Autism Spectrum Disorder Therapy: Analysis of Artificial Intelligence integrated Robotic Approach

Ananya Ananth Rao, Shaun Qien Yeau Tan, Raghavi R, Archit Srivastava, C H Renumadhavi
Department of Electronics and Instrumentation Engineering, R V College of Engineering, Bengaluru, India
E-mail: ananyaananthrao.ei17@rvce.edu.in, shaunqytan@gmail.com, raghaviramrao@gmail.com, architsrivastava3115@gmail.com, renumadhavi@rvce.edu.in

Abstract. Autism Spectrum Disorder is a developmental disorder that may manifest in a myriad of ways such as difficulties in social interaction and a tendency to engage in repetitive patterns of behaviour. Over the years, several kinds of treatment protocols have been proposed and implemented. One such area that is attracting the attention of researchers in the field is a robot-based approach in the treatment of children diagnosed with the disorder. Here we propose a viable method via the integration of apex technological methods like Artificial Intelligence, Machine Learning and Medical Robotics, coupling it with problem specific algorithms in OpenCV along with principles of Applied Behavioural Analysis to help possibly alleviate a key symptom displayed by children in terms of level of social interaction - that of eye-contact. This would be achieved via an AI-integrated Robotic Framework. The project also considers the possibility of inclusion of the growing research field of Quantum Computing to realize the process and investigates its viability as a potential source of innovation in the future.

1. Introduction
Pervasive Developmental Disorders is a broad term that is applied to various developmental disorders such as Rett’s Disorder, Childhood Disintegrative Disorder, Pervasive Developmental Disorder-Not Otherwise Specified (PDD-NOS) and Asperger’s Syndrome along with Autism [1].

The World Health Organisation puts the worldwide ASD prevalence estimate at 1 in 160 children [2].

India reports a rate of approximately 1 out of 500 or 0.20% of population [1]. Symptoms typically show up in the first two years of life although one can possibly be diagnosed at any age. These symptoms can potentially restrict a person’s ability to function in social settings such as classrooms and workplaces as per the ‘Diagnostic and Statistical Manual of Mental Disorders (DSM-5)’ by the American Psychiatric Association [3].

1.1. Signs and diagnosis of ASD

The word ‘spectrum’ is attached to autism because of the diversity of symptoms - both in terms of type and severity - experienced by those with the disorder [3]. This means that a person diagnosed with ASD need not necessarily have every symptom associated with it. Those who do show the same symptoms might do so in varying degrees of severity. ASD symptoms are generally life-long [4]. While some people learn to live independently, others need life-long care.
Having ASD might preclude a person from making eye contact, pointing at an object of interest or direct attention to an object being pointed at, communicating their needs and feelings in typical actions or phrases, easily adapting to a routine change or reacting to things in a way deemed “conventional” [4]. Echolalia is a rather common symptom [5]. There is also a rather prevalent tendency of wanting to be alone, a reduced interest in or understanding of others and a preference to not engage in any display of affection such as cuddling unless they want to. Therefore on one hand they may not respond to people, while on the other hand, they might not know how to, even if they were interested [6].

Due to the lack of a standardised diagnostic test, such as a blood test, diagnosis of ASD comes with its challenges. Doctors therefore rely on the behaviour and development of a child to arrive at a reliable conclusion which typically occurs at around 2 years of age. There are chances that a diagnosis may not be made until the child is much older. At such times, the child is deprived of critical early intervention. Diagnosis for various age groups is as follows [3]:

- **Children** - consists of general developmental screening and additional screening. The children that show developmental abnormalities during the general developmental screening are advised additional screening. Any anomalies in the behaviour of older children are noticed by people in their places of social interaction such as schools, play areas and at home.

- **Adults** - diagnosis among adults is comparatively more challenging as the signs and symptoms shown may be similar to other underlying disorders like anxiety or attention deficit hyperactivity disorder (ADHD).

The final diagnosis may require a consolidation of the conclusions of several different types of healthcare experts such as child psychologists or psychiatrists, developmental paediatricians, speech-language pathologists and neuropsychologists. Blood and hearing tests along with thinking, language and skills required to independently carry out day-to-day activities weigh in on the final evaluation.

Some studies suggest that children with ASD tend to prefer geometric shapes over social figures [6]. Eye tracking to evaluate this preference has been touted as a method of early diagnosis along with neuroimaging [6].

### 1.2. Causes

All causes of ASD are not yet known. There are, however, certain possible contributing factors that may increase the likelihood for developing the disorder. A unanimous opinion among researchers is that autism has a strong hereditary factor. Siblings of those with autism have a risk of ASD that is fifty times more than the general population [7]. The risk is at 60–90% and 0–5% among identical twins and fraternal twins respectively [7]. Individuals also have a heightened risk of ASD if they have tuberous sclerosis or fragile X syndrome. When it comes to pregnancy and birth, there is a link seen between the prescription drugs valproic acid and thalidomide taken during pregnancy and a higher risk of ASD [8][9]. ASD risk in a child may also increase with increasing age of the parents at the time of conception and/or birth of the child [4]. Environmental factors such as vaccines, mercury, viruses, and low Vitamin D levels are suspected to play a role [9].

### 1.3. Treatments

ASD currently does not have a cure. Since the display of symptoms among those affected is so diverse, treatments are generally aimed at the specific symptoms seen, on a case by case basis. The earlier the intervention, the greater the chances for a child’s improvement [4]. Several treatment methods have been proposed. Care and rehabilitation of these children is usually multi-faceted.

