

MỤC LỤC

I. Lí do chọn đề tài.....	Trang 1
II. Tìm hiểu vấn đề nghiên cứu.....	Trang 1
1. Câu hỏi nghiên cứu	Trang 1
1.1 Ngôn ngữ ký hiệu.....	Trang 1
1.2. Các giải pháp hiện tại.....	Trang 2
2. Giải pháp đề ra	Trang 2
3. Ý nghĩa khoa học và cộng đồng của dự án	Trang 3
3.1 Ý nghĩa khoa học	Trang 3
3.2 Ý nghĩa cộng đồng	Trang 3
III. Phần cứng.....	Trang 3
1. Trung tâm xử lí (LattePanda)	Trang 3
2. Leap Motion	Trang 4
3. Kính AR (Augmented Reality - Thực tế ảo tăng cường).....	Trang 5
4. Các thiết bị ngoại vi khác.....	Trang 5
IV. Phần mềm.....	Trang 5
1. Dịch ngôn ngữ ký hiệu.....	Trang 5
1.1. Ngôn ngữ ký hiệu.....	Trang 5
1.2. Dữ liệu.....	Trang 6
1.3. Mô hình nhận diện ký hiệu	Trang 9
1.4. Đánh giá mô hình.....	Trang 11
2. Chuyển văn bản thành giọng nói và nhận diện giọng nói.....	Trang 12
3. Giao diện	Trang 12
V. Kết luận	Trang 13
1.Ưu điểm.....	Trang 13
2.Khuyết điểm	Trang 13
3.Hướng phát triển	Trang 14
VI. Tài liệu tham khảo.....	Trang 14

Dự án:

THIẾT BỊ THÔNG DỊCH NGÔN NGỮ KÍ HIỆU

I. Lí do chọn đề tài

Ngôn ngữ là phương tiện giao tiếp chủ yếu và hiệu quả nhất giữa người với người trong xã hội. Ngôn ngữ đưa con người đến gần nhau hơn, xóa bỏ mọi rào cản về sắc tộc, màu da,... Thế nhưng, đối với cộng đồng người Chăm di cư (Cộng đồng người Chăm di cư là những người hoàn toàn không “nghe, nói” được) hoạt động giao tiếp bằng ngôn ngữ của họ gặp nhiều khó khăn bởi họ không thể nói, nghe được bằng ngôn ngữ thông thường, họ sử dụng ngôn ngữ ký hiệu riêng làm phương tiện giao tiếp chính.

Theo thống kê của Tổng cục Thống kê Dân số và nhà ở, Việt Nam có 2.5 triệu người di cư và khiếm thính. Vì giao tiếp bằng ngôn ngữ ký hiệu riêng nên cộng đồng người Chăm di cư khó đạt được mục đích giao tiếp trong mối quan hệ với người bình thường; họ còn gặp khó khăn trong việc sử dụng các dịch vụ công cộng như bệnh viện, ngân hàng, xe buýt,...; họ dễ gặp những bất trắc rủi ro do không tiếp nhận được âm thanh phát ra từ môi trường xung quanh.

Để giúp cộng đồng người Chăm di cư hòa nhập với xã hội, để cuộc sống của họ tốt đẹp hơn, công bằng hơn, nhiều thiết bị hỗ trợ giao tiếp dành cho người câm di cư ra đời. Theo tìm hiểu chúng tôi biết được những thiết bị này vẫn tồn tại nhiều hạn chế và thiếu tính khả thi nên ít được cộng đồng biết đến và sử dụng.

Vì những lý do trên, nhóm tác giả đã nghiên cứu và phát triển “**Thiết bị thông dịch ngôn ngữ ký hiệu**” với mong muốn đem đến cho cộng đồng người Chăm di cư một thiết bị hữu ích giúp họ hòa nhập với cộng đồng, xã hội trong giao tiếp bằng ngôn ngữ.

* Trong báo cáo này, nhóm tác giả xin phép dùng từ “người bình thường” cho những người không thuộc Cộng đồng người câm di cư và từ “khiếm thính” để chỉ những người thuộc Cộng đồng người Chăm di cư.

II. Tìm hiểu vấn đề nghiên cứu

1. Câu hỏi nghiên cứu

- Thứ nhất: Hiện nay người khiếm thính giao tiếp bằng cách thức nào ?
- Thứ hai: Trên thị trường hiện nay đã có những thiết bị nào hỗ trợ người khiếm thính ?
- Thứ ba: Nghiên cứu, phát triển để tạo ra thiết bị hỗ trợ người khiếm thính giao tiếp bằng ngôn ngữ ?

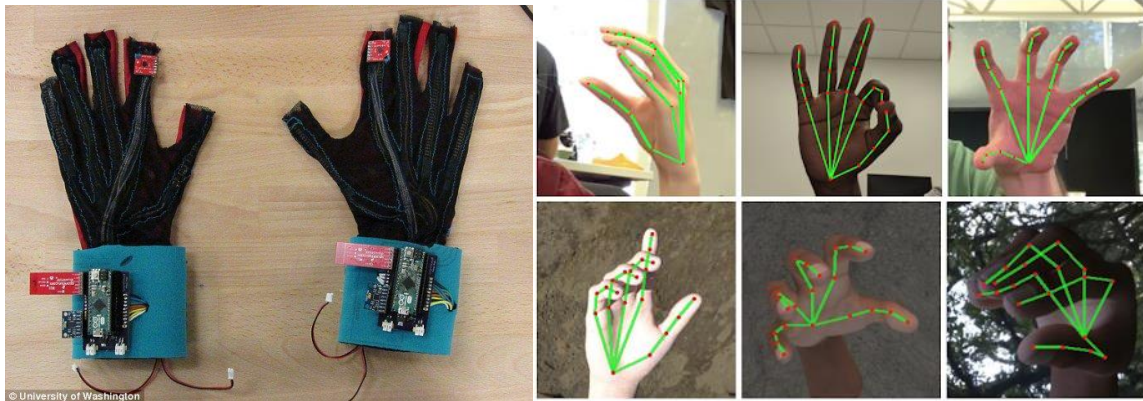
1.1 Ngôn ngữ ký hiệu

Hiện nay, Cộng đồng người Chăm di cư sử dụng ngôn ngữ ký hiệu để giao tiếp trong cuộc sống hằng ngày.

