ISSN: 1859-2171
e-ISSN: 2615-9562
TNU Journal of Science and Technology 225(06): 135 - 140

ỨNG DỤNG THỊ GIÁC MÁY TÍNH VÀ CÔNG NGHỆ TRÍ TUỆ NHÂN TẠO TRONG HỆ THỐNG ĐIỀU KHIỂN CÁNH TAY ROBOT

Roãn Văn Hóa*, Đinh Thọ Long
Trường Đại học Kinh tế - Kỹ thuật Công nghiệp

TÓM TẮT

Trong bài báo này, tác giả trình bày một hệ thống điều khiển cánh tay robot bằng cách nhận dạng cử chỉ tay từ người điều khiển. Hệ thống dựa trên ba bước chính: xác định vị trí cử chỉ tay trên hình ảnh nhận được, xác định đường viền của cử chỉ tay và nhận diện cử chỉ này sử dụng mạng thần kinh nhân tạo và công nghệ học sâu (Deep Learning). Việc sử dụng trích xuất vùng quan tâm và phát hiện đường viền giúp giảm khối lượng tính toán, từ đó tăng tốc quá trình nhận dạng cử chỉ tay, giúp cánh tay robot có thể thực hiện thao tác theo thời gian thực. Kết quả thực nghiệm cho thấy hiệu quả tích cực của phương pháp được đề xuất.

Từ khóa: *Trí tuệ nhân tạo; công nghệ học sâu; hệ thống điều khiển cánh tay robot; thị giác máy tính; phát hiện cạnh.*

Ngày nhận bài: 17/3/2020; Ngày hoàn thiện: 27/4/2020; Ngày đăng: 11/5/2020

ROBOTIC ARM CONTROL BY USING COMPUTER VISION ALGORITHMS WITH CONVOLUTIONAL NEURAL NETWORK

Roan Van Hoa*, Dinh Tho Long
University of Economics - Technology Industrial

ABSTRACT

In this paper, we present a robotic arm control system by recognizing hand gestures from the operator. The system is based on three main steps: locating hand gestures on received images from webcam, determining the contours of hand gestures and recognizing these gestures using artificial neural networks and Deep Learning technology. The use of area ripping and contour detection reduces the computational weight, thereby speeding up the hand gesture recognition process, enabling the robotic arm to perform real-time operations. Experimental results show the positive effect of the proposed method.

Keywords: *Artificial intelligence; deep learning technology; robot arm control system; computer vision; edge detection.*

Received: 17/3/2020; Revised: 27/4/2020; Published: 11/5/2020

* Corresponding author. *Email: rvhoa@uneti.edu.vn*

1. Giới thiệu

Trong thời đại công nghệ 4.0, công nghệ Trí tuệ nhân tạo ngày càng phát triển với rất nhiều các ứng dụng trong đời sống thực tế. Một trong các ứng dụng điển hình của công nghệ này, đó là trong lĩnh vực Thị giác máy tính, xử lý hình ảnh và nhận diện hình thái. Các hệ thống như vậy, khi được tích hợp vào thao tác điều khiển robot, được sử dụng rộng rãi trong các hoạt động lắp ráp tự động, hình thành các hệ thống có thể hoạt động trong môi trường có cấu trúc và không cấu trúc, thông qua việc sử dụng các cơ chế phản hồi cảm giác tiên tiến. Các hệ thống này cũng có thể tự động đưa ra quyết định thông qua việc sử dụng các thuật toán tự học (learning phase) và lý luận.

Một trong những hệ thống phổ biến, được tập trung nghiên cứu trong thời gian gần đây đó là các cánh tay robot tích hợp hệ thống điều khiển chuyển động có kiểm soát, thông qua cử chỉ tay. Các hệ thống này được tích hợp chức năng phân tích tọa độ, xử lý trong thời gian thực để tăng hiệu quả của hệ thống. Phương pháp được chọn và triển khai là chụp và phát hiện các vùng quan tâm trong khung hình, được thực hiện bằng kỹ thuật kết hợp tính năng điểm (Point Feature Matching). Bên cạnh đó, tác giả cũng kết hợp giảm nhiễu trong quá trình thu nhận hình ảnh được bằng việc sử dụng và so sánh bốn kỹ thuật lọc hình ảnh: Canny, Sobel, Prewitt và Roberts. Bước cuối cùng đó là thực hiện phân loại hình ảnh bởi công nghệ Trí tuệ nhân tạo, bao gồm một mạng thần kinh nhân tạo Convolutional Neural Network (CNN).