- **Nutrition**: There exists some evidence - albeit limited - for the ‘Opioid-Excess Theory’, its underlying mechanism and treatment method. This theory postulates that a deficit in the production of gluten- and casein-related digestive enzymes leads to insufficient metabolization of gluten and casein-related peptides. There is an ensuing disruption in the central nervous system owing to the peptides getting attached to opioid neuro-receptors instead of crossing the brain–blood barrier. This
is said to manifest in the form of ASD symptoms. This theory spurred the popularity of a gluten-free or casein-free diet. Vitamin D and Folic acid as dietary supplements are also popular as a treatment method [6]. However, it is to be noted that in general, there is no strong or sufficient evidence to support dietary intervention for treatment of ASD [10]. Any changes in diet is a decision taken by caregivers - who feel that there may be an alleviation of symptoms - in consultation with the concerned healthcare practitioners [4].

- Pharmacological: People with ASD usually also tend to develop epileptic activity, irritability, aggression and hyperactivity. Suitable pharmacological intervention, therefore, becomes necessary for which there exist two FDA approved medications, risperidone and aripiprazole. While there is evidence to support short-term benefits in terms of behaviour of those with ASD, long-term benefits, and shortcomings, if any, are yet to be sufficiently concluded. Although clinical trials haven’t yet established their effectiveness, SSRIs (selective serotonin reuptake inhibitor) are also often used to treat ASD’s comorbid symptoms [6].

- Behavioural and communication oriented: Applied Behavioural Analysis (ABA), speech therapy, assistive technology, occupational therapy and social skills training are some approaches that fall under this category [4]. As robots have displayed a certain potential for bridging the gap between technology and medical evidence - based treatment protocols, researchers and healthcare experts all over the world have been engaged in coming up with robot-based treatments for those with ASD. Interactive robots have been touted as a possible means of mitigating problems related to eye contact, emotion recognition and so on [11]. Since such a treatment for children would require a simulation of intelligence in behaviour, action and basic autonomy in decision making, an integration of artificial intelligence and machine learning would be vastly favourable [11]. In this text, the relevant principles of ABA and assistive technology would be delved upon as they form the basis on which our project was implemented albeit, in a software based approach.

1.4. Applied Behavioural Analysis

Applied Behaviour Analysis (ABA) is a type of therapy aimed at improving deficiencies in various social, communication and adaptive learning skills through positive reinforcement [12,13]. It began with the thought - that was later confirmed - from B. F Skinner that behaviour is determined by selection by consequences [14]. Skinner showed that positive consequences to a behaviour encourage the recurrence of the behaviour while those met with negative responses would eventually die out. The earliest treatment for ASD based on ABA, developed by Ivar Lovaas, was the early and intensive behavioural intervention (EIBI). EIBI was originally developed as an intensive treatment that would span several hours over 5-7 days a week in a one-on-one format and relied heavily on what is known as Discrete Trial Teaching (DTT). Discrete Trial teaching works on repetitively and briefly teaching a child by providing a specific instruction. Data are collected across each of the trials or on a subset of the trials [15]. Studies have also been carried out that show how combining the principles of ABA with that of a socially assistive robot (SAR) that was designed to adapt based on the interactions and responses from a child and effectively implement treatment procedures without necessitating the presence of a human caretaker allowed greater flexibility in treatment locations [16].
2. Related Work

Due to the gradual synergism of technology and medicine, progress has been made on the development of robots suitable for the treatment of autistic children. In this regard, positive effects of such robots have already been reported in literature such as in [18,19].

As seen in [20] robots may also be used in robot mediated therapeutic systems for skill learning through imitation. Effective ASD therapy was experimented between interactive intervention based on Robots as well as under normal classroom scenarios, by calculating and analysing the average amount of time ASD diagnosed children could maintain eye contact in each of those environments. It was noted that the eye contact maintaining performance was performed better in an interactive-robot intervened environment [21].

A low cost robot design that is capable of interacting with ASD diagnosed children was implemented wherein an object detector is trained and tested based on Haar-like features for the robot to recognise the actions of the subject [22].

Thus, as seen from such work, research into robot mediated treatment protocols have seen gradual growth, with artificial intelligence playing a pivotal role though there is plenty of room - a definite need - for more such studies. Therefore, a software implementation similar to the ABA integrated framework as given in [11], will be described in the subsequent sections.

3. Methodology and Implementation

The main focus of the basic software implementation carried out by us was on the accurate detection and recognition of the eye contact displayed by the subjects, with audio feedback being used in an attempt to reinforce positive behaviour; the behaviour being improved eye contact in this case. Thus, the main input the system received was a video feed recorded using a webcam at real-time with the primary output the system gave in response to the subject eye contact being an audio response.

The basic project outline is defined in the immediate subsection following which explanation of the project model structure will be undertaken.
3.1. Basic Project Outline
The basic project outline is defined in the block diagram as shown in figure 2.

Here we have used a generic webcam with a resolution of around 1280x720 pixels which transfers the real-time video feed to the eye contact detection and recognition model, where the model looks to identify a subject’s face and eye regions and then based on the predetermined algorithm determines whether the subject was displaying eye contact or not.

Also included are a timer and counter. The timer is used to determine for how long the model should be actively detecting eye contact. Its value is predetermined before the model is run and may be tweