International Journal of Engineering &Technology, 8 (4) (2019) 603-610

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper



Augmented reality using artificial neural networks –a review

Sreekumar Narayanan [1] *, Srinath Doss [2]

[1] Assistant Professor, Faculty of Computing, Botho University, Botswana
[2] Professor, Faculty of Computing, Botho University, Botswana
**Corresponding author E-mail: sreekumar.narayanan@bothouniversity.ac.bw*

Abstract

The present paper reviews the areas where Augmented Reality (AR) has been used in Artificial Neural Networks (ANN) (Artificial Neural Networks). The focus on systems based on AR is largely on enhancing technologies in diverse application areas such as; defense, robotics, medical, manufacturing, education, entertainment, assisted driving, maintenance and mobile assistance. However, AR is now finding much usage in ANN. The research considered a review based methodology wherein most studies conducted in the past on AR and ANN were reviewed. AR with ANN has profound applications in various sectors and has been developed in an extended way but still has some distance to go afore industries, the military and the common public will receive it as a accustomed user interface. AR would modernize the way people animate and the way industries endeavor by effective utilization. There is an incredible potential in fields such as construction, art, architecture, repair and manufacturing with mediated reality and well-organized visualization through AR.

Keywords: *Augmented Reality; Artificial Neural Networks; Tracking Tracking Technologies; Gesture Recognition.*

1. Introduction

The spate of developments in the processing capacity of computers, innovative technologies pertaining to 3D and capable technologies for image processing have paved the way for real-time systems that have thorough capabilities for computing like the augmented reality (AR) system (ARS) that warrants the need for extensive input data from visual sensors and offers a level of user interaction that is responsive. The focus on systems based on AR is largely on enhancing technologies in diverse application areas such as; defense, robotics, medical, manufacturing, education, entertainment, assisted driving, maintenance and mobile assistance. However, AR is now finding much usage in artificial neural networks (ANN). AR simulation is one area where ANN is being extensively utilized. ANN is being used in AR simulation in order to classify electromyography (EMG) signals with a view to locate a movement that corresponds to it. The use of AR in ANN in several areas are now being widely explored and quick progression in mechanism, the use of AR towards in ANN will have profound applications.
The objective of this paper is to ascertain and review the areas where AR has been used in ANN. This paper will review most studies conducted in the past with similar backgrounds or themes. At the outset, the paper will delve into the definitions of AR and ANN which would be followed by an extensive review of the usage of the two in tandem.

2. Literature review

2.1. Artificial neural network (ANN)

A basic neural network (NN) is known to comprise of a multitude of basic processors that are connected and are collectively referred to as neurons where each neurons are known to generate a series of activations of real-value. Activation of input neurons are through sensors that perceive the environment and at the same time activation of other neurons occur through weighted connections from neurons that were active earlier. The environment can be influenced by certain neurons when specific actions are triggered. To render the NN to project a specific behavior would require the need to find weights which is known as learning or credit assignment. A simple example would be driving a car. On the basis of the problem at hand and the manner in which neurons are connected, behavior of such kind would warrant a computational stage that is long. In such scenarios, each stage mostly in manners that are non-linear, would convert network activation in an aggregate manner.
Though there is vast literature available that pertains to ANNs, there is a dearth of ANN definitions that present a clear-cut idea about the concept. Most of the definitions pertaining to ANN elucidate the concept by drawing reference mostly to figures rather than presenting explanations that are substantiated from a mathematical perspective. Definitions that are largely utilized within applications pertain to units of processing as is evidenced in [41] [28] [38] and algorithms for learning which are exemplified in [41] [28] [38] [64]. The definition of learning according to [41] [28] [38] states that information capture through an alteration in synaptic weights. At the same time, [41] [38] also state that during the process of learning it is possible for ANN to alter its own topology. A rather practical definition is

Copyright © Sreekumar Narayanan, Srinath Doss. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

presented, which states that "ANNs are distributed, adaptive, generally nonlinear learning machines built from many different processing elements (PEs). Each PE receives connections from other PEs and/or itself. The interconnectivity defines the topology. The signals flowing on the connections are scaled by adjustable parameters called weights, wij. The PEs sum all these contributions and produce an output that is a nonlinear (static) function of the sum. The PEs' outputs become either system outputs or are sent to the same or other PEs."

2.2. Elements of ANN

Much emphasis is laid down upon the algorithm of learning and the processing elements by [28][38][64]. It has been observed by Rojas (1996) since the mechanism for computing a biological neuron is yet to be fully understood the author prefers PE rather than opting to use an artificial neuron. Inspiration for PEs is drawn from the neurons that are known to exist in the nerve system of animals. Stimulus is obtained by real neurons which is then modified through synaptic weights and then amalgamate them following which an output with a singular response that is different from the combination is produced. Akin to this, three primary elements of PE were identified by [41]as a summing function for synaptic weight which joins input that matches the corresponding weight and a function for activation in order to generate an output. According to [28] the output generated by a PE is outlined as a function's function where the process of summing is executed to join input, later a function of activation is utilized to compute the result. Likewise, it has been observed by that the four structures (axon, synapses, dendron and cell body) of a genetic neuron is responsible for moulding the slight structure that would be selected from models that are biologically derived[37].

2.3. Neural network architecture

Several neural network architectures are known to be in existence. The variation amongst these architectures arise due to several reasons which would comprise of aspects related to the quantum of layers that are concealed, techniques for learning and functions of activation etc. The three most commonly utilized neural network architectures are as follows: -

2.3.1. Feed forward neural networks (FFNN)

Systems of FFNN permits only signal flow in a single direction which is from the input nodes, passing through concealed nodes and eventually reaching the output nodes. The feed forward neural network is devoid of any loops or feedback. Akin to other neural networks, networks of this kind have their perceptions organized as layers. It is imperative for a link to exist among the inner layer and the hidden layer where it is also necessary for the hidden layer to be linked with another output or hidden layers. Since the flow of information is 'feed forward' between layers, this kind of network is termed as feed forward neural network.

