

VisionWalk: Trợ lý Điều hướng Sử dụng AI cho Người Khiếm Thị

Lê Đại Hòa¹, Nguyễn Tường Bách Hỷ¹, Liêu Hải Lưu Danh¹ và Lê Hoàng Vũ¹

Giáo viên hướng dẫn: PGS.TS Đinh Điền²³, TS Nguyễn Hồng Bửu Long²³, TS Lương An Vinh²³

Đại học Khoa học Tự nhiên, Đại học Quốc gia Thành phố Hồ Chí Minh

¹22TNT1, Ngành Trí tuệ Nhân tạo, Khoa Công nghệ Thông tin,

²Bộ môn Công nghệ tri thức, Khoa Công nghệ thông tin, ³Trung tâm Ngôn ngữ học Tính toán

Tóm tắt nội dung— VisionWalk là ứng dụng AI trên nền tảng di động, hỗ trợ người khiếm thị di chuyển an toàn và tự tin hơn. Sử dụng các mô hình học sâu như YOLOv8, ứng dụng nhận diện biển báo giao thông, vật cản và nguy hiểm tiềm ẩn theo thời gian thực. Các tính năng chính bao gồm cảnh báo biển báo giao thông, phát hiện vật cản tĩnh và di động, cùng ước lượng khoảng cách chính xác. VisionWalk tận dụng camera điện thoại và công nghệ chuyển văn bản thành giọng nói, đảm bảo chi phí thấp và dễ tiếp cận. Ứng dụng giúp tăng cường tính độc lập, an toàn và hòa nhập xã hội cho người khiếm thị.

Từ khóa— Điều hướng hỗ trợ bằng AI, Hỗ trợ người khiếm thị, Nhận diện biển báo giao thông, Phát hiện vật cản, YOLOv8, Điều hướng thời gian thực, Công nghệ chuyển văn bản thành giọng nói.

I. GIỚI THIỆU

Di chuyển trên các con phố đông đúc hoặc trong các không gian công cộng là một thách thức lớn đối với những người khiếm thị. Việc không thể phát hiện ra các vật cản, biển báo giao thông hoặc thay đổi trong môi trường xung quanh thường dẫn đến rủi ro cao về tai nạn và giảm sự độc lập trong việc di chuyển. Khi công nghệ ngày càng phát triển, các giải pháp sử dụng Trí tuệ Nhân tạo (AI) đã xuất hiện như một công cụ tiềm năng trong việc hỗ trợ người khiếm thị di chuyển an toàn và hiệu quả.

Dự án VisionWalk nhằm giải quyết những khó khăn này bằng cách phát triển một trợ lý dẫn đường sử dụng Trí tuệ Nhân tạo, được thiết kế đặc biệt để hỗ trợ người khiếm thị trong việc di chuyển trong thời gian thực. Bằng cách sử dụng các kỹ thuật thị giác máy tính và các mô hình AI, VisionWalk sẽ nhận diện các biển báo giao thông quan trọng, các vật cản và các nguy cơ tiềm ẩn trên đường đi, đồng thời cung cấp cảnh báo âm thanh ngay lập tức để hướng dẫn người dùng.

Ứng dụng này sẽ tập trung vào ba chức năng chính: Nhận diện biển báo giao thông trong thời gian thực, Phát hiện các vật cản tĩnh và di động, và cuối cùng là cung cấp cảnh báo và chỉ dẫn thời gian thực giúp người dùng di chuyển an toàn. Sử dụng camera của điện thoại di động, hệ thống sẽ nhận diện và phân loại các biển báo giao thông như biển

cấm người đi bộ, vạch qua đường dành cho người đi bộ, cùng với các vật cản như cột đèn, xe cộ, hoặc khu vực công trình thi công. Ngoài ra, ứng dụng còn tính toán khoảng cách và hướng di chuyển của các vật cản và cảnh báo người dùng tương ứng.

Hệ thống sẽ được tối ưu hóa để hoạt động trên các nền tảng di động phổ biến, giúp người dùng dễ dàng tiếp cận mà không cần thiết bị phần cứng đặc biệt. Thông qua dự án này, chúng tôi không chỉ muốn cải thiện sự độc lập và an toàn cho người khiếm thị mà còn góp phần xây dựng một xã hội hòa nhập hơn, nơi công nghệ AI có thể được ứng dụng để nâng cao chất lượng sống cho nhóm người yếu thế.

