**SỬ DỤNG TÍNH NĂNG NHẬN DẠNG CỬ CHỈ ĐỂ ĐIỀU kHIỂN ỨNG DỤNG 3DVIEWER.**

Mục đích: Tài liệu phân tích và mô tả hệ thống nhận dạng cử chỉ bàn tay để điều khiển ứng dụng 3DViewer

Người viết: Trần Văn Quang Huy

Email: [tvq\_huy@brycen.com.vn](mailto:tvq_huy@brycen.com.vn)

Huế, tháng 8 – năm 2020.

LỊCH SỬ SỬA ĐỔI

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| # | Người viết | Phiên bản | Ngày | Ghi chú |
| 1 | Tran Van Quang Huy  [tvq\_huy@brycen.com.vn](mailto:tvq_huy@brycen.com.vn) | 1.0 | 2020/08/10 | Bắt đầu viết tài liệu |
| 2 | Tran Van Quang Huy  [tvq\_huy@brycen.com.vn](mailto:tvq_huy@brycen.com.vn) | 1.1 | 2020/08/11 | Bố trí lại cấu trúc tài liệu và thêm nội dung cơ bản. |
| 3 | Tran Van Quang Huy  [tvq\_huy@brycen.com.vn](mailto:tvq_huy@brycen.com.vn) | 1.2 | 2020/08/12 | Viết phần giới thiệu chung. |
| 4 | Tran Van Quang Huy  [tvq\_huy@brycen.com.vn](mailto:tvq_huy@brycen.com.vn) | 1.3 | 2020/08/17 | Thêm biểu đồ tuần tự, biểu đồ phân rã chức và và biểu đồ luồng. |
| 5 | Tran Van Quang Huy  [tvq\_huy@brycen.com.vn](mailto:tvq_huy@brycen.com.vn) | 1.4 | 2020/08/18 | Thêm bảng mô tả cử chỉ và mockup giao diện |
| 6 | Tran Van Quang Huy  [tvq\_huy@brycen.com.vn](mailto:tvq_huy@brycen.com.vn) | 1.5 | 2020/09/14 | Bổ sung mục công nghệ sử dụng |

MỤC LỤC

***Trang***

[LỊCH SỬ SỬA ĐỔI 1](#_Toc49784635)

[MỤC LỤC 2](#_Toc49784636)

[DANH MỤC HÌNH ẢNH 4](#_Toc49784637)

[PHẦN 1: GIỚI THIỆU CHUNG 5](#_Toc49784638)

[1.1 Đặt vấn đề 5](#_Toc49784639)

[1.1.1 Đặt vấn đề 5](#_Toc49784640)

[1.1.2 Lý do 5](#_Toc49784641)

[1.2 Mục tiêu của hệ thống 5](#_Toc49784642)

[1.2.1 Mục tiêu tổng quát 5](#_Toc49784643)

[1.2.2 Mục tiêu cụ thể 5](#_Toc49784644)

[1.3 Phạm vi thực hiện 6](#_Toc49784645)

[1.3.1 Đơn vị chủ quản 6](#_Toc49784646)

[1.3.2 Đối tượng phục vụ 6](#_Toc49784647)

[1.4 Khảo sát, phác thảo hệ thống 6](#_Toc49784648)

[PHẦN 2: PHÂN TÍCH THIẾT KẾ 8](#_Toc49784649)

[2.1 Yêu cầu chức năng 8](#_Toc49784650)

[2.1.1 Phân tích chức năng 8](#_Toc49784651)

[2.1.2 Biểu đồ phân rã chức năng 9](#_Toc49784652)

[2.1.3 Các phím tắt và hành động chuột muốn điều khiển 10](#_Toc49784653)

[2.2 Kiến trúc hệ thống 14](#_Toc49784654)

[2.3 Kiến trúc thành phần 15](#_Toc49784655)

[2.4 Biểu đồ tuần tự 16](#_Toc49784656)

[2.5 Giao diện ứng dụng 18](#_Toc49784657)

[2.6 Công nghệ sử dụng 18](#_Toc49784658)

[2.6.1 Ngôn ngữ lập trình Python 18](#_Toc49784659)

[2.6.2 Các thư viện chính 18](#_Toc49784660)

[2.6.2.1 Keras 18](#_Toc49784661)