2.3.2. Feedback neural networks

With the help of loops within the network, a feedback neural network permits signals to travel in both directions. The signals pertain to feedback which is known to traverse between layers. Also known as recurrent networks, such networks are said to have a dynamic characteristic which means that they are able to continuously alter their state till such time that a satisfactory response is generated.

2.3.3. Feed forward back propagation neural network

This kind of network is made up of two algorithms pertaining to neural networks. 'Feed forward' as such indicates the technique which allows the network to identify a specific pattern whereas, 'back propagation' elucidates a procedure that provides training for the network. Simply said, 'feed forward' elucidates the manner in which neural networks works and remembers patterns [26].
Easy way to comply with the paper formatting requirements is to use this document as a template and simply type your text into it.

3. Augmented reality system (ARS)

Caudel and Mizell in 1992 were the ones who first utilized the term 'augmented reality' to draw reference to a see-through, heads-up display in combination with sensing for head position and systems that facilitated workplace registration [15]: "This technology is used to 'augment' the visual field of the user with information necessary in the performance of the current task, and therefore we refer to the technology as 'augmented reality' (AR)." A formal definition of the term AR was later provided by [50] to have the features mentioned herein (based on [6]: (a) the capability to join objects both virtual and real in an actual environment, (b) the capability that allows alignment among virtual and real objects in tandem, and (c) the capability that facilitates an interaction during process, that too in three extents and in real time. From this definition, it is said that AR is not restricted to any particular technology and is aptly backed up with additional definitions. For instance, according to [49] [5] it is possible to apply AR to every senses which would also include hearing, smell and touch. From a general perspective, AR is executed through a kind of anchor in the actual world to enable navigation in the AR system. Target images can be used to execute anchors. By linking target images and virtual images, the virtual objects can be oriented to match real world objects [61].
With regards to the taxonomy presented by Milgram pertaining to mixed virtual reality displays [61], the tenets of AR hangs somewhere amidst virtuality and reality. It is almost real as elements that are virtual is combined with an environment that is real. As opposed to that, there is an augmented virtuality which is also considered as a kind of mixed reality akin to AR, and is as similar to virtual reality as users get totally engrossed in environments which are simulated. The association with reality, virtuality and varied reality is well-defined by Milgram as the continuum of reality-virtuality [60].

3.1. Tracking technologies

Tracking techniques based on sensors like inertial, magnetic, acoustic, mechanical and/or optical sensor. Each of these comes with their own merits and demerits. For instance, sensors that are magnetic are light and possess a high update rate at the same time there is also a scope for distortion from metallic substances that is close by. A good review of sensor based tracking is presented.
A huge number of techniques for tracking with computer vision can be split into two types viz., feature based and model based. The logic behind methods that are feature based lies on seeking a match amongst 2D image features and their frame coordinates for a 3D world.

The pose of the camera can be observed through a projection of 3D coordinates of the component within 2D image coordinates that have been observed while reducing the distance to its matching 2D features.
Within certain applications for AR, the vision from computer on its own might not offer tracking solutions that are robust which resulted in the development of hybrid methods that conjoin diverse technologies for sensing. For instance, it was proposed by [4] that there was a need for tracking systems that help the development of AR systems that can function outdoors, which would be GPS based and would also include computer vision and inertial sensing.

3.2. Display technology

This type of displays exhibit a trait where they have the capacity to see beyond the medium of display into the real world that envelops the person observing it, thus helping to realize an extent of presence to the highest level and the supreme grade of 'real space imaging' [65]. Augmentation in display is largely realized with the use of mirrors that are utilized to superimpose graphics generated by computers with an optical medium directly to scenes in the real environment. Displays of this kind are considered as a technology that is mature in aviation systems for the military, which is mostly in the form of head mounted displays (HMDs) or panel mounted. Nonetheless, They are finding innovative applications when interposed with VR based technology, particular in environments like maintenance and manufacturing [15] [43] [61].

3.3. Development tools

Tools for AR development can be applied across diverse platforms. Certain tools pertain to AR applications that are desktop based whereas others would refer to AR application in mobiles. Some tools for development enable tracking based on computer vision to develop the AR application which is commonly supposed to be used indoors, at the same time sensors in mobile devices are deployed to develop applications for external usage [13].

3.4. Input and interaction technology

Taking into account the intricate association amongst an overlaid AR content and the real world, it is believed that there is a scope for creating an innovative AR interface metaphor by amalgamating the improved display potential of AR by physically manipulating tangible user interfaces. This type of amalgamation is termed as tangible AR [46].
Interfaces of AR utilize a single input modality in order to interact with content that is virtual. Nonetheless, interfaces within tangible AR have certain restrictions, as such it only permits user interaction with observable virtual content. In order to circumvent such restrictions gesture and speech interactions within AR environments are being widely explored by [12].

3.5. Evaluation of AR system

Evaluation of AR systems presents diverse challenges. Research pertaining to AR systems is still in its infancy hence the experience with diverse design factors are quite limited. Certain methods for evaluation refers to evaluation techniques that are traditional and evaluations based on guideline, expert and those that are based on users such as aloud method or heuristic evaluation (system usability), multimodal system output, after-effect, time (task completion time) accuracy (e.g. error rates), visual feedback, [27].

4. Application of ANN in AR

A growing interest in AR applications has been witnessed of late to establish distinct settings for education. The paper presented by [18] presents a literature review on AR in educational settings where factors such as utility, benefits, traits and efficiency of AR in an educational context are considered. The key findings presented from the review give an in-depth insight into the existing status of research pertaining to AR. In addition, the paper also delves into the vision and trends pertaining to the scope for additional research and future for the application of AR.

