

Xác định hướng của người cho hệ thống Robot di động hỗ trợ người khiếm thị di chuyển trong nhà

Human Direction Determination for Visually Impaired Support Indoor Mobile Robot

Tổng Thị Lý¹, Lương Thanh Tuấn², Trần Minh Quang², Hoàng Kim Nga², Đỗ Trọng Hiếu^{2*},
 Dương Minh Đức²

¹Đại học Công nghiệp Hà Nội,

²Đại học Bách khoa Hà Nội

*Corresponding author's email: hieu.dotrong@hust.edu.vn

Abstract

Independent mobility is an essential part of the daily life of many visually impaired individuals. To assist them in navigation and improve their quality of life, various designs have been developed, including portable devices such as electronic canes, navigation suitcases, and guided robotic dogs. However, when it comes to indoor environments, constantly carrying an object alongside is not practical. Therefore, there have been numerous studies on mobile robot systems for navigation and assistance in indoor environments. One of the requirements for such a non-contact robot system is to determine the direction of the visually impaired individual to provide accurate assistance. This article focuses on developing algorithms for determining the direction of the object that needs assistance in indoor environments. The object's information is collected using a 3D camera, and then techniques such as object tracking, skeleton framework construction, and Gaussian filtering of depth images are used to calculate the object's direction angle. The proposed algorithm was tested with a representative sample of visually impaired individuals in a laboratory setting. The results showed high accuracy for static objects and acceptable performance for dynamic objects. This forms the basis for applying this algorithm to mobile robot systems that assist visually impaired individuals.

Keywords: Visual Impaired, Mobile Robot, Object Tracking, Pose Estimation, Gauss Filter.

Tóm tắt

Việc di chuyển độc lập là một phần không thể thiếu trong cuộc sống hàng ngày của nhiều người khiếm thị. Để giúp người khiếm thị điều hướng, cải thiện chất lượng cuộc sống, đã có những thiết kế không chỉ ở dạng mang theo như gậy điện tử và vali dẫn đường, chó robot. Tuy nhiên, với môi trường trong nhà, việc luôn phải mang theo một vật bên cạnh sẽ thực sự không phù hợp, vì thế đã có nhiều nghiên cứu về các rô bốt di động điều hướng và hỗ trợ họ trong môi trường trong nhà. Một trong những yêu cầu của hệ thống robot không tiếp xúc này là cần xác định hướng của người khiếm thị để ra được các hỗ trợ chính xác. Bài báo tập trung vào việc phát triển các thuật toán xác định hướng đối tượng cần hỗ trợ di chuyển trong môi trường trong nhà. Các thông tin của đối tượng được thu thập bằng camera 3D, sau đó các kỹ thuật bám đối tượng, xây dựng khung xương kết hợp với bộ lọc Gauss lọc ảnh độ sâu được sử dụng để tính toán ra góc hướng của đối tượng. Thuật toán đề xuất đã được thử nghiệm với tập mẫu đại diện cho người khiếm thị trong phòng thí nghiệm. Các kết quả thu được cho thấy thuật toán đem lại độ chính xác cao với đối tượng tĩnh và chấp nhận được với đối tượng động. Đây là cơ sở để cho phép ứng dụng thuật này cho hệ thống robot di động trợ giúp người khiếm thị.

1. Giới thiệu

Những người bị khuyết tật thị giác, mất khả năng cảm nhận thị giác một phần hoặc toàn phần luôn đối mặt với những thách thức lớn trong việc di chuyển và tương tác trong môi trường xung quanh. Theo Tổ chức y tế thế giới (WHO), trên toàn cầu hiện có khoảng 314 triệu người mù hoặc có thị lực yếu, trong đó có khoảng 45 triệu người mù. Người trên 50 tuổi chiếm tỷ lệ 80% trong số này, và có tới 90% người mù sống ở các nước nghèo và đang phát triển, nơi mà việc tiếp cận dịch vụ y tế gặp khó khăn [1]. Do đó, vấn đề hỗ trợ người khiếm thị trong các công việc di chuyển sinh hoạt hàng ngày luôn thu hút các nỗ lực của các nhà khoa học. Đã có nhiều nghiên cứu đã phát triển các sản phẩm hỗ trợ để giúp người khiếm thị trong quá trình di chuyển, chủ yếu tập trung vào việc thu thập thông tin về môi trường xung quanh và hỗ trợ người khiếm thị trong việc ra quyết định di chuyển và tránh va chạm với các vật thể [2]. Các công cụ này thường được phân loại thành hai nhóm chính là "định hướng" hay còn gọi là "tìm đường", và nhóm phát hiện và tránh các chướng ngại vật gần đó gọi là "di chuyển" [3]. Các công cụ này có thể sử dụng công nghệ định vị như GPS để xác định vị trí hiện tại của người khiếm thị và

cung cấp hướng dẫn chi tiết để đi đến được đích đến mong muốn. Các công cụ xử lý ảnh cũng được sử dụng rất rộng rãi trong việc nhận biết môi trường xung quanh để từ đó đưa ra các khuyến cáo hay chỉ dẫn cho người khiếm thị [4], [5].

Trong nhiều thập kỷ qua, gậy dò đường và chó dẫn đường là những cách thức rất phổ biến mà người mù đã sử dụng để có thể tự đi lại trong cuộc sống hàng ngày. Vì lý do đó, các nghiên cứu cho người khiếm thị cũng tập trung vào xây dựng các loại gậy dẫn đường điện tử hoặc robot chó dẫn đường [6-8]. Gậy dẫn đường cung cấp cho người dùng thông tin về môi trường xung quanh thông qua giác quan xúc giác, giúp họ khám phá và phát hiện các chướng ngại vật trên mặt đất như đồ đạc, bậc thang, ...[6], [7]. Chó dẫn đường có khả năng hướng dẫn con người đi qua các không gian hẹp và lộn xộn, mang lại lợi ích lớn cho người khiếm thị [8]. Các sản phẩm trên hầu hết chỉ xét đến tương tác một chiều giữa đối tượng và robot, nghĩa là robot chủ yếu đo lường và thu thập các thông tin môi trường mà không thu thập thông tin của đối tượng. Thêm vào đó, các nghiên cứu và sản phẩm hiện có đối với người khiếm thị trong việc di chuyển thường yêu cầu người dùng phải tương tác với thiết bị đó như mang theo các thiết bị cầm nắm các thiết bị. Điều này tạo ra khó khăn cho người khiếm thị trong cuộc sống hàng ngày, đặc biệt là trong môi trường trong nhà [9], [10]. Thêm vào đó, các đối tượng khác khi di chuyển trong cùng một không gian với người khiếm thị cũng có thể bị cản trở bởi các thiết bị này.