Ngôn ngữ ký hiệu là ngôn ngữ của Cộng đồng người Chăm di cư. Nó là ngôn ngữ bẩm sinh và tự nhiên của người di cư, được Cộng đồng người Chăm

điều sử dụng rộng rãi. Ngôn ngữ ký hiệu sử dụng hình dạng bàn tay, chuyển động cơ thể, cử chỉ điệu bộ và sự thể hiện trên khuôn mặt để giao tiếp, trao đổi kinh nghiệm, suy nghĩ, nhu cầu và cảm xúc. Giao tiếp bằng ngôn ngữ ký hiệu giúp họ cảm thấy thoải mái, tự nhiên và tự tin hơn.

1.2. Các giải pháp hiện tại



Găng tay phiên dịch (trái) và thiết bị phiên dịch bằng camera (phải)

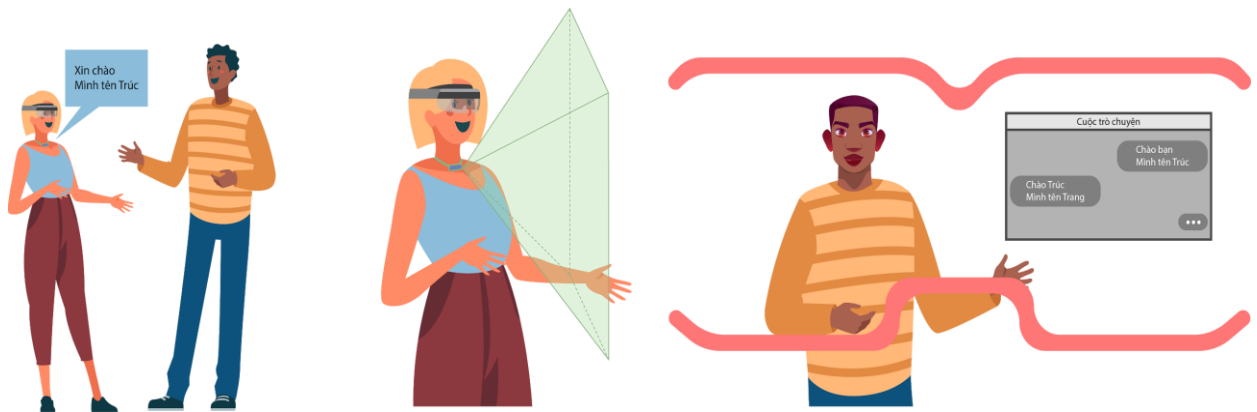
Hiện nay, đã có những sản phẩm hỗ trợ trong việc phiên dịch ngôn ngữ ký hiệu bằng nhiều phương pháp khác nhau.

Hạn chế:

- Thiết bị sử dụng camera: độ chính xác thấp, phụ thuộc nhiều vào ánh sáng môi trường, độ phân giải camera, ảnh hưởng tới quyền riêng tư.
- Thiết bị sử dụng cảm biến (găng tay): cồng kềnh, kém thẩm mỹ, gây bất tiện trong sinh hoạt hằng ngày.
- Chỉ dịch một chiều.
- Ngoài ra vẫn có những sản phẩm của nước ngoài đã nghiên cứu và phát triển. Nhưng những sản phẩm này không phổ biến, khó tìm mua. Hơn thế sản phẩm hướng tới dịch ASL (American sign language), không phải dịch ngôn ngữ ký hiệu Việt Nam.

2. Giải pháp đề ra

Dự án nghiên cứu nhằm tạo ra thiết bị hỗ trợ cho người khiếm thính, thiết bị có khả năng dịch ngôn ngữ ký hiệu sang ngôn ngữ thường, hỗ trợ người khiếm thính biểu đạt suy nghĩ, đồng thời cung cấp thông tin từ lời nói của đối phương tới đối tượng sử dụng. Dự án tạo ra với mong muốn ứng dụng hỗ trợ Cộng đồng người Câm điếc trong việc giao tiếp với xã hội với hai quá trình: nói và nghe.



Cách hoạt động của thiết bị

Gồm các chức năng chính :

- Nhận diện ngôn ngữ ký hiệu.
- Hiển thị thông tin lời nói từ đối phương.

3. Ý nghĩa khoa học và cộng đồng của dự án

3.1 Ý nghĩa khoa học

Đề tài nghiên cứu ra cách dịch ngôn ngữ khác với các sản phẩm đi trước: Dùng Leap motion để lấy thông tin của cánh tay trong không gian ba chiều. Tiếp đến, dự án cũng nghiên cứu phương pháp dịch ngôn ngữ ký hiệu đơn giản. Kết hợp với công nghệ thực tế ảo tăng cường giúp hiển thị thông tin trước mắt người dùng.

3.2 Ý nghĩa cộng đồng

Dự án giúp đỡ Cộng đồng người Câm điếc đến gần với xã hội. Giúp xã hội tiếp nhận với ngôn ngữ ký hiệu một cách tự nhiên, và thân thiện hơn. Tạo nên công cụ dịch hoàn chỉnh - 2 chiều. Ngoài ra còn giúp nhiều người trong việc tự học ngôn ngữ ký hiệu.