4.1. Gaming sector

A collaborative handheld called Virtuoso that was based on an educational game was developed in 2006. Apart from AR version, the game likewise found application on a desktop PC in Macromedia flash as well as a paper based game that required no mechanism at all. The objective of the game was to segregate an array of artwork on the basis of their creation date within a timeline having three diverse circumstances which included; a PDA, a paper and a PC. The findings indicated that though the players were evaluated with the three diverse conditions, no substantial variation in educational outcomes was observed. A noteworthy fact that emerged was that more attention was accorded by players to the conditions with regards to the utilization of space and the number of players who chose PDA and paper versions of the game as it enabled mutual collaboration as compared to the PC version. Further, the players found it more satisfying and enjoyable to play the game on a PDA interface within the said three conditions.
A studiers tube was introduced as a framework by [74] to develop an AR that was handheld. With this framework the authors were able to develop two different mobile AR games viz., 'expedition schatzsuche' and 'medien welten'. The case study conducted on behalf of 'meiden welten' was structured for a group of students from high school within the age bracket of 12 – 15. For 'expedition schatzsuche' all participants were presented with their individual handhelds to enable them to perform the game. The notion of the game was to solve a quest which was puzzle based and other tasks associated with exhibits. Interviews, data logging and observation was done to evaluate the outcomes and involved a total of 12 students all falling in the age group of 12 years who were known to utilize handheld devices to discover historical artifacts.

4.2. Education

Arvanitis et al. (2007) used a MAR technology based system was utilized to create the CONNECT project that offered support to students to learn science in formal as well as informal learning environment. In this project students were expected to wear an HMD and

other learning platforms mediated with computers to help them in physical and intellectual interaction and visualization within a learning environment that related to materials for instructions. This was facilitated through a hands-on experimentation and reflection that was 'minds on'. Further, it also allowed students to assess the efficacy and usability of the CONNECT project.

4.3. Training

The paper presented by [81], offers an argument that an innovative manner to handle information was essential to effectively utilize AR for facilitating a training amongst operators to learn tasks that were new. The authors proposed an innovative approach to use dynamic information where information is modified automatically for individual operators and their progress in learning for improving efficiency within shorter learning durations. A rule-based skilled system that is aware of human cognition was executed to find out the content of information within the AR. The authors are of the opinion that by using this approach, it is possible for the operator to learn new tasks with more efficacy as compared to utilizing a common approach of static information content that was predetermined.
A method that was suggested by [60] could be deployed in three diverse contexts. These contexts involve the use of a single camera which is supported with a two-camera setup by a sphere based on mirrors, gazing sphere and association of a two-camera setup with fish eye lens. The arrangements above mentioned warrant the need for extra hardware that has the potential to limit the capability of the AR application. A proposed a technique using mirror ball capture. The consistence of illumination by means of fish eye lens is a method accomplished by [44] which increases the requirement of special hardware. [33] recommends a method which had a related provision of having an accepted object in the operational environment. Another method in the literature is illumination assessment by means of a video stream of color images and fractional 3D information engendered by a stereophonic camera [55]. The drawback of this tactic is the requirement of special hardware.

5. Industrial maintenance

In the past, researchers had proposed augmented reality interactive system with voice control and natural language to control the cranes. For instance, the study by [56] developed a prototype based on human machine speech based interfaces with augmented reality and artificial neural network. Various algorithms have been used for different stages such as RANSAC method, dot product of vectors, point feature histograms [42] [17] [40].

6. Tracking technologies

Forward neural network is a method for perceiving the steering light in an environment using a solo light source was introduced. Major design issues faced in this approach is designing the neural network, data set construction and selection of appropriate training model. On basis of the light effect generation, production of the neural network is done. In paper [52] the author developed the clothoid road model using laser camera, digital map and fuse measurements of a camera. Predicting distance from the ego-vehicle and the sensor's range is reliant on the Sugeno-fuzzy system to determine the appropriate weighting for each of the different sensors. Ability to model cubic road curvatures is only through the clothoid model. Joining several clothoids together can exemplify complex curvature which is common in arbitrary rural roads. However, it would significantly increase the parameter space and is not demonstrated by the aforementioned method. In [25] sensor fusion system is not reliant on the digital map and the clothoid model. Sensor fusion system associates to the group of systems that records a routine map comprising landmarks and to confine the car later on using sensor data. Synchronizes from a DGPS sensor are recorded to landmarks extracted from frontward and regressive by detecting cameras and laser scanner of a tenancy grid. Those in the database the systematic GPS location is advanced by matching simultaneously removed landmarks. The design of a landmark database desires to be accomplished in progress though a procedure. The preservation of the database needs to be achieved on a systematic basis to eliminate landmark blunders such as building construction, etc.
Using a convolutional neural network, road scene segmentation from solo images was described by [2]. The CNN is qualified on wide accessible manifest road scenes that are non-essential images from the application of the camera used. The CNN grouping is merged with the color intensity circulation from a ROI leading to overcome and permit adaptions to instant situations of the vehicle through a Bayesian framework. Latest progresses of CNN classifiers display an extraordinary transferability features [62]. The input data among the sensors is analogous, a CNN can be trained before on a vast dataset e.g. Image Net [24] or Cityscapes [21] and improved to the current sensor with only segment of the data desirable to progress it from scrape.
A context that combines a radar sensor, a digital map and a NIR camera sensor was described by [66]. Usage of particle filter achieves a long-term localization by combining a digital map and a radar grid map. By fusing the longitudinally by combination of the long-term recorded digital map is accomplished by precise lateral localization using an visual lane acknowledgment algorithm in the camera image. In the successive works introduced a Bayesian fusion system that accomplishes the last road course valuation. The clothoid is not a road course model but moderately lists of associated 2D point's specimen the left and right borders of the lane in both the systems. The above-mentioned methods rely on the lane pattern indicators for the assessment of an optical map.Based on intensity distribution thresh hold from camera images designated a road boundary estimator. The threshold intensity circulation is removed from a region of interest (ROI) on the converse perception plotted camera image. By using a Bayes filter removed road boundaries are traced over time. For grayscale night vision images, the thresholding method is failed though it is an efficient approach and simple for identifying road pixels in color images.
Using the training decision trees, the road detection was performed by [30] by using a ground plane detection and a disparity features of a stereo camera. Several other handcrafter color and texture features to categorize super pixels divided by a watershed transform. This method intensely trusts on structures not available for grayscale NIR monocular camera images.