Robot di động với các trang bị cảm biến cần thiết ngày nay có thể di chuyển mọi vị trí trong môi trường hoạt động của nó, tự xác định vị trí và lập kế hoạch đi giữa các vị trí bằng thuật toán SLAM (Simultaneous Localization And Mapping) [11]. Robot di động có thể thu thập thông tin không chỉ về môi trường xung quanh mà còn về các đối tượng, từ đó cung cấp các chỉ dẫn nhanh chóng và chính xác, hướng dẫn người khiếm thị điều hướng và tương tác với các vật thể trong khoảng cách nhỏ, đáp ứng một cách linh hoạt. Robot di động vì vậy có thể cung cấp hướng dẫn chi tiết cho người khiếm thị, bao gồm thông báo khi cần rẽ trái hoặc rẽ phải, và đưa ra chỉ dẫn đến các địa điểm cụ thể [12]. Các kết quả nghiên cứu và thử nghiệm đã cho thấy robot điều hướng di động có khả năng hỗ trợ hiệu quả cho việc di chuyển độc lập của người khiếm thị trong môi trường thực tế [13]. Tuy nhiên, tương tự như gậy điện tử và chó robot, người dùng cần tương tác trực tiếp với robot di động để nhận được sự trợ giúp. Điều này trong nhiều trường hợp gây chiếm dụng không gian và cản trở sự di chuyển của các đối tượng khác trong cùng một không gian hoạt động.

Để khắc phục nhược điểm nói trên của hệ thống robot di động hỗ trợ người khiếm thị, một hệ thống không yêu cầu tiếp xúc trực tiếp giữa robot và người dùng cần được phát triển. Và trong hệ thống này tương tác hai chiều giữa robot và người khiếm thị cần phải được thiết lập. Robot hướng dẫn người di chuyển đồng thời phải nắm được phương hướng di chuyển thực tế của người khiếm thị.

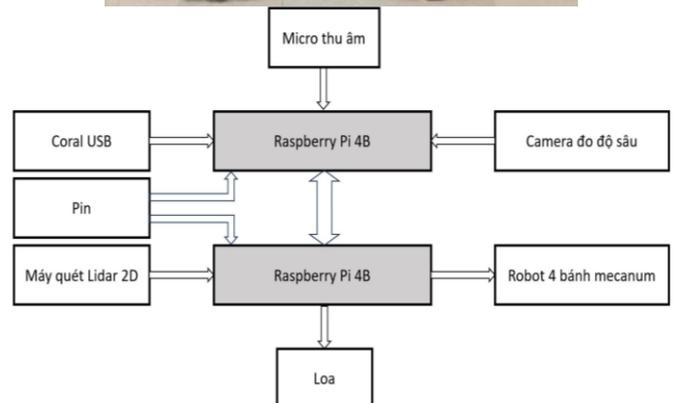
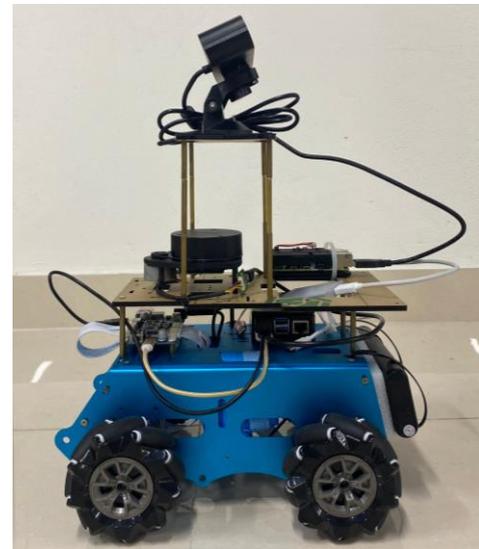
Bài báo này tập trung vào việc xác định vị trí và phương hướng di chuyển thực tế của người khiếm thị sử dụng camera 3D và các kỹ thuật xử lý ảnh. Thêm vào đó, để nâng cao chất lượng tín hiệu phản hồi hình ảnh và hướng của người khiếm thị, nhóm nghiên cứu đã sử dụng một phương pháp kết hợp giữa bộ lọc Gauss cho ảnh độ sâu và thuật toán ước tính hướng di chuyển. Việc sử dụng bộ lọc Gauss giúp tăng tính đồng nhất của tín hiệu độ sâu đồng thời giảm thiểu nhiễu và thông tin

không cần thiết. Tiếp theo, thuật toán ước tính hướng di chuyển được áp dụng để xác định hướng của người khiếm thị phản hồi trạng thái đáp ứng của đối tượng với các lệnh chỉ dẫn yêu cầu từ robot. Nghiên cứu cũng đã tiến hành các thử nghiệm và đánh giá hiệu quả của phương pháp được đề xuất trên một số tập mẫu đại diện cho người khiếm thị tại phòng thí nghiệm.

Bài báo được cấu trúc thành các phần chính như sau: Phần một trình bày tổng quan về nội dung nghiên cứu, giới thiệu mục tiêu và phạm vi của nghiên cứu. Phần hai trình bày cấu trúc phần cứng và phần mềm cho robot di động. Phần ba trình bày thuật toán xác định hướng của đối tượng. Phần bốn trình bày các kết quả thực nghiệm với tập mẫu. Cuối cùng, phần năm là kết luận, tổng hợp lại các kết quả và nhấn mạnh đóng góp và hạn chế của nghiên cứu, đồng thời đề xuất hướng phát triển trong tương lai.

2. Hệ thống robot trợ giúp người khiếm thị

2.1. Cấu hình phần cứng



Hình 1. Hình ảnh và cấu trúc phần cứng của Mobile Robot.

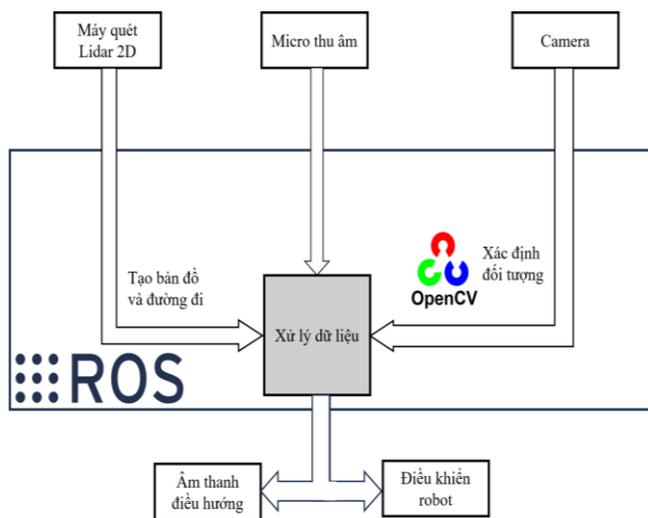
Robot di động được phát triển trong nghiên cứu này là robot di động 4 bánh mecanum với cấu hình bao gồm các thiết bị chính như Hình 1. Trong đó, phần cứng được sử dụng để phản hồi nhận diện đối tượng là camera 3D có khả năng chụp ảnh và xác định độ sâu của ảnh được phát triển bởi công ty Orbbec. Ngoài ra các thiết bị khác như 4 động cơ cùng driver được trang bị cho 4 bánh mecanum cho phép robot có thể di chuyển trong môi trường làm việc của robot. Cảm biến lidar 2D cho

phép thu thập thông tin khoảng cách từ robot tới các điểm trong môi trường xung quanh robot. Các thông tin này sẽ được sử dụng để thực hiện các thuật toán SLAM cho robot. Robot được điều khiển bởi bộ điều khiển trung tâm là 02 máy tính Raspberry Pi IV. Ngoài ra Coral USB được sử dụng để tăng tốc các tác vụ xử lý ảnh, đảm bảo các tác vụ xử lý ảnh có thể thực hiện theo thời gian thực. Các thiết bị như loa hay micro cũng được trang bị cho robot cho phép người khiếm thị và robot có thể giao tiếp với nhau qua phương tiện âm thanh.