Phương pháp này đảm bảo nhận dạng toàn bộ hình ảnh trong khung hình, thỏa mãn các giả định được đặt ra trong mô phỏng robot. Ngoài ra, cấu trúc được phát triển cho phép cánh tay robot có thể duy trì hoặc thay đổi sự hình thành các quỹ đạo xác định và thực hiện các nhiệm vụ thao tác riêng lẻ.

Bài báo này được chia thành 5 phần, trong đó phần 2 sẽ trình bày vấn đề được thảo luận và các giải pháp đã được thực hiện. Vấn đề này được giải quyết bằng thuật toán giới thiệu trong phần 3 cùng với phương pháp thống kê để xác minh độ tin cậy. Các kết quả sau khi áp dụng đề xuất được trình bày trong phần 4 và kết luận trong phần 5.

2. Tình hình nghiên cứu trong và ngoài nước

Trong [1], tác giả đã trình bày việc ứng dụng mạng thần kinh nhân tạo trong việc điều khiển cánh tay robot – một đối tượng động học phi tuyến. Bài báo cũng giới thiệu các bước và bản chất của việc thiết kế bộ điều khiển bằng mạng nơ-ron theo mô hình mẫu. Các kết quả mô phỏng đã thể hiện sự đúng đắn của phương pháp và mở ra khả năng ứng dụng vào thực tiễn.

Việc giúp máy tính nhận ra và hiểu ngôn ngữ cơ thể người, từ đó điều khiển các thành phần robot đã khá phổ biến, kỹ thuật này được đề cập và sử dụng bởi các tác giả trong bài báo [2]. Các tác giả đã trình bày một phương pháp để điều khiển robot, sử dụng cử chỉ tay, trong đó các cử chỉ được một mạng thần kinh nhân tạo dạng CNN nhận ra từ hình ảnh được chụp bằng camera gắn tại một vị trí cố định. Trong nghiên cứu của Rautary [4], tác giả đã được thực hiện một phân tích so sánh về việc sử dụng cử chỉ tay như sự tương tác giữa con người và máy móc. Tác giả nói rằng việc sử dụng cử chỉ tay mang đến một sự thay thế hấp dẫn và tự nhiên cho sự tương tác giữa máy tính và con người. Nhận dạng cử chỉ cũng được sử dụng để điều hướng các robot bốn chân, chẳng hạn như trong [6]. Tác giả sử dụng phân đoạn theo ngưỡng (Threshold Segmentation), biến đổi trung bình thích nghi liên tục (Continuously Adaptive Mean-Shift) và Restricted Boltzmann Machines để phân loại cử chỉ trong thời gian thực, từ đó đưa ra mệnh lệnh điều khiển cho các robot bốn chân.

Trong nghiên cứu được thực hiện bởi Parada [7], các tác giả đã sử dụng kỹ thuật nhận dạng cử chỉ tay trong lĩnh vực điều khiển ô tô bằng cách tạo ra một hệ thống giao diện cho phép

sử dụng các thiết bị tự động mà không bị phân tâm và do đó giảm số vụ tai nạn giao thông liên quan đến việc mất tập trung khi lái xe. Trong bài báo [8], tác giả Gupta cũng sử dụng kỹ thuật nhận dạng cử chỉ tay cho các giao diện trực quan trong ô tô. Trong bài báo này, các tác giả chỉ ra rằng các cử chỉ tay được thực hiện bằng tay trên vô lăng bánh xe hoặc gần với nó dẫn đến sự mất tập trung vật lý thấp.

Từ những nghiên cứu trên, bài báo này đã phát triển một phương pháp để di chuyển một cánh tay robot bằng cử chỉ tay, trong thời gian thực. Phương pháp này sẽ được trình bày chi tiết trong phần tiếp theo.

3. Thuật toán điều khiển robot thông qua hình ảnh

Trong phần này, cấu trúc của hệ thống được giới thiệu. Bước đầu tiên là thu thập và lưu trữ các cử chỉ tay, sau đó các hình ảnh này được phân tích bởi thuật toán và phân loại bằng các mạng thần kinh nhân tạo.