7. Interaction

A supervised machine learning based on quaternions allowing gaze estimation was proposed by [63] for foveated rendering in AR devices. A categorization of the algorithm is based on the Ensemble Neural Networks (ENN) projected by[17] and the algorithm was pragmatic Augmented Reality- Question answering System (AR-QAS) in mobile accessories. An interactive user interface was offered in AR-

QAS which automatically replies according to the user queries. The trial results demonstrate a planned algorithm suggesting a huge accuracy on comparison by logistic regression method as well as the method of ANN.
The Helping Hands remote improved reality system, wearable, transportable, hands-free, low price system included of two networked laptops, a head-mounted display damaged by the recipient and a display screen used tenuously by the instructor was developed by [59]. It found that Helping Hands system did not negotiation learning outcomes.

8. Tracking

Due to the complications in tracking general objects, using markers is a concession in AR for indirectly tracking purposes. Markers can be both passive or active, and they can track by traditional cameras, infrared cameras, or lasers. Viva paper uses passive markers, and HMD cameras monitor them. One of the earliest and most famous visual fiducial systems is ARTool Kit[46], which used a square tag comprising a wide black border with an image pattern inside the square. The four corners of the square are used to assesst the pose of the square, while the inner pattern is harmonized against a database of valid patterns for identification. ARToolKit's template matching method however can suffer from high false positives [31]. Many later square-based fiducial systems practice binary codes to remedy the problem of inter tag confusion.
An sample of a system that uses a binary coding scheme to correct bit errors and progress revealing robustness is ARTag [32]. ARToolkit Plus [75] and Studierstube Tracker [57] are other related systems. Apriltag [67] enhances the earlier approaches by guaranteeing a minimum Hamming distance between a tag and its four rotations. Though, the algorithm for detecting quads that is every quad represents the black border around a tag candidate is computationally intensive for deriving image gradients, fitting segments, building lines, and searching quads. Our experimental result (see Table 2) displays that the frame rate is at 23fps on average running with a modern CPU, and thus not appropriate for our real-time detection requirement. The detector has a lower false negative rate, but a relatively higher false positive rate due to a huge number of quad-candidates. (Figure 5(a) in Section 3.3 compares various schemes).
Another prevalent open source library used to track visual fiducial markers is ArUco [31]. The author emphasis on configurable marker generation and occlusion resolution. They practice a contour- based method to speed up quad detection for marker detection. On comparison to the edge- based approach used in Apriltag [67], this method may prime to location error of quad corners, triggering virtual book jitter. The method is based on ArUco, and improvements are made to decrease the aforementioned error (details in Section 3). In addition to widely used square tags, other tag encoding schemes exist.
Blob detection introduced by d-touch [22] is based on RectTIVision [8]. RUNE Tags [9] practice circular dot patterns to make up markers, but do not offer adequate correspondence points to achieve pose estimation. Since Vive paper relies on accurate pose estimation, employing square markers is our logical choice. There are still very problematic tasks to accomplish in real time while direct object tracking and marker less tracking may be desirable [58] by providing 3D models of the tracked objects [82] or by comforting to planar scenes [48]. The work by [20] uses preceding model constructions of the targeted objects as search candidates.
SIFT feature monitored by a RANSAC based homography valuation to track planar targets uses Digilog books [48]. A deep learning approach to produce and decode pictorial markers, which are color descriptions with artistic stylization proposed by the work of [36]. The presence of these methods is more "human-friendly" than regular visual markers but suffers from being tremendously extremely computationally intensive [36][82]. Additionally, unavoidable object recognition mistakes can decrease tracking robustness [20][48]. The Aristo platform is designed to work with any object without prior model information and must support extremely reliable and efficient tracking. Therefore, the marker-based methodology is existing choice. [1] proposed a convolutional neural network to solve the detection issue in camera-based tracking for AR.

9. Gesture recognition for tracking

The first step of our gesture recognition algorithm is hand segmentation, which is been widely proposed in literature. To improve segmentation accuracy different methods utilize different types of visual features and, in some cases, a combination of them (Rautaray & Agrawal, 2015). Textures, motion, skin color, shapes and anatomical models of the hand are some of the popular features. For instance, the study by [14] proposed ARAMIS, a novel interaction approach based on gesture. The authors used hybrid approach such as wearable, pervasive computing paradigm, brute force, fuzzy, ML methods, virtual, non-optical sensing technologies. Further the study by [29] proposed 3D EMoSIFT feature for gesture recognition and authors developed the recognition stage in five steps especially with DTQ, 3D EMoSIFT, SOMP, and neural network classifier for recognition. [52] also proposed features for hand recognition using BRIEF, SIFT, HOG, and the Gabor filter response. Further various classifiers have been proposed such as adapted boosting classifier, hidden Markov model, support vector machine, and conditional random field to discriminate hand gestures.
To produce a hand probability map [54] developed a pixel-to-pixel hand segmentation network by using a binary hand masks and raw images. Features with convolutional layers and maps them directly to a segmentation mask with a fully connected layer was extracted. The proposed CNN-based image alignment scheme [16] achieves greater performance compared to numerous state-of-the-art algorithms. Although CNN-based methods perform well, a trained model is typically very large, and most mobile strategies and wearable are resource-strained to compute fast enough to fulfill the real-time requirement of hand segmentation. To RGB images input images for hand segmentation are not imperfect. With the introduction of service depth cameras like Kinect, researchers can make use of the depth information. A large blob detection performed by [59] in a depth range with median filter and depth normalization as post-processing.
GPU-accelerated implementation proposed by [10] and [11] and formalize a multi-model hand-segmented grounded on haphazard forest and Kth-nearest neighbors. These are achieved by random forest-based methods and state-of-the-art performance, but they entail pixel-level hand. To use deep information with random forest was proposed by [45] and add bilateral filtering and decision adjustment for hand segmentation. A random regression forest using depth context features was by [53] for randomly appraised pixels and fuse the information from both color and depth. Two monochromatic IR cameras and three infrared LEDs was used by the [51] controller used to capture hand and finger motions in the commercial field. By using depth sensors, the problem of changes in brilliance can be moderately alleviated. For the current generation of depth sensors, the depth information still contains substantial noise. There are methods focusing on dealing with special backgrounds instead of learning features from the foreground.
A pool of widely used local appearance features and global appearance features was examined by the work of [52]. The authors suggest that a sparse set of structures achieves the best performance, and global presence models are useful for adapting variations in illumination of hand. To deal with different illumination conditions Gabor filters, super pixel and histograms to perform hand segmentation and uti-

