A Survey On Augmented Reality Applications Using Deep Learning

Cengiz Güngör
Gaziosmanpasa University

GE-International Journal of Engineering Research
ISSN(O): 2321-1717, ISSN(P): 2394-420X
Impact Factor- 5.613, Volume 5, Issue 12, December 2017
Website- www.aarf.asia, **Email** : editor@aarf.asia , editoraarf@gmail.com

A SURVEY ON AUGMENTED REALITY APPLICATIONS USING DEEP LEARNING

Cengiz Güngör[1], Kenan Zengin[1]

[1]Gaziosmanpasa University, Faculty of Natural Sciences and Engineering, Computer Engineering, Tokat, Turkey.

ABSTRACT

Unlike the totally artificial structure of the virtual reality, the augmented reality works by placing virtual objects on user's vision while he/she watching the real world. In the first applications of augmented reality, a pre-defined pattern is used to do this, which is sufficient to place objects on the pattern. The position of the pattern in the camera was also giving direction to the object. It was a simple process of training the pattern and capturing it in the image. Afterwards owing to feature vectors used in systems which are trained by artificial neural networks, complex patterns can be used. Deep learning is a revolutionary technique in artificial intelligence applications. According to neural networks which are working with high error rates, this technique exceeds the recognition properties of the human brain. Instead of using a pattern, augmented reality applications that are use deep learning need to be educated an object which an information shown on it. For example, when the source object is a human hand, recognition can be performed in all of the different positions of any human hand. It is clear that deep learning will be effective in the future of the augmented reality.

KEYWORDS - Augmented Reality, Virtual Reality, Deep Learning, Markers

1. INTRODUCTION

Augmented Reality (AR) is a technique for capturing predefined markers in a real world image and placing virtual objects on these markers [1], [2]. Researchers have been working on AR technique for a long time. In 1968, Ivan Sutherland developed a prototype, often considered the first Virtual Reality (VR) system and the AR system. The prototype was intended to help the user with the three-dimensional information displayed when the user looked in the correct direction. However, this system, which was very heavy, had to be mounted on the ceiling [3]. The applications of this technology have shown a remarkable increase in recent years. The reason for this is that cheaper technologies such as mobile phones have reached wider uses [1]. With increasing interest, technology companies are increasingly investing in AR.

The most important difference between AR and VR is that AR has 10% to 20% virtual contribution, but VR's world is completely virtualized like computer games [1]. The most important constraint of AR is that it needs a special marker to make this contribution.

In order to make AR without markers, it is necessary to capture the objects with their natural structures. For example, if a developed system recognizes a person in the crowd and is asked to write his name and information on it, the system must move from many views of that person. This can be done with deep learning, a popular application of recent days.

Deep learning is advanced form of Artificial Neural Networks (ANN). ANN was first created by McCulloch and Pitts [4]. ANN, an imitation of the human nervous system, has evolved through various stages and has been used in many areas. The ANN systems perform their tasks close to the capacity of the human brain. But the deep learning systems can make more successful decisions [5].

2. CURRENT TECHNOLOGIES

Usage of AR and Deep Learning has new opportunities. The technologies which are using currently are given in this chapter.

2.1. Augmented Reality

A marker based AR scans video images for the pre-defined markers, estimates the distance and direction of each markers, and places the virtual objects or text messages on it according to this direction [1]. (Figure. 1)

AR applications have been widely studied for educational, entertainment and sales purposes. As a result of this, mobile device manufacturers interest in AR developments.

