

# Application of the Augmented Reality Goggle for a Muting Procedure with Convolutional Neural Networks Algorithm

Krzysztof Lalik<sup>[0000-0003-4277-9801]</sup>, Mateusz Kozek<sup>[0000-0001-7222-2970]</sup>, and Ireneusz Dominik<sup>[0000-0002-2304-7763]</sup>

AGH University of Science and Technology, Faculty of Mechanical Engineering and Robotics, Kraków, Poland

**Abstract.** The use of augmented reality glasses is one of many elements of the standard known as Industry 4.0. This article shows the use of Microsoft HoloLens headset to support the work of the semi-automatic machine operator. As the quality of collecting cloud points for these glasses is not enough, it was decided to implement a neural algorithm that was used to filter the online image obtained from augmented reality headset.

## 1 Introduction

Mixed reality initially found its application mainly in gaming industry. However, its use in industry, especially automotive is increasing in past years [1–3]. For example, with QR codes attached to machines, you can enhance employee performance in repair work by displaying the necessary information, such as assembly instructions or possible machine errors. The operator equipped with such headset can also have a new, innovative interface for communication with control systems. This creates a new cyberphysical system connecting humans with machines Fig. 1.

Industry 4.0 is a general concept, which means integrating intelligent machines, systems and making changes to production processes. This integration is designed, among other things, to improve the efficiency of the manufacturing process and to allow to make changes in much more flexible way than it has been before [4,5].

The use of information technology, communication, system integration and networking started the transition to Industry 4.0. Thanks to the proposed interface, machines and IT systems exchange the necessary information in the production process (Fig. 2). This information may also be presented to the operator in the form of a hologram imposed on the actual device. An employee present on station has the possibility to display relevant data on the manufactured product.

Creating a universal production line allows you to create different variants and combinations of products without having to modify machines. This approach significantly affects the ability to adapt to individual customer inquiries, which increases market competitiveness.



**Fig. 1.** Machine operator equipped with Microsoft HoloLens glasses.



**Fig. 2.** Hologram applied to the actual station.

## 2 Mixed Reality platform – Microsoft HoloLens

Augmented reality is a system to support the real world with additional computer-generated animations, images or information. An essential part of augmented reality technology research is the use of the real world and the placement of virtual elements, such as computer-generated graphics. In [6] they presented a pattern of continuity of reality. The four points on it reflect the different degrees of the virtual and real-world combinations. The real world therefore consists only of real elements. Augmented reality (AR) is made up of elements of the real world, but is extended to include virtual elements, such as holograms. This is possible, among other things, with Microsoft HoloLens headset. Extended Virtuality (AV), on the other hand, consists of elements of the virtual world, extended with real elements. The virtual world consists only of computer-generated virtual elements.

Microsoft HoloLens (Fig. 3) used for the research presented in this article have several sensors, thanks to which it is possible to orient the space, create holograms and communicate with the user. The device consists of an inertial measuring sensor (IMU),

containing an accelerometer, gyroscope and a magnetometer. In addition, four environment recognition cameras, a 120° x120° depth camera, a light intensity sensor, and a 2.4-megapixel camera that is responsible for taking photos and capturing image.



**Fig. 3.** Microsoft HoloLens 1 AR headset [7].

Headset display 16:9 HD images to create 2.3M light points. Their holographic density is 2.5k light points per radian. The rendering process is based on vision, the glasses have automatic calibration based on measurement of the pupil distance. HoloLens comes with four microphones, ambient light sensor, depth camera, 2MP/HD video camera and four "environment understanding" cameras.

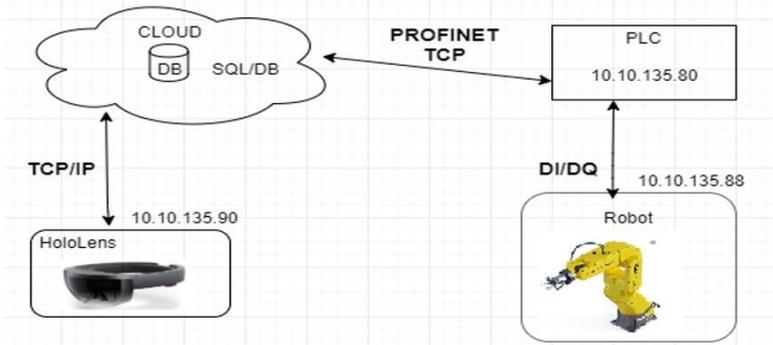
With four cameras mapping the environment, the space in which the user is located is accurately presented. Once you have collected an ambient image, the HPU processor processes the information, which results in an interpretation of the environment. Infrared emitters send light beams that, when reflected from surrounding objects, go to the camera depth, which determines the distance between objects based on the intensity of light streams. The information collected is again passed to the processor that creates the ambient image. Thanks to the inertial measuring sensor, it is possible to adapt the holograms created to the user's current location.