II. NHỮNG CÔNG TRÌNH NGHIÊN CỨU LIÊN QUAN

Trong những năm gần đây, các nghiên cứu và ứng dụng công nghệ AI cho người khiếm thị đã đạt được nhiều tiến bộ đáng kể, đặc biệt là trong lĩnh vực hỗ trợ di chuyển và phát hiện vật cản. Một số hệ thống và nghiên cứu nổi bật trong lĩnh vực này bao gồm:

- 1) Wayfinder - Hệ thống hỗ trợ người khiếm thị di chuyển trong môi trường không gian công cộng: Hệ thống Wayfinder sử dụng công nghệ GPS và cảm biến để giúp người khiếm thị di chuyển trong các khu vực công cộng như trung tâm thương mại hoặc sân bay. Tuy nhiên, hệ thống này chủ yếu sử dụng tín hiệu âm thanh để cung cấp thông tin và không thực sự tập trung vào việc nhận diện các vật cản hoặc biển báo giao thông trong môi trường xung quanh.
- 2) NAVCog - Ứng dụng dẫn đường cho người khiếm thị sử dụng công nghệ Bluetooth Low Energy (BLE): NAVCog là một ứng dụng hỗ trợ người khiếm thị trong việc tìm đường trong các khu vực nội bộ, chẳng hạn như các tòa nhà hoặc bệnh viện. Hệ thống này sử dụng BLE để xác định vị trí và cung cấp hướng dẫn qua tín hiệu âm thanh. Tuy nhiên, ứng dụng này không cung cấp khả năng nhận diện và cảnh báo về các vật cản di chuyển hoặc biển báo giao thông, mà chỉ tập trung vào dẫn đường.
- 3) BlindSquare - Hệ thống dẫn đường cho người khiếm thị sử dụng GPS: BlindSquare là một ứng dụng di

động được thiết kế để giúp người khiếm thị điều hướng môi trường ngoài trời thông qua hệ thống định vị GPS. Ứng dụng cung cấp thông tin về các địa điểm xung quanh và cảnh báo về các vật thể cố định. Tuy nhiên, BlindSquare không tích hợp khả năng phát hiện vật cản động, như xe cộ hoặc người đi bộ khác, trong khi các mối nguy hiểm này thường xuyên gây ra sự cố trong quá trình di chuyển của người khiếm thị.

- 4) **Sónar-based Systems** - Hệ thống sử dụng sóng siêu âm và cảm biến khoảng cách: Một số hệ thống hỗ trợ người khiếm thị sử dụng cảm biến siêu âm để phát hiện vật cản trước mặt và cảnh báo cho người sử dụng qua âm thanh. Ví dụ, hệ thống Ultracane sử dụng cảm biến siêu âm để phát hiện các vật cản tĩnh và di chuyển trong phạm vi gần. Tuy nhiên, các hệ thống này có hạn chế trong việc nhận diện các vật cản nhỏ hoặc không thể phát hiện vật cản di chuyển như xe cộ, và độ chính xác của chúng có thể bị ảnh hưởng bởi môi trường xung quanh.
- 5) **Deep Learning for Object Detection**: Việc áp dụng các mô hình học sâu (Deep Learning) trong nhận diện đối tượng đã cho thấy hiệu quả rõ rệt trong các bài toán nhận diện vật cản và biển báo giao thông. Các mô hình như YOLO (You Only Look Once) và Faster R-CNN đã được sử dụng trong các ứng dụng nhận diện biển báo và vật thể trong nhiều lĩnh vực khác nhau. Các nghiên cứu như của **Redmon et al. (2016)** và **Ren et al. (2015)** đã chứng minh khả năng của YOLO và Faster R-CNN trong việc phát hiện các đối tượng trong ảnh với độ chính xác cao và tốc độ xử lý nhanh.
- 6) **Vision-based Assistive Technologies for Blind**: Các nghiên cứu về công nghệ hỗ trợ người khiếm thị dựa trên thị giác sử dụng camera điện thoại và các mô hình học máy để phát hiện biển báo giao thông và vật cản. Ví dụ, nghiên cứu của **Pustokhina et al. (2016)** giới thiệu một hệ thống phát hiện biển báo giao thông bằng cách sử dụng mô hình học sâu với dữ liệu ảnh từ camera di động. Tuy nhiên, những hệ thống này chưa hoàn thiện trong việc nhận diện đa dạng các vật cản di chuyển và tính toán khoảng cách chính xác.