[2.6.3 Mô hình nhận dạng cử chỉ 19](#_Toc49784662)

[PHẦN 3: TRIỂN KHAI 20](#_Toc49784663)

[3.1 Các phần mềm cần cho việc triển khai hệ thống 20](#_Toc49784664)

[3.1.1 Tác vụ thực hiện 20](#_Toc49784665)

[3.2 Đánh giá và lưu ý cho người dùng 20](#_Toc49784666)

[DANH MỤC TÀI LIỆU THAM KHẢO 21](#_Toc49784667)

DANH MỤC HÌNH ẢNH

***Trang***

[Hình 1. Mô tả cách thức hoạt động mức tổng quát. 7](#_Toc48659422)

[Hình 2. Biểu đồ thể hiện các chức năng chính có trong hệ thống. 9](#_Toc48659423)

[Hình 3. Bảng phím tắt và điều khiển chuột của ứng dụng 3DViewer. 10](#_Toc48659424)

[Hình 4. Biểu đồ luồng thực thi của hệ thống. 14](#_Toc48659425)

[Hình 5. Module nhận dạng hình ảnh. 15](#_Toc48659426)

[Hình 6. Module điều khiển của hệ thống. 15](#_Toc48659427)

[Hình 7. Biểu đồ tuần tự người dùng giao tiếp với 3DViewer thông qua cử chỉ. 16](#_Toc48659428)

[Hình 8. Biểu đồ tuần tự người dùng giao tiếp với màn điều chỉnh hành vi và hiển thị thông tin. 17](#_Toc48659429)

1. GIỚI THIỆU CHUNG
   1. Đặt vấn đề
      1. Đặt vấn đề

Ứng dụng 3DViewer của dự án FPGA đang sử dụng phím tắt từ bàn phím và chuột cho việc điều khiển chế độ hiển thị của hình ảnh 3D, và hiện tại dự án đang muốn mở rộng thêm một tính năng trong việc điều khiển ứng dụng thông qua cử chỉ tay từ người dùng, được trích xuất thông qua webcam trong thời gian thực.

* + 1. Lý do
* Phục vụ và hỗ trợ tốt cho công việc của dự án FPGA.
* Mở rộng thêm tính năng, tạo điều kiện thuận lợi, khả thi để triển khai dự án thành công sau này.
  1. Mục tiêu của hệ thống
     1. Mục tiêu tổng quát
* Sử dụng tính năng nhận dạng cử chỉ tay để thay đổi giao diện dữ liệu đám mây điểm – Point Cloud Data – trong 3DViewer thay cho việc sử dụng chuột và bàn phím, nhằm cung cấp thêm một cách thức điều khiển ứng dụng.
  + 1. Mục tiêu cụ thể
* Ứng dụng được mô hình nhận dạng cử chỉ vào thực tế, cụ thể là sử dụng nhận dạng cử chỉ phục vụ cho mục đích điều khiển.
* Phối hợp và kết nối giữa những module riêng lẻ lại với nhau.
* Mở rộng khả năng tương tác giữa người và ứng dụng 3DViewer.
  1. Phạm vi thực hiện
     1. Đơn vị chủ quản

Công ty TNHH MTV Brycen Việt Nam.

* + 1. Đối tượng phục vụ

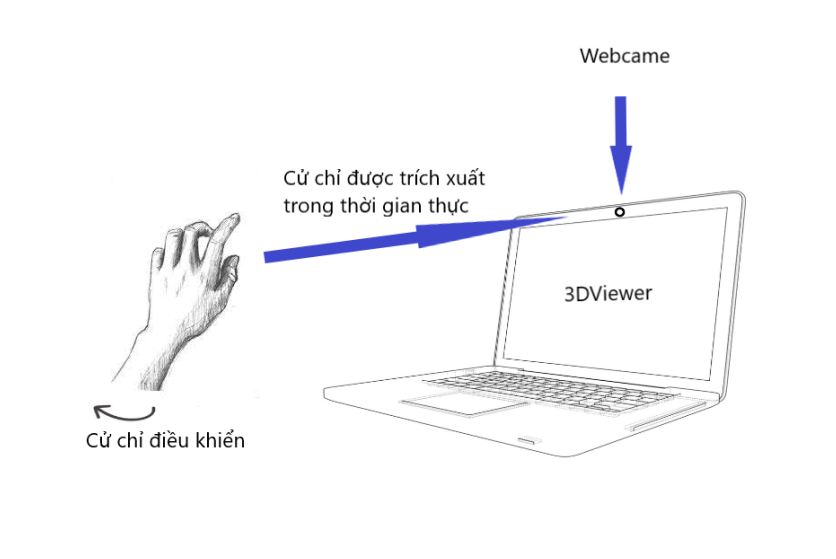
Phục vụ cho ứng dụng của dự án FPGA và các phòng ban có nhu cầu liên quan.