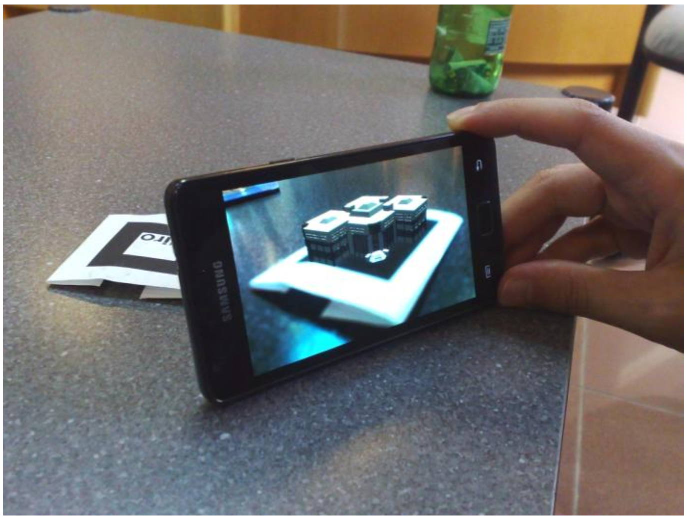

Figure. 1. Using Marker For AR

The markerless AR systems are successfully working with combining visual information that are received from environment and GPS signals or in-door position device signals. In Figure. 2, the image gives the position of the furniture in the room without a marker [6]. Furniture can be rotated in the application, moved and tested in the room. Glasses such as Google Glass are also being developed for markerless AR applications. But today they are not so popular. Mobile AR is more preferred on mobile devices.

Figure. 2. Sample AR Application [6]

2.2. Deep Learnıng

The deep learning is based on Convolutional Neural Networks (CNN) architecture. The neural network class is a huge class of machine learning algorithms including CNN, fully connected neural networks (FCNN) and many more [24]. In 90's neural nets was equal meaning to FCNN. This net structure has got a large set of parameters. So, FCNN is not scalable well. But CNN (newer type of neural nets) is not fully connected. CNN layers only connected to a few previous layers (or neurons) [24]. CNN is much more successful than older nets.

A basic CNN architecture is given in Figure. 3 [7]. This sample network outputs; there is a traffic sign in given image and writes 60 on this sign.

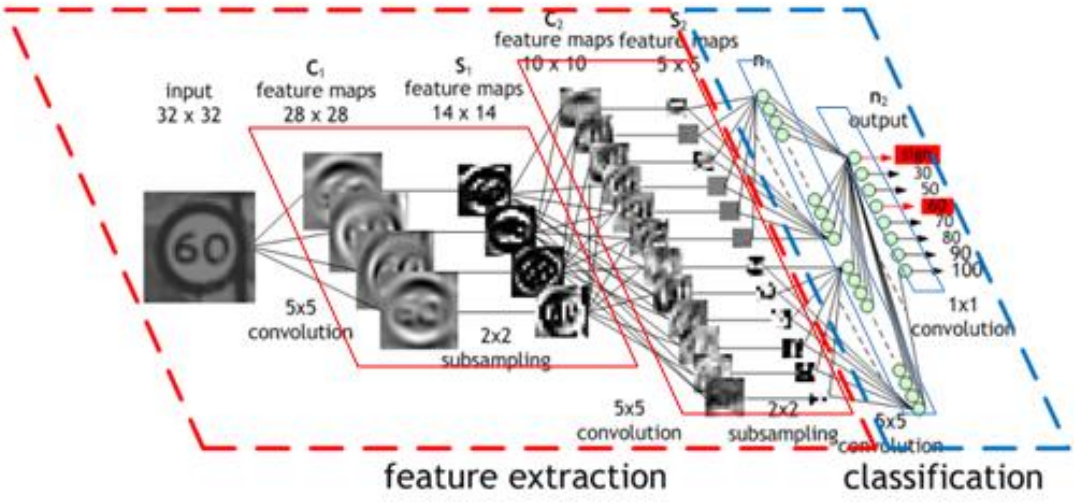

Figure. 3. Basic CNN Model (Image by Maurice Peemen) [7]

In this structure, firstly some layers designed by couples of a convolution and after sub-sampling (or pooling) layers. Last layers are fully connected, and decisions are generated by this layer [8]. The system in Figure. 3 can find correct outputs at 95% rates.

CNN’s are constructed by trainable sections. In this structure, the difference from desired result propagates to back, and used for updating all weightings. These back propagation algorithm are shown in Figure. 4. The result of this process will be decreasing the error rate.