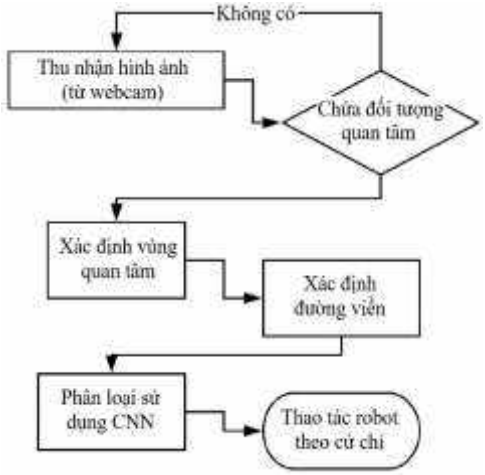
3.1. Cấu trúc hệ thống

Hệ thống sử dụng một webcam để thu và gửi hình ảnh về cho hệ thống máy tính. Sau đó, phương pháp nhận dạng sẽ dựa trên một chuỗi các bước sẽ cho phép theo dõi thời gian thực. Các bước này bắt đầu bằng các thuật toán để thiết lập một vùng quan tâm, phát hiện các cạnh và so sánh với các đặc điểm được xác định trong các phân loại.

Sau đó, từ cử chỉ được nhận diện, hệ thống sẽ truyền tín hiệu để điều khiển cánh tay robot thực hiện một số hành động nhất định dựa trên cử chỉ này. Một số thuật toán cũng được phát triển để hỗ trợ độ tin cậy hệ thống. Phương thức đề xuất được giới thiệu trên (hình 1).

3.2. Xác định khu vực quan tâm trên ảnh

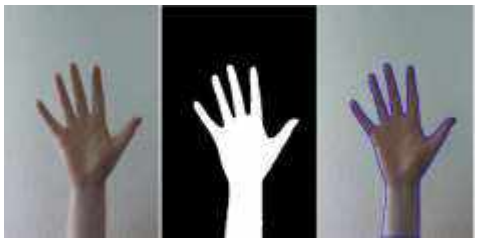
Phương pháp xác định các cử chỉ tay từ hình ảnh nhận được từ webcam được lập trình thử nghiệm sử dụng thư viện OpenCV và Node.js.



Hình 1. *Sơ đồ cấu trúc hệ thống*

Trước hết, hình ảnh được xử lý để tạo một mặt nạ nhị phân (binary mask) và đường viền của bàn tay. Sau đó, các hình ảnh được phân đoạn dựa trên màu da tay bằng cách sử dụng thao tác theo bậc. Từ đó, các khung hình được chuyển đổi từ định dạng BGR mặc định trong OpenCV sang không gian màu HLS (Hue, Lightness, Saturation). Kênh Hue mã hóa thông tin màu thực tế. Bằng cách này, phạm vi giá trị Hue thích hợp của da được tính toán và sau đó sử dụng để điều chỉnh các giá trị cho Độ bão hòa (Saturation) và Độ sáng (Lightness).

Cuối cùng, các hàm của OpenCV được áp dụng để tìm các đường viền của mặt nạ nhị phân (binary mask) của bàn tay. Một ví dụ của thuật toán này được mô tả trên (hình 2).



Hình 2. *Thuật toán xác định vùng quan tâm, mặt nạ nhị phân và đường viền của các cử chỉ tay. Thuật toán được lập trình và thử nghiệm sử dụng OpenCV*

Bốn kỹ thuật phổ biến trong phát hiện đường viền được sử dụng: Sobel, Roberts, Prewitt, và Canny [3]. Các bộ lọc của các thuật toán này được giới thiệu trong tài liệu [3] và đều được tích hợp trong OpenCV.

3.3. Tổng quan về cánh tay robot

Cánh tay robot như trên (hình 3) được sử dụng gồm bốn bậc tự do, mỗi liên kết có biên độ 180 độ. Để điều khiển cánh tay robot, tám cử chỉ tay được sử dụng như trên (bảng 1).