❑ Mặc dù các hệ thống hiện tại đã mang lại nhiều lợi ích cho người khiếm thị trong việc di chuyển và nhận diện vật cản, nhưng hầu hết vẫn thiếu tính năng phát hiện các vật cản động (như xe cộ và người đi bộ) hoặc khả năng nhận diện biển báo giao thông chính xác và thời gian thực. VisionWalk dựa trên sự kết hợp của các mô hình học sâu như YOLOv8 và Faster R-CNN để cung cấp khả năng nhận diện chính xác biển báo giao thông và vật cản, đồng thời sử dụng các mô hình đo chiều sâu để tính toán khoảng cách và hướng di chuyển của các vật cản. Điều này giúp người dùng nhận được cảnh báo chính xác và kịp thời về các nguy cơ, từ đó cải thiện khả năng tự di chuyển và an toàn cho người khiếm thị.

III. Ý TƯỞNG TIẾP CẬN

Để thực hiện ứng dụng VisionWalk, một trợ lý điều hướng thông minh cho người khiếm thị, chúng tôi sẽ sử dụng sự kết hợp của các mô hình học máy và xử lý hình ảnh thời gian thực để nhận diện biển báo giao thông và vật cản, từ đó đưa ra cảnh báo âm thanh cho người sử dụng. Dưới đây là những bước chính trong quá trình tiếp cận và xây dựng hệ thống:

- 1) **Thu thập Dữ liệu** Ứng dụng sử dụng camera điện thoại làm thiết bị thu thập hình ảnh/video từ môi trường xung quanh người dùng. Để đảm bảo hiệu quả trong việc nhận diện và phát hiện vật cản, hệ thống sẽ liên tục thu thập dữ liệu từ camera khi người dùng di chuyển.
- 2) **Xử lý Dữ liệu và Nhận diện Biển báo**:
 - **Biển báo giao thông**: Mô hình học sâu, chẳng hạn như YOLO hoặc Faster R-CNN, sẽ được sử dụng để nhận diện các loại biển báo giao thông trong ảnh. Các biển báo như "*Cấm người đi bộ*", "*Trẻ em qua đường*", và "*Đường trơn trượt*" sẽ được ưu tiên nhận diện. Kết quả trả về bao gồm tên biển báo và vị trí của nó trong hình ảnh, từ đó có thể đưa ra cảnh báo âm thanh cho người dùng.
 - **Phát hiện vật cản**: Các mô hình YOLO (hoặc EfficientDet) sẽ giúp phát hiện các vật cản trên đường như cột đèn giao thông, biển báo, gốc cây, hay thậm chí các phương tiện giao thông như xe máy, ô tô và người đi bộ. Mô hình Depth Estimation (Dự đoán chiều sâu) sẽ giúp tính toán khoảng cách từ người sử dụng đến vật cản, từ đó quyết định mức độ ưu tiên của cảnh báo.
- 3) **Dự đoán khoảng cách và hướng di chuyển của vật cản**: Một trong những điểm quan trọng trong ứng dụng là khả năng dự đoán khoảng cách và hướng di chuyển của các vật cản. Điều này sẽ giúp người dùng hiểu rõ hơn về mối nguy hiểm hoặc sự thay đổi môi trường xung quanh. Các mô hình như MonoDepth2 (Dự đoán chiều sâu từ camera đơn) hoặc Stereo Depth Estimation (Dự đoán chiều sâu từ camera kép) sẽ được sử dụng để xác định khoảng cách từ người sử dụng đến các vật cản.
- 4) **Cảnh báo và Điều hướng**: Sau khi các biển báo và vật cản được phát hiện, ứng dụng sẽ sử dụng thư viện Text-to-Speech (TTS) như Google TTS hoặc Amazon Polly để phát âm thanh cảnh báo. Các cảnh báo âm thanh này sẽ giúp người dùng nhận thức rõ ràng các tình huống nguy hiểm hoặc hướng dẫn họ di chuyển an toàn, ví dụ: "*Có biển báo Cấm người đi bộ cách bạn 3m*" hay "*Có xe máy cách bạn 1.5m bên phải*".
- 5) **Tối ưu hóa và triển khai trên thiết bị di động**:
 - **Tối ưu hóa mô hình**: Để ứng dụng hoạt động mượt mà trên các thiết bị di động phổ thông, các mô hình học máy sẽ được tối ưu hóa bằng các công nghệ như TensorFlow Lite hoặc PyTorch Mobile. Các kỹ thuật như Quantization và

Pruning cũng sẽ được sử dụng để giảm kích thước mô hình và tăng tốc độ xử lý.

