GESTURE COMMANDS IN HOME ENVIRONMENT

S. Stankov
Technical university of Gabrovo

S. Ivanov
Technical university of Gabrovo

Abstract
For disabled people it is difficult to reach a place at home for to do something specific. Using a surveillance camera we capture the gesture of a person's hand for activating a specific command.

Keywords: DNN, Gesture Recognition, Linux

INTRODUCTION

Nowadays many places are automated using sensors, artificial intelligence and other technologies. In home environment disabled people are having difficulties in reaching some areas for switching a light or some other common action. This papers presents a home automation using a camera for command recognition.

TECHNOLOGIES FOR GESTURE RECOGNITION

Many technologies exist for gesture recognition – using kinetic sensor, brain activity, skin recognition and others. The most modern and reliable method is using a camera and DNN (deep neural networks) for detecting human gesture and execute based command type. Systems like this are used in areas such as automation for different objects or areas, virtual realities, distance control and others.

SYSTEM HARDWARE

The hardware is based on a Raspberry PI linux board. The motherboard features Quad Core 1.2GHz Broadcom BCM2837 64bit CPU, 1GB RAM it also has Bluetooth, Wireless LAN, 40 GPIO for general purposes, 4 USB ports, camera slot and a DSI display port. With the current specification various tests can be ran with different cameras such as USB or IP and the input data can be viewed on a HDMI or DSI display. The board supports various OS types such as : Ubuntu, Raspbian, Windows 10 IoT [1].

The video controller can support modern television resolution standards such as HD and Full HD as well as higher and lower resolutions of older camcorders. Can also generate 576i and 480i composite video signals for PAL-BGHID, PAL-M, PAL-N, NTSC and NTSC-J.

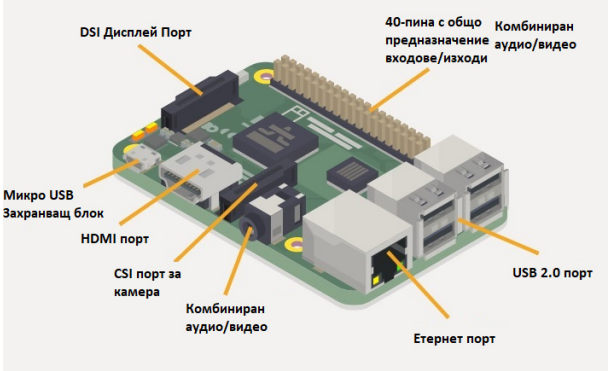

Fig. 1. *RASPBERRY PI 3 MODEL B*

PI NOIR CAMERA V2

The camera module can meet for high-definition video playback, as well as for still pictures. Support 1080p30, 720p60 and VGA90 video modes.
Video camera support h.264

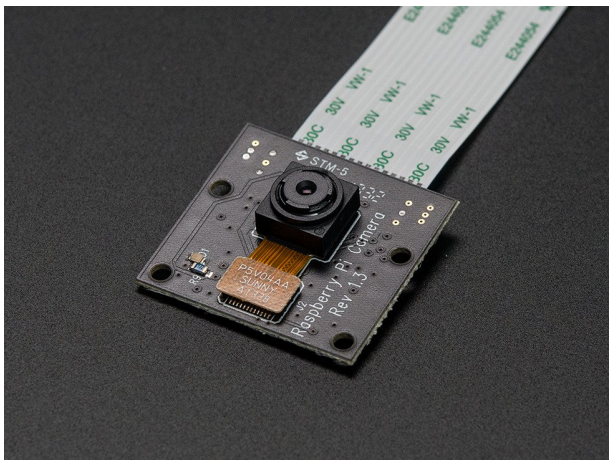

Fig. 2. *Camera Raspberry Pi Noir*

VPU UNIT

The Movidius stick is Vision Processing Unit(VPU) that uses specialized processor with high computing capacity to perform complex operations on static and dynamic data using artificial neural networks. For real-time video processing, it can derive and compare every information in the series of frames, which means that it performs calculations depending on values of image pixels. Thus, creating communication between the Raspberry Pi and the VPU as the Raspberry Pi transmitting data to the VPU for further processing it is received a significant acceleration of performance.

Fig. 3. *Movidius VPU*

TRAINING THE NEURAL NETWORK

A neural network has been trained using the MobileNet architecture which provides high speed for detecting objects and also it can be used on different platforms.

The training requires a preliminary database of photos from which it is necessary to extract the characteristics of the objects which have to be found after that.

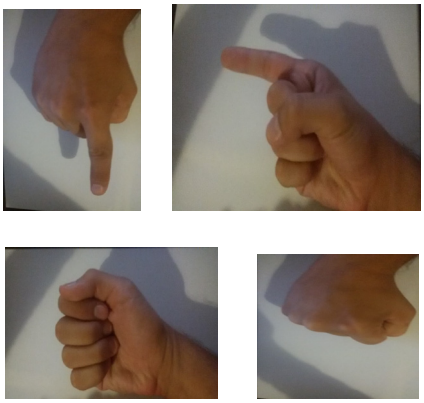

Fig. 4. *Examples of training images for neural network*

How high accuracy and depending on the architecture on the neural network different amount of images are requirered for reaching high accuracy, usually around 200 images per object. For training were used the machine learning library TensorFlow and python programming language, The neural network is able to recognize multiple objects and by doing software filtering which object is a gesture, human head or body it is possible to process commands based on the image data.