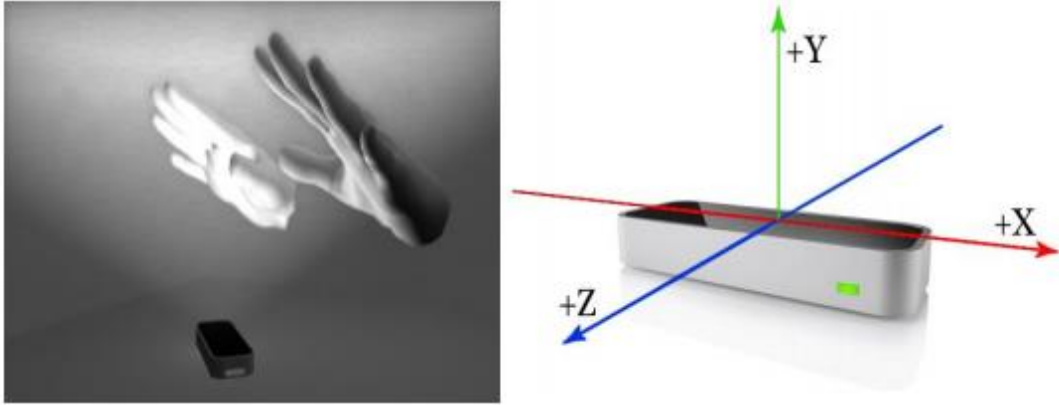
III. Phần cứng

1. Trung tâm xử lý (LattePanda)

Máy tính LattePanda: sử dụng hệ điều hành windows 10 hoàn chỉnh, dễ dàng cài đặt. Mang lại tính di động và khả năng hoạt động độc lập cho thiết bị.

- CPU: intel Z8350 (4 nhân, 1.44 GHz - 1.92 GHz)
- Hệ điều hành: windows 10
- RAM: 2GB DDR3L
- Nguồn: 5V - 2A
- Kích thước: 110mm x 94mm x 30mm
- Trọng lượng: 100g
- Cổng kết nối và mạng: HDMI, âm thanh 3.5mm, USB 3.0 x 1, USB 2.0 x 2, LAN, Wi-Fi 802.11n 2.4G, Bluetooth 4.0.
- Tích hợp Arduino Leonardo

2. Leap Motion



Leap motion (nguồn: www.leapmotion.com)

Leap Motion là một thiết bị ghi nhận hình ảnh về sự chuyển động của đôi tay trong không gian ba chiều. Kích thước của Leap Motion nhỏ gọn chỉ với 13mm x 13mm x 76mm và nặng 45 gam. Thay vì các cảm biến chiều sâu (depth sensor) trước đây, Leap Motion sử dụng hai camera cảm biến hồng ngoại và ba LED hồng ngoại, cho phép theo dõi ánh sáng có bước sóng 850nm. Ba LED cho phép tái tạo lại một mẫu vật 3D dưới ánh sáng hồng ngoại theo trục Y để tương tác với bàn tay người trong một khoảng không tầm 60cm bên trên mặt thiết bị và 30cm sang hai bên.

Leap Motion thu các hình ảnh vào bộ nhớ dưới dạng ảnh đen trắng, các ảnh này được lưu trong bộ nhớ và được xử lý bằng phần mềm loại bỏ các hình ảnh không phù hợp cũng như tăng cường các ảnh thiếu sáng sau đó được chuyển tới lớp ứng dụng cho lập trình viên. Leap Motion có thể cung cấp thông tin về:

- Vị trí cả hai tay và 10 ngón tay riêng biệt trong không gian.
- Hướng, vị trí và tốc độ từng đối tượng quan sát.
- Theo dõi được cả các khớp xương bàn tay, ngón tay.

Với độ chính xác 0,01mm ở tốc độ 200 khung hình trên giây và tập API rất tiện dụng, hỗ trợ nhiều ngôn ngữ lập trình và thường xuyên được cập nhật, Leap Motion cho phép phát triển rất nhiều ứng dụng trong các lĩnh vực khác nhau.

Nhờ khả năng theo dõi chuyển động rất nhanh và chính xác. Leap Motion được sử dụng để ghi nhận thông tin từ tay người dùng, từ đó tính toán, xử lý dữ liệu nhằm xác định câu nói người dùng muốn diễn đạt.

3. Kính AR (Augmented Reality - Thực tế ảo tăng cường)



Kính AR

Công nghệ thực tế tăng cường AR là sự kết hợp của thế giới thật với thông tin ảo, giúp hiển thị thông tin trước mắt người dùng mà không làm ảnh hưởng đến tầm nhìn.

Được thiết kế bằng 2 thành phần chính:

- Gương một chiều chuyên dụng.
- Màn hình 3.5 inch (độ phân giải 480 x 320 pixel) .

4. Các thiết bị ngoại vi khác

- Micro thu âm - Ghi lại lời nói của đối phương.
- Loa - Phát ra âm thanh biểu đạt lời nói của người dùng.

IV. Phần mềm

1. Dịch ngôn ngữ ký hiệu

1.1. Ngôn ngữ ký hiệu

1.1.1 Đặc tính của ngôn ngữ ký hiệu

Ngôn ngữ ký hiệu là ngôn ngữ chủ yếu được Cộng đồng người câm điếc sử dụng, bởi vì họ không có khả năng giao tiếp bằng lời nói nên họ tận dụng mọi cơ quan cơ thể, thông qua cử chỉ, điệu bộ, nét mặt thay cho lời nói nhằm truyền tải thông tin.

Do đó ta có thể rút ra các thành tố của một ký hiệu bao gồm:

- Vị trí làm ký hiệu. (1)
- Hình dạng bàn tay. (2)
- Chuyển động của tay. (3)
- Chiều hướng của lòng bàn tay. (4)
- Biểu hiện của nét mặt. (5)

Nhà ngôn ngữ học William Stokoe đã chỉ ra các yếu tố quan trọng của ngôn ngữ ký hiệu:

- Vị trí tương quan của tay và cơ thể.
- Hình dạng bàn tay.
- Hướng và chuyển động bàn tay.

(5): Chỉ thật sự cần thiết trong những thông tin lời nói mang tính cảm xúc, và nó không thật sự quan trọng. Do đó, ở phạm vi dự án này, nhóm tác giả chỉ

ngiên cứu và dịch những ký hiệu được mô tả hoàn toàn bởi các yếu tố (1), (2), (3), (4).

Qua đó ta có thể thấy Leap Motion có thể cung cấp những thông tin cơ bản, cần thiết để nhận diện ngôn ngữ ký hiệu. Leap Motion ghi lại tọa độ các khớp xương và hướng của ngón tay, lòng bàn tay và cẳng tay, đó là những thông tin quan trọng và đặc trưng cho ngôn ngữ ký hiệu (hình dạng bàn tay, chuyển động của bàn tay và vị trí tương đối của hai tay). Tuy nhiên những dữ liệu này vẫn là dữ liệu thô, cần phải qua xử lý trước khi đưa vào mô hình máy học.