- Hỗ trợ đa nền tảng: Ứng dụng sẽ được phát triển trên nền tảng Flutter hoặc React Native, cho phép ứng dụng hoạt động trên cả Android và iOS với khả năng tương thích cao.

6) Đảm bảo thời gian thực: Để cung cấp cảnh báo kịp thời và hiệu quả, ứng dụng cần xử lý và phân tích dữ liệu trong thời gian thực. Đảm bảo rằng độ trễ khi phân tích hình ảnh dưới 1s sẽ giúp người dùng nhận cảnh báo nhanh chóng và chính xác.

7) Phát triển và triển khai: Sau khi mô hình được huấn luyện và tối ưu hóa, ứng dụng sẽ được phát triển, thử nghiệm, và triển khai trên các thiết bị di động. Quá trình thử nghiệm sẽ tập trung vào việc cải thiện độ chính xác của các mô hình phát hiện và cảnh báo, cũng như tối ưu hóa trải nghiệm người dùng.

□ Tóm lại, ý tưởng tiếp cận của VisionWalk tập trung vào việc kết hợp công nghệ AI, nhận diện hình ảnh và âm thanh để tạo ra một ứng dụng hỗ trợ người khiếm thị trong việc di chuyển an toàn và độc lập. Cải tiến và tối ưu hóa các mô hình học sâu là yếu tố quan trọng giúp ứng dụng hoạt động hiệu quả trên các thiết bị di động phổ thông.

IV. MÔ HÌNH HOẠT ĐỘNG VÀ CÔNG NGHỆ

A. Mô hình cơ bản

Đầu tiên quan sát một mô hình hoạt động đơn giản như sau mà hầu như tất cả các dự án về định hướng cho người khiếm thị đều sử dụng:

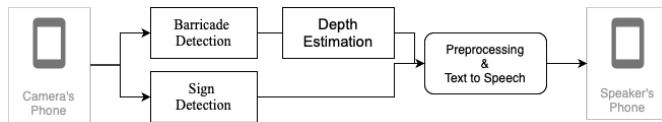


Figure 1. Mô hình cơ bản cho một hệ thống định hướng

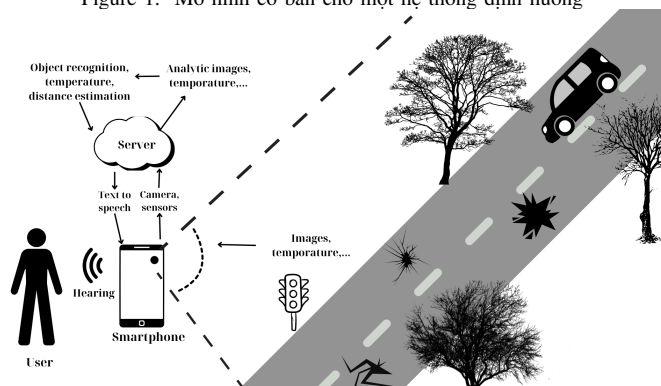


Figure 2. Minh họa mô hình

B. Các công nghệ sử dụng

Để làm hoàn thiện hơn mô hình cơ bản, ta cần phải tăng cường bằng những công nghệ hỗ trợ như sau:

1) Framework phát triển AI:

- TensorFlow/Keras hoặc PyTorch: Xây dựng và huấn luyện mô hình học sâu.
- OpenCV: Xử lý hình ảnh cơ bản, phát hiện vật cản hoặc kết nối với camera.

2) Mô hình phát hiện đối tượng:

- YOLOv8: Nhẹ, nhanh, phù hợp cho ứng dụng di động.
- Faster R-CNN: Độ chính xác cao, dùng để huấn luyện nhận diện biển báo.

3) Mô hình đánh giá khoảng cách:

- MonoDepth2: Dự đoán chiều sâu từ camera đơn (RGB).
- Stereo Depth Estimation: Sử dụng camera kép nếu có.

4) Phát âm thanh cảnh báo:

- Tích hợp thư viện Text-to-Speech (TTS) như Google TTS hoặc Amazon Polly.
- Thông báo giọng nói dễ nghe và rõ ràng.