* 1. Khảo sát, phác thảo hệ thống

Đáp ứng được tác vụ thay thế điều khiển ứng dụng 3DViewer bằng cử chỉ tay thông qua tín hiệu hình ảnh được thu nhận từ webcam máy tính, đảm bảo thuận tiện và dễ sử dụng cho người dùng.

Chức năng của hệ thống có thể được mô tả phác thảo như sau:

1. Tín hiệu cử chỉ của người dùng được thu nhận trực tiếp thông qua webcam máy tính trong vùng được chỉ định (các mô tả về cử chỉ tương ứng với chức năng thực hiện sẽ được thể hiện trên màn hình cho người dùng dễ nắm bắt và sử dụng)
2. Hình ảnh nhận được sau khi trích xuất sẽ đi qua một vài bước tiền xử lý cơ bản rồi được gửi đến mô hình dự đoán cử chỉ để đưa ra dự đoán.
3. Dự đoán được ra sẽ tương ứng với một hành vi đã được quy định trước đó, sẽ được hệ thống đọc và tiến hành điều khiển chuột/bàn phím tương ứng đối với ứng dụng 3DViewer.



Hình . Mô tả cách thức hoạt động mức tổng quát.

Khó khăn của dự án:

* Việc tương thích giữa những module khác nhau thường dễ phát sinh lỗi.
* Cần đảm bảo tốc độ gửi nhận tín hiệu hợp lý để không tạo cảm giác mất kiểm soát cho người dùng.
* Nắm bắt và kiểm soát tốt cử chỉ tay là nhiệm vụ tương đối khó trong việc xác định điểm bắt đầu và điểm kết thúc của một cử chỉ trong thời gian thực.
* Việc truyền tải và thực thi quá nhiều tầng xử lý sẽ dẫn đến độ trễ trong lúc thực thi của hệ thống.

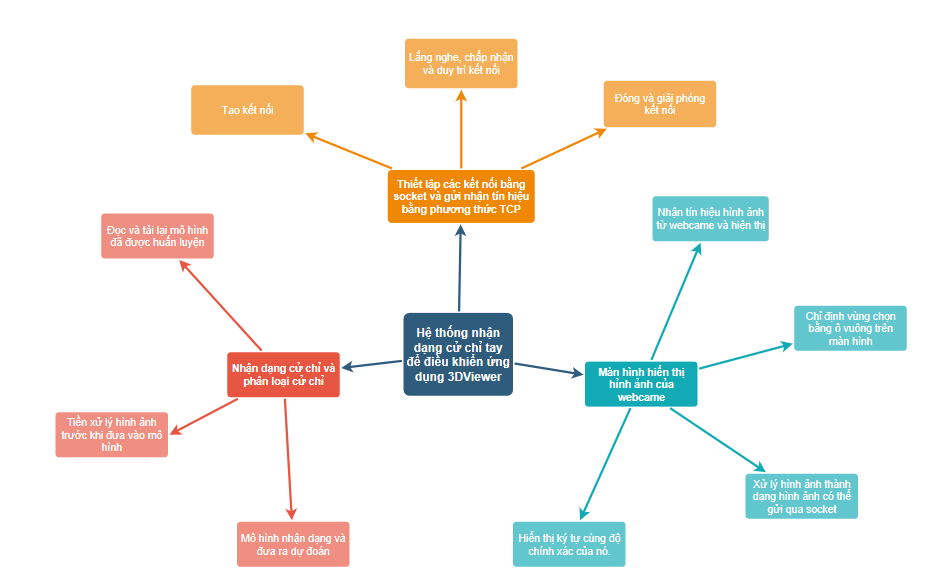
Thuận lợi:

* Mô hình đã dự đoán đã làm việc rất tốt trong môi trường thử nghiệm thực tế.
* Có nhiều thư viện hỗ trợ trong việc liên kết gửi nhận tín hiệu điều khiển thông qua kết nối socket.

