Master Thesis

Study on application of artificial intelligence in design of a computer vision system for Polish sign language recognition

Katarzyna Sawicka, Kamila Sosnowska
Automation and Robotics Engineering, Biomechanics and Biorobotics
Academic year 2018/2019

Supervisor: dr hab. inż. Cezary Rzymkowski, prof. PW

1. Introduction

Communication through gestures is one of the basic forms of human communication. Human hands, due to their 27 degrees of freedom and the ability to perform many different gestures, are the basic element of the body involved in this communication.

A group of people who could benefit from the development of technologies that allow reading hand gestures are deaf people who communicate using sign language on a daily basis.

The purpose of this diploma thesis was to attempt to apply artificial intelligence, in particular convolution neural networks, to develop a computer vision system for recognizing selected signs of the Polish sign language. The work is limited to the classification of twenty-two characters of the sign alphabet. The completed task was divided into three important issues: segmentation task, detection task and image classification task.

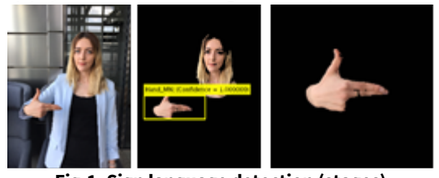

Fig.1. Sign language detection (stages)

2. Segmentation and detection task

The segmentation task was to recognize selected image elements (in this case, elements of the human skin - face and right hand) and extract them from the photo. The filter model was based on checking a group of conditions imposed on the image represented in two color models - RGB and YCbCr according to an iterative algorithm. In addition, a segmentation stage was introduced, by selecting from the extracted elements those that will be relevant in the detection process. A similar procedure was also applied to photos after hand detection from filtered images to reduce the occurrence of excess elements in the output images.

The detection task involved recognizing the image of a human hand using a convolutional neural network and cutting out this element from the whole image to enable its classification. The algorithm was based on the transfer learning technique, using the ready CIFAR-10 Net network, which was initially taught on a collection of 50,000 images of various objects from the CIFAR-10 database. The last step of the detection stage was an algorithm that, based on the coordinates of the windows proposed by the detection network, determined the most advantageous square area of hand excision from the photo.

3. Classification task

In the classification task, which was the last stage implemented by the designed system, a revolutionary neural network based on the popular AlexNet network was used, which was adapted through the learning process transferred to the problem being solved. The classification was based on analyzing the cut-out image element and assigning a hand gesture to the appropriate category, recognized by the network as the most likely. The classification was carried out in two ways:

- using a single neural network,
- using a set of neural networks by classifying the sign in two stages.

4. Program testing

As part of the work, the impact of each of the program components on the overall system effectiveness was investigated, possible sources of error were identified, and methods to improve system performance were proposed.

Research on the effectiveness of the developed program on a set of test data resulted in approximately 45% accuracy in the case of one-stage classification and 40% in the case of two-stage classification.

Due to the complexity of the system, the program was sensitive to inaccuracies in the operation of individual stages. The efficiency of the segmentation and detection stage for the test set reached approximately 84%, while the classification was 53% and 47%, respectively, for one- and two-stage classification.

5. Results

- The system designed as part of this work can provide a good basis for developing a computer vision system based on artificial intelligence and deep learning algorithms with high efficiency, which can be successfully used for practical applications.
- One of the factors that could affect the effectiveness of character classification was the fact that the set of characters considered was very extensive (22 classes), and individual characters often did not differ much, which can be a significant problem in the case of projects using networks neural.
- In order to improve the effectiveness of character recognition, further work on program development may focus on improving the training data set in terms of quantity and quality, modification of filter parameters and / or processed image, as well as optimization of neural network models used for detection and classification.
- The developed sign language recognition program, after appropriate modifications, could be used as a sign language translator, sign language learning application, or gesture control system.