### 3 Methodology

The creation of the proposed cyberphysical system will allow for specific security measures for the operator without the need for additional activities. An example of such Fail-Safe function could be, for example, muting function. This is a temporary and deliberate deactivation of a specific technical protective measure, which may be, for example, a safety curtain. With this function, it is possible to perform a specific technological operation by using appropriate SENSORS or AR glasses. The subsequent positions and direction of movement of the specified object (mainly its dimensions) are detected from the moment it appears in the protection zone until it leaves the zone. Presence sensors are usually used for this purpose.

In this case, the glasses scan the robot's environment and automatically attach muting functions to the operator, without the need for additional actions, such as the press the appropriate button or other user interface.

In order to implement such an algorithm, it is necessary to build a system which architecture is shown on Fig. 4. AR glasses connect wirelessly via WiFi router to a server located in the cloud. Also, in the cloud there is a convolutional neural algorithm (CNN), which analyzes the image transmitted by AR glasses on an ongoing basis. If the algorithm decides that the operator is too close to the robot, then the corresponding flag is sent from the cloud to the PLC. The controller based on its software and the state of the flag received from CNN decides to send information about the muting procedure to the robot.



**Fig. 4.** Headset communication diagram.

An example cloud of points derived from the AR goggles is shown in Fig. 5. This image explains why CNN was needed. Although AR glasses can also transmit actual images using their cameras, the point cloud gives you better spatial location information relative to your objects and operator. The innovation of this solution is that the cloud of points is used, instead of the classic 2D vision methods.

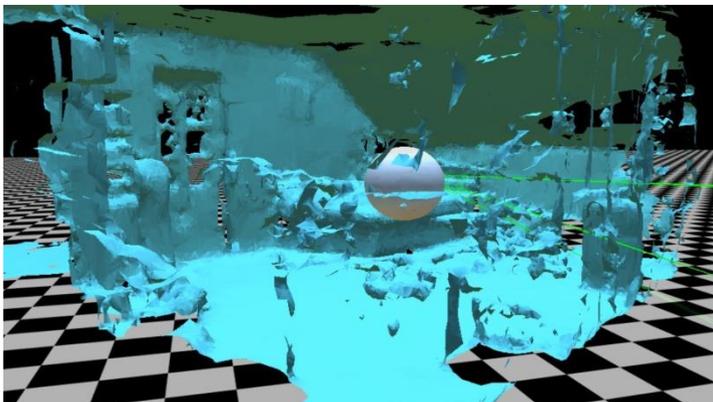
The technology of the AR goggles, despite its significant advances in technical development, is still not flawless. Hence, images require post-processing. The main advantage of convolutional networks is their native ability to filter images at low computing cost. CNN therefore sharpens every frame of the uploaded video and classifies every shape on it. The network task is to extract (classify) the outline of the robot and to evaluate the distance from the operator.

In image processing, deep machine learning method is used specifically convolutional neural networks (CNN) that improves the work of employee. This algorithm works well in automatic extraction of objects from the point cloud and allows for an excellent classification of processed images.

## 4 Results

Convolutional networks through training can learn what special features of the image help in its classification. Their advantage over standard deep networks is greater effectiveness in detecting intrusive dependencies in images [8–10]. This is made possible by using filters to study the relationship between adjacent pixels. Each point cloud is a matrix of values which are proportional to its width, depth, and pixel height. Each pixel is represented by three values. CNN reduces the size of the cloud of points which minimizes computational cost without losing valuable data, that is, those that carry the most relevant information for classification.

This paper uses a deep learning convolutional network to detect the position of the robot.