1. PHÂN TÍCH THIẾT KẾ
   1. Yêu cầu chức năng
      1. Phân tích chức năng

Chức năng chính của hệ thống là đọc và nhận dạng đúng cử chỉ người dùng gửi đến, từ đó suy ngược ra hành động người dùng muốn tác động lên ứng dụng 3DViewer và thực thi hành động tương ứng.

* Ràng buộc:
* Chỉ hoạt động hiệu quả đối với tay trái của người dùng.
* Cử chỉ được thu nhận trong vùng cho phép và đúng góc nhìn trực diện.
* Cử chỉ sử dụng chỉ là cử chỉ đơn, không nên sử dụng một lúc nhiều cử chỉ.
* Chỉ nên sử dụng các cử chỉ trong bảng mô tả, không nên sáng tạo cử chỉ để điều khiển ứng dụng.
* Nên sử dụng trong môi trường đủ ánh sáng.
* Yêu cầu:
* Tốc độ nhận dạng và thực thi hành động phải tương thích với hành động trong thực tế, không quá nhanh và cũng không quá chậm so với cách sử dụng bàn phím và chuột thông thường.
* Các phím hay cử chỉ nên chọn sao cho có thể miêu tả một phần hành động muốn thực thi, làm cho người dùng dễ sử dụng, dễ liên tưởng.
* Không nên sử dụng các cử chỉ phức tạp khó biểu diễn.
  + 1. Biểu đồ phân rã chức năng

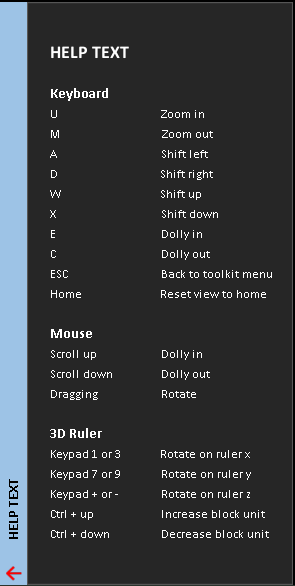


Hình . Biểu đồ thể hiện các chức năng chính có trong hệ thống.

Các chức năng của hệ thống gồm:

* Hiển thị và cung cấp thông tin cho người dùng thông qua màn hình và hình ảnh được thu nhận từ webcam máy tính.
* Mô hình dự đoán sẽ có nhiệm vụ tiếp nhận hình ảnh đã được tiền xử lý và đưa ra dự đoán tương ứng. Mô hình này đã được huấn luyến để có thể nhận biết phát hiện cử chỉ trước đó
* Chức năng kết nối và gửi nhận tín hiệu với ứng dụng 3DViewer sẽ đóng vai trò cầu nối giao tiếp giữa hệ thống và ứng dụng theo mô hình client-server kết nối thông qua socket sử dụng phương thức TCP để truyền nhận tín hiệu. Trong đó 3DViewer đóng vai trò là client và server là nơi thực thi phần đưa ra dự đoán từ hình ảnh mà nhận được từ 3DViewer.
  + 1. Các phím tắt và hành động chuột muốn điều khiển

Các phím tắt được ứng dụng 3DViewer sử dụng trong việc điều khiển hiển thị hình ảnh 3D.



Hình . Bảng phím tắt và điều khiển chuột của ứng dụng 3DViewer.

Bộ ngôn ngữ cử chỉ của người câm điếc có thể biểu diến được tất cả các chữ cái trong bảng chữ cái thành cử chỉ, bảng dưới đây sẽ mô tả các cử chỉ tương ứng biểu diễn cho các chữ cái cần sử dụng:

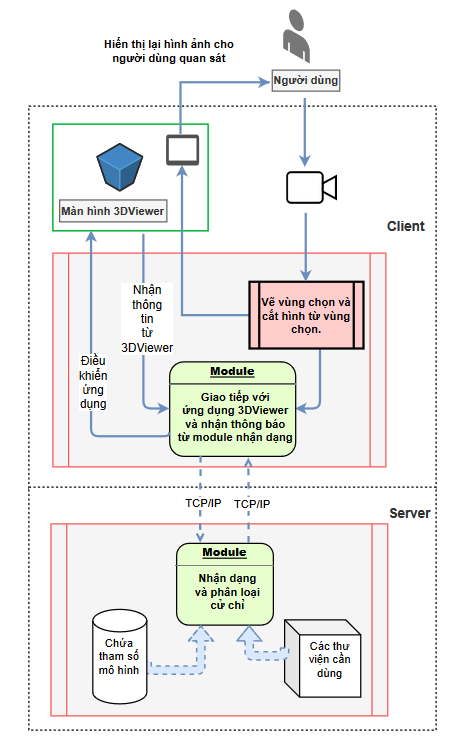
Bảng 1. Hình ảnh cử chỉ tương ứng với chữ cái.

|  |  |
| --- | --- |
| **Phím tắt và hành động chuột** | **Hình ảnh cử chỉ tương ứng** |
| + Phím U: phóng to hình ảnh giữ nguyên góc nhìn. |  |
| + Phím M: thu nhỏ hình ảnh giữ nguyên góc nhìn. |  |
| + Phím A: di chuyển hình ảnh sang trái. |  |
| + Phím D: di chuyển hình ảnh sang phải. |  |
| + Phím W: di chuyển hình ảnh lên trên. |  |
| + Phím X: di chuyển hình ảnh xuống dưới. |  |
| + Phím E: phóng to hình ảnh theo phạm vi quan sát |  |
| + Phím C: thu nhỏ hình ảnh theo phạm vi quan sát. |  |
| + Phím ESC: được thay thế bằng phím B (back), sẽ mở lại lại menu toolkit ban đầu. |  |
| + Phím H: đưa ứng trở về màn hình Home, màn hình khi bắt đầu của ứng dụng 3DViewer. |  |
| + Phím L: nhấn chuột trái và di chuyển qua bên trái một khoảng 5 pixel trên màn hình. |  |
| + Phím R: nhấn trái và di chuyển qua bên phải một khoảng 5 pixel trên màn hình. |  |
| + Phím V: nhấn chuột trái di chuyển xuống một khoảng 5 pixel trên màn hình. |  |
| + Phím Y: nhấn chuột trái di chuyển lên một khoảng 5 pixel trên màn hình. |  |

*Lưu ý :*

* Có hai hành động Dolly in và Dolly out đã được điều khiển tốt thông qua cử chỉ tương tác đối với bàn phím nên không cần phải lặp lại trường hợp đó ở chuột.
* Hành động nhấn và kéo chuột sẽ được tái hiện lại theo bốn hướng cố định là lên xuống, trái, phải tương ứng với bốn phím Y, V, L, R.
  1. Kiến trúc hệ thống

Kiến trúc tổng quan của hệ thống:



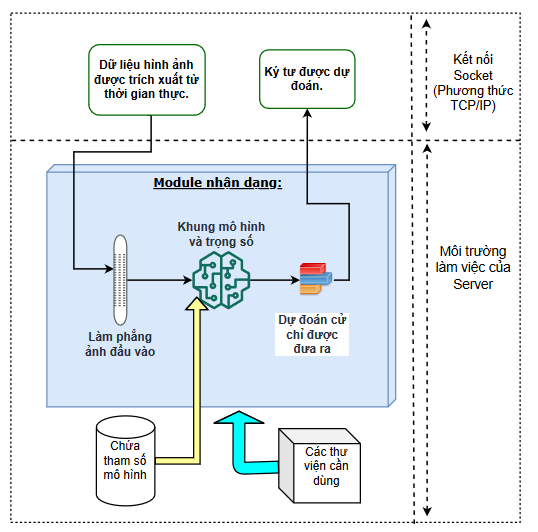
Hình . Biểu đồ luồng thực thi của hệ thống.

Hệ thống sẽ được phân làm 2 phần chính theo mô hình client-server, giao tiếp thông qua socket bằng phương thức TCP/IP. Phía server sẽ nắm giữ phần lớn những duyết định đưa ra, còn ở cài client (hay 3DViewer) sẽ được tích hợp thêm module điều khiển và giao tiếp, sẽ là nơi gửi nhận dữ liệu với server và thực thi quyết định được đưa ra ở server.