2.2. Nền tảng phần mềm

Trong nghiên cứu này, ROS (Robot Operating System) được sử dụng để nghiên cứu môi trường và không gian làm việc của robot [14]. ROS là một phần mềm mã nguồn mở phổ biến trong lĩnh vực robot và tự động hóa, cung cấp các công cụ, thư viện và hệ thống để phát triển, kiểm thử và điều khiển robot. Các gói phần mềm SLAM có sẵn trong ROS [15] được sử dụng để xác định vị trí của robot, cũng như xây dựng bản đồ của môi trường xung quanh. Thông tin này rất quan trọng trong việc điều hướng và điều khiển robot trong môi trường không gian phức tạp.

Trong nghiên cứu này, một thư viện mã nguồn mở khác là OpenCV [16] cũng được sử dụng cho xử lý ảnh và thị giác máy tính. Mô hình MoveNet hiệu suất cao [17] được sử dụng để phát hiện và theo dõi các điểm đặc trưng (keypoint) trên cơ thể người.



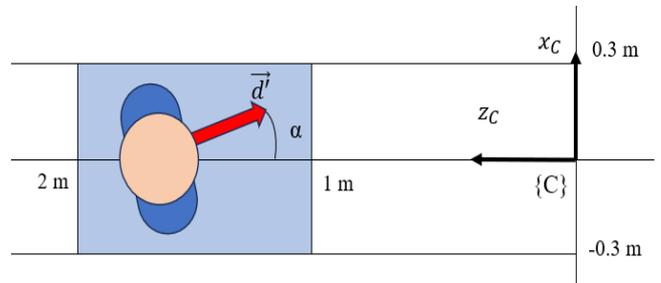
Hình 2. Nền tảng phần mềm.

2.3. Nguyên lý làm việc của hệ thống robot

Robot di động sẽ di chuyển trong môi trường hoạt động để dựng lên bản đồ của môi trường hoạt động. Khi được yêu cầu đưa người khiếm thị đến một vị trí nào đó trong môi trường hoạt động, robot sẽ xây dựng quỹ đạo di chuyển từ vị trí hiện tại đến vị trí đích và bắt đầu di chuyển đồng thời hướng dẫn hay điều hướng người khiếm thị đi theo quỹ đạo đã được xây dựng. Việc điều hướng người khiếm thị sẽ được thực hiện bằng việc phát ra các chỉ thị di chuyển thích hợp dưới dạng âm thanh (đi thẳng, dừng, lệch trái, lệch phải, quay trái, quay phải) dựa vào vị trí tương đối giữa người và robot và quan trọng nhất là hướng di chuyển hiện tại của người khiếm thị. Để điều hướng đối tượng, việc xác định thông tin về hướng và vị trí hiện tại của đối tượng so với robot là cần thiết và lệnh

điều hướng phụ thuộc vào thông tin này của đối tượng. Trong nghiên cứu này, không gian làm việc được chia làm 3 vùng như trong Hình 3:

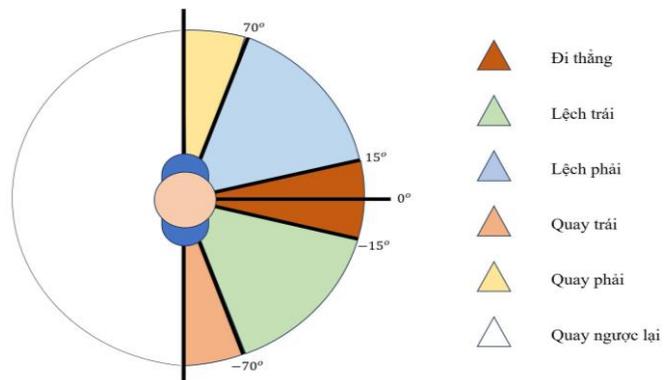
- Vùng 1: Khoảng cách giữa người và robot <1m, robot sẽ ra tín hiệu cho người dừng lại, robot tiếp tục đi
- Vùng 2: Khoảng cách giữa người và robot >2m, robot sẽ dừng lại, người đi tiếp
- Vùng 3: Trong phạm vi 1-2m tính từ camera {C} người sẽ được điều hướng.



Hình 3. Vùng điều hướng.

Trong hình 3, $\{C x_c y_c z_c\}$ là hệ tọa độ của robot (camera), $\vec{d'}$ là véc tơ hình chiếu của véc tơ hướng của đối tượng lên mặt phẳng XZ, và α là góc hướng của đối tượng.

Khi đối tượng ở trong vùng điều hướng thì các chỉ thị định hướng sẽ dựa vào hướng của đối tượng như trong Hình 4.



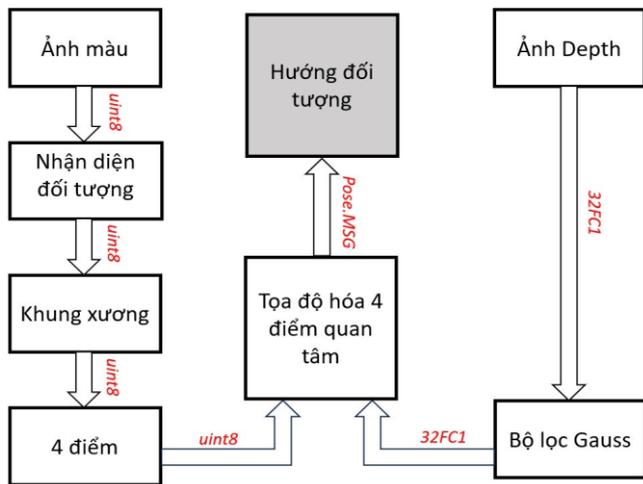
Hình 4. Vùng điều hướng.

Như vậy để robot ra chỉ thị điều hướng một đối tượng, robot cần thu thập thông tin về vị trí và hướng của đối tượng đó. Thông qua việc tính toán chính xác góc quay của người, robot có thể điều hướng một cách chính xác và nhận biết mức độ tuân thủ các lệnh điều hướng từ đối tượng.

