The Intelligent Robotic System for Text Typing based on the Image Processing

Victor Popa, Irina Cojuhari

Abstract

In this paper it is described the designed intelligent multifunctional robotic system for text typing on a computer. The text recognition is performed by the neural networks and then through digital image processing and photometry technologies the robotic arm performs the typing of each letter.

Keywords: machine vision, neural networks, photometry technologies, image processing, pinhole camera model.

1. Introduction

Nowadays, rapid progress of industrial technologies imposes new requirements and challenges to the different domains of human activity. In so way, the robotic systems, artificial intelligence, cloud computing and machine vision are becoming more integrated into different kinds of technological systems [5].

Due to the artificial intelligence domain, the control and implementation of different operations has become more effective, with possibility of further reusing.

As a research study, it was proposed to develop an intelligent system, based on visual detection of the keyboard and it is calculating the distance between the end point of the arm and the key corresponding to the required letter. All this is implemented by digital processing of images, that are captured by the video camera, the letters on the keyboard are being recognized and interpreted by the neural network. The distance and calculation of the movement trajectory of the robotic arm is done based on algorithms which implement photometry techniques.

2. Description of the Technologies Used in the Designed System

2.1 Machine Vision

Machine Vision is a scientific direction in the field of digital image processing technologies, this processing is performed by intelligent algorithms (neural networks), with the ultimate goal of using the processed data to solve various automation problems [1-2].

The Machine Vision areas of use:

- Large industrial production (maritime, automotive, chemical)
- Industrial security systems
- Visual control and management systems (bar reading metering)
- Automatic control of vehicles
- Product quality control
- Automotive Industry
- ADAS (Advanced driver-assistance systems).

2.2 Pinhole camera model

The **pinhole camera model** defines mathematical relationship between a 3D point and its projection onto the image plane of an ideal pinhole camera, where the camera aperture is described as a point and no lenses are used to focus light (Figure 1) [3-4].

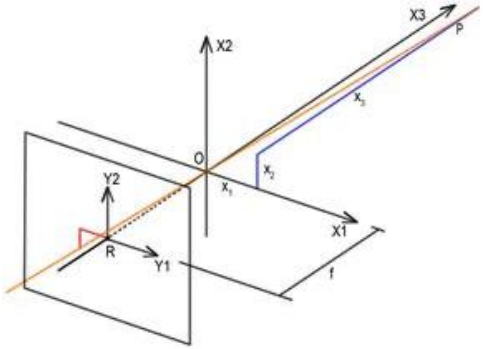

Figure 1.The geometry of pinhole camera model

2.3 Neural networks

A neural network is a network of circuit of neurons that acts based on the series of algorithms that aim to recognize underlying relationships in a set of data through a process that imitates the way the human brain operates (Figure 2) [6].

The use of artificial intelligence for the purpose of interpreting images has conditioned the popularization and reduction of the prices of computer vision systems, which has led to their widespread use in almost all branches of industry. The main cause is the high degree of reconfiguration of hardware and software itself. In turn, the possibility of self-learning of neural networks offers a wide range of possibilities depending on the appearance of each technological process of production.

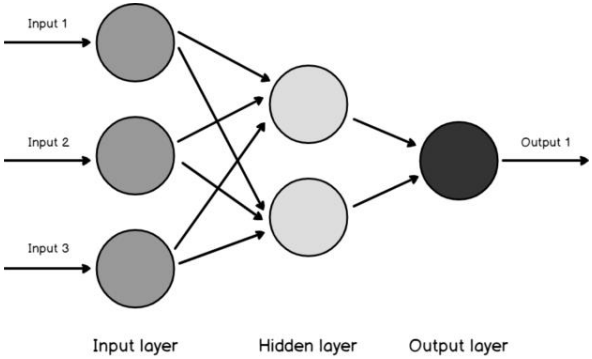

Figure 2. Example of neural network

3. Description of the Designed System

The designed system offers the possibility to automate the process of typing on the computer the text extracted from the image, where the acquisition of the text is done by a video camera, and the extraction and interpretation of the text is done with neural networks, the keyboard and letters are detected in real time by various techniques of digital image processing (Figure 3).

In order to increase the working speed of the program and to minimize the overload of the computer operating system, the interpretation of the image is performed on the web server of the "Microsoft" company. After the text is extracted from the image, the program starts the typing cycle of the text on the keyboard and the letters on the keyboard are also interpreted by the local neural network in real time. The program checks after each keystroke performed the current distance between the point arm terminal and the required letter on the keyboard, by the digitally processing of the image and converting the number of pixels into the

distance that is measured in millimetres, the algorithm generates the trajectory of the arm's movement from the current position point to the required position of the letter on the keyboard, the cycle is repeated until all letters are typed.