lize a collection of random forests adopted by [7]. Further introduced a shape mask with structured forests to better utilize shape information.

Elderly Care:

Study by [35] proposed augmented coaching ecosystem for non-obtrusive adaptive personalized elderly care. Authors have used machine learning, augmented reality and multimodal user interface, machine-to-machine interaction along with internet of things.

Feature detection and Segmentation:

In the [23] proposed a novel for image augmented reality using two axes in the process. In the detection part, a machine learning step is added, and the registration was done using a statistical appearance models and covariance matrices of dense image descriptions [23]. An arbitrary quadrangle-shaped panel as a background in the Visual Panel system and use dynamic background subtraction to segment hands out. [34] model the background of markers by a mixture of two Gaussian and compute a constriction mask by color segmentation. Use of a black and white background by [39] and [57] to compute a histogram for setting the threshold of division.

10. Conclusion

AR has developed an extended way but still has some distance to go afore industries, the military and the common public will receive it as a common and a familiar user interface. In everyday life through social acceptance, it would be conceivable for prevalent use of AR systems. AR technology can be introduced in schools to enable education at all stages like toddlers educating about shapes or college students picturing and understanding about different mechanisms or elements of a machine. It would be even simpler and more precise for navigation. AR would be extremely useful for personal reminders interpreted through and would accelerate communication. AR would modernize the way people animate and the way industries work by effective utilization. There is an incredible potential in fields such as construction, art, architecture, repair and manufacturing with mediated reality and well-organized visualization through AR. If AR has been oppressed fully it also has the possibility to facilitate the areas of simulation, education, training, military and medical applications. All the limitations must be overwhelmed in order to achieve this.

References

[1] Akgul, O., Penekli, H.I. & Genc, Y. (2016). "Applying Deep Learning in Augmented Reality Tracking." In: 2016 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS). [Online]. 2016, IEEE, pp. 47–54. Available from: http://ieeexplore.ieee.org/document/7907444/. https://doi.org/10.1109/SITIS.2016.17.

[2] Alvarez, J.M., Gevers, T., LeCun, Y. & Lopez, A.M. (2012). "Road Scene Segmentation from a Single Image". In: In Proceedings of the European Conference on Computer Vision. [Online]. 2012, Berlin Heidelberg: Springer, pp. 376–389. Available from: http://link.springer.com/10.1007/978-3-642-33786-4_28. https://doi.org/10.1007/978-3-642-33786-4_28.

[3] Arvanitis, T.., Petrou, A., Knight, J.., Savas, S., Sotiriou, S., Gargalakos, M. & Gialouri, E. (2007). "Human factors and qualitative pedagogical evaluation of a mobile augmented reality system for science education used by learners with physical disabilities." Personal and Ubiquitous Computing. [Online]. 13 (3). pp. 243–250. Available from: https://www.sciencedirect.com/science/article/pii/S1877042813038305. https://doi.org/10.1007/s00779-007-0187-7.

[4] Azuma, R. (1997). "A Survey of Augmented Reality". California: Hughes Research Laboratories. https://doi.org/10.1162/pres.1997.6.4.355.

[5] Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S. & MacIntyre, B. (2001). "Recent Advances in Augmented Reality." Computer Graphics and Applications. [Online]. 26 (6). pp. 34–47. Available from: https://www.sciencedirect.com/science/article/pii/S2212827115011920. https://doi.org/10.1109/38.963459.

[6] Azuma, R.T., Hoff, B.R., Neely, H.E., Sarfaty, R., Daily, M.J., Bishop, G., Vicci, L., Welch, G., Neumann, U., You, S., Nichols, R. & Cannon, J. (1999). "Making augmented reality work outdoors requires hybrid tracking." In: IWAR '98 Proceedings of the international workshop on Augmented reality. [Online]. 1999, Natick, MA, USA: A. K. Peters, Ltd, pp. 219–224. Available from: https://dl.acm.org/citation.cfm?id=322709.

[7] Baraldi, L., Paci, F., Serra, G., Benini, L. & Cucchiara, R. (2014). "Gesture Recognition in Ego-centric Videos Using Dense Trajectories and Hand Segmentation." In: 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops. June 2014, IEEE, pp. 702–707. https://doi.org/10.1109/CVPRW.2014.107.

[8] Bencina, R., Kaltenbrunner, M. & Jorda, S. (2005). "Improved Topological Fiducial Tracking in the reacTIVision System." In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops. 2005, Washington, DC, USA: IEEE, pp. 99–99. https://doi.org/10.1109/CVPR.2005.475.

[9] Bergamasco, F., Albarelli, A., Cosmo, L., Rodola, E. & Torsello, A. (2016). "An Accurate and Robust Artificial Marker Based on Cyclic Codes." IEEE Transactions on Pattern Analysis and Machine Intelligence. 38 (12). pp. 2359–2373. https://doi.org/10.1109/TPAMI.2016.2519024.