3. Thuật toán xác định hướng của người so với robot

Quá trình xác định hướng di chuyển của đối tượng bao gồm việc phát hiện và nhận dạng hình dạng đối tượng (người) trong hình ảnh hoặc video. Khi vị trí đã được xác định trong khung hình, bước tiếp theo là xác định các điểm đặc trưng trên cơ thể người để từ đó có thể tính toán hướng di chuyển. Để xác định được tọa độ của các điểm đặc trưng trong hệ tọa độ robot, ta cần kết hợp ảnh từ camera thường và ảnh từ camera chiều sâu. Bộ lọc Gauss được đề xuất để cải thiện chất lượng xử lý ảnh chiều sâu. Bằng cách theo dõi sự thay đổi vị trí của các điểm đặc trưng này theo thời gian, chúng ta có thể tính toán và xác

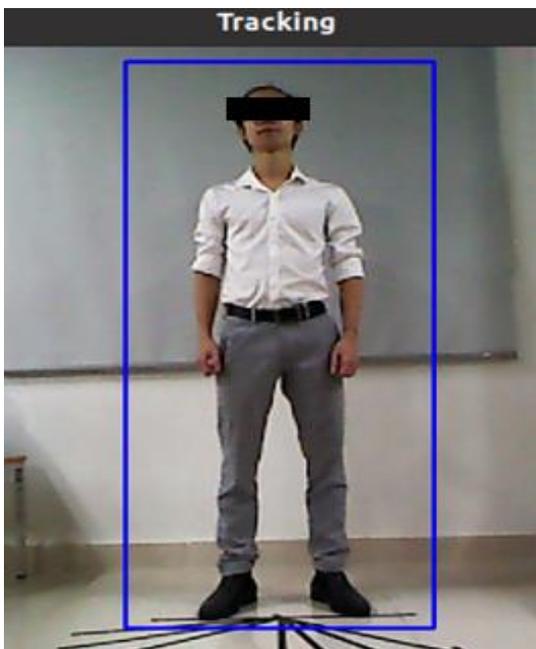
định hướng di chuyển của người. Cấu trúc quá trình xác định hướng được đề xuất như trong Hình 5.



Hình 5. Sơ đồ cấu trúc quá trình xác định hướng đề xuất.

3.1. Thuật toán nhận diện và bám đối tượng

Trong nghiên cứu này sử dụng thuật toán MOSSE (Minimum Output Sum of Squared Error) là một phương pháp để theo dõi các đối tượng trong video [18]. Thuật toán này tập trung vào việc xác định vị trí của một đối tượng bằng cách sử dụng bộ lọc tuyến tính. Ý tưởng cơ bản của thuật toán áp dụng là huấn luyện một bộ lọc tối ưu để phản ánh đặc điểm cụ thể của đối tượng mục tiêu

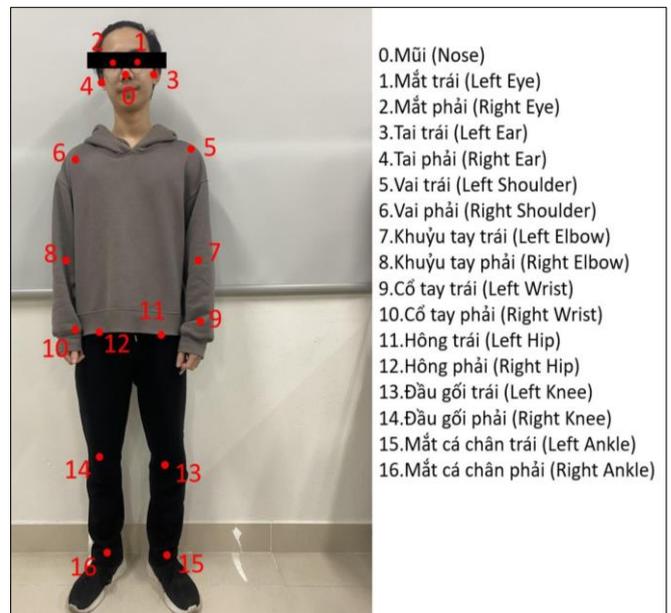


Hình 6. Xác định đối tượng bằng thuật toán MOSSE.

Kết quả thực nghiệm áp dụng MOSSE cho đối tượng thí nghiệm như Hình 6. Để làm điều này, thuật toán sử dụng kỹ thuật tối ưu hóa để điều chỉnh các trọng số của bộ lọc sao cho tổng bình phương sai số đầu ra là nhỏ nhất khi áp dụng bộ lọc đó lên các khung hình trong video. Một ưu điểm của việc áp dụng thuật toán MOSSE là tốc độ xử lý nhanh do sử dụng bộ lọc tuyến tính và kỹ thuật tối ưu hoá hiệu quả.

3.2. Thuật toán xác định khung xương đối tượng

Để xác định khung xương của đối tượng, nghiên cứu sử dụng MoveNet, một trong những thuật toán xác định tư thế (pose estimation) tiên tiến sử dụng deep learning, được phát triển bởi Google Research [19]. MoveNet được thiết kế để ước lượng vị trí của các điểm quan trọng (keypoint) trên cơ thể người từ dữ liệu video hoặc hình ảnh. MoveNet sử dụng mô hình neural network để dự đoán vị trí của các điểm quan trọng trên cơ thể như mắt, mũi, đầu vai, cổ tay, và đầu gối. Thay vì sử dụng các mô hình phức tạp, MoveNet tập trung vào việc tối ưu hóa hiệu suất và tốc độ, đặc biệt là khi được triển khai trên các thiết bị có tài nguyên hạn chế như điện thoại di động. Kết quả sử dụng MoveNet cho đối tượng thí nghiệm như Hình 7, cho ta 17 điểm quan trọng trên cơ thể người. Một trong những điểm mạnh của MoveNet là khả năng hoạt động với tốc độ nhanh và hiệu suất tốt trong việc ước lượng tư thế trên video thời gian thực hoặc ảnh tĩnh. Tuy nhiên, như các thuật toán xác định tư thế khác, MoveNet có thể gặp khó khăn trong việc nhận diện tư thế chính xác nếu đối tượng bị che khuất, quá xa hoặc góc nhìn không tốt [20].



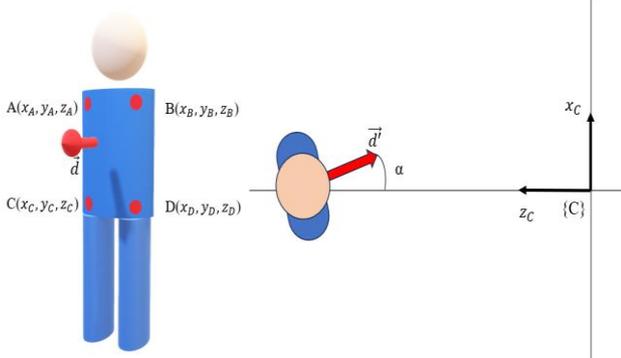
Hình 7. 17 điểm quan trọng trên cơ thể người trên MoveNet.