1.1.2. Ngữ pháp ngôn ngữ ký hiệu

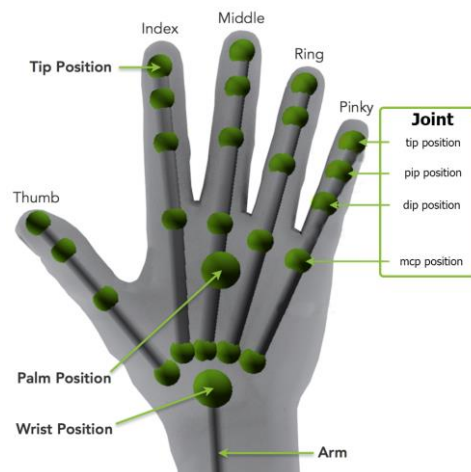
Người khiếm thính là một bộ phận trong cộng đồng xã hội, họ có hai đối tượng giao tiếp chính: người khiếm thính và người bình thường. Nếu những người khiếm thính giao tiếp với nhau thì họ sẽ sử dụng đúng cú pháp của ngôn ngữ ký hiệu do có sự tương đồng về tư duy, về cách nhìn nhận thế giới. Song nếu người khiếm thính giao tiếp với người bình thường thì họ lại có xu hướng cố gắng biểu đạt sao cho người bình thường có thể hiểu được. Khi đó, cú pháp của ngôn ngữ ký hiệu gần với cú pháp của ngôn ngữ nói tự nhiên. Vì vậy, việc xử lý thêm phần chuyển đổi ngữ pháp của ngôn ngữ ký hiệu sang ngôn ngữ thường là không cần thiết.

1.2. Dữ liệu

Leap Motion không được thiết kế để nhận diện ngôn ngữ ký hiệu. Mặc dù Leap Motion có thể ghi lại các tọa độ khớp xương của ngón tay, lòng bàn tay và cẳng tay, nhưng những đặc trưng của ký hiệu (hình dạng bàn tay, chuyển động của bàn tay và vị trí tương đối của hai tay) vẫn bị ẩn chứa trong dữ liệu thô. Để giải quyết vấn đề này, dự án đã xử lý dữ liệu qua nhiều bước.

1.2.1. Giảm dữ liệu nhiễu từ Leap Motion

Do Leap Motion không được cố định trước ngực, nên việc rung lắc khi hoạt động là điều không thể tránh khỏi, khiến dữ liệu ghi nhận được bị nhiễu. Thông tin nhận được từ Leap Motion là một chuỗi các khung hình. Do đó, tác giả sử dụng một bộ lọc của Savitzky-Golay để giải quyết vấn đề này:



Các khớp xương

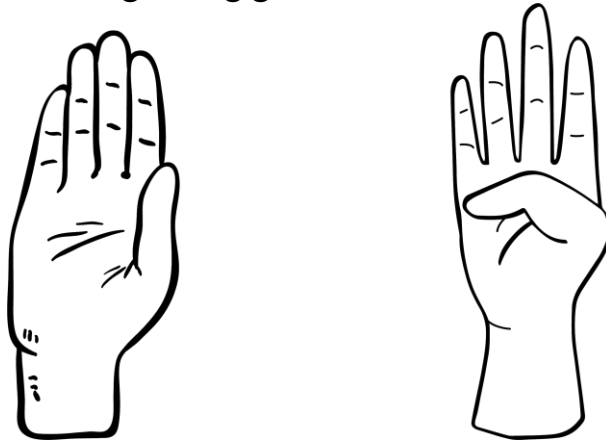
Gọi $J_{i,j,t} = (x_{i,j,t}, y_{i,j,t}, z_{i,j,t})$ với $i = \{ \text{Left}, \text{Right} \}$, $j = \{1, \dots, N\}$,

$t = \{1, \dots, T\}$ biểu thị thông tin của khung hình thứ t của chuỗi thời gian tọa độ 3D của các khớp xương ngón tay, lòng bàn tay và cẳng tay của một ký tự ngôn ngữ ký hiệu. Trong đó x, y, z là tọa độ của khớp xương trong không gian, i là loại cánh tay, j là thứ tự của các khớp xương, t là thứ tự của khung hình.

$$\tilde{J}_{i,j,t} = (-3J_{i,j,t-2} + 12J_{i,j,t-1} + 17J_{i,j,t} + 12J_{i,j,t+1} - 3J_{i,j,t+2})/35$$

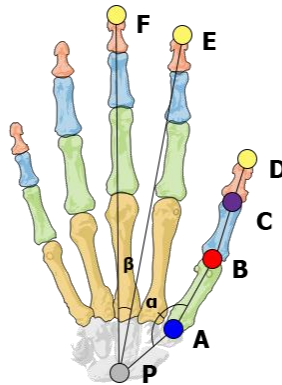
1.2.2. Hình dạng và hướng của bàn tay

Dữ liệu ban đầu chỉ đơn thuần là thông tin tọa độ của các khớp tay trong không gian ba chiều Oxyz. Thông tin này không thích hợp cho việc nhận diện hình dạng bàn tay. Vì vậy ta cần một phương pháp để trích xuất (biểu diễn hình dạng bàn tay) những thông tin cần thiết về hình dạng bàn tay mà không làm ảnh hưởng đến vị trí của nó trong không gian.



Ký hiệu “B” (bên trái) và “4” (bên phải)

Vì hình dạng bàn tay là thông tin độc lập hoàn toàn với vị trí của bàn tay trong không gian. Qua đó, dự án đã sử dụng các góc giữa các khớp ngón tay và các ngón tay trong cùng bàn tay nhằm miêu tả hình dạng bàn tay. Các góc này được tính qua tọa độ trong không gian ba chiều.