[10] Betancourt, A., Marcenaro, L., Barakova, E., Rauterberg, M. & Regazzoni, C. (2016). "GPU Accelerated Left/Right Hand-Segmentation in First Person Vision." In: G. Hua & H. Jégou (eds.). Computer Vision – ECCV 2016 Workshops. ECCV 2016. Lecture Notes in Computer Science, vol 9913. 2016, Springer, Cham, pp. 504–517. https://doi.org/10.1007/978-3-319-46604-0_36.

[11] Betancourt, A., Morerio, P., Barakova, E., Marcenaro, L., Rauterberg, M. & Regazzoni, C. (2017). "Left/right hand segmentation in egocentric videos." Computer Vision and Image Understanding. 154. pp. 73–81. https://doi.org/10.1016/j.cviu.2016.09.005.

[12] Billinghurst, M., Clark, A. & Lee, G. (2015). "A Survey of Augmented Reality." Foundations and Trends® in Human–Computer Interaction. [Online]. 8 (2–3). pp. 73–272. Available from: http://www.nowpublishers.com/article/Details/HCI-049. https://doi.org/10.1561/1100000049.

[13] Billinghurst, M., Kato, H. & Myojin, S. (2009). "Advanced Interaction Techniques for Augmented Reality Applications." In: International Conference on Virtual and Mixed Reality. [Online]. 2009, Berlin: springer, pp. 13–22. Available from: http://link.springer.com/10.1007/978-3-642-02771-0_2. https://doi.org/10.1007/978-3-642-02771-0_2.

[14] Carrino, S., Mugellini, E., Khaled, O.A. & Ingold, R. (2011). "Gesture-based hybrid approach for HCI in ambient intelligent environmments." In: 2011 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2011). [Online]. June 2011, IEEE, pp. 86–93. Available from: http://ieeexplore.ieee.org/document/6007691/. https://doi.org/10.1109/FUZZY.2011.6007691.

[15] Caudel, T. & Mizell, D. (1992). Augmented Reality: "An Application of Heads-Up Display Technology to Manual Manufacturing Processes." In: In: Proc. of the Twenty-Fifth Hawaii International Conference on System Sciences. [Online]. 1992, Kauai: IEEE, pp. 659–669. Available from: http://ieeexplore.ieee.org/document/183317/. https://doi.org/10.1109/HICSS.1992.183317.

[16] Chang, C.-H., Chou, C.-N. & Chang, E.Y. (2017). CLKN: "Cascaded Lucas-Kanade Networks for Image Alignment." 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/CVPR.2017.402.

[17] Chen, C.-H., Wu, C.-L., Lo, C.-C. & Hwang, F.-J. (2017a). "An Augmented Reality Question Answering System Based on Ensemble Neural Networks." IEEE Access. 5. pp. 17425–17435. https://doi.org/10.1109/ACCESS.2017.2743746.

[18] Chen, P., Liu, X., Cheng, W. & Huang, R. (2017b). "A Review of using Augmented Reality in Education from 2011 to 2016." In: Innovations in Smart Learning. [Online]. Singapore: Springer Science+Business Media, pp. 13–18. Available from: http://link.springer.com/10.1007/978-981-10-2419-1_2. https://doi.org/10.1007/978-981-10-2419-1_2.

[19] Chen, T., Bin, D., Liu Daxue, Zhang Bo & Qixu, L. (2011). "3D LIDAR-based ground segmentation." In: The First Asian Conference on Pattern Recognition. [Online]. November 2011, IEEE, pp. 446–450. Available from: http://ieeexplore.ieee.org/document/6166587/.

[20] Comport, A.I., Marchand, E., Pressigout, M. & Chaumette, F. (2006). "Real-time markerless tracking for augmented reality: the virtual visual servoing framework." IEEE Transactions on Visualization and Computer Graphics. 12 (4). pp. 615–628. https://doi.org/10.1109/TVCG.2006.78.

[21] Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S. & Schiele, B. (2016). "The cityscapes dataset for semantic urban scene understanding." In: In Proceedings of the Conference on Computer Vision and Pattern Recognition, 2016. [Online]. 2016, CVPR. Available from: http://sci-hub.cc/10.1109/itsc.2016.7795862. https://doi.org/10.1109/CVPR.2016.350.

[22] Costanza, E. & Robinson, J. (2003). "A Region Adjacency Tree Approach to the Detection and Design of Fiducials." Video Vision and Graphics. pp. 63–69.

[23] Dandachi, G., Assoum, A., Elhassan, B. & Dornaika, F. (2015). "Machine learning schemes in augmented reality for features detection." In: 2015 Fifth International Conference on Digital Information and Communication Technology and its Applications (DICTAP). [Online]. April 2015, IEEE, pp. 101–105. Available from: http://ieeexplore.ieee.org/document/7113179/. https://doi.org/10.1109/DICTAP.2015.7113179.

[24] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K. & FeiFei, L. (2009). "Imagenet: A large-scale hierarchical image database." In: In Proceedings of the Conference on Computer Vision and Pattern Recognition, 2009. [Online]. 2009, CVPR. Available from: http://sci-hub.cc/10.1109/itsc.2016.7795862.

[25] Deusch, H., Wiest, J., Reuter, S., Nuss, D., Fritzsche, M. & Dietmayer, K. (2014). "Multi-sensor self-localization based on Maximally Stable Extremal Regions." In: 2014 IEEE Intelligent Vehicles Symposium Proceedings. [Online]. June 2014, IEEE, pp. 555–560. Available from: http://ieeexplore.ieee.org/document/6856413/. https://doi.org/10.1109/IVS.2014.6856413.

[26] Dharwal, R. & Kaur, L. (2016). "Applications of Artificial Neural Networks: A Review". Indian Journal of Science and Technology. [Online]. 9 (47). Available from: http://www.indjst.org/index.php/indjst/article/view/106807. https://doi.org/10.17485/ijst/2015/v8i1/106807.