3.3. Xác định hướng của người sử dụng 4 điểm quan trọng từ khung xương

Việc xác định hướng đi của một người dựa trên vị trí và hướng của các điểm thường là một bài toán phức tạp trong lĩnh vực thị giác máy tính và xử lý ảnh. Điều này phụ thuộc rất lớn vào số lượng điểm và cách lựa chọn, có thể thay đổi tùy thuộc vào phương pháp cụ thể và mục đích ứng dụng. Việc lấy 3 điểm (hai vai và một điểm trọng tâm) mặc dù có thể xác định được ảnh 2D, tuy nhiên có thể chưa đủ để xác định hướng đi chính xác trong một số trường hợp. Ví dụ, nếu chúng ta giả định rằng người di chuyển theo mặt phẳng ngang, ta có thể sử dụng 3 điểm để xác định hướng di chuyển. Điểm ở thân có thể đại diện cho trọng tâm của thân, và việc theo dõi sự thay đổi vị trí của các điểm vai và điểm thân có thể dẫn đến việc xác định hướng di chuyển.

Tuy nhiên, trong một số trường hợp phức tạp hơn, ví dụ như khi người di chuyển trong không gian 3D hoặc có các thay đổi

hướng đi, việc sử dụng 4 điểm trên cơ thể giúp theo dõi chuyển động của vai và hông cùng một lúc, từ đó cung cấp thông tin chi tiết về cách mà cơ thể di chuyển và xoay. Điều này đặc biệt hữu ích trong các tình huống phức tạp như di chuyển trong không gian 3D hoặc khi có các rẽ nhánh trong hướng đi.



Hình 8. Tọa độ hóa 4 điểm quan tâm.

Mặt khác với việc sử dụng từ 5 điểm trở lên trong việc xác định hướng đi của cơ thể có thể giúp xác định chính xác hướng đi chuyển của người hơn trong một số trường hợp. Nhưng điểm của việc sử dụng từ 5 điểm trở lên là tăng phức tạp và tốn thời gian trong quá trình thu thập dữ liệu và xử lý dữ liệu đặc biệt với việc sử dụng Raspberry.

Trong nghiên cứu này, chúng tôi đề xuất một phương án sử dụng bốn điểm ít di chuyển của thân trên như trong Hình 8. Các điểm A, B, C, D tương ứng với các điểm 5, 6, 11, 12 trong Hình 7 lần lượt là các tọa độ vai trái, vai phải, hông trái, hông phải. Khi đã xác định được 4 điểm đặc trưng trên cơ thể thông qua xác định tư thế, việc tọa độ hóa các điểm này là quá trình quan trọng để thu được thông tin về hướng đi chuyển. Tọa độ hóa 4 điểm đặc trưng này đặt nền móng cho việc tạo ra một véc tơ hướng. Bằng việc xác định được tọa độ 4 điểm A, B, C, D, véc tơ góc hướng của người so với camera \vec{d}' có thể được xác định một cách dễ dàng vì là hình chiếu của véc tơ \vec{d} trên mặt phẳng Oxz với $\vec{d}' = \vec{AD} \times \vec{BC}$.

a/ Xác định tọa độ các điểm trong không gian sử dụng camera chiều sâu

Ở đây, cần xác định tọa độ vị trí P (trong ví dụ Hình 9) trong hệ tọa độ 3D với tâm đặt tại vị trí đặt camera có thông số trong Bảng 1 và Bảng 2.

Bảng 1. Thông số Camera Astra S.

Product Name	Astra S
Range	0.4m – 2m
FOV	60°H x 49.5°V x 73°D
Size	165mm x 30mm x 40mm
RGB Image Resolution	640 x 480 @30fps
Depth Image Resolution	640 x 480 @30fps
Accuracy	+/- 1 – 3mm @1 m

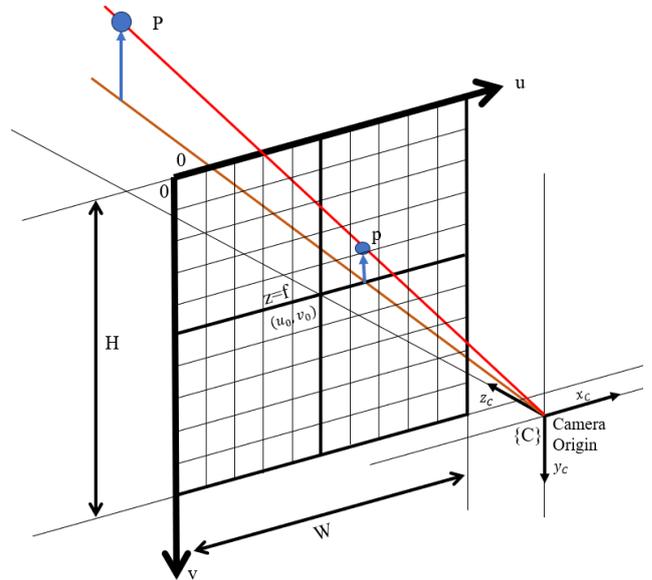
Trong hệ trục tọa độ thuần nhất (homogeneous coordinates), véc tơ $[u \ v \ 1]^T$ được sử dụng để thể hiện vị trí điểm 2D trong hệ tọa độ khung ảnh (image coordiante system), $[X \ Y \ Z \ 1]^T$ được sử dụng để thể hiện vị trí điểm 3D trong tọa độ không gian. Khi đó mỗi quan hệ của hai tọa độ này như sau (với z_c là thông số độ sâu của ảnh lấy từ camera độ sâu):

$$z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & \gamma & c_x & 0 \\ 0 & f_y & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R_{3 \times 3} & T_{3 \times 1} \\ 0_{1 \times 3} & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (1)$$

$$z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (2)$$

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} z_c \cdot u - c_x \cdot z_c \\ f_y \\ f_x \\ z_c \end{bmatrix} \quad (3)$$

Phép chiếu (1) được sử dụng để ánh xạ từ tọa độ không gian sang tọa độ khung ảnh. Giả sử camera không bị quay hay tịnh tiến ta được (2). Sử dụng (3) ta có thể tính được tọa độ vị trí quan tâm $[X \ Y \ Z]^T$ một cách chính xác. Nhưng bên cạnh đó vẫn còn nhiều như nhiễu hình ảnh bởi các điểm ảnh không mong muốn hoặc nhiễu có thể xuất hiện trong hình ảnh từ camera độ sâu. Điều này có thể do nhiễu điện từ, nhiễu tín hiệu hoặc do nhiễu quang học, các yếu tố đó ảnh hưởng đến kết quả làm việc nên chúng tôi đề xuất kết hợp bộ lọc ở phần tiếp theo.



