Recall = 81.32\%$, thời gian $t = 1.34s$. Biểu đồ đây minh họa so sánh 2 phương pháp đề xuất phát hiện vật cản:

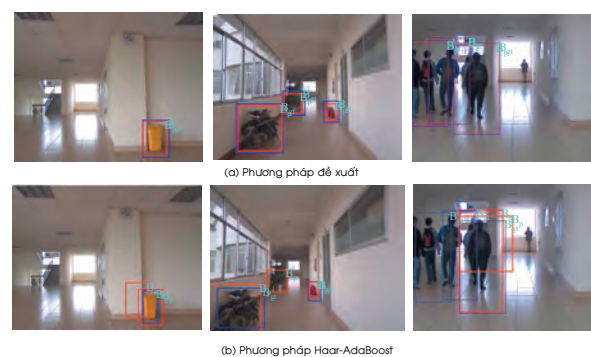
Hình 14 minh họa kết quả đánh giá so sánh 2 phương pháp, trong đó các lớp đối tượng của phương pháp đề xuất cao hơn so với phương pháp Haar-AdaBoost và thời gian tính toán là



Hình 14. Biểu đồ so sánh hai phương pháp phát hiện đối tượng

nhanh hơn.

Một số kết quả phát hiện đối tượng, trong đó hình 15(a) minh họa giải thuật phát hiện vật cản đề xuất, trong đó hình chữ nhật màu xanh được khoanh vùng và đánh dấu từ trước chông khít lên hình chữ nhật màu đỏ là kết quả của giải thuật phát hiện trả về. Hình 15(b) minh họa kết quả phương pháp phát hiện vật cản sử dụng Haar-AdaBoost, trong đó tại một số khung hình xuất hiện trường hợp phát hiện nhầm.



Hình 15. Một số hình ảnh minh họa phát hiện đối tượng

3) *Kết quả đánh giá ước lượng khoảng cách:* kết quả chi tiết có trong bảng dưới đây:

Đối với vật cản động (người) sai số tiêu chuẩn $RMSE \sim 0.4m$ giải thuật HoG-SVM phát hiện người phát huy hiệu quả nhất ở khoảng cách từ vị trí $[9.85m \rightarrow 13.96m]$ so với gốc tọa độ. Ngoài vùng quan sát này như quá gần hoặc quá xa không phát hiện được.

Đối với các vật cản tĩnh lớp chậu hoa cho kết quả tốt hơn các lớp vật cản khác, lớp bình cứu

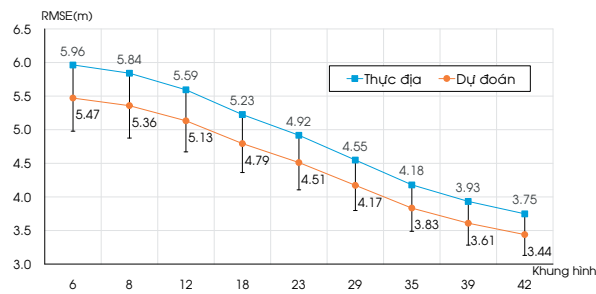
Bảng III
KẾT QUẢ ĐÁNH GIÁ SAI SỐ ƯỚC LƯỢNG KHOẢNG CÁCH
VẬT CẢN

Tên lớp	Sai số tiêu chuẩn RMSE(m)	Khoảng cách phát hiện(m)
Chậu hoa	0.41	2.22
Bình cứu hỏa	0.65	3.75
Thùng rác	0.47	4.04
Người	0.44	4.12

hỏa có độ sai số lớn nhất vì số lượng số điểm đặc trưng được đối sánh giữa hai ảnh chưa đủ quyết định vùng chứa đối tượng, đối với lớp thùng rác cho kết quả trung bình.

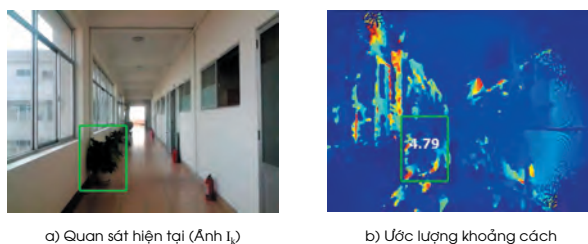
Phần dưới đây trình bày chi tiết kết quả đánh giá ước lượng khoảng cách các lớp vật cản.