* 1. Kiến trúc thành phần

Kiến trúc của mỗi module đơn lẻ trong hệ thống:

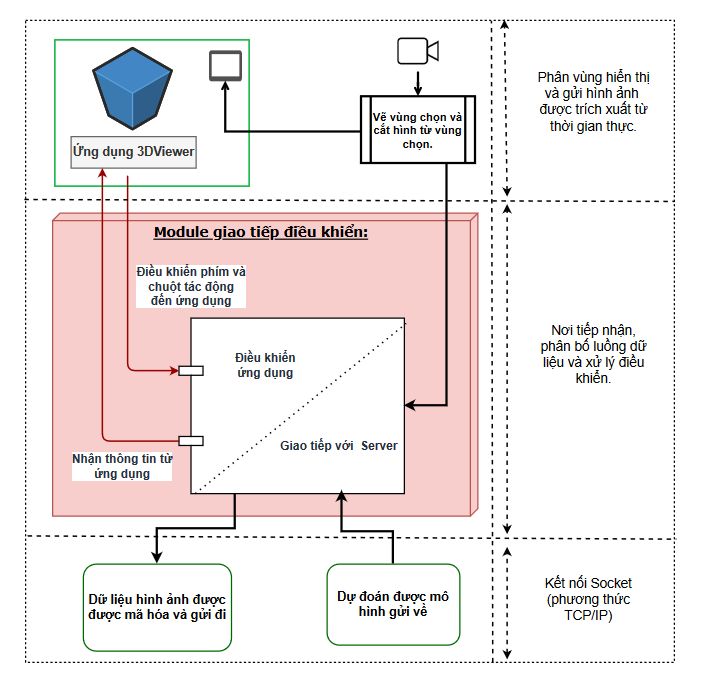
* Module nhận dạng ở server:



Hình . Module nhận dạng hình ảnh.

Đây sẽ là module chính của hệ thống, là nơi xử lý các nhiệm vụ phức tạp về nhận dạng và phát hiện các đặc trưng của hình ảnh để có thể đưa ra dự đoán chính xác cho hệ thống. Mọi tín hiệu gửi đến và truyền đi đều được thực hiện trong thời gian thực, nên cần quan tâm đến tốc độ thực thi và sự toàn vẹn của dữ liệu trong suốt quá trình trao đổi.

* Module giao tiếp và điều khiển:

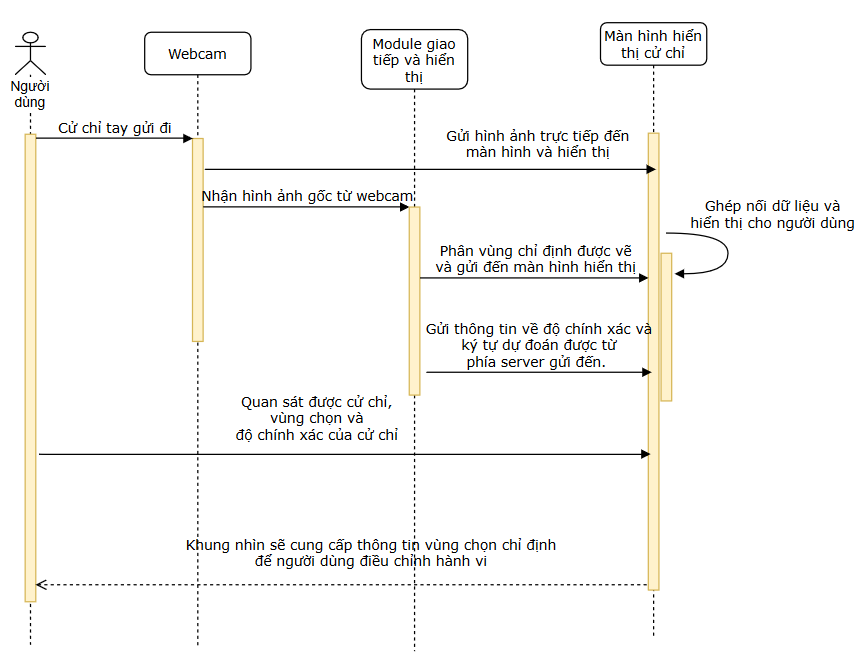


Hình . Module điều khiển của hệ thống.