Hình 9. Chuyển đổi tọa độ so với khung camera.

Bảng 2. Các thông số khung chuyển đổi.

H	Kích thước chiều dọc pixel của khung ảnh
W	Kích thước chiều ngang pixel của khung ảnh
u, v	Tọa độ trong hệ tọa độ ảnh
z_c	z _c = Z (hiện tại)
f_x	Tiêu cự theo Ox
f_y	Tiêu cự theo Oy
c_x, c_y	Tọa độ tâm của khung ảnh trong trong hệ tọa độ ảnh
X, Y, Z	Tọa độ vị trí trong không gian

b/ Xây dựng bộ lọc Gauss cho ảnh độ sâu

Nhiều xuất hiện trong quá trình tính toán độ sâu này là điều không tránh khỏi. Do vậy, để tăng độ chính xác trong quá trình sử dụng dữ liệu từ camera độ sâu để xác định hướng của người so với camera, chúng tôi đề xuất sử dụng bộ lọc Gauss [21]. Ý tưởng chung là giá trị mỗi điểm ảnh sẽ phụ thuộc nhiều vào các điểm ảnh ở gần hơn là các điểm ảnh ở xa. Trọng số của sự

phụ thuộc được lấy theo hàm Gauss (cũng được sử dụng trong quy luật phân phối chuẩn) [22]. Việc lọc ảnh dựa trên hàm phân phối Gauss hai chiều biểu diễn dưới dạng:

$$h(m, n) = \frac{1}{2\pi\sigma^2} e^{-\frac{m^2 + n^2}{2\sigma^2}} \quad (4)$$

với σ là độ lệch chuẩn của phân bố Gauss.

Khi đó giá trị các điểm ảnh đầu ra $g[i, j]$ sẽ được tính bằng tích chập giữa các điểm ảnh đầu vào $f[i, j]$ và hàm phân phối Gauss có kích thước $[x, y]$ như sau:

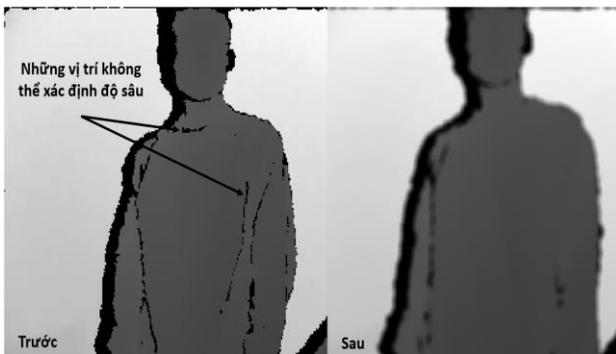
$$g[i, j] = \sum_{m=-x'}^{x'} \sum_{n=-y'}^{y'} f[i - m, j - n] \cdot h[m, n] \quad (5)$$

hay

$$g[i, j] = \frac{1}{2\pi\sigma^2} \sum_{m=-x'}^{x'} \sum_{n=-y'}^{y'} e^{-\frac{m^2 + n^2}{2\sigma^2}} f[i - m, j - n] \quad (6)$$

với x' và y' là phần nguyên của $x/2$ và $y/2$.

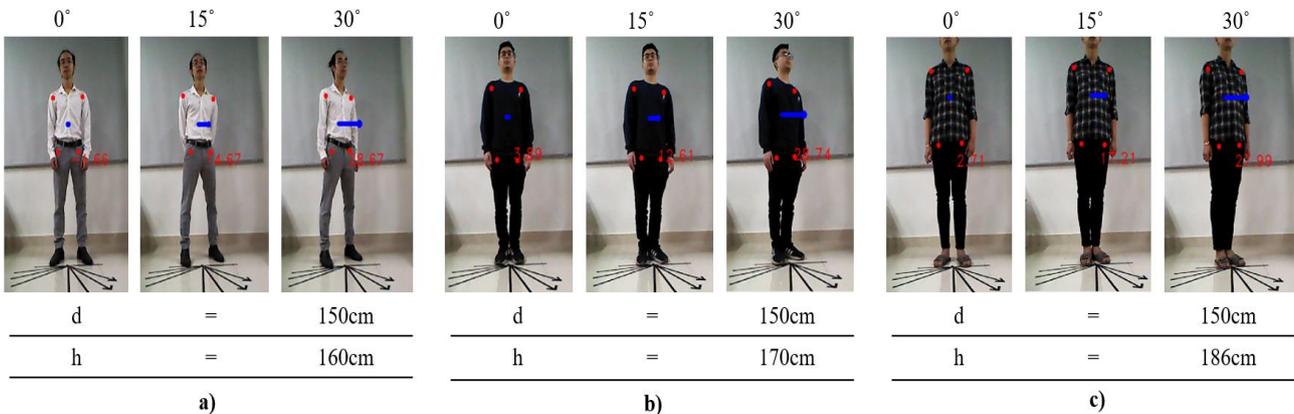
Kết quả thực nghiệm sử dụng Gauss áp dụng cho việc nhận dạng người khiếm thị như Hình 10. Có thể thấy với việc sử dụng bộ lọc Gauss cho ảnh độ sâu là có hiệu quả, đảm bảo loại bỏ những mấp mô hay nếp gấp do quần áo, nhiễu ảnh hướng tới tính toán tọa độ các điểm quan tâm trên ảnh.



Hình 10. So sánh trước và sau khi dùng bộ lọc Gauss.

4. Đánh giá thực nghiệm

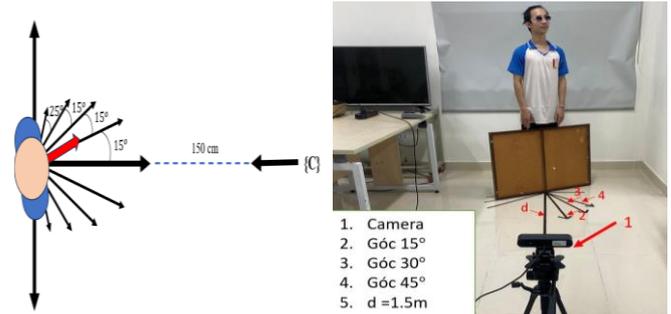
4.1. Mô tả thực nghiệm



Hình 12. Thiết lập thực nghiệm xác định hướng trên 3 đối tượng có chiều cao khác nhau.

Để kiểm tra các đề xuất về việc nhận dạng và xác định hướng của người khiếm thị, một mô hình thực nghiệm đã được thiết lập. Trong mô hình này, robot được trang bị một camera Astra S đặt trên robot có khả năng đo khoảng cách trong phạm vi 0,4-2 m. Độ phân giải ảnh màu và ảnh độ sâu là 640 x 480 / 30fps. Sai số của camera trong khoảng $\pm(1-3\text{mm})/1\text{m}$.