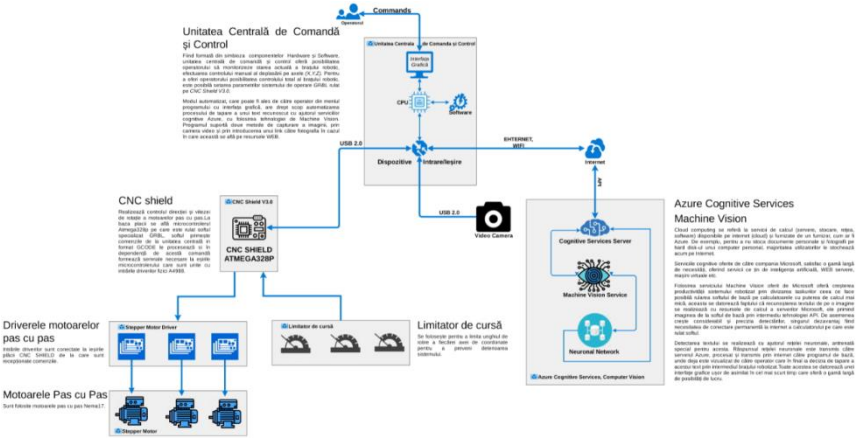

Figure 3. System overview

4. The Process of Acting the Intelligent Robotic System

4.1 Graphical User Interface (GUI) of the program

To make possible the interaction between the user and the intelligent robotic system, the graphical interface (Figure 4) has been developed, which facilitates monitoring and management of the process and it provides the main functionalities for a comfortable interaction.

The graphical interface has the following advantages and possibilities:

- Facilitates monitoring and management of the process of acquisition, detection and typing of the text
- Arrangement of the palettes in the intuitive and clear way
- Possibility to view the detected text
- Categorization of processes
- Possibility to control the whole process of detection the objects and web images.

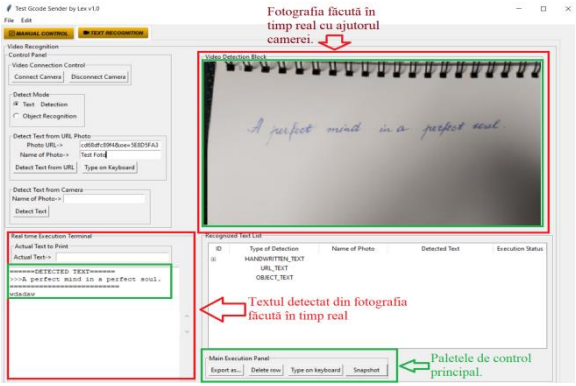

Figure 4. Graphical User Interface of the system

4.2 Text detection and interpretation by the Azure Cognitive Services

For the text to be typed by the robotic arm, it must first be acquired, processed, and then transmitted to the neural network for interpretation. In order to benefit from cognitive services (Figure 5) Microsoft needs to create a personal account on the Azure.com portal, after successful authentication the resource group is created that offers the possibility to detect and interpret the text on the image, the authentication between client and server is done by the API technology with the indication (ssh key, ssh endpoint) (Figure 6).

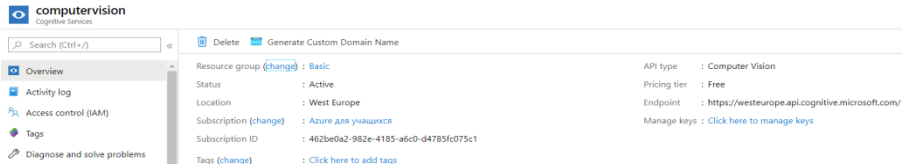

Figure 5. Azure Cognitive Services main control panel

```
In [1]: #----------IMPORTAREA BIBLIOTECILOR NECESARE----------
        from azure.cognitiveservices.vision.computervision import ComputerVisionClient          #Bibliotecile Azure
        from azure.cognitiveservices.vision.computervision.models import TextOperationStatusCodes#Bibliotecile Azure
        from azure.cognitiveservices.vision.computervision.models import TextRecognitionMode
        from azure.cognitiveservices.vision.computervision.models import VisualFeatureTypes      #Bibliotecile Azure
        from msrest.authentication import CognitiveServicesCredentials

        from array import array                                                                  #Biblioteca pentru lucrul cu vectorii
        import os                                                                                #Biblioteca pentru lucrul cu SO
        from PIL import Image                                                                    #Biblioteca pentru prelucrarea imaginii
        import sys
        import time                                                                              #Biblioteca de lucrul cu timpul
```

Figure 6. Import libraries in order to use Azure Services

After the text has been extracted (Figure 7) from the image, it is transmitted to the basic program, so that it can be typed by the robotic arm.

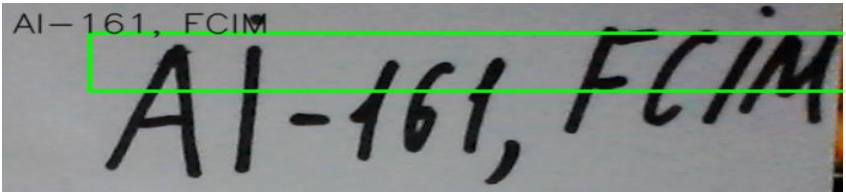

Figure 7. Extracted text by the neural network

The next step is detection of letters on the keyboard, using artificial intelligence.