Để hình dung được việc áp dụng các công nghệ này như thế nào, hãy quan sát biểu đồ hoạt động sau:

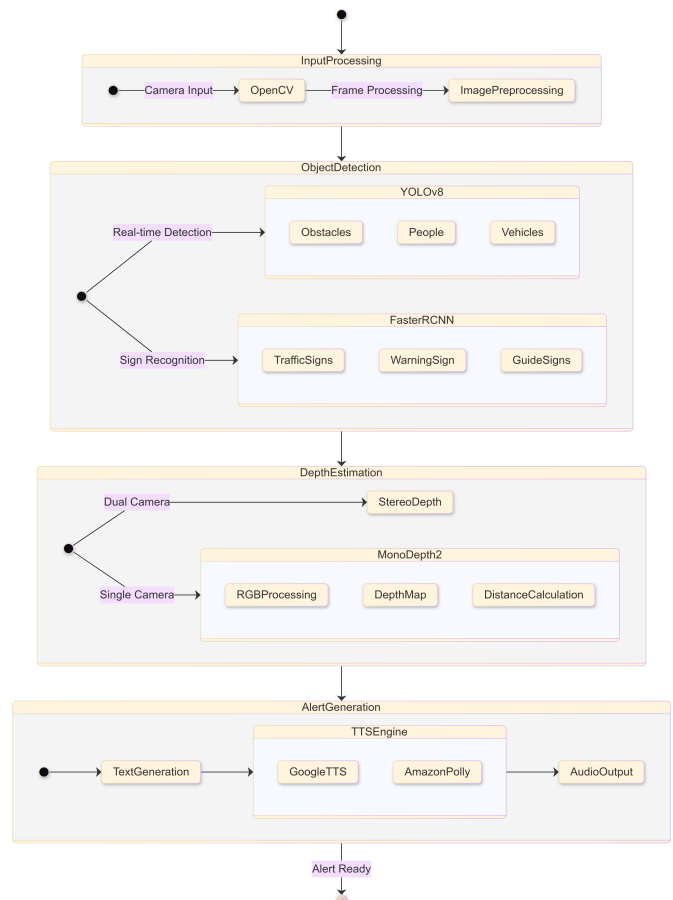


Figure 3. Các mô hình học máy và công nghệ sử dụng trong VisionWalk

C. Mô hình VisionWalk

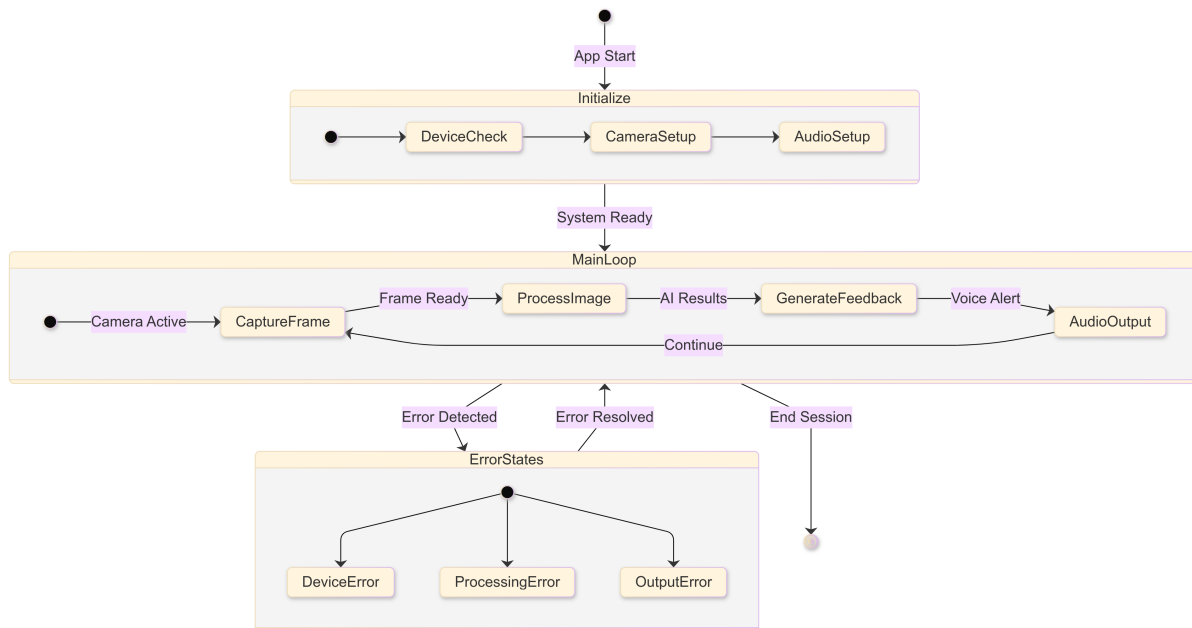


Figure 4. Luồng chính của ứng dụng

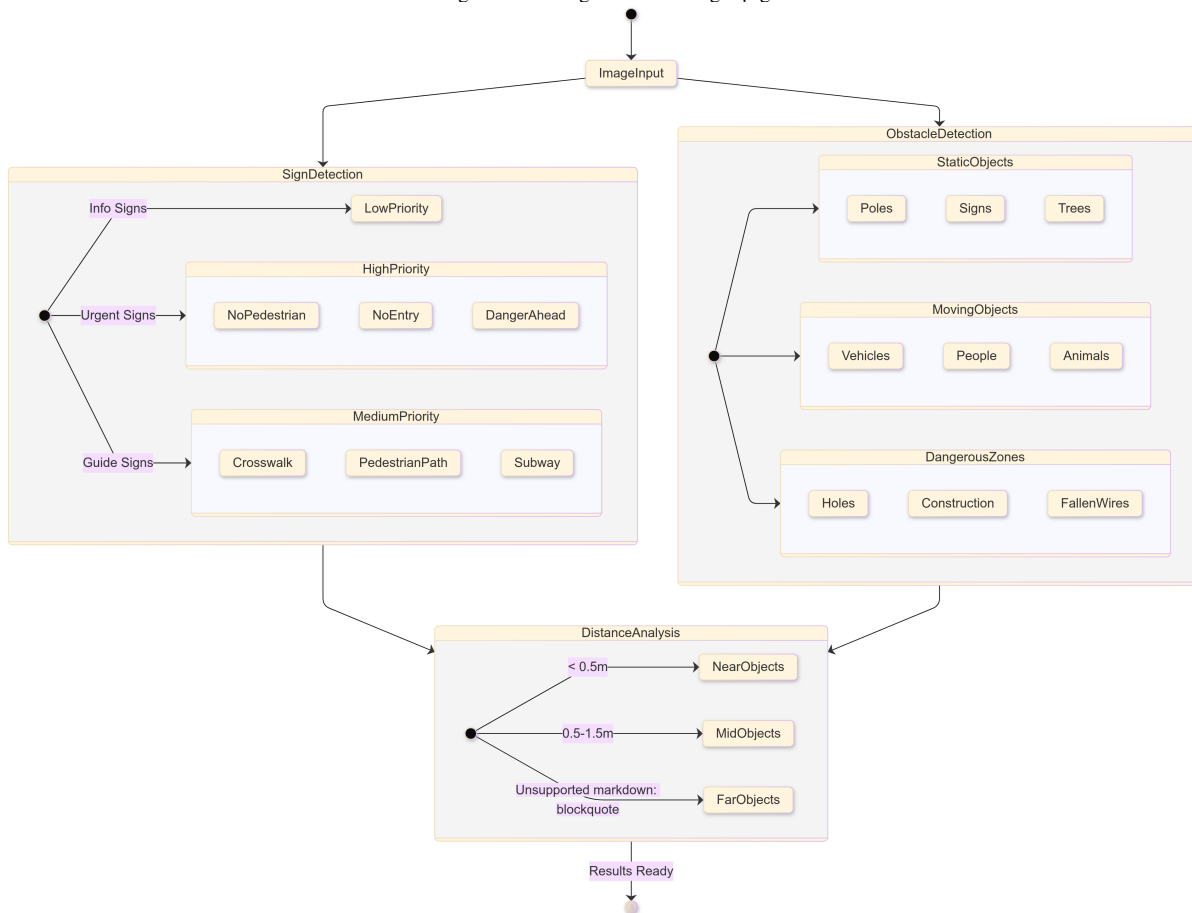


Figure 5. Quy trình xử lý AI

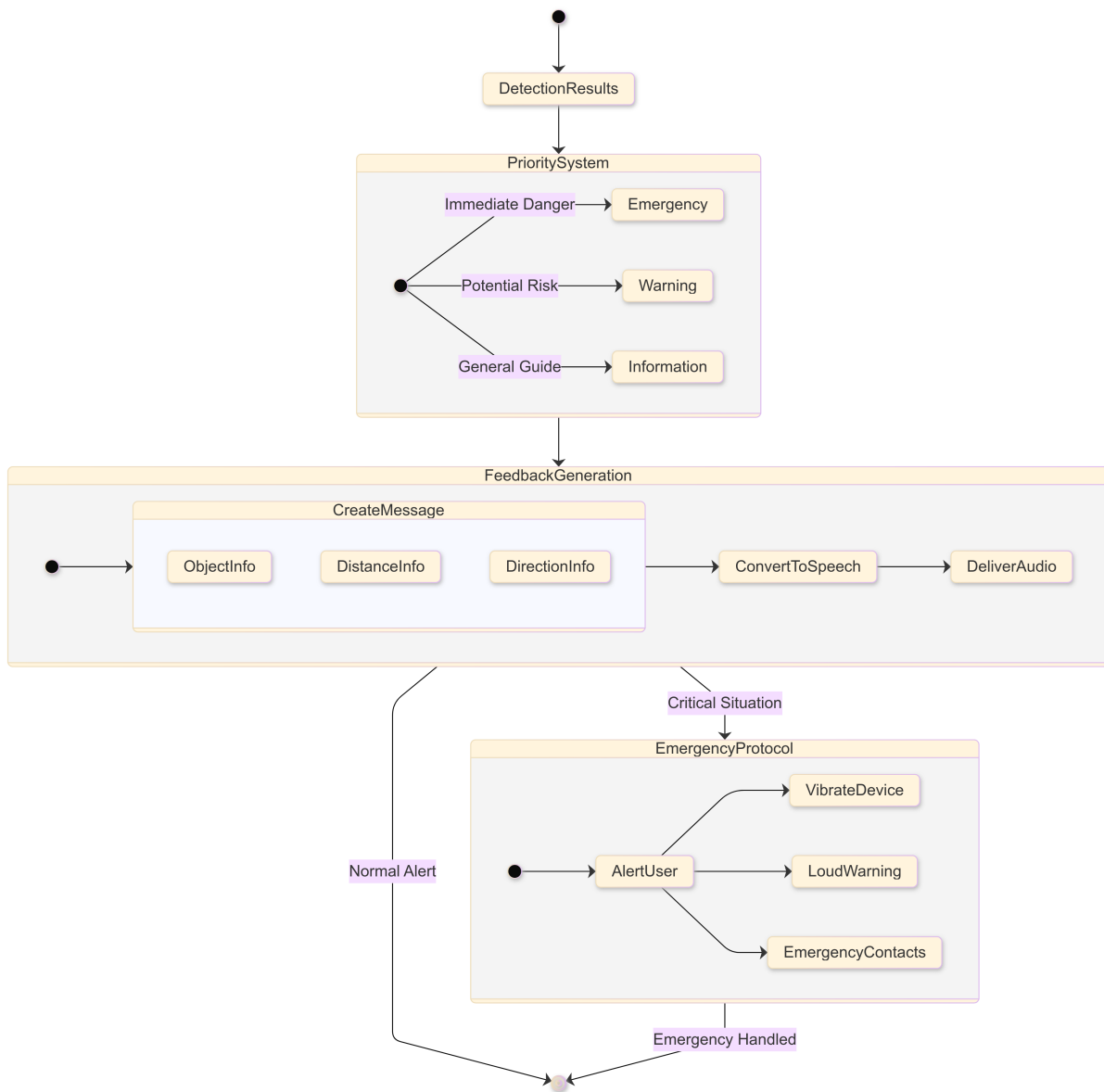


Figure 6. Hệ thống phản hồi và cảnh báo

V. MỤC TIÊU VÀ KẾT QUẢ MONG ĐỢI

A. Mục tiêu

Ứng dụng VisionWalk nhằm mục tiêu hỗ trợ người khiếm thị trong việc di chuyển an toàn và tự tin hơn trong môi trường giao thông. Cụ thể, ứng dụng cung cấp các cảnh báo thời gian thực về biển báo giao thông, vật cản và các nguy hiểm tiềm ẩn trên đường, giúp người khiếm thị nhận thức và tránh được các chướng ngại vật, như xe cộ, vật cản hoặc khu vực không an toàn. Ngoài ra, ứng dụng cũng giúp người dùng giảm sự phụ thuộc vào người đi kèm hoặc các thiết bị trợ giúp truyền thống như gậy dò đường, hỗ trợ họ tự tìm đường, nhận diện không gian, và đưa ra quyết định khi tham gia giao thông. Mục tiêu lâu dài là tăng cường an toàn cho người khiếm thị, đặc biệt ở các khu vực đông đúc hoặc có nguy cơ cao, và giúp họ hòa nhập tốt hơn vào xã hội thông qua việc nâng cao khả năng di chuyển độc lập.