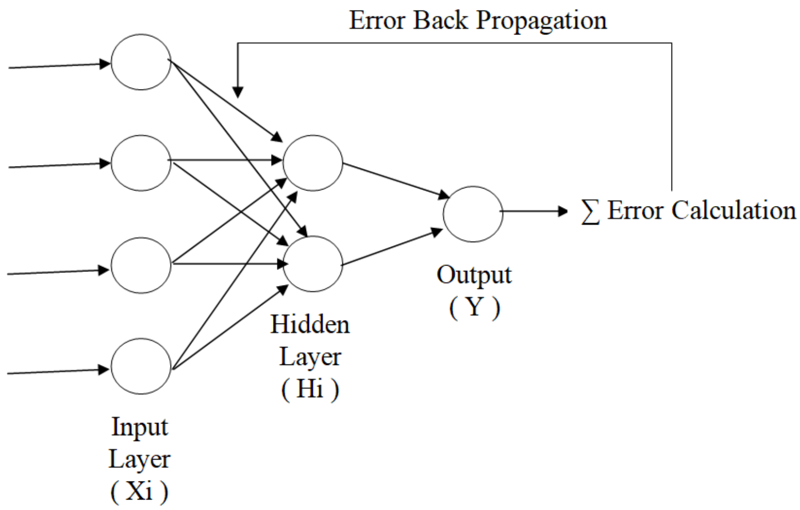

Figure. 4. Sample Error Back Propagation

Input signal of CNN can be any types of signal like image, video, sound or source text from any language which will be convert to any language. CNN learns features about signals by itself. While the process is continuing, one specific feature stored in every section of CNN by hierarchically [9]. A sample of Lee's study is given in Figure. 5.

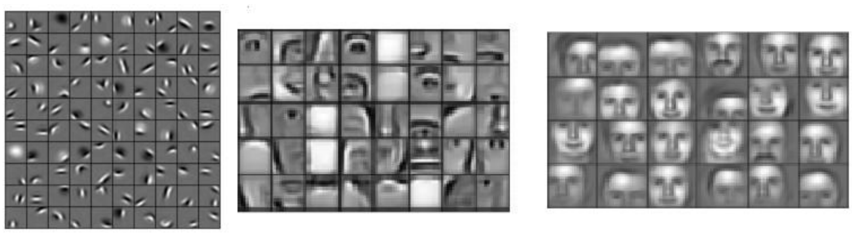

Figure. 5. Learned Hierarchical Features [9]

Deep learning training stage is the same as older net's stages. There are always many data to be learned, so this process can be named as big-data problem. If a big-data process is needed in any project, parallel processing using is must too. Parallel processing can be executed on any Graphic Processing Units (GPU) which are 10 to 100 times faster than Central Processing Unit (CPU) [10]. The famous GPU companies like NVidia fund these type researches very much. Also, NVidia has a web site named as 'Deep Learning Institute'. As a result of this power, especially autonomous driving firms prefer developing their deep learning based applications on NVidia's sets.

In 2016, NVidia launched a autonomous drive system which were used deep learning technology [23][24]. In this project they used three RGB cameras, respectively left, right and central cameras. System generates random steering rotation, and gear shift value, and calculates differences (errors) from user behavior. These errors back propagated to CNN layers, and the system learns to drive day by day. After training of the system, it tested in simulator before real world driving test. All driving data stored in a solid state disc. In real world system already controlled by user for incorrect using. Because Tesla's car was involved in a fatal accident in 2016. The driving report shown that neither car nor driver was used brakes. The system of Tesla defines higher objects as traffic symbols in bright weather conditions. The result of this error, a truck cannot be seen by system, and they crushed to back of the truck. Nvidia's car reached 98% correct decisions. But already 2% wrong decisions can be kill passengers. This is very important ethic problem [22]. That's why, autonomous driving cars still testing.