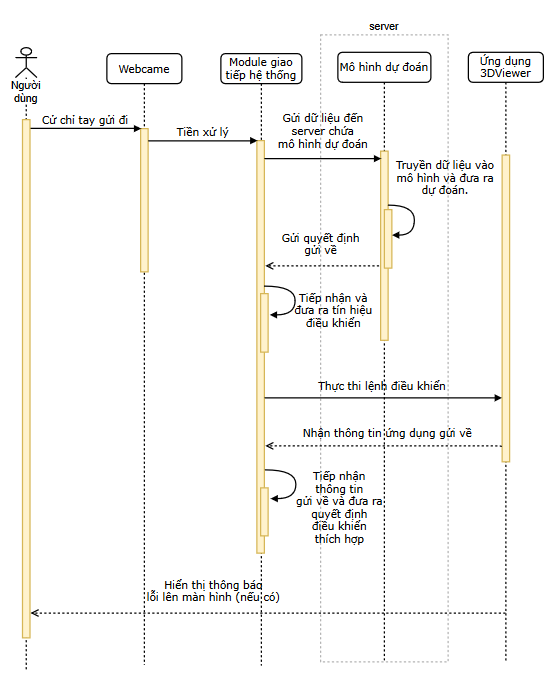
Module là nơi xử lý hầu hết các sự kiện điều khiển tác động đến ứng dụng 3DViewer, luồng dữ liệu chính được đi qua, được đón nhận và xử lý thay đổi trong thời gian thực.

* 1. Biểu đồ tuần tự

Mô tả hoạt động của người dùng tác động lên hệ thống được hiển thị qua màn hình chính của 3DViewer trong đó có vùng hiển thị cử chỉ người dùng giúp người điều khiển có thể tự điều chỉnh hành vi và vị trí của bàn tay sao cho hợp lý.



Hình . Biểu đồ tuần tự người dùng giao tiếp với màn điều chỉnh hành vi và hiển thị thông tin.

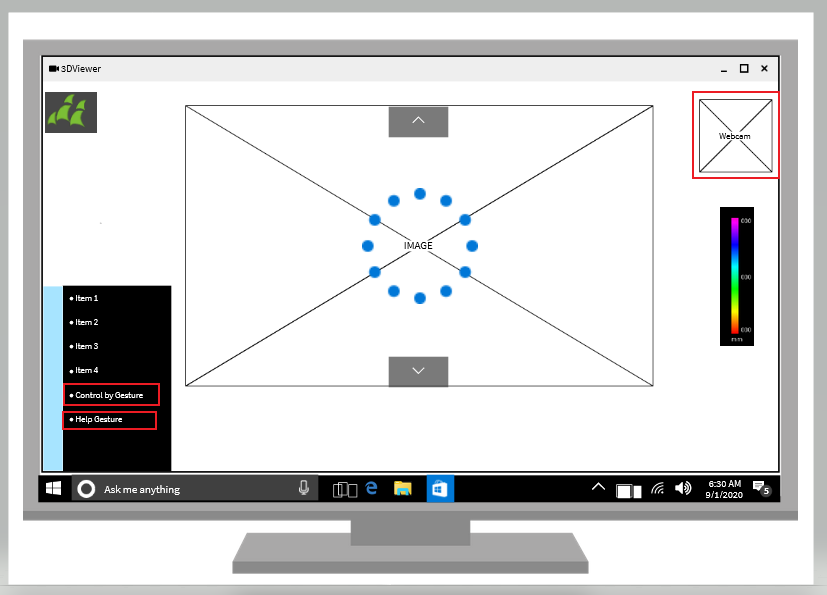


Hình . Biểu đồ tuần tự người dùng giao tiếp với

3DViewer thông qua cử chỉ.

Biểu đồ trên miêu tả những sự kiện và hiển thị mà người dùng giao tiếp với màn hình 3DViewer, luồng đi của dữ liệu và việc các quyết định điều khiển được đưa ra và điều chỉnh trong suốt quá trình thực thi của hệ thống.

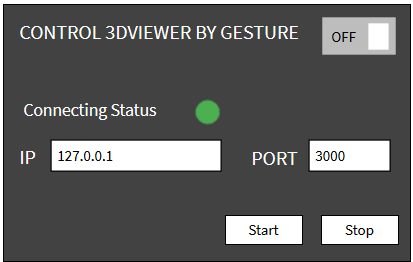
* 1. Giao diện ứng dụng
     1. Giao diện tổng quan



Hình . Giao diện 3DViewer được tích hợp thêm chức năng điều khiển bằng cử chỉ tay.