[27] Fernandez, C., Izquierdo, R., Llorca, D.F. & Sotelo, M.A. (2015). "A Comparative Analysis of Decision Trees Based Classifiers for Road Detection in Urban Environments." In: 2015 IEEE 18th International Conference on Intelligent Transportation Systems. [Online]. September 2015, IEEE, pp. 719–724. Available from: http://ieeexplore.ieee.org/document/7313214/. https://doi.org/10.1109/ITSC.2015.122.

[28] Fiala, M. (2005a). "ARTag, a Fiducial Marker System Using Digital Techniques." In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). 2005, Washington, DC, USA: IEEE, pp. 590–596.

[29] Fiala, M. (2005b). "Comparing ARTag and ARToolkit plus fiducial marker systems." In: IREE International Worksho on Haptic Audio Visual Environments and their Applications, 2005. 2005, Ottawa, Ont., Canada, Canada: IEEE, pp. 147–152.

[30] Fournier, A., Gunawan, A.S. & Romanzin, C. (1992). "Common Illumination between Real and Computer Generated Scenes." In: Technical Report Common Illumination between Real and Computer Generated Scenes. 1992, Vancouver, BC, Canada, Canada: University of British Columbia.

[31] Garrido-Jurado, S., Muñoz-Salinas, R., Madrid-Cuevas, F.J. & Marín-Jiménez, M.J. (2014). "Automatic generation and detection of highly reliable fiducial markers under occlusion." Pattern Recognition. 47 (6). pp. 2280–2292. https://doi.org/10.1016/j.patcog.2014.01.005.

[32] Gordienko, Y., Stirenko, S., Alienin, O., Skala, K., Sojat, Z., Rojbi, A., Lopez Benito, J.R., Artetxe Gonzalez, E., Lushchyk, U., Sajn, L., Llorente Coto, A. & Jervan, G. (2017). "Augmented Coaching Ecosystem for Non-obtrusive Adaptive Personalized Elderly Care on the basis of Cloud-Fog-Dew computing paradigm." In: 2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). [Online]. May 2017, IEEE, pp. 359–364. Available from: http://ieeexplore.ieee.org/document/7973449/. https://doi.org/10.23919/MIPRO.2017.7973449.

[33] Grinchuk, O., Lebedev, V. & Lempitsky, V. (2016). "Learnable Visual Markers." NIPS. (Nips). pp. 1–9.

[34] Hamrol, A., Ciszak, O., Legutko, S. & Jurczyk, M. (2017). "Advances in Manufacturing." [Online]. Berlin: Springer. Available from: https://books.google.co.in/books?id=1f06DwAAQBAJ&dq=Human-Machine+Speech-Based+Interfaces+with+Augmented+Reality+and+Interactive+Systems+for+Controlling+Mobile+Cranes&source=gbs_navlinks_s.

[35] Haykin, S. (1999). Neural Networks: "A Comprehensive Foundation." New Jersey: Prentice Hall.

[36] Himmelsbach, M., Luettel, T. & Wuensche, H.-J. (2009). "Real-time object classification in 3D point clouds using point feature histograms." In: 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems. [Online]. October 2009, IEEE, pp. 994–1000. Available from: http://ieeexplore.ieee.org/document/5354493/. https://doi.org/10.1109/IROS.2009.5354493.

[37] Janin, A.L., Mizell, D.W. & Caudell, T.P. (1993). "Calibration of head-mounted displays for augmented reality applications." In: Proceedings of IEEE Virtual Reality Annual International Symposium. [Online]. 1993, IEEE, pp. 246–255. Available from: http://ieeexplore.ieee.org/document/380772/.

[38] Kakuta, T., Oishi, T. & Ikeuchi, K. (2004). "Virtual Kawaradera: Fast Shadow Texture for Augmented Reality." Proc. of Intl. Society on Virtual Systems and MultiMedia. (October). pp. 141–150.

[39] Kang, B., Tan, K.-H., Tai, H.-S., Tretter, D. & Nguyen, T.Q. (2016). "Hand Segmentation for Hand-Object Interaction from Depth map.", 2016. https://doi.org/10.1109/GlobalSIP.2017.8308644.

[40] Kato, H. & Billinghurst, M. (1999). "Marker tracking and HMD calibration for a video-based augmented reality conferencing system." In: Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR'99). 1999, San Francisco, CA, USA, USA: IEEE Comput. Soc, pp. 85–94. https://doi.org/10.1109/IWAR.1999.803809.

[41] Kim, K., Lepetit, V. & Woo, W. (2010). "Scalable real-time planar targets tracking for digilog books.", The Visual Computer. 26 (6–8). pp. 1145–1154. https://doi.org/10.1007/s00371-010-0490-6.

[42] Kipper, G. & Rampolla, J. (2012). "Augmented Reality: An Emerging Technologies Guide to AR.", Italy: Syngress.

[43] Krevelen, D. & Poelman, R. (2010). "A Survey of Augmented Reality Technologies Applications and Limitations.", International Journal of Virtual Reality. [Online]. 9 (2). pp. 1–20. Available from: https://www.sciencedirect.com/science/article/pii/S2212827115011920. https://doi.org/10.20870/IJVR.2010.9.2.2767.

[44] Leap Motion (2017). "Leap Motion.", 2017.

[45] Li, C. & Kitani, K.M. (2013). "Pixel-Level Hand Detection in Ego-centric Videos.", In: 2013 IEEE Conference on Computer Vision and Pattern Recognition. June 2013, Portland, OR, USA: IEEE, pp. 3570–3577. https://doi.org/10.1109/CVPR.2013.458.

[46] Liang, H., Wang, J., Sun, Q., Liu, Y.-J., Yuan, J., Luo, J. & He, Y. (2016). "Barehanded music. In: Proceedings of the 20th ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games - I3D '16.", 2016, New York, New York, USA: ACM Press, pp. 87–94. https://doi.org/10.1145/2856400.2856411.

[47] Malik, S., McDonald, C. & Roth, G. (2002). "Hand tracking for interactive pattern-based augmented reality.", In: Proceedings. International Symposium on Mixed and Augmented Reality. 2002, Darmstadt, Germany, Germany: IEEE Comput. Soc, pp. 117–126. https://doi.org/10.1109/ISMAR.2002.1115080.