Deep learning's decisions so correct that given outputs in Figure. 6 are enough to show why many deep learning applications so successful.

Figure. 6. Samples for output (decisions) of deep learning systems [10]. At left: "A black and white dog jumps over a bar", and at right: "A girl in purple dress jumps in the air".

3. AR SYSTEMS USING DEEP LEARNING

AR based on catch pre-defined markers in the real world images. It is the easy way to do this, catch the marker and put the object on it. But if AR uses deep learning, it doesn't use a specific image of the markers. It learns many forms of the markers, so we can call it as the markerless AR. And deep learning systems more successful than old neural nets. In example an old net looks a kitchen, and defines cups and kettles as cups. But deep learning system will be catch cups over 95% rate correctly [10].

Many examples can be given to deep learning AR applications. One of them is beta version of tool for the blind persons, android BlindTool. The application can be downloaded freely from PlayStore. While tool is running, if the mobile phone camera directed to a laptop, tool guess it is a laptop with maximum rate 35 or 40%, writes and says “laptop” as a best guess (Figure. 7).

Figure. 7. Screen Of BlindTool Beta Application

When AR uses deep learning, its basic task which is putting info or objects on real world images will be same. Because of this, automotive companies’ autonomous drive systems are perfect example for using deep learning with AR. A sample screen output is given in Figure. 8 which uses NVidia’s Tegra X1 kit developed for mobile devices [12]. Deep learning application developers mostly use NVidia’s Titan series GPUs for desktop uses [13][14].

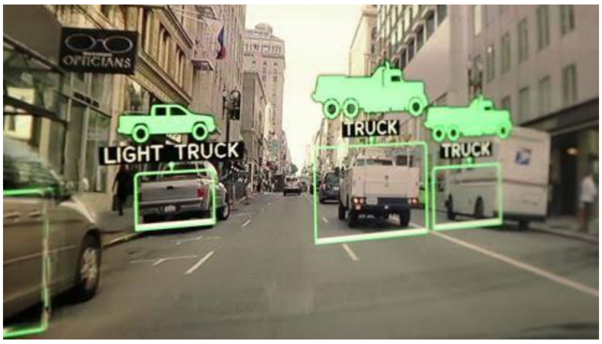

Figure. 8. Screen Sample For Autonomous Drive [11]

The big companies like Facebook, Google and Baidu are funding on deep learning-AR projects. For example, Google's Tango project started in 2014 and a special 5-inch phone was built. This device was aimed at bringing a new dimension to mobile devices by doing 3D sensing and mapping [15]. This limited production device was able to map your home when you walked around your home. Not just Google interested this project, Asus ZenFone AR went on the market with support of Google Tango in 2017.

A successful facial recognition is an important feature for AR. Some companies compete with their face recognition algorithms. Figure. 9 is showing that Baidu has passed Google [10], [16]. Baidu has his own special parallel processing systems to create artificial intelligence. The vertical axis of Figure. 9 is the percentage error rate encountered for 6000 matches attempted for face recognition, and the dashed line is human's threshold. The most important thing to notice is the fact that the face recognition succession of AR is better than human's face recognition abilities now.

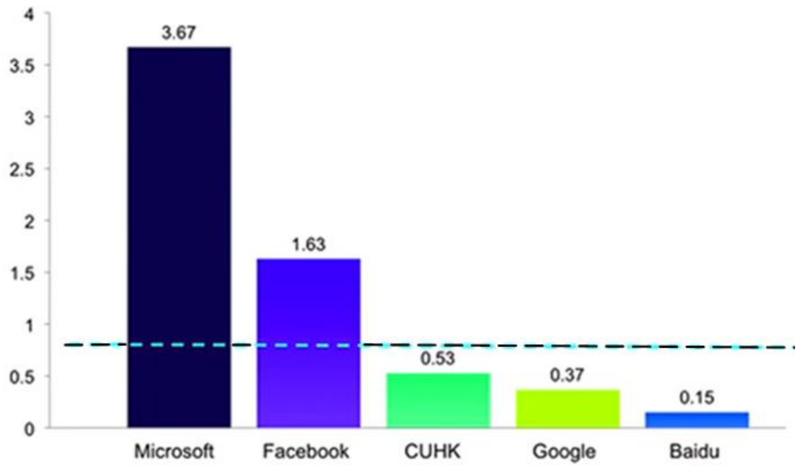

Figure. 9. Face Recognition Thresholds Of Companies [10].