Để kiểm nghiệm độ chính xác của thuật toán chúng tôi đưa ra kịch bản thực nghiệm với những góc độ khác nhau. Các góc được sử dụng 15°, 30°, 45° và góc 70° như Hình 11. Đối tượng đứng cách camera 1,5m với các hướng khác nhau. Kiểm nghiệm thực hiện với hai kịch bản. Trong kịch bản 1, các đối tượng tham gia đo góc có chiều cao khác nhau. Trong kịch bản 2, đối tượng xoay người từ từ với tốc độ 7,5 độ/ giây.



Hình 11. Ảnh setup kiểm nghiệm trên phòng thí nghiệm.

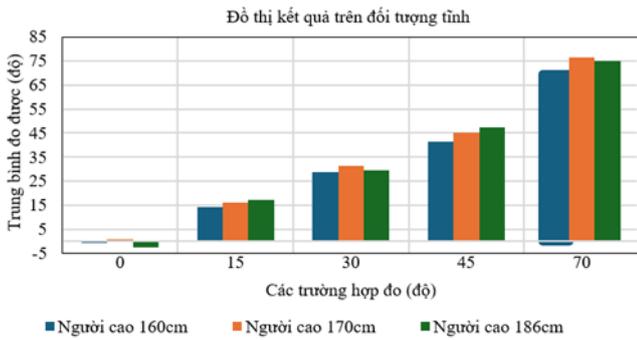
4.2. Kết quả đánh giá thực nghiệm

Kịch bản 1: Áp dụng cho các đối tượng có chiều cao khác nhau đứng cố định và được tiến hành theo các bước sau:

Bước 1: Với các góc đã được thiết lập sẵn: 0°, 15°, 30°, 45°, 70°, cho đối tượng đứng có hướng trùng với hướng của các vạch góc.

Bước 2: Sử dụng camera để đo góc của người so với camera. Kết quả thu được cho từng trường hợp đối tượng có chiều cao khác nhau được trình bày trong Hình 12.

Kịch bản 1 được thực hiện lặp đi lặp lại nhiều lần cho ba đối tượng sau đó lấy kết quả trung bình ta thu được kết quả thực nghiệm như trong Hình 13. Có thể thấy thuật toán có độ chính xác khá cao (sai số <5%), và sai lệch giữa các đối tượng là không lớn. Điều đó cho thấy thuật toán đề xuất có thể áp dụng với các đối tượng có chiều cao phổ thông trong thực tế.



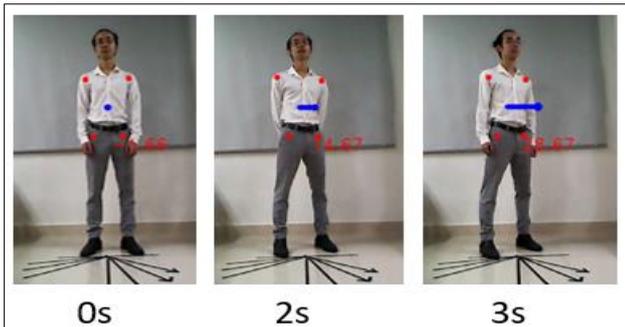
Hình 13. Kết quả thực nghiệm đối tượng đứng cố định.

Kịch bản 2: Áp dụng cho đối tượng dịch chuyển quay

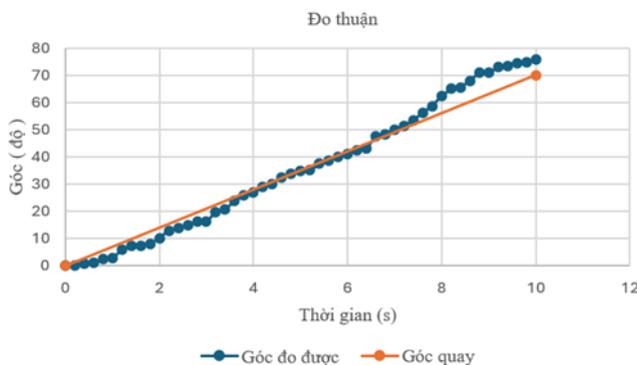
Bước 1: Đối tượng có chiều cao 1m60 quay từ góc ban đầu 0° đến góc cuối cùng là 30° đồng thời ghi lại thời gian của đối tượng bằng cách sử dụng đồng hồ bấm giờ để xác định thời gian tại các vị trí 0° , 15° và 30° .

Bước 2: Camera ghi hình đối tượng và thuật toán xác định hướng tính toán góc của người so với camera.

Kết quả thực nghiệm được biểu diễn trong Hình 14 và Hình 15. Kết quả thực nghiệm cho thấy với tốc độ di chuyển trung bình của đối tượng thì thuật toán đề xuất vẫn đảm bảo xác định được góc quay của đối tượng. Tuy kết quả có độ chính xác chưa cao do sai số của camera và nhiễu từ quá trình quay không thể chính xác lựa chọn điểm, nhưng trong khoảng thời gian ngắn, kết quả thu được vẫn đáng tin cậy và ở mức chính xác chấp nhận được cho ứng dụng hỗ trợ người khiếm thị.



Hình 14. Góc quay theo các hướng từ 0° đến 30° .



Hình 15. Kết quả thực nghiệm đối tượng quay.

Các thí nghiệm với trường hợp khoảng cách giữa đối tượng và camera là 1,2 m và 1,8 m cũng đã được tiến hành và các kết quả thu được cũng có độ chính xác tương tự như trường hợp khoảng cách là 1,5 m. Ngoài ra, các thí nghiệm với trường hợp đối tượng không nằm giữa khung hình cũng đã được tiến hành và kết quả thu được tương tự như trường hợp đối tượng ở chính giữa khung hình.

5. Kết luận

Bài báo trình bày một thuật toán xác định hướng cho đối tượng sử dụng ảnh camera 3D phục vụ phản hồi thông tin người khiếm thị trong hệ thống robot di động trợ giúp di chuyển người khiếm thị. Thuật toán đã tích hợp các công nghệ xử lý ảnh tiên tiến bao gồm nhận diện và bám đối tượng, xác định khung xương đối tượng, kết hợp sử dụng bộ lọc Gauss để tăng chất lượng ảnh chiều sâu từ đó tính góc giữa camera và người khiếm thị sử dụng 4 điểm quan tâm là vai trái, vai phải, hông trái, hông phải. Kết quả thực nghiệm cho thấy đối với các đối tượng đứng cố định, thuật toán cho kết quả chính xác với sai số dưới 5%. Với các thí nghiệm cho đối tượng di chuyển với tốc độ trung bình phù hợp với tốc độ di chuyển của người khiếm thị, kết quả cũng là chấp nhận được cho yêu cầu của thuật toán trợ giúp di chuyển. Các kết quả thực nghiệm tiên hành trong phòng thí nghiệm đã chứng minh tính đúng đắn và chính xác của thuật toán được đề xuất. Thêm vào đó, các thiết bị sử dụng đều có chi phí thấp hứa hẹn khả năng thương mại hóa cao.