Gọi α và β lần lượt góc giữa các khớp trong cùng ngón và góc giữa các ngón tay trong cùng bàn tay. Ta có thể tính được α và β theo công thức:

$$\alpha_i = \arccos\left(\frac{\overrightarrow{AB} \cdot \overrightarrow{AP}}{|\overrightarrow{AB}| \cdot |\overrightarrow{AP}|}\right)$$

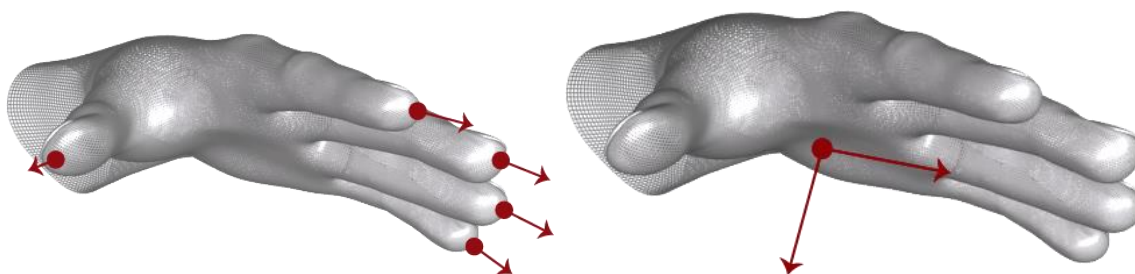
Tương tự với các góc còn lại.

Tuy nhiên, thông tin về các góc là vẫn chưa đủ để diễn tả hình dạng bàn tay trong không gian. Có những từ tuy rằng thông tin về các góc giống nhau nhưng lại mang nghĩa hoàn toàn khác nhau.



Ký hiệu “màu đỏ” (bên trái) và “màu đen” (bên phải)

Nhận thấy sự thiếu sót này, nhóm tác giả đã trích xuất thêm thông tin về hướng của ngón tay và hướng của lòng bàn tay được cung cấp từ Leap Motion được biểu diễn dưới dạng vector chỉ hướng trong không gian ba chiều.



Hướng của các ngón tay và hướng của lòng bàn tay

1.2.3. Vị trí trong không gian

Thông tin về vị trí của bàn tay trong không gian là một thông tin quan trọng. Có nhiều ký tự có hình dạng giống nhau nhưng vị trí trong không gian hoàn toàn khác nhau.



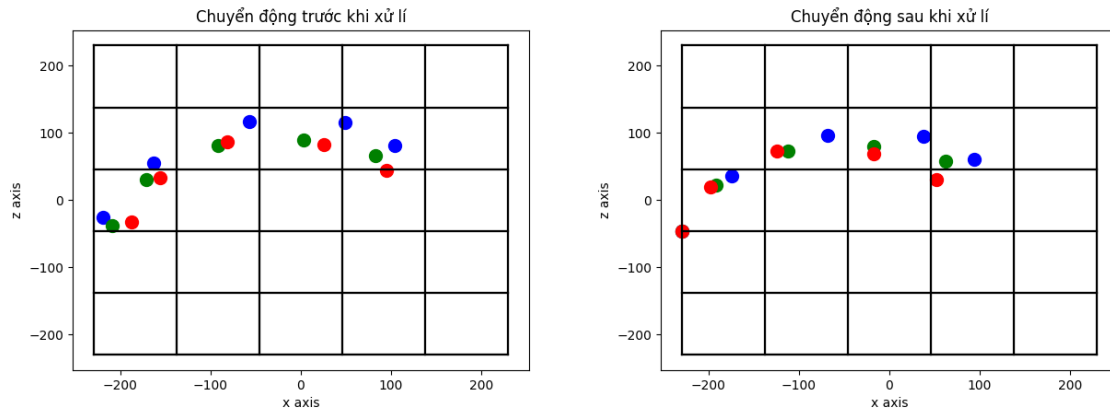
Ký hiệu “thấp” (bên trái) và “cao” (bên phải)

Tọa độ của các ngón tay trong không gian có thể cho ta biết về vị trí. Tuy nhiên, thông tin này không tốt vì phụ thuộc vào ký tự mà bàn tay đang diễn tả. Do đó, ta sử dụng thêm thông tin về vị trí của lòng bàn tay.

Hầu hết các ký tự đều có một phạm vi diễn tả nhất định. Tuy nhiên, trong thực tế, vị trí của ký hiệu trong mỗi lần thực hiện không hoàn toàn giống nhau, tạo nên sự khác biệt lớn về dữ liệu. Để giải quyết vấn đề này, nhóm tác giả sử dụng phương pháp chia vùng hoạt động nhằm giảm bớt sự chênh lệch không cần thiết trong cùng một ký tự.

Gọi $Palm[t]$ với $t = \{0, \dots, 4\}$ là thông tin trong không gian của lòng bàn tay tại khung hình thứ t , gọi $Coordinate[i]$ với $i = \{0, \dots, 24\}$ là tọa độ của vùng chứa $Palm[0]$.

$$Palm[t] = Palm[t] - Palm[0] + Coordinate[i]$$



Chuyển động của lòng bàn tay trước và sau khi xử lí (từ “mọi người”)

1.2.4. Thu thập dữ liệu

Một hành động diễn tả một từ trong ngôn ngữ ký hiệu là chuỗi các hành động được miêu tả bằng tay, nên 1 khung hình không thể diễn tả được đầy đủ thông tin.

Trung bình một hành động được miêu tả trong khoảng từ 1 giây đến 2 giây (tùy vào tốc độ mỗi người). Trong phạm vi dự án này, mỗi ký tự được giới hạn trong 1.7 giây, được miêu tả qua 5 khung hình (1 frame / 0.4s).

Dựa trên dữ liệu thu thập được, ta trích xuất đặc điểm của các ký tự bao gồm:

- Góc giữa các khớp tay trong cùng ngón (có phạm vi $[0, 180]$)
- Góc giữa các ngón tay trong cùng bàn tay (có phạm vi $[0, 180]$)
- Hướng của các ngón tay (có phạm vi $[-1, 1]$)
- Vị trí của bàn tay trong không gian (có phạm vi $[-230, 230]$)

Để dàng nhận thấy sự chênh lệch lớn giữa các phạm vi, gây sự sai sót khi tính hàm mất mát (loss) trong quá trình huấn luyện. Vì vậy thông tin về hướng của các ngón tay được nhân 100 nhằm làm tương đối về phạm vi dữ liệu.