- *Lớp chậu hoa*: Với khoảng cách 2.22m có 09 khung hình phát hiện được trong tổng số 37 khung hình thu nhận, do vậy trong thực tế robot di chuyển 24cm thì thực hiện 1 lần (vận tốc robot $v = 300\text{mm/s}$).



Hình 16. Vị trí ước lượng khoảng cách thuộc lớp chậu hoa

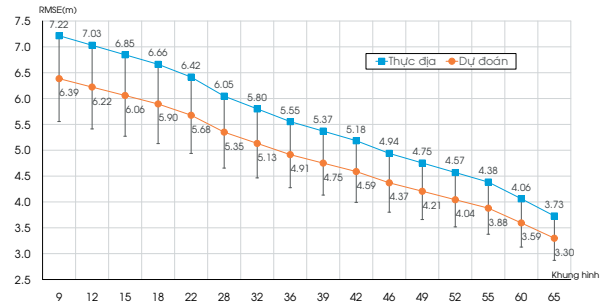
Kết quả phát hiện ước lượng khoảng cách lớp chậu hoa tại khung hình số 18 minh họa hình dưới đây.



Hình 17. Minh họa ước lượng khoảng cách lớp chậu hoa

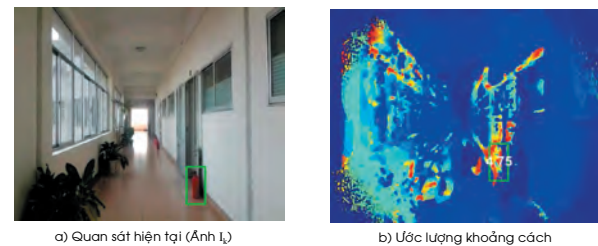
- *Lớp bình cứu hỏa*: với khoảng cách 3.75m robot quan sát được 61 khung hình trong đó có 16 khung hình phát hiện và dự đoán

khoảng cách vật cản. Trong thực tế, robot di chuyển được 26cm thì thực hiện phát hiện và ước lượng khoảng cách một lần.



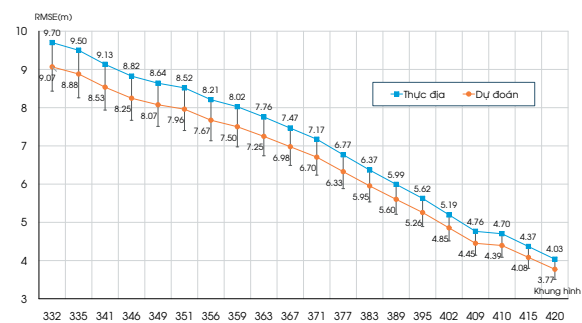
Hình 18. Vị trí ước lượng khoảng cách lớp bình cứu hỏa

Kết quả ước lượng khoảng cách tại khung hình số 39 được minh họa hình 19.



Hình 19. Minh họa ước lượng khoảng cách lớp bình cứu hỏa

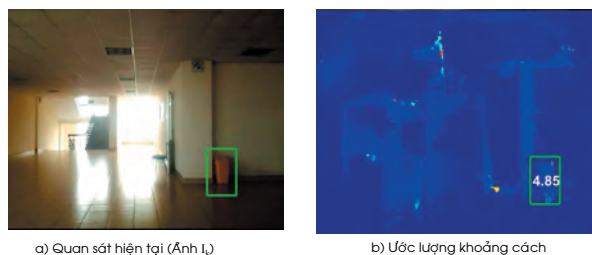
- *Lớp thùng rác*: với khoảng cách 4.04m robot quan sát được 89 khung hình, trong đó có 20 khung hình phát hiện và ước lượng khoảng cách, trong thực tế cứ robot di chuyển 22cm thì thực hiện phát hiện và ước lượng khoảng cách một lần.