Trong tương lai, các tác giả đề ra mục tiêu là cải thiện thuật toán cho kết quả chính xác hơn trong các tình huống đối tượng động và cử động nhanh. Điều này sẽ cung cấp một trải nghiệm tốt hơn cho người dùng. Việc hoàn thiện và kiểm nghiệm hệ thống robot trợ giúp người khiếm thị sử dụng thuật toán đề xuất trong bài báo này cũng sẽ được hoàn thành trong thời gian tới.

Tài liệu tham khảo

- [1] "Vision impairment and blindness." last access: Jan. 14, 2024. [Online]. Available: <https://www.who.int/news-room/factsheets/detail/blindness-and-visual-impairment>
- [2] B. Kuriakose, R. Shrestha, and F. E. Sandnes, "Tools and Technologies for Blind and Visually Impaired Navigation Support: A Review," *IETE Technical Review*, vol. 39, no. 1, pp. 3–18, Jan. 2022, doi: 10.1080/02564602.2020.1819893.
- [3] S. Real and A. Araujo, "Navigation Systems for the Blind and Visually Impaired: Past Work, Challenges, and Open Problems," *Sensors*, vol. 19, no. 15, Art. no. 15, Jan. 2019, doi: 10.3390/s19153404.
- [4] B.-S. Lin, C.-C. Lee, and P.-Y. Chiang, "Simple Smartphone-Based Guiding System for Visually Impaired People," *Sensors*, vol. 17, no. 6, Art. no. 6, Jun. 2017, doi: 10.3390/s17061371.
- [5] Y. Zhao, E. Kupferstein, D. Tal, and S. Azenkot, "It Looks Beautiful but Scary": How Low Vision People Navigate Stairs and Other Surface Level Changes," in *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*, in ASSETS '18. New York, NY, USA: Association for Computing Machinery, Oct. 2018, pp. 307–320. doi: 10.1145/3234695.3236359.
- [6] A. Dernayka, M.-A. Amorim, R. Leroux, L. Bogaert, and R. Farcy, "Tom Pouce III, an Electronic White Cane for Blind People: Ability to Detect Obstacles and Mobility Performances," *Sensors*, vol. 21, no. 20, Art. no. 20, Jan. 2021, doi: 10.3390/s21206854.
- [7] A. D. P. dos Santos, F. O. Medola, M. J. Cinelli, A. R. Garcia Ramirez, and F. E. Sandnes, "Are electronic white canes better than traditional canes? A comparative study with blind and blindfolded participants," *Univ Access Inf Soc*, vol. 20, no. 1, pp. 93–103, Mar. 2021, doi: 10.1007/s10209-020-00712-z.
- [8] A. Xiao, W. Tong, L. Yang, J. Zeng, Z. Li, and K. Sreenath, "Robotic Guide Dog: Leading a Human with Leash-Guided Hybrid Physical Interaction," May 2021, pp. 11470–11476. doi: 10.1109/ICRA48506.2021.9561786.
- [9] S. Kayukawa, T. Ishihara, H. Takagi, S. Morishima, and C. Asakawa, "Guiding Blind Pedestrians in Public Spaces by Understanding Walking Behavior of Nearby Pedestrians," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 4, no. 3, p. 85:1-85:22, Sep. 2020, doi: 10.1145/3411825.

- [10] S. Kayukawa, T. Ishihara, H. Takagi, S. Morishima, and C. Asakawa, "BlindPilot: A Robotic Local Navigation System that Leads Blind People to a Landmark Object," in *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, in CHI EA '20. New York, NY, USA: Association for Computing Machinery, Apr. 2020, pp. 1–9. doi: 10.1145/3334480.3382925.
- [11] Z. An, L. Hao, Y. Liu, and L. Dai, "Development of mobile robot SLAM based on ROS," *International Journal of Mechanical Engineering and Robotics Research.*, vol. 5, Jan. 2016, doi: 10.18178/ijmerr.5.1.47-51.
- [12] M. Bamdad, D. Scaramuzza, and A. Darvishy, "SLAM for Visually Impaired Navigation: A Systematic Literature Review of the Current State of Research." arXiv, Sep. 22, 2023. doi: 10.48550/arXiv.2212.04745.
- [13] Q.-H. Nguyen *et al.*, "A Visual SLAM System on Mobile Robot Supporting Localization Services to Visually Impaired People," in *Computer Vision - ECCV 2014 Workshops*, L. Agapito, M. M. Bronstein, and C. Rother, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2015, pp. 716–729. doi: 10.1007/978-3-319-16199-0_50.
- [14] M. M. Basheer and A. Varol, "An Overview of Robot Operating System Forensics," in *2019 1st International Informatics and Software Engineering Conference (UBMYK)*, Nov. 2019, pp. 1–4. doi: 10.1109/UBMYK48245.2019.8965649.
- [15] D. Talwar and S. Jung, "Particle Filter-based Localization of a Mobile Robot by Using a Single Lidar Sensor under SLAM in ROS Environment," in *2019 19th International Conference on Control, Automation and Systems (ICCAS)*, Oct. 2019, pp. 1112–1115. doi: 10.23919/ICCAS47443.2019.8971555.
- [16] I. Culjak, D. Abram, T. Pribanic, H. Dzapo, and M. Cifrek, "A brief introduction to OpenCV," in *2012 Proceedings of the 35th International Convention MIPRO*, May 2012, pp. 1725–1730.
- [17] G. Goyal, F. Di Pietro, N. Carissimi, A. Glover, and C. Bartolozzi, "MoveEnet: Online High-Frequency Human Pose Estimation With an Event Camera," presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 4024–4033.
- [18] Rumpal Kaur Sidhu, "Tutorial on Minimum Output Sum of Squared Error Filter", Master thesis, Colorado State University, 2006.
- [19] Chung, Jen-Li, Lee-Yeng Ong, and Meng-Chew Leow, "Comparative Analysis of Skeleton-Based Human Pose Estimation" *Future Internet* 14, no. 12: 380. 2022. <https://doi.org/10.3390/fi14120380>
- [20] E. P. Washabaugh, T. A. Shanmugam, R. Ranganathan, and C. Krishnan, "Comparing the accuracy of open-source pose estimation methods for measuring gait kinematics," *Gait & Posture*, vol. 97, pp. 188–195, Sep. 2022, doi: 10.1016/j.gaitpost.2022.08.008.
- [21] I. T. Young and L. J. van Vliet, "Recursive implementation of the Gaussian filter," *Signal Processing*, vol. 44, no. 2, pp. 139–151, Jun. 1995, doi: 10.1016/0165-1684(95)00020-E.
- [22] K. Ito and K. Xiong, "Gaussian filters for nonlinear filtering problems," *IEEE Transactions on Automatic Control*, vol. 45, no. 5, pp. 910–927, May 2000, doi: 10.1109/9.855552.