B. Kết quả mong đợi

Kết quả của ứng dụng VisionWalk là khả năng phát hiện chính xác hơn 90% các biển báo giao thông và vật cản thường gặp, bao gồm các biển báo cấm, biển chỉ dẫn, cảnh báo nguy hiểm và các vật cản tĩnh hoặc di động như xe cộ, người đi bộ khác và động vật. Ứng dụng sẽ cung cấp cảnh báo thời gian thực với độ trễ dưới 1 giây, giúp người sử dụng có thể phản ứng kịp thời với các nguy hiểm. Ứng dụng được thiết kế để hoạt động tốt trong mọi điều kiện môi trường, từ ban ngày đến ban đêm và trong các điều kiện thời tiết xấu. Ngoài ra, ứng dụng sẽ được tối ưu hóa để hoạt động mượt mà trên các thiết bị di động phổ thông, không yêu cầu phần cứng đặc biệt, giúp tiết kiệm chi phí và mở rộng khả năng tiếp cận cho người khiếm thị.

TÀI LIỆU THAM KHẢO

- [1] Ahmetovic, D., Gleason, C., Kitani, K. M., Takagi, H., & Asakawa, C. (2016). NavCog: turn-by-turn smartphone navigation assistant for people with visual impairments or blindness. In *Web for All Conference (W4A '16)*. ACM, Article 9, 1-2.
- [2] Wiegand, K., Schmitz, B., & Kurschl, W. (2015). BlindSquare: A location-based application to support visually impaired people. In *International Conference on Computers Helping People with Special Needs* (pp. 165-172). Springer.
- [3] Fallah, N., Apostolopoulos, I., Bekris, K., & Folmer, E. (2013). Indoor human navigation systems: A survey. *Interacting with Computers*, 25(1), 21-33.
- [4] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 779-788).
- [5] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems*, 28, 91-99.
- [6] Pustokhina, I. V., Pustokhin, D. A., Gupta, D., Khanna, A., Shankar, K., & Nguyen, G. N. (2016). An effective training scheme for deep neural network in edge computing enabled Internet of Medical Things (IoMT) for smart healthcare system. *Neural Computing and Applications*, 32, 15897-15911.
- [7] Hoyle, B., & Waters, D. (2008). Mobility AT: The Ultracane does work! *Access Journal*, 9(4), 15-20.
- [8] Benjamin, J. M., Ali, N. A., & Schepis, A. F. (1973). A laser cane for the blind. *Proceedings of the San Diego Biomedical Symposium*, 12, 53-57.
- [9] Golledge, R. G., Marston, J. R., Loomis, J. M., & Klatzky, R. L. (2004). Stated preferences for components of a personal guidance system for nonvisual navigation. *Journal of Visual Impairment & Blindness*, 98(3), 135-147.
- [10] Wilson, J., Walker, B. N., Lindsay, J., Cambias, C., & Dellaert, F. (2007). SWAN: System for wearable audio navigation. In *11th IEEE International Symposium on Wearable Computers* (pp. 91-98).
- [11] Manduchi, R., & Coughlan, J. (2012). (Computer) vision without sight. *Communications of the ACM*, 55(1), 96-104.
- [12] Dakopoulos, D., & Bourbakis, N. G. (2010). Wearable obstacle avoidance electronic travel aids for blind: A survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 40(1), 25-35.
- [13] Abidi, M. H., Siddiquee, A. N., Alkhalefah, H., & Srivastava, V. (2023). A comprehensive review of navigation systems for visually impaired individuals.
- [14] Khan, S., Nazir, S., & Khan, H. U. (2021). Analysis of Navigation Assistants for Blind and Visually Impaired People: A Systematic Review. *IEEE Access*, 9, 26712-26730. DOI: 10.1109/ACCESS.2021.3052415
- [15] Qureshi, M. S., Khan, I. U., Qureshi, S. B., Khan, F. M., & Aleshaiker, S. (2023). Empowering the Blind: AI-Assisted Solutions for Visually Impaired People.