Hình 20. Vị trí ước lượng khoảng cách lớp thùng rác

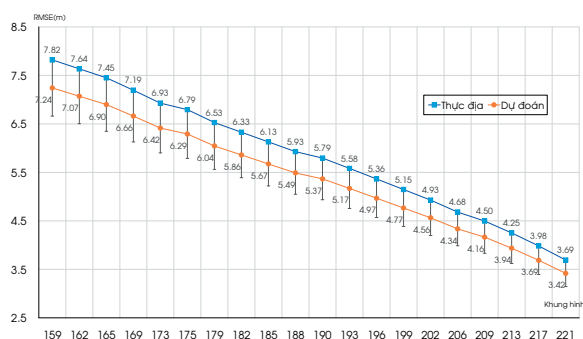
Kết quả ước lượng khoảng cách tại khung hình 409 được minh họa hình 21, trong đó hình 21(a) quan sát hiện tại thu thập hình ảnh

I_k , hình 21(b) ước lượng khoảng cách từ vật cản tới robot.



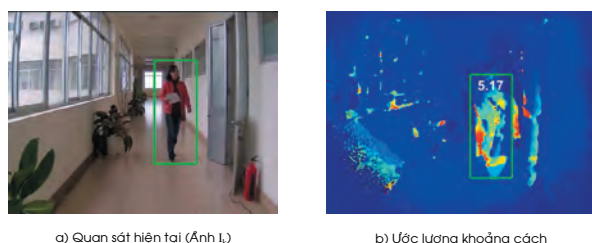
Hình 21. Minh họa ước lượng khoảng cách lớp thùng rác

- *Lớp người*: Ở khoảng cách 4.12m robot quan sát được 63 khung hình trong đó có 20 khung hình phát hiện ước lượng khoảng cách.



Hình 22. Vị trí ước lượng khoảng cách lớp Người

Kết quả ước lượng khoảng cách tại khung hình 193 minh họa ở hình 23, trong đó hình 23(a) quan sát hiện tại thu thập hình ảnh I_k , hình 23(b) ước lượng khoảng cách từ vật cản tới robot.



Hình 23. Minh họa ước lượng khoảng cách lớp Người

V. KẾT LUẬN

Bài báo trình bày một phương pháp phát hiện và ước lượng khoảng cách vật cản dựa vào kỹ thuật xử lý ảnh sử dụng một camera (thông thường) duy nhất. Công việc chính là nghiên cứu các phương

pháp phát hiện nhanh vật cản tĩnh sử dụng phương pháp đối sánh mẫu trên bộ dữ liệu vị trí quan trọng đã được đánh dấu các vị trí huấn luyện từ trước. Kết quả của phần này làm nền tảng để dự đoán khoảng cách trên vùng phát hiện được bằng phương pháp xây dựng bản đồ chênh lệch từ hai quan sát chuyển động tịnh tiến. Phương pháp đề xuất đã được đánh giá là khả thi giúp cảnh báo cho NKT các loại vật cản phía trước, kết quả này cũng góp phần quan trọng trong hệ thống định vị hình ảnh trợ giúp dẫn đường cho NKT sử dụng robot.

LỜI CẢM ƠN

Cảm ơn đề tài “Trợ giúp định hướng người khiếm thị sử dụng công nghệ đa phương thức” mã số: **ZEIN2012RIP19** - Hợp tác quốc tế các trường Đại học tại Việt - Bỉ (VLIR) đã hỗ trợ trong quá trình thực hiện bài báo này.