Bảng 1. *Tám cử chỉ tay sử dụng cho cánh tay robot*

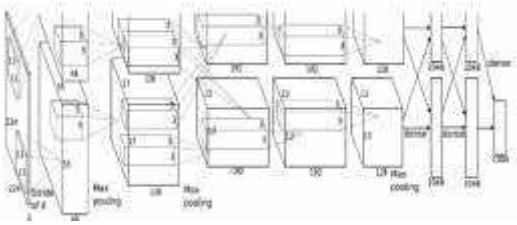
Cử chỉ	Khớp nối	Góc xoay (độ)
G1	1	0 - 180
G2	1	180 - 0
G3	2	0 - 160
G4	2	160 - 0
G5	3	0 - 120
G6	3	120 - 0
G7	4	0 - 150
G8	4	150 - 0

Hình 3. *Cánh tay robot bốn bậc tự do*

3.4. Bộ phân loại

Bước cuối cùng của hệ thống là đào tạo các bộ phân loại chịu trách nhiệm thực hiện nhận dạng cử chỉ. Thuật toán mạng thần kinh nhân tạo CNN được sử dụng vì các thuật toán này đã chứng minh được tính ưu việt trong các vấn đề phức tạp. Chính vì vậy, công nghệ học sâu với các mạng thần kinh nhân tạo ngày càng trở nên phổ biến, đặc biệt là trong lĩnh vực thị giác máy tính.

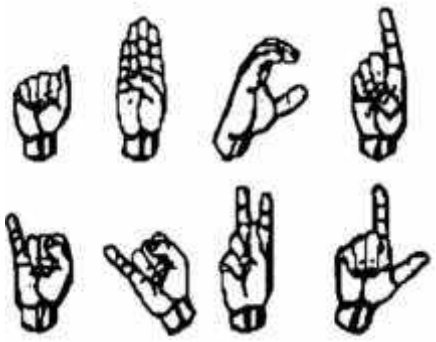
Kiến trúc CNN được sử dụng là AlexNet [5]. AlexNet nhận được đầu vào 227 x 227 pixel mỗi kênh. Trong lớp chập đầu tiên, nó sử dụng bộ lọc 11 x 11 x 3, trong lớp thứ hai là 5 x 5 x 3 và trong lớp thứ ba, 3 x 3 x 3. Ngoài ra lớp thứ ba, thứ tư và thứ năm được kết nối mà không sử dụng lớp gộp pooling. Cuối cùng, mạng có hai lớp được kết nối đầy đủ Fully Connected với 2048 nơ-ron mỗi lớp và một lớp đầu ra có 1000 nơ-ron, cũng chính là số lớp phân loại.



Hình 4. *Kiến trúc mạng nơ-ron nhân tạo AlexNet (Nguồn: [5])*

Đối với công việc này, kỹ thuật học chuyển giao **(**transfer learning**)** đã được áp dụng để tăng tốc quá trình đào tạo, sử dụng cấu trúc của mạng Alexnet, thay đổi lớp đầu ra thành 08 nơ-ron theo các loại cử chỉ được phân loại, do đó tác giả không phải đào tạo tất cả các trọng số của các lớp mạng. Nếu như vậy thì đây sẽ là một quá trình tốn kém, vì AlexNet là mô hình trọng số sử dụng tập hợp con gồm 1000 danh mục ImageNet (đây là một cơ sở dữ liệu về hình ảnh rất phổ biến hiện nay).

Tổng cộng có 800 hình ảnh đã được sử dụng, trong đó 60 phần trăm (480) trong số này được sử dụng trong quá trình đào tạo CNN, trong khi 40 phần trăm còn lại (320) được sử dụng để thử nghiệm và kiểm tra độ chính xác của bộ phân loại. Các hình ảnh được sử dụng cho đào tạo và kiểm tra CNN được điều chỉnh từ các công trình điều khiển sử dụng cử chỉ tay phổ biến, các cử chỉ này được hiển thị trong (hình 5).



Hình 5. *Một số cử chỉ tay được sử dụng trong bài báo này. Theo thứ tự tư trái qua phải, từ trên xuống dưới lần lượt là các cử chỉ từ G1 đến G8*

3.5. Thỏa thuận phân loại

Để phân tích chất lượng của việc phân loại, cần phải phân loại đối tượng này nhiều lần.

Giống như một công cụ phân loại, ma trận nhầm lẫn (confusion matrix) được sử dụng để cung cấp cơ sở để mô tả tính chính xác của phân loại và mô tả các lỗi, giúp tinh chỉnh phân loại. Từ một ma trận nhầm lẫn có thể rút ra một số biện pháp để tính toán độ chính xác của phân loại, và ở đây chỉ số Kappa được sử dụng để giải quyết vấn đề này.

Ma trận nhầm lẫn được hình thành bởi một mảng các ô vuông được sắp xếp theo hàng và cột biểu thị số lượng đơn vị mẫu của một loại, phân loại được suy ra từ thuật toán và so sánh với dữ liệu phân loại chính xác. Thông thường, bên dưới các cột là dữ liêu tham chiếu, các dữ liệu này được so sánh với dữ liệu phân loại được thể hiện trên các dòng.