There are many other companies working on AR. For example, TeraDeep are helping developers with a hardware (FPGA kit) thar are named "learning camera". Yann LeCun, director of Facebook's AR department, is also a consultant for TeraDeep [17], [18].

A very important example is the DeepHand project. This study conducted at Purdue University for user interface interaction with the AR system [19]. The outcome point of the prolect is the basic necessity in the human-computer interaction interfaces of the intelligent hand position estimation and the AR. Laboratory manager Professor Donald W. Feddersen emphasizes the importance of the project that:

> *"If your hands can not interact with the virtual world, you can not do anything. That's why hands are so important." [20]*

In Figure. 10 is shown some images of this work. The systems in which the hands are used can be used for repair and maintenance purposes [21].

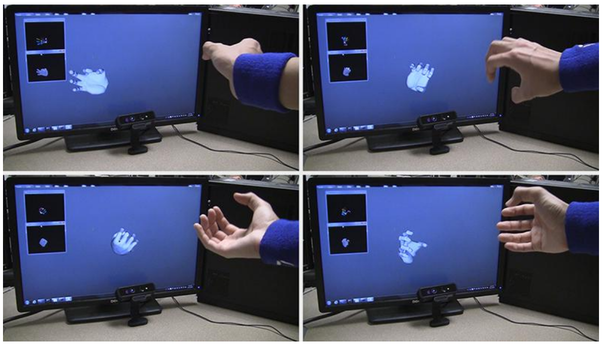

Figure. 10. Images Of DeepHand Project [20].

In other academic work, the estimation of the intelligent hand position has been slightly overlooked. It is difficult to deal with an organ where every finger moves freely. That is why all body tracking studies are more extensive [20].

A single depth camera is used for DeepHand and GPU power is not utilized. Position estimate starts by taking inputs from the depth and RGB cameras (Figure. 11). A cropped depth hand image is obtained from these images. A CNN layer estimates the direction of the wrist. The gathered information of hand is given as input into the fingers evaluation phase. In this phase, there are 5 CNN layers that estimate the direction of each finger. Finally, a hand model is used to combine the estimates and a virtual hand image is displayed on the screen [20].

A model with 21 degrees of freedom (DOF) was used to model the hand, and 18 joint angles and x, y, z positions were used for position estimation. They used 18-layer CNN [20]. The researchers made successful estimates with low error rates (average 1.6 cm).