TÀI LIỆU THAM KHẢO

- [1] Q.-H. Nguyen, H. Vu, T.-H. Tran, and Q.-H. Nguyen, “Developing a way-finding system on mobile robot assisting visually impaired people in an indoor environment,” *Multimedia Tools and Applications*, pp. 1–25, 2016.
- [2] N. Dalal and B. Triggs, *Histograms of oriented gradients for human detection*, 2005, vol. 1.
- [3] J. Kim and Y. Do, “Moving obstacle avoidance of a mobile robot using a single camera,” *Procedia Engineering*, vol. 41, pp. 911–916, 2012.
- [4] T. Taylor, S. Geva, and W. W. Boles, “Monocular vision as a range sensor.” *International Conference on Computational Intelligence for Modelling, Control and Automation*, 2004.
- [5] E. Einhorn, C. Schroeter, and H.-M. Gross, “Monocular obstacle detection for real-world environments,” in *Autonome Mobile Systeme 2009*. Springer, 2009, pp. 33–40.
- [6] L. Nalpantidis, I. Kostavelis, and A. Gasteratos, “Stereovision-based algorithm for obstacle avoidance,” in *Intelligent Robotics and Applications*, 2009, vol. 5928, pp. 195–204.
- [7] M. Bai, Y. Zhuang, and W. Wang, “Stereovision based obstacle detection approach for mobile robot navigation,” in *Intelligent Control and Information Processing (ICICIP), 2010 International Conference on*. IEEE, 2010, pp. 328–333.
- [8] R. A. Hamzah, H. N. Rosly, and S. Hamid, “An obstacle detection and avoidance of a mobile robot with stereo vision camera,” in *Electronic Devices, Systems and Applications (ICEDSA), 2011 International Conference on*. IEEE, 2011, pp. 104–108.
- [9] R. Lagisetty, N. Philip, R. Padhi, and M. Bhat, “Object detection and obstacle avoidance for mobile robot using stereo camera,” in *Control Applications (CCA), 2013 IEEE International Conference on*. IEEE, 2013, pp. 605–610.

- [10] D. S. O. Correa, D. F. Sciotti, M. G. Prado, D. O. Sales, D. F. Wolf, and F. S. Osório, "Mobile robots navigation in indoor environments using kinect sensor," in *Critical Embedded Systems (CBSEC), 2012 Second Brazilian Conference on*. IEEE, 2012, pp. 36–41.
- [11] S. Nissimov, J. Goldberger, and V. Alchanatis, "Obstacle detection in a greenhouse environment using the kinect sensor," *Computers and Electronics in Agriculture*, vol. 113, pp. 104–115, 2015.
- [12] B. Peasley and S. Birchfield, "Real-time obstacle detection and avoidance in the presence of specular surfaces using an active 3d sensor," in *Robot Vision (WORV), 2013 IEEE Workshop on*. IEEE, 2013, pp. 197–202.
- [13] Q.-H. Nguyen, H. Vu, T.-H. Tran, and Q.-H. Nguyen, "A vision-based system supports mapping services for visually impaired people in indoor environments," in *Control Automation Robotics & Vision (ICARCV), 2014 13th International Conference on*. IEEE, 2014, pp. 1518–1523.
- [14] M. Muja and D. G. Lowe, "Scalable nearest neighbor algorithms for high dimensional data," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 36, 2014.
- [15] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [16] P. S. Heckbert, "Fundamentals of texture mapping and image warping," Master's thesis, University of California, 1989.
- [17] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.
- [18] R. C. Browning, E. A. Baker, J. A. Herron, and R. Kram, "Effects of obesity and sex on the energetic cost and preferred speed of walking," *Journal of Applied Physiology*, vol. 100, no. 2, pp. 390–398, 2006.
- [19] T. Santhanam, C. Sumathi, and S. Gomathi, "A survey of techniques for human detection in static images," in *Proceedings of the Second International Conference on Computational Science, Engineering and Information Technology*, 2012, pp. 328–336.
- [20] M. A. Mahammed, A. I. Melhum, and F. A. Kochery, "Object distance measurement by stereo vision," *International Journal of Science and Applied Information Technology (IJSAIT) Vol*, vol. 2, pp. 05–08, 2013.
- [21] J. Mrovlje and D. Vrancic, "Distance measuring based on stereoscopic pictures," in *9th International PhD Workshop on Systems and Control, Young Generation Viewpoint. Izola, Slovenia*, 2008.
- [22] A. J. Woods, T. Docherty, and R. Koch, "Image distortions in stereoscopic video systems," in *IS&T/SPIE's Symposium on Electronic Imaging: Science and Technology*. International Society for Optics and Photonics, 1993, pp. 36–48.
- [23] A. Coste, "3d computer vision-stereo and 3d reconstruction from disparity," Technical report, Tech. Rep., 2013.
- [24] Y.-J. Zhang, *Advances in image and video segmentation*. IGI Global, 2006.
- [25] M. Pollefeys, R. Koch, and L. Van Gool, "A simple and efficient rectification method for general motion," in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, vol. 1. IEEE, 1999, pp. 496–501.
- [26] P. Kamencay, M. Breznán, R. Jarina, P. Lukac, and M. Zachariasova, "Improved depth map estimation from stereo images based on hybrid method," *Radioengineering*, vol. 21, no. 1, pp. 70–78, 2012.
- [27] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International journal of computer vision*, vol. 88, no. 2, pp. 303–338, 2010.
- [28] P. Viola and M. J. Jones, "Robust real-time face detection," *International journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004.