Các chỉ số thu được từ ma trận nhầm lẫn là: độ chính xác tổng quan, độ chính xác của từng lớp, chỉ số Kappa và một số chỉ số khác. Tổng độ chính xác được tính bằng cách chia tổng đường chéo chính của ma trận lỗi x_{ii}, cho tổng số mẫu được thu thập *n*. Theo phương trình (1).

$$T = \frac{\sum_{i=1}^{a} x_{ii}}{n} \quad (1)$$

Phân phối độ chính xác trên các lớp riêng lẻ không được hiển thị trong tổng độ chính xác, tuy nhiên độ chính xác của một lớp riêng lẻ có thể có được bằng cách chia tổng số mẫu được phân loại chính xác trong danh mục đó cho tổng số mẫu trong danh mục đó.

Trong bài báo này, biện pháp Kappa được sử dụng để mô tả cường độ của thỏa thuận, dựa trên số lượng phản hồi phù hợp. Kappa là thước đo của thỏa thuận interobserver và đo lường mức độ thỏa thuận vượt quá những gì có thể xảy ra. Đây là một kỹ thuật đa biến rời rạc được sử dụng để đánh giá độ chính xác theo chủ đề và sử dụng tất cả các yếu tố của ma trận nhầm lẫn trong tính toán của nó. Hệ số Kappa (*K*) là thước đo của thỏa thuận thực tế (được biểu thị bằng các yếu tố đường chéo của ma trận nhầm lẫn) trừ đi thỏa thuận cơ hội (được biểu thị bằng tổng các tích của hàng và cột, không bao gồm các mục không được nhận dạng). Hệ số Kappa có thể được tính từ phương trình 2:

$$K = \frac{n\sum_{i=1}^{a} x_{ii} - \sum_{i=1}^{a} x_{+i}x_{ii}}{n^2 - \sum_{i=1}^{a} x_{+i}x_{ii}} \quad (2)$$

Thước đo thỏa thuận này có giá trị tối đa 1, trong đó giá trị 1 này đại diện cho tổng thỏa thuận và các giá trị gần bằng 0, cho biết không có thỏa thuận nào, hoặc thỏa thuận được tạo ra một cách chính xác do tình cờ. Giá trị cuối cùng của Kappa nhỏ hơn 0, âm, cho thấy rằng thỏa thuận được tìm thấy ít hơn mong đợi. Do đó, nó cho thấy sự bất đồng. Việc giải thích các giá trị Kappa được thể hiện trên (bảng 2).

Bảng 2. *Bảng giá trị Kappa*

Giá trị Kappa	Cấp độ thỏa thuận
< 0	Không thỏa thuận
0-0,19	Thỏa thuận kém
0,20-0,39	Thỏa thuận công bằng
0,40-0,59	Thỏa thuận vừa phải
0,60-0,79	Thỏa thuận đáng kể
0,80-1,00	Thỏa thuận gần như hoàn hảo

4. Kết quả thực nghiệm

Bảng 3. *Độ chính xác của phương pháp và thời gian đào tạo*

Phương pháp	Độ chính xác trung bình	Thời gian đào tạo (training) theo giây
Truyền thống	99,60%	161,30
Prewitt	95,60%	45,59
Roberts	94,00%	39,45
Sobel	94,00%	43,58
Canny	98,10%	44,26

Trong phần này, độ chính xác của phương pháp đề xuất được đánh giá bằng cách sử dụng kỹ thuật xác nhận tiêu chuẩn, trong đó độ chính xác của CNN được đo bằng phương pháp trình bày trong công thức 1. Hiệu suất của thuật toán CNN cho từng phương pháp đề xuất được trình bày trong (bảng 3). Quá trình học và đào tạo mạng thần kinh nhân tạo CNN được thực hiện trên card đồ họa NvidiaGeforce RTX 2080, với bộ xử lý 3072 lõi CUDA. Các tham số tương tự được sử dụng trên CNN trong tất cả các phương pháp

phát hiện cạnh. Tham số MaxEpochs (một Epoch tương ứng với một lần hoàn thành toàn bộ dữ liệu) được đặt là 15. Tham số Mini-Batch Sizetương ứng với số lượng quan sát được thực hiện ở mỗi lần lặp được đặt là 80.