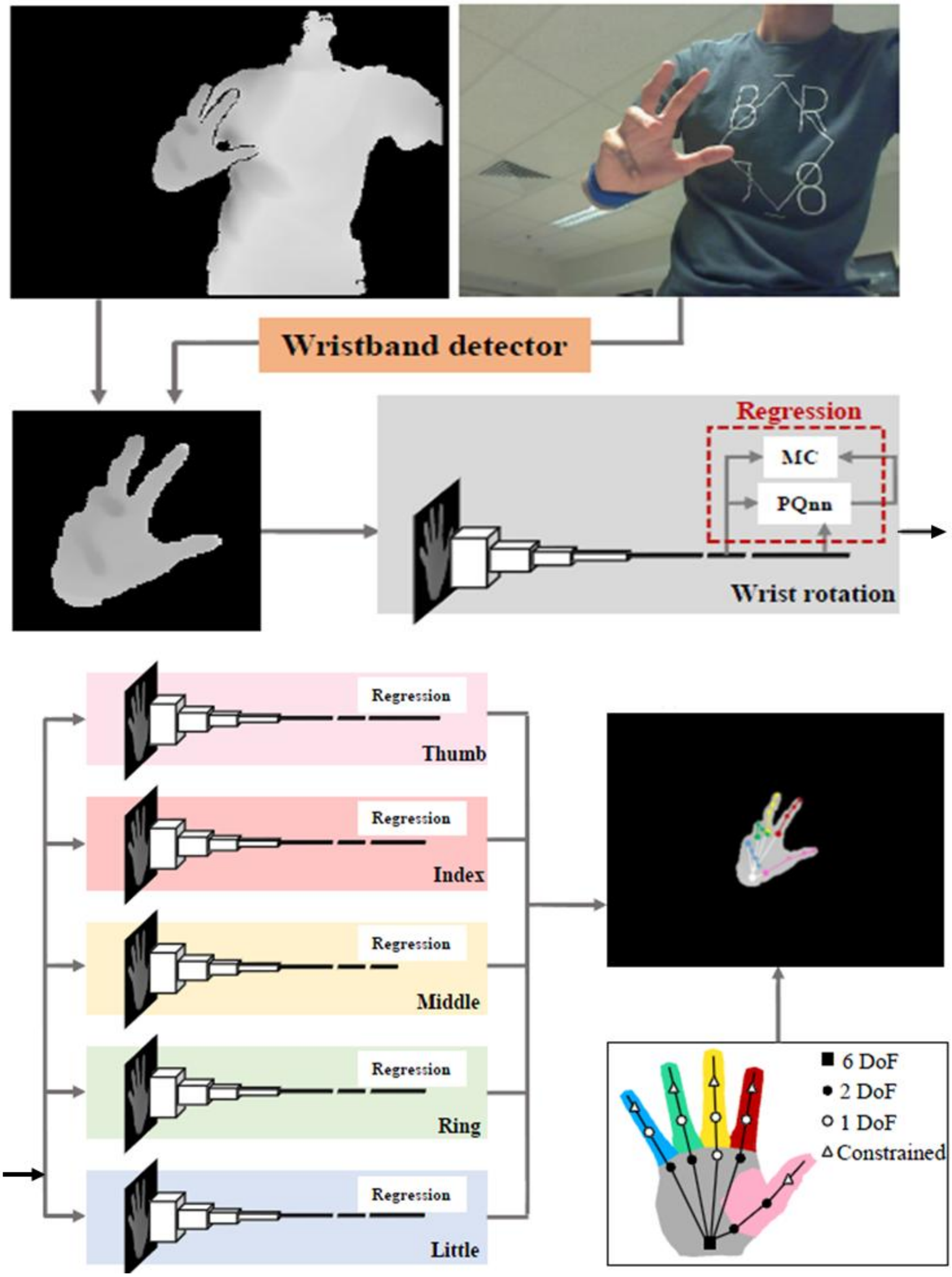

Figure. 11. DeepHand CNN Structures. Top: Estimating Wrist Position, Bottom: Estimating Fingers Positions [20].

4. DISCUSSION AND CONCLUSIONS

While deep learning and AR are used by researchers as hot topics in various fields, the unification of the two methods opens new horizons. It is estimated that many of these ideas have not been published yet.

In this study, deep learning and AR issues are explained, and the research results of the combination of the two are examined. Despite the fact that people's ability to see and classify has not been passed for many years, new project issues have been started after the success of deep learning. The use of deep learning in important issues such as autonomous driving, which we can entrust to human life, shows how confident the method is. The fact that the number of AR applications that use deep learning will increase, and more and more artificial intelligence applications will process visual information and help us in the near future.

5. ACKNOWLEDGEMENTS

This work presented as an oral presentation at the ISMSIT 2017 Symposium [25].

6. REFERENCES

1. C. Güngör ve M. Kurt, “Improving Visual Perception of Augmented Realityon Mobile Devices with 3D Red-Cyan Glasses”, *22nd Signal Processing and Communications Applications* (*SIU-14), Trabzon,* April 23-25, 2014.