OBSTACLE DETECTION AND DISTANCE ESTIMATION USING MONOCULAR CAMERA IN NAVIGATION SERVICES FOR VISUALLY IMPAIRED PEOPLE

Abstract - In this paper, we propose a method for obstacle detection and distance estimation using monocular camera mounted on a mobile robot. The proposed system aims to support visually impaired people navigating in indoor environment. The obstacles include static and dynamic objects on that encumber human mobility. For static objects, supporting information such as type of object, positions, and corresponding images in relevant scenes are stored in database (DB). To detect them, the images captured during robot's movements are compared with the corresponding images through a localization algorithm proposed in [1]. Then the existing objects in DB will be identified and distances from them to current robot's position is estimated. For dynamic objects, such as movements of people in scenes, we use HOG-SVM algorithm [2]. To estimate distance from camera to detected obstacles, we utilize a disparity map which is built from consecutive frames. The experiments are evaluated in the hall of building floor of 60 meters under different lighting conditions. The results confirm that the proposed method could exactly detect and estimate both static and dynamic objects. This shows the feasibility to help visually impaired people avoiding obstacles in navigation.



Nguyễn Quốc Hùng nhận bằng thạc sĩ ngành Khoa học máy tính tại Đại học Thái Nguyên năm 2010, hiện là nghiên cứu viên của Viện nghiên cứu quốc tế MICA. Hướng nghiên cứu hiện tại là: phân tích và nhận dạng hình ảnh/video; điều khiển robot di động; định vị, xây dựng bản đồ môi trường, tương tác của người-robot.



Vũ Hải nhận bằng tiến sĩ ngành công nghệ thông tin tại Trường đại học Osaka - Nhật Bản năm 2009. Hiện đang là giảng viên, nghiên cứu viên tại Viện nghiên cứu quốc tế MICA, Trường Đại học Bách Khoa Hà Nội. Hướng nghiên cứu hiện tại phân tích nhận dạng hình ảnh/video, xử lý ảnh Y tế, điều khiển mạng camera, kinect.



Trần Thị Thanh Hải nhận bằng tiến sĩ ngành công nghệ thông tin tại Pháp năm 2006. Hiện đang là giảng viên, nghiên cứu viên tại Viện nghiên cứu quốc tế MICA, Trường Đại học Bách Khoa Hà Nội. Hướng nghiên cứu hiện tại phân tích và nhận dạng hình ảnh/video, tương tác người - máy.



Nguyễn Quang Hoan nhận bằng tiến sĩ ngành hệ thống thông tin tại Liên Xô năm 1973. Hiện đang là giảng viên tại khoa công nghệ thông tin - Đại học sư phạm Hưng Yên. Hướng nghiên cứu hiện tại mạng nơ ron, mờ trong điều khiển.



Hoàng Văn Nam nhận bằng kỹ sư ngành Kỹ thuật điều khiển và tự động hóa tại Đại học Bách Khoa Hà Nội năm 2014. Hiện tại anh đang học thạc sĩ chuyên ngành Môi trường cảm thụ, đa phương tiện và tương tác tại viện nghiên cứu quốc tế MICA, Đại học Bách Khoa Hà Nội. Hướng nghiên cứu hiện tại là: xử lý ảnh, học máy.