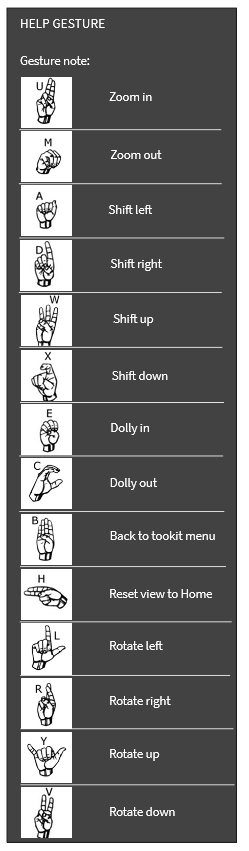
* + 1. Giao diện bảng điều khiển kết nối với Server

Đây là bảng điều khiển xuất hiện sao khi nhấn vào chức năng **Control by Gesture:**



Hình . Bảng thiết lập kết nối với Server.

* + 1. Giao diện bảng hiển thị cử chỉ trợ giúp



Kéo thả chuột

Hình . Bảng Help Gesture hiển thị cử chỉ trợ giúp.

* 1. Công nghệ sử dụng
     1. Ngôn ngữ và phần mềm hỗ trợ

Việc phát triển và tích hợp chức năng được phát triển dựa trên 2 ngôn ngữ lập trình Python và C# cùng với các công cụ được nêu ở bảng bên dưới:

|  |  |
| --- | --- |
| **Ngôn ngữ/ Công cụ** | **Phiên bản** |
| Python | 3.7.6 (x64) |
| C# | 7.3 (x64) |
| Hệ điều hành | Window 10 Pro (x64) |
| Unity | 2019.3.9f1 (x64) |
| Microsoft Visual Studio Community 2019 | 16.7.2 (x64) |
| Visual Studion Code | 1.49.0 (x64) |
| Anconda | 4.8.4 |
| Jupyter NoteBook | 6.1.1 |
| Spyder | 4.1.4 |

Ngoài ra, trong quá trình huấn luyện mô hình có sử dụng Colab – là một dịch đám mây miễn phí, cho phép viết và thực thi code Python ngay trên trình duyệt và có hỗ trợ GPU.

* + 1. Các thư viện chính

Một số thư viện chính được sử dụng cho mô hình nhận dạng cử chỉ thông qua hình ảnh:

|  |  |
| --- | --- |
| **Thư viện** | **Phiên bản** |
| Tensorflow | 2.1.0 |
| Keras | 2.3.1 |
| Numpy | 1.18.5 |
| OpenCV (opencv-python) | 3.4.2.16 |

* + 1. Mô hình nhận dạng cử chỉ

Mô hình nhận dạng hình hay mô hình phân loại hình ảnh được xây dựng trên mô hình mạng nơ-ron tích hợp chập và triển khai bằng kiến trúc VGG16 – một trong những kiến trúc mạng nơ-ron hiện đại và hoạt động hiệu quả đối với nhiệm vụ phân loại ảnh.



Hình . Kiến trúc mạng VGG16.

Quá huấn luyện luyện sử dụng bộ dữ liệu từ Kaggle và một số hình tự chụp, được điều chỉnh theo nhiệm vụ hướng đến và mục tiêu muốn đạt được khi dùng cử chỉ để điều khiển ứng dụng 3DViewer.

1. TRIỂN KHAI
   1. Các phần mềm cần cho việc triển khai hệ thống

(Các bước cái đặt từ tổng quan đến chi tiết)

Dự trù sẽ tích hợp thẳng các thư viện, module vào ứng dụng 3DViewer.

* + 1. Tác vụ thực hiện

Người dùng nên sử dụng tay trái để điều khiển ứng dụng 3DViewer vì mô hình được huấn luyện và hoạt động tốt đối với cử chỉ được biểu diễn bằng tay trái

* 1. Đánh giá và lưu ý cho người dùng

Mô hình vẫn có thể bị kích hoạt nếu gặp những ảnh có các đặc trưng thích hợp

DANH MỤC TÀI LIỆU THAM KHẢO