[48] Mather, C., Barnett, T., Broucek, V., Saunders, A., Grattidge, D. & Huang, W. (2017)." Helping Hands: Using Augmented Reality to Provide Remote Guidance to Health Professionals.", Studies in Health Technology and Informatics. 241. pp. 57–62.

[49] Maung, P.P. (2012). "Augmented Reality using a Neural Network.", [Online]. Ripon College. Available from: http://www.micsymposium.org/mics2012/submissions/mics2012_submission_27.pdf.

[50] Milgram, P. & Kishino, F. (1994). "A Taxonomy of Mixed Reality Visual Displays.", IEICE Transactions on Information Systems. [Online]. 77 (12). pp. 1321–1329. Available from: http://www.micsymposium.org/mics2012/submissions/mics2012_submission_27.pdf.

[51] Montenegro, J.M.F. & Argyriou, V. (2016). "Gaze estimation using EEG signals for HCI in augmented and virtual reality headsets. [Online].", 2016. Kingston. Available from: http://eprints.kingston.ac.uk/38202/1/Argyriou-V-38202-AAM.pdf. [Accessed: 9 November 2017]. https://doi.org/10.1109/ICPR.2016.7899793.

[52] Muller, B., Reinhardt, J. & Strickland, M.. (M.5M."Neural Networks an Introduction.", Berlin Heidelberg: Springer-Verlag.

[53] Naimark, M. (1991). "Elements of realspace imaging: A proposed taxonomy.", Stereoscopic Displays and Applications II. [Online]. 1457. Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.83.6861&rep=rep1&type=pdf.

[54] Nowrouzezahrai, D., Geiger, S., Mitchell, K., Sumner, R., Jarosz, W. & Gross, M. (2011). "Light factorization for mixed-frequency shadows in augmented reality.", In: 2011 10th IEEE International Symposium on Mixed and Augmented Reality. October 2011, IEEE, pp. 173–179. https://doi.org/10.1109/ISMAR.2011.6092384.

[55] Olson, E. (2011). AprilTag: "A robust and flexible visual fiducial system.", In: 2011 IEEE International Conference on Robotics and Automation. May 2011, Daejeon, South Korea: IEEE, pp. 3400–3407. https://doi.org/10.1109/ICRA.2011.5979561.

[56] Pemasiri, A., Wijebandara, C., Wijayarathna, S., Perera, A. & Gamage, C. (2015). "An online lighting model estimation using neural networks for augmented reality in handheld devices." In: 2015 Fifteenth International Conference on Advances in ICT for Emerging Regions (ICTer). [Online]. August 2015, IEEE, pp. 4–8. Available from: http://ieeexplore.ieee.org/document/7377658/. https://doi.org/10.1109/ICTER.2015.7377658.

[57] Pressigout, M. & Marchand, E. (2006). "Hybrid tracking algorithms for planar and non-planar structures subject to illumination changes.", In: 2006 IEEE/ACM International Symposium on Mixed and Augmented Reality. [Online]. October 2006, IEEE, pp. 52–55. Available from: http://ieeexplore.ieee.org/document/4079256/. https://doi.org/10.1109/ISMAR.2006.297794.

[58] Risack, R., Klausmann, P., Krüger, W. & Enkelmann, W. (1998). "Robust lane recognition embedded in a real-time driver assistance system.", In: IEEE Intelligent Vehicles Symposium. [Online]. 1998, IEEE, pp. 35–40. Available from: http://sci-hub.cc/10.1109/itsc.2016.7795862.

[59] Schule, F., Schweiger, R. & Dietmayer, K. (2013b). "Augmenting night vision video images with longer distance road course information.", In: 2013 IEEE Intelligent Vehicles Symposium (IV). [Online]. June 2013, IEEE, pp. 1233–1238. Available from: http://ieeexplore.ieee.org/document/6629635/. https://doi.org/10.1109/IVS.2013.6629635.

[60] Seo, Y. & Rajkumar, R.R. (2014). "Detection and tracking of boundary of unmarked roads", In: International Conference on Information Fusion. 2014, Spain: IEEE.

[61] Shelhamer, E., Long, J. & Darrell, T. (2015). "Fully Convolutional Networks for Semantic Segmentation.", In: In Proceedings of the Conference on Computer Vision and Pattern Recognition. [Online]. 2015, CVPR, pp. 1–10. Available from: https://people.eecs.berkeley.edu/~jonlong/long_shelhamer_fcn.pdf.

[62] Sinha, A., Choi, C. & Ramani, K. (2016). "DeepHand: Robust Hand Pose Estimation by Completing a Matrix Imputed with Deep Features.", In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). June 2016, IEEE, pp. 4150–4158. https://doi.org/10.1109/CVPR.2016.450.

[63] Supan, P., Stuppacher, I. & Haller, M. (2006). "Image Based Shadowing in Real-Time Augmented Reality.", The International Journal of Virtual Reality. 5 (3). pp. 1–7. https://doi.org/10.20870/IJVR.2006.5.3.2692.

[64] Syberfeldt, A., Danielsson, O., Holm, M. & Wang, L. (2016). "Dynamic Operator Instructions Based on Augmented Reality and Rule-based Expert Systems.", Procedia CIRP. [Online]. 41. pp. 346–351. Available from: http://linkinghub.elsevier.com/retrieve/pii/S2212827115011920. https://doi.org/10.1016/j.procir.2015.12.113.

[65] Tsogas, M., Floudas, N., Lytrivis, P., Amditis, A. & Polychronopoulos, A. (2011). "Combined lane and road attributes extraction by fusing data from digital map, laser scanner and camera." Information Fusion. [Online]. 12 (1). pp. 28–36. Available from: http://linkinghub.elsevier.com/retrieve/pii/S1566253510000199. https://doi.org/10.1016/j.inffus.2010.01.005.