SOFTWARE TOOLS USED FOR IMPLEMENTATION OF ALGORITHM

To implement the presented algorithm are used several software technologies. For processing of images is used OpenCV [2]. This is an open source computer vision and machine learning library. It is designed to provide a common computer vision application infrastructure and accelerate machine perception. The library has more than 2,500 optimized algorithms that include a comprehensive set of classical and advanced computer visions and machine learning algorithms[3]. It has C ++, Python, Java and MATLAB interfaces and supports Windows, Linux, Android and Mac OS. OpenCV mostly relies on real-time vision applications. There are over 500 algorithms and about 10 times more features that compose or maintain these algorithms.

TensorFlow - TensorFlow is an end-to-end open source platform for machine learning. It has a comprehensive, flexible ecosystem of tools, libraries and community resources that lets researchers push the state-of-the-art in ML

and developers easily build and deploy ML powered applications.

ALGORITHM

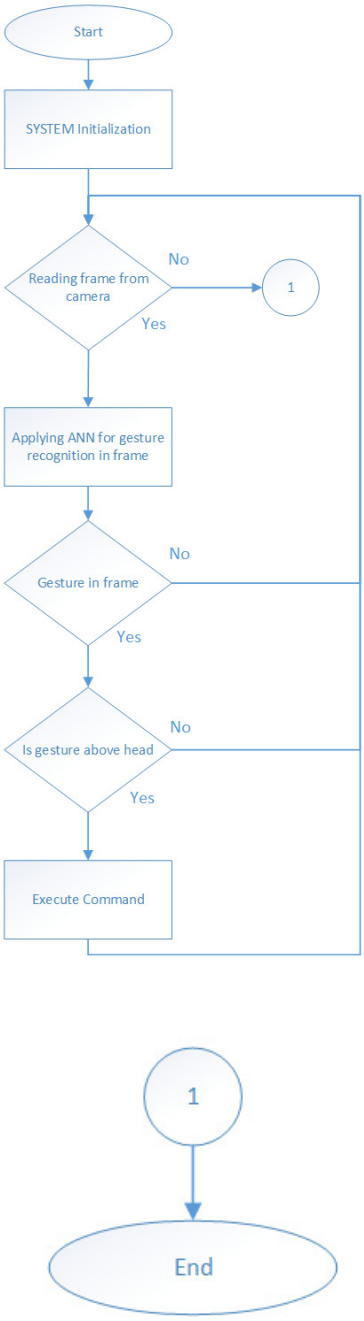

Fig. 5. *Algorithm for gesture detection*

The algorithm starts with importing the specified libraries and initializing the USB camera. After reading each frame it has been scanned applying the trained artificial neural network for gesture recognition. If the gesture is found and it is above the person's head a command is being executed.

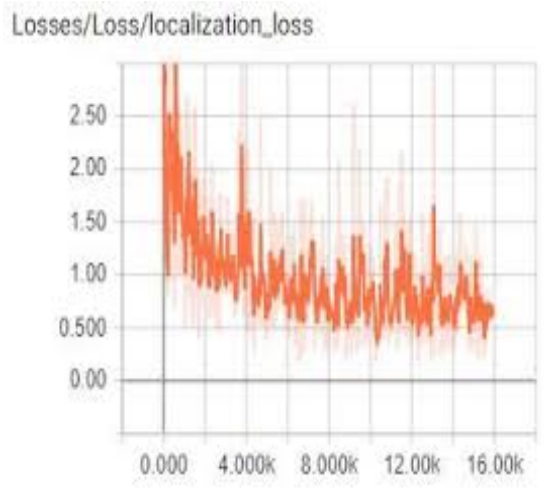

Fig. 6. *Results in training the neural network using TensorBoard*

EXPERIMENTAL RESULTS

In industrial environment it is a common problem that works may not be able to reach a control terminal for different reasons. Using artificial neutal networks a solution is provided for automated gesture recognizing system that is able to execute multiple commands based on a human gesture (if the gesture is above his head). The reason for processing the command after the gesture is above his head is because if it is below a person's head he may do that gesture by accident causing an unwanted process to execute.

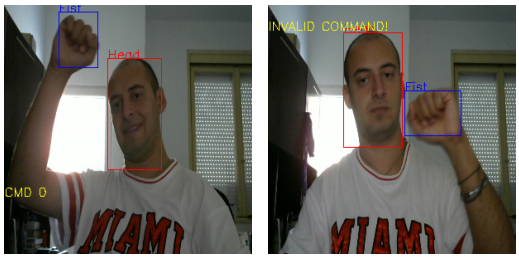

Fig. 7. *Attempting to perform first command invalid / valid state*

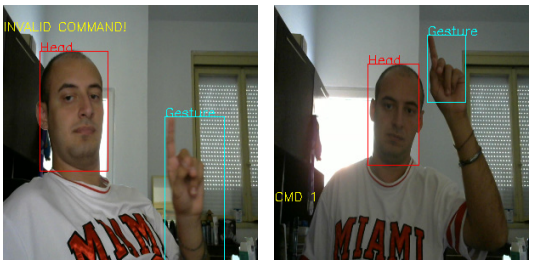

Fig. 6. *Attempting to perform second command invalid / valid state*

For training of the neural network for finding dangerous objects 300 images were used extracting their characteristics and training an neural network with MobileNet architecture for 200 000 cycles for reaching the maximum accuracy. The system can detect objects at a different instance. The neural networks find the objects and places a label with their name. The speed for finding objects is around 100ms and it is highly dependent on the input camera resolution. These results were achieved with input resolution 800x600 px. Using a higher resolution, the mentioned speed will be slower while using a lower resolution it will be faster.

CONCLUSIONS

The proposed vision system uses low-cost components and commercially available technologies for detection of gestures in a industrial zone.

The system is effective in recognition of human head, gestures and other objects if needed.

REFERENCE

[1] - https://www.raspberrypi.org
[2] - https://opencv.org/
[3] Packt Publishing, OpenCV 3 Computer Vision Application Programming Cookbook 2017.