4.3 Detection of letters on the keyboard

Visual detection of letters on the keyboard is performed using specially trained neural networks to detect typefaces. The neural network used in this project has the role of interpreting the image (Figure 8) already processed by the digital image processing technologies. The image segmentation is performed by a separate script that is not part of the neural network, it is assumed that the image transmitted for interpretation to the neural network can only be a letter that corresponds to a single class of letters.

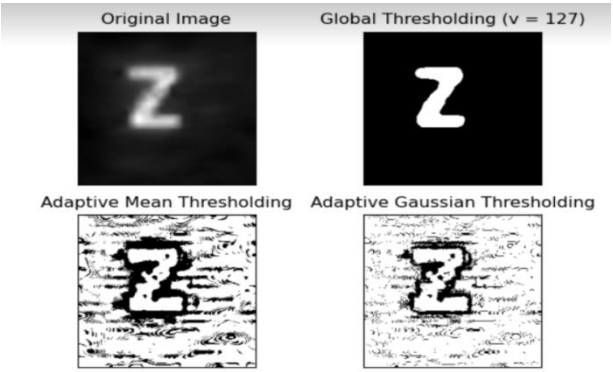

Figure 8. Digital processing of the image, which is sent to the neural network for interpretation

The digital image processing algorithm keeps the position of each letter transmitted to the neural network for interpretation, after the interpretation has been performed successfully, the algorithm assigns to each interpreted letter a position in the 2D axis system and draws this information on the real time screen view (Figure 9).

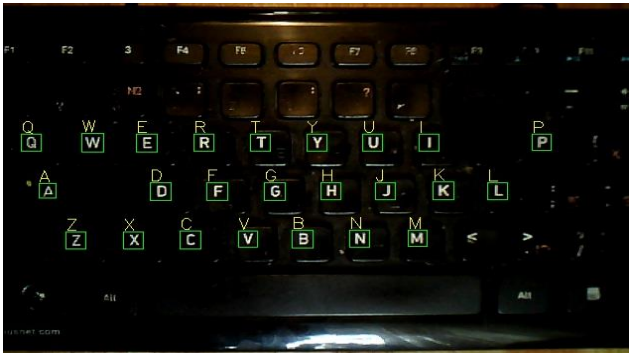

Figure 9. Result of letter identification by the keyboard detection algorithm

4.4 Calculating the distance between the letter and the robotic arm

In order to calculate the distance between the letters on the keyboard and the robotic arm, the pinhole model (Figure 1) was implemented, which offers the possibility to translate the position of the objects from the 3D space in the image plane (Figure10).

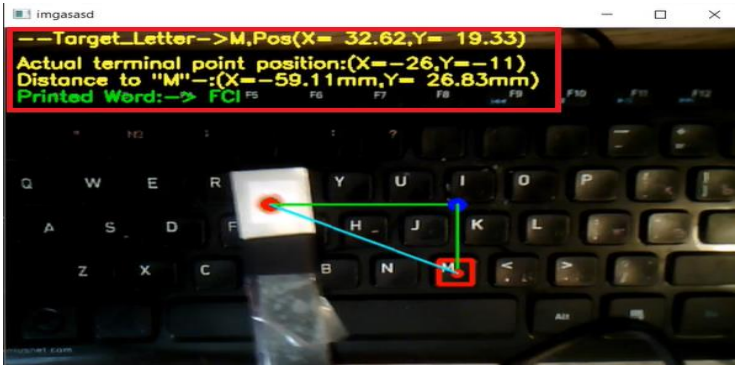

Figure 10. Calculating the distance between the letter and the robotic arm

5. Conclusion

This study proves the enormous possibilities of artificial intelligence in digital image processing, offering a wide range of possible variants of use depending on the wishes and needs of the engineer. The only one disadvantage is that it requires a modern technical material base and large computing resources. Machine vision is a branch that is increasingly used in all areas of life, which involves the development of this branch and as a result is an area of growing interest for subsequent research studies, which in relation to this study will focus on solving real problem.

References

[1] R. Szeliski. *Computer Vision: Algorithms and Applications*. Springer, September 3, 2010.

[2] D. Dechow. *The Fundamentals of Machine Vision*. FANUC America Corporation 3900.

[3] P. Sturm. *Pinhole Camera Model*. In: Ikeuchi K. (eds) Computer Vision. Springer, Boston, MA, 2014.

[4] R. Hartley, A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge, UK: Cambridge UP, 2003.

[5] K. M. Lynch, F. C. Park. *Modern Robotics Mechanics, Planning and Control.* Park May 3, 2017.

[6] D. Janglová. *Neural Networks in Mobile Robot Motion*. Inernational Journal of Advanced Robotic Systems Volume 1 Number 1 (2004), pp. 15-22.

Victor Popa [1], Irina Cojuhari[2]

[1]Politehnica University Timișoara
E-mail: victor.popa1@student.upt.ro

[2]Technical University of Moldova
E-mail: irina.cojuhari@ati.utm.md