Nhóm tác giả đã tiến hành thu thập dữ liệu của 14 từ. Mỗi từ là một bộ dữ liệu có kích thước 1200, được chia theo tỉ lệ 80:20 lần lượt là số bộ dữ liệu dùng để huấn luyện và đánh giá.

1.3. Mô hình nhận diện ký hiệu

Với cách xử lý dữ liệu ở trên, mỗi ký hiệu được nhận biết qua 5 frame hình. Với input có dạng là sequence (chuỗi), nhóm đề xuất giải pháp sử dụng mô hình LSTM thường được sử dụng cho các bài toán với dữ liệu là sequence hay time-series. LSTM có thể mang thông tin của frame (ảnh) từ state trước tới các

state sau, rồi ở state cuối là sự kết hợp của tất cả các ảnh để dự đoán hành động. LSTM còn tránh được hiện tượng thông tin ở càng xa càng dễ mất (vanishing gradient). Kiểu bài toán được chọn là “Many to one” - bài toán có nhiều input nhưng chỉ có 1 output.

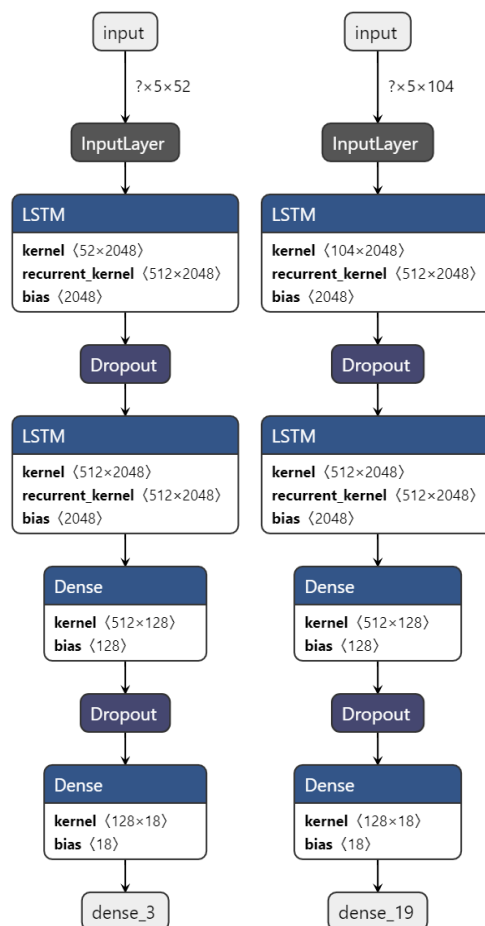
Xác định kích thước đầu vào là dữ liệu của 5 frame hình và kết quả là từ được dự đoán. Mô hình phải phù hợp cho các thiết bị di động.

Nhóm tác giả đã thử và xây dựng nhiều mô hình khác nhau với các kích thước, tầng và độ phức tạp của mô hình. Nhằm xác định đâu là mô hình hoạt động tốt nhất với bộ dữ liệu hiện tại. Mô hình tốt nhất thu được với bộ dữ liệu có kiến trúc tương đối đơn giản.

Kiến trúc của mô hình:

- Kích thước đầu vào `input_shape = (5, 52)` cho một tay và `(5, 104)` cho hai tay.
- 2 tầng LSTM với 512 hidden neurons mỗi tầng, batch size là 32.
- 2 tầng fully connected.
- Sử dụng hàm tối ưu “Adam”, loss là “sparse categorical crossentropy”.
- Ở các tầng có sử dụng Dropout nhằm tránh over-fit bằng việc bỏ qua một vài unit trong suốt quá trình train trong mô hình, những unit bị bỏ qua được lựa chọn ngẫu nhiên

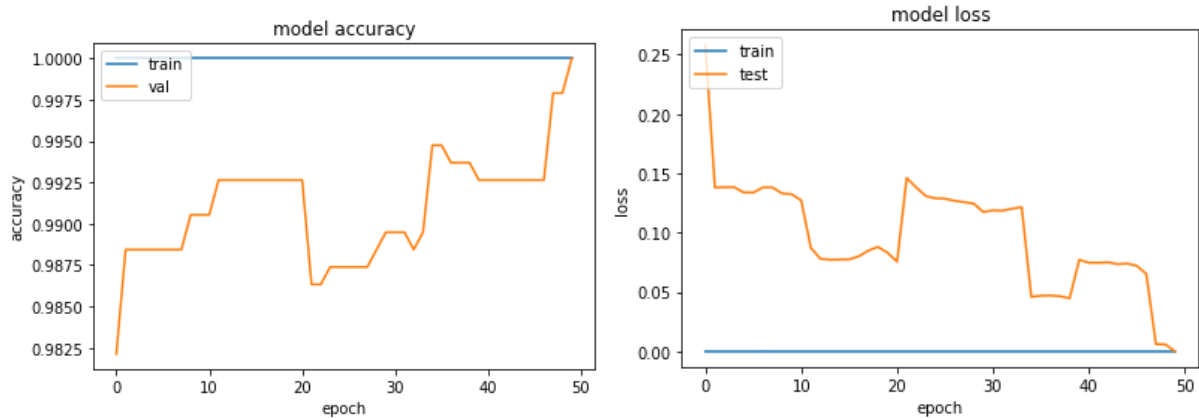
Mô hình được huấn luyện với 50 epochs trên môi trường Google Colab (card đồ họa Tesla K80). Toàn bộ quá trình được xây dựng và huấn luyện sử dụng thư viện Keras.