Trong bảng 3, có thể xác minh rằng các ảnh gốc (có màu) cho độ chính xác tốt nhất là 99,60% trong thí nghiệm này. Tuy nhiên, thời gian đào tạo của mạng CNN lại cao hơn khoảng bốn lần so với toán tử Canny, ở vị trí thứ hai với độ chính xác 98,10%, tiếp theo là các phương pháp trích xuất cạnh Prewitt, Roberts và Sobel, với thời gian xử lý ngắn nhất cho phương pháp Roberts.

Trong (bảng 4), tác giả thể hiện kết quả phân tích phù hợp với chỉ số Kappa, trong đó giá trị *K* cho mỗi phương pháp thể hiện sự đồng thuận gần như hoàn hảo cho tất cả các phương pháp được sử dụng. Có thể thấy, phương pháp Canny mặc dù có độ chính xác và thỏa thuận thấp hơn không đáng kể so với phương pháp truyền thống nhưng có thời gian đào tạo thấp hơn rất nhiều.

Bảng 4. *Giá trị chỉ số Kappa*

Phương pháp	Thỏa thuận phân loại *K*
Truyền thống	0,9915
Prewitt	0,9523
Roberts	0,9571
Sobel	0,9523
Canny	0,9843

5. Kết luận

Bài báo này trình bày một hệ thống điều khiển cánh tay robot thông qua việc nhận diện các cử chỉ tay của người điều khiển bằng phương pháp trí tuệ nhân tạo. Độ chính xác của phương pháp được chứng minh là giảm khoảng bốn lần thời gian đào tạo của CNN nhờ việc giảm khối lượng dữ liệu thông qua ứng dụng bộ lọc để nhận biết đường viền, với độ chính xác giảm tương đối ít so với việc nhận dạng thông qua ảnh gốc.

Thuật toán cũng đưa ra giải pháp xử lý video trong thời gian thực, giúp việc đánh giá và chuyển các hành động của người vận hành thành hành động cho cánh tay robot được hiệu quả hơn. Trong thời gian tới, tác giả dự định tăng số lượng cử chỉ và hình ảnh, nghiên cứu sử dụng các cử chỉ tay gần với các ứng dụng trong công nghiệp. Bên cạnh đó, tác giả cũng dự định thử nghiệm các phương pháp xác định đường viền khác, để cải thiện độ chính xác của phương pháp.

TÀI LIỆU THAM KHẢO/ REFERENCES

[1]. H. C. Nguyen, “Research on the application of Neural Networks in identification and control of robotic arms-A nonlinear dynamic object,” *Journal of Science and Technology – University of Da Nang*, vol. 5, pp. 14-18, 2016.

[2]. A. Saraiva, R. Melo, V. Filipe, J. Sousa, N.M Fonseca Ferreira, and A. Valente, “Mobile multirobot manipulation by image recognition,” *International Journal of Systems Applications, Engineering Development*, vol. 12, pp. 63-68, 2018.

[3]. V. A. Nguyen, “Comparison of Edge Detection Techniques,” *Vietnam National University Journal of Science: Natural Sciences and Technology*, vol. 31, no. 2 pp. 1-7, 2015.

[4]. S. S. Rautaray and A. Agrawal, “Vision based hand gesture recognition for human computer interaction: a survey,” *Artificial Intelligence Review*, vol. 43, no. 1, pp. 1-54, 2015.

[5]. A. Krizhevsky, S. Ilya, and G. E. Hinton, “ImageNet Classification with Deep ConvolutionalNeural Networks,” *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097-1105, 2012.

[6]. G. B. Choudhary, and C. B. V. Ram, “Real time robotic arm control using hand gestures,” in High Performance Computing and Applications (ICHPCA), 2014 International Conference on. IEEE, 2014, pp. 1-3.

[7]. F. Parada-Loira, E. Gonz´alez-Agulla, and J. L. Alba-Castro, “Hand gestures to control infotainment equipment in cars,” in Intelligent Vehicles Symposium Proceedings, 2014 IEEE. IEEE, 2014, pp. 1-6.

[8]. S. Gupta, P. Molchanov, X. Yang, K. Kim, S. Tyree, and J. Kautz, “Towards selecting robust hand gestures for automotive interfaces,” in Intelligent Vehicles Symposium (IV), 2016 IEEE. IEEE, 2016, pp. 1350-1357.