2. M. F. Akbaş ve C. Güngör, “A Survey on Maarker-Based Tracking Systems in Augmented Reality and Design of Multi-Layer Marker Model”, *DEU Journal of Science and Engineering, Volume 19, Issue 56,* May 2017, pp: 599-619.

3. I. E. Sutherland, “A Head-Mounted Three Dimensional Display”, *Proceedings of Fall Joint Computer Conference*, 1968, 757-764.

4. W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The Bulletin of Mathematical Biophysics, vol. 5,* pp. 1943, 115–133.

5. devblogs.nvidia.com. (2017). Deep Learning for Julia. Available: https://devblogs.nvidia.com/parallelforall/mocha-jl-deep-learning-julia/

6. www.zmescience.com (2017). Available: https://www.zmescience.com/other/did-you-know/difference-virtual-augmented-reality/

7. devblogs.nvidia.com (2) (2017). Available: https://devblogs.nvidia.com/parallelforall/deep-learning-nutshell-core-concepts/

8. C. Szegedy, W. Liu, Y. Q. Jia, P. Sermanet, S. Reed, D. Anguelov, et al., "Going Deeper with Convolutions," *2015 Ieee Conference on Computer Vision and Pattern Recognition (Cvpr)*, 2015, pp. 1-9.

9. H. Lee, R. Grosse, R. Ranganath and A. Ng. "Unsupervised Learning of Hierarchical Representations with Convolutional Deep Belief Networks". *Communications of the ACM, vol. 54, no. 10*, 2011, pp. 95-103.

10. www.derinogrenme.com (2017). Available: http://www.derinogrenme.com/2015/07/21/derin-ogrenme-deep-learning-nedir/

11. www.pcworld.com (2017). Available: https://www.pcworld.com/article/2864243/nvidia-launches-tegra-x1-bringing-deep-neural-learning-to-self-driving-cars.html

12. www.greenbot.com (2017). Available: https://www.greenbot.com/article/2879437/everything-you-need-to-know-about-nvidias-new-tegra-x1-chip.html

13. timdettmers.com (2017). Available: http://timdettmers.com/2017/04/09/which-gpu-for-deep-learning/

14. www.nvidia.com (2017). Available: https://www.nvidia.com/en-us/geforce/products/10series/titan-x-pascal/

15. www.webrazi.com (2017). Available: https:/webrazzi.com/2014/02/21/google-project-tango

16. www.technologyreview.com (2017). Available: www.technologyreview.com/s/537436/baidus-artificial-intelligence-supercomputer-beats-google-at-image-recognition/

17. www.teradeep.com (2017). Available: https:// www.teradeep.com

18. www.webrazi.com (2) (2017). Available: https://webrazzi.com/2015/12/26/2016-derin-ogrenme-alaninda-ses-getirecek-bes-girisim/

19. A. Sinha, C. Choi and K. Ramani, ”DeepHand: Robust Hand Pose Estimation by Completing a Matrix Imputed with Deep Features”, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, Las Vegas, USA.

20. www.purdue.edu (2017). Available: https://www.purdue.edu/newsroom/releases/2016/Q2/new-tool-for-virtual-and-augmented-reality-uses-deep-learning.html

21. M. Schröder and H. Ritter, “Deep Learning for Action Recognition in Augmented Reality Assistance Systems”, *Proceeding SIGGRAPH’17 Posters Article: 75*, LA

22. E. Welling and M. Oppelt, “Convolutional Neural Networks In Autonomous Vehicle Control Systems” (2017). Available: http://www.pitt.edu/~budny/papers/77.pdf

23. B. Firner, B. Flepp, K. Zieba, L. Jackel, M. Bojarski, and U. Muller. "End-to-End Deep Learning for Self-Driving Cars" (2017). Available: https://devblogs.nvidia.com/parallelforall/deep-learning-self-driving-cars

24. Quora (2017). Available: https://www.quora.com/Is-there-a-difference-between-neural-networks-and-convolutional-neural-networks

25. Ismsit (2017)). Available: http://www.ismsitconf.org/