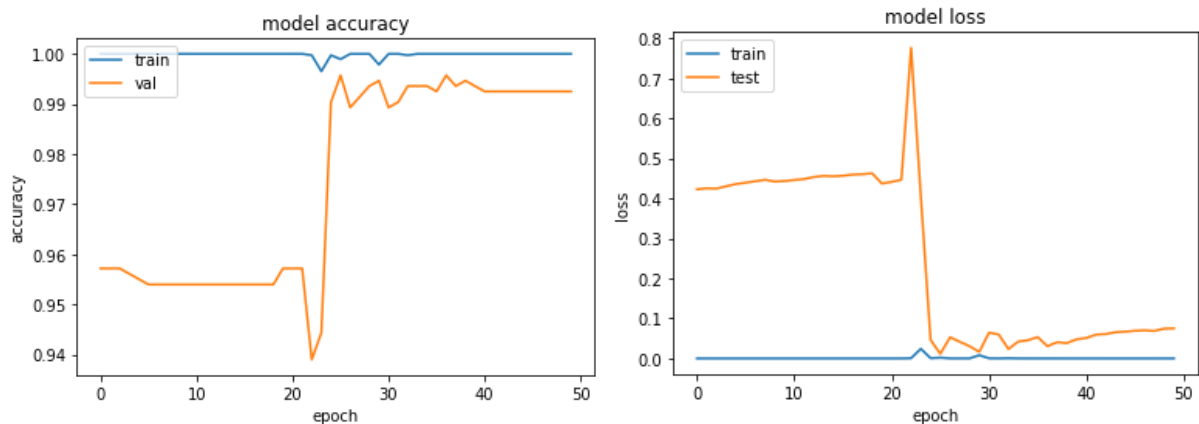
Kiến trúc mô hình một tay (trái) và mô hình hai tay (phải).

* Ban đầu, nhóm tác giả sử dụng một mô hình cho việc nhận diện ký tự diễn tả bằng hai tay và một tay. Qua thử nghiệm cho thấy, đối với ký hiệu sử dụng một tay, thông tin từ bàn tay còn lại là không cần thiết. Điều này có nghĩa là dữ liệu đó chỉ làm nhiễu tập dữ liệu, nhóm tác giả đã loại bỏ tất cả các tọa độ từ bàn tay còn lại và độ chính xác thu được tăng lên đáng kể.

1.4. Đánh giá mô hình

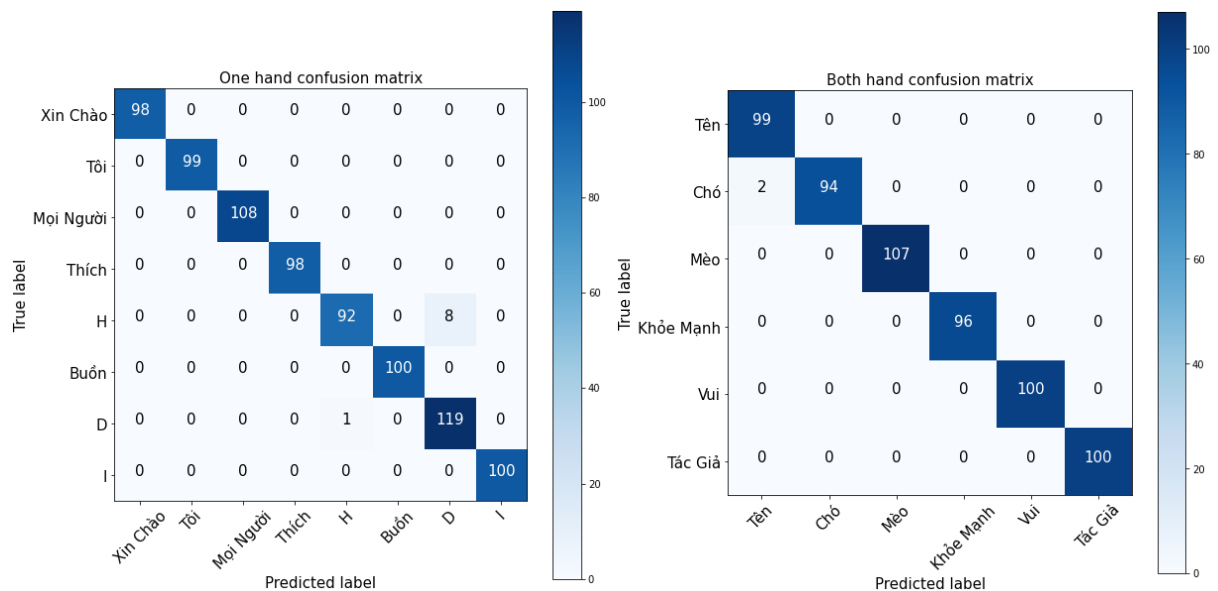


Biểu đồ accuracy và loss của model một tay



Biểu đồ accuracy và loss của model hai tay

Mô hình được thử nghiệm trên tập dữ liệu test (khoảng 100 bộ mỗi từ) đạt được độ chính xác tương đối tốt với độ chính xác 0.9987 (đối với model một tay) và 0.9966 (đối với model hai tay).



Confusion matrix

2. Chuyển văn bản thành giọng nói và nhận diện giọng nói

Text to speech và Speech to text là hai phần quan trọng đóng vai trò như phương thức giao tiếp giữa chương trình và người dùng. Sau quá trình chọn lọc thì dự án sử dụng thư viện được hỗ trợ bởi Python cung cấp trình tạo giọng nói và nhận diện giọng nói tốt với độ ổn định.

- Text to speech giúp biểu đạt thông tin về ký hiệu mà người sử dụng muốn truyền đạt tới đối phương.
- Speech to text giúp chuyển thông tin lời nói của đối phương thành văn bản, đưa vào giao diện người dùng (UI) và hiển thị lên kính.

3. Giao diện

Dự án sử dụng thư viện Tkinter được hỗ trợ bởi Python để thiết kế giao diện cho người sử dụng. Giao diện bao gồm 2 thông tin cơ bản: Ký tự người sử dụng muốn diễn tả và thông tin lời nói từ đối phương.