**Fig. 5.** Sample cloud of points obtained from AR headset.

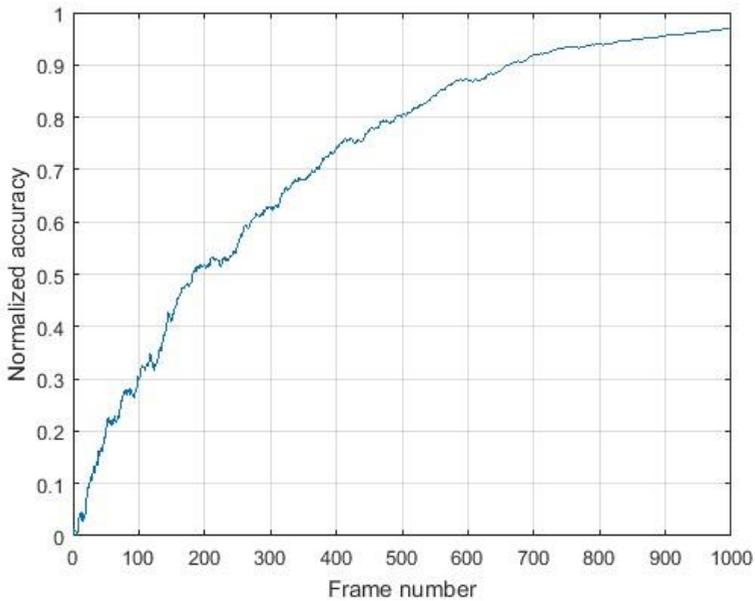
Training conventional neural networks is like that of CNN. Convolutional neural network, on the other hand, processes data between layers. In the dedicated layers of convolution, the number of scales is reduced because part of the scales is shared for the calculation of the output.

Multiple cells are taught in a specific convolutional layer. Hence, there are many images at the output of a cell. Due to the reproduction of incoming data, layers typical of convolutional neural networks are used, for example, there are several types of reducing layers, so called polling layers [11,12], which are responsible for data reduction. While in the reducing layer, statistical filtration is carried out within the mask of the desired size. The CNN contains several reducing layers and after Cross-channel Normalization data are processed using a nonlinear activation function.

The headset was placed on a tripod and then the video was uploaded to CNN. The glasses were sending a 20 FPS video. The effectiveness of the recognition of the outline of the robot during learning is shown in Fig. 6. Frames from the stream were a set of input data in the training process. During the training process, the position of the device relative to the robot was changed. An accuracy of 96.5% has been achieved. To test the operation of the algorithm, the user put on AR glasses and made three laps around the robot. Headset was ensured to be positioned at different angles and distances relative to the robot. Each frame in the video was then tested for the effectiveness of the algorithm. Every frame on which CNN's algorithm recognized the robot was classified as positive. The effectiveness in this case was 87.1%. The poor performance was mainly due to the fact that the learning process was carried out for quasi-static conditions, while learning should take place for dynamic Headset position parameters.

## 5 Conclusions

This article presents an innovative cyberphysical system to support the work of the device operator. It complies with Industry 4.0 standards. The use of AR glasses allowed the creation of an intelligent operator protection system. The muting process presented in the article is just one of the possible functions in robotics.



**Fig. 6.** Training efficiency of CNN.

The cloud of points sharpening system used has proven to be satisfactorily effective. Cloud calculations and less compute cost of the algorithm allowed it to be implemented in a real industrial installation.

Further work will proceed towards the application of the solution for the operator's cooperation with the robot. This will be possible to transform standard robots into collaborative robots. AR glasses will allow you to influence the range and type of movement of the robot in which the operator is located. This requires the creation of special interfaces between the control unit and the robot. Nevertheless, this is a very developmental course of action.

## References

1. I. Dominik *Journal of Intelligent Material Systems and Structures*, 27, 1917–1926 (2016).
2. G. Reinhard, V. Jesper, S. Stefan 2016 *Global Industry 4.0 Survey*, 46, 35–50 (2016).
3. R.Y. Zhong, X. Xu X, E. Klotz, S.T. Newman *Engineering*, 3, 616–630 (2017).
4. J. Lee, B. Bagheri, H.A. Kao *Manufacturing Letters*, 3, 18–23 (2015).
5. M. Rüssmann, M. Lorenz, P. Gerbert, M. Waldner, J. Justus, P. Engel, M. Harnisch *Industry 4.0: The Future of Productivity and Growth in Manufacturing Industries Tech. Rep.* (2015).
6. M. Garon, P.O. Boulet, J.P. Doironz, L. Beaulieu, J.F. Lalonde *Adjunct Proceedings of the 2016 IEEE International Symposium on Mixed and Augmented Reality, ISMAR-Adjunct 2016* (2017).
7. HoloLens (1st gen) hardware URL <https://docs.microsoft.com/en-us/hololens/hololens1-hardware> (2019).

8. Raycad <https://medium.com/@raycad.seedotech/convolutional-neural-network-cnn-8d1908c010ab> (2017).
9. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Preprint 1409.4842 (2015)
10. G. Zhao, Z. Zhang, J. Wang, H. Guan arXiv (2017).
11. P. Mazurek, D. Oszutowska-Mazurek 23rd International Conference on Methods and Models in Automation and Robotics, MMAR 2018 (2018).
12. D. Oszutowska-Mazurek, O. Knap Advances in Intelligent Systems and Computing (2017).