Hình ảnh của giao diện khi đưa lên kính

V. Kết luận



Hình ảnh người sử dụng thiết bị

Hiện tại thiết bị chỉ ở mức thử nghiệm. Tuy nhiên, thiết bị có rất nhiều tính năng hữu ích cho người khiếm thính. Các nghiên cứu và việc ứng dụng các nghiên cứu vào thiết bị với mục tiêu giúp ích cho xã hội nói chung và người khiếm thính nói riêng bằng những công nghệ của thời đại 4.0, thuần Việt, cho người Việt.

Thiết bị sẽ không ngừng phát triển, ngày càng chuyên sâu và tối ưu hơn, hoàn thiện hơn để có thể giúp ích cho họ, giúp họ dễ dàng hòa nhập với cuộc sống thường ngày hơn. Mã nguồn sẽ được mở cho cộng đồng cùng phát triển và xây dựng vì lợi ích chung của những người khiếm thính Việt Nam nói riêng và cộng đồng nói chung.

1.Ưu điểm

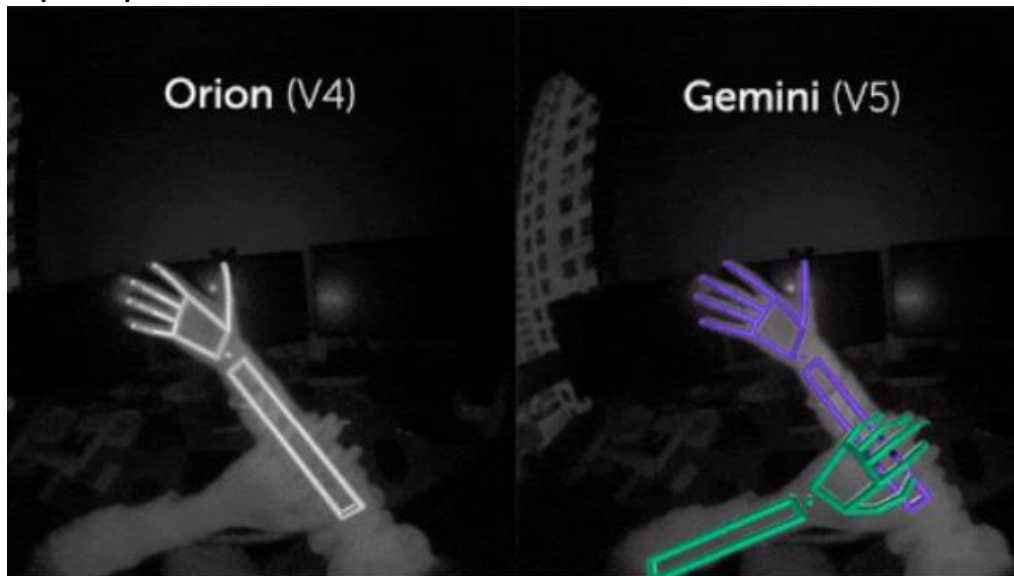
- Thiết bị đơn giản, dễ sử dụng.
- Tỷ lệ nhận diện và hiệu quả tương đối.
- Hạn chế được những nhược điểm của các dự án đi trước.
- Không sử dụng camera do đó không ảnh hưởng tới quyền riêng tư và bảo mật.

2.Khuyết điểm

- Vẫn còn tỉ lệ sai sót trong quá trình hoạt động vì hạn chế về dữ liệu và sự bất ổn định của Leap Motion.
- Vẫn còn hạn chế về dữ liệu nên không nhận diện được nhiều ký tự.
- Thiết bị chưa thật sự nhỏ gọn.
- Giá thành chưa phù hợp với số đông.
- Việc nhận diện các chuyển động phức tạp của Leap Motion trong ngôn ngữ ký hiệu không được tối ưu, gây sự sai sót.
- Hình ảnh hiển thị lên kính quá gần (10cm), không phù hợp với điểm cực cận của mắt người bình thường (37,5 cm).

3. Hướng phát triển

- Thu nhập thêm dữ liệu, mở rộng danh sách ký hiệu của thiết bị. Đồng thời cải thiện độ chính xác trong việc nhận diện.
- Cải tiến về mặt phần cứng, giúp thiết bị nhỏ gọn hơn.
- Thêm tính năng biên tập cơ sở dữ liệu. Người dùng có thể bổ sung thêm từ mới, hay chỉnh sửa quy định chuyển đổi theo ngôn ngữ địa phương. Làm đa dạng về dữ liệu, tăng độ chính xác.
- Mở rộng phạm vi thiết bị dùng cho việc học và dạy học ngôn ngữ ký hiệu.
- Nghiên cứu tích hợp thấu kính giúp đưa hình ảnh từ kính ra xa, giúp người sử dụng nhìn rõ hơn.
- Cập nhật phần mềm từ phía nhà sản xuất Leap Motion, giúp việc nhận diện được chính xác hơn.



Sự cải thiện về độ chính xác giữa “v4” và “v5”

VI. Tài liệu tham khảo

1. Giáo trình ngôn ngữ ký hiệu thực hành - Nhà xuất bản đại học sư phạm
2. Mấy vấn đề về cú pháp của ngôn ngữ ký hiệu Việt Nam - Thạc sĩ Nguyễn Thị Phương và Giáo sư tiến sĩ Nguyễn Đức Tôn.
3. Từ điển ngôn ngữ ký hiệu online: tudienngonngukyhieu.com
4. Bộ lọc nhiễu: Abraham Savitzky and Marcel JE Golay. 1964. Smoothing and differentiation of data by simplified least squares procedures. Analytical chemistry 36, 8 (1964), 1627–1639.
5. Tài liệu của Keras : keras.io/documentation/
6. Tài liệu về TensorFlow: tensorflow.org/api_docs/
7. Tổng hợp báo cáo khoa học: paperswithcode.com
8. Lưu trữ báo cáo khoa học Online của Đại học Cornell: arxiv.org/
9. Nguyễn Thanh Tuấn - Deep Learning cơ bản
10. Nhận diện hành động sử dụng LSTM machinelearningmastery.com/how-to-develop-rnn-models-for-human-activity-recognition-time-series-classification
11. API hỗ trợ Leap Motion : developer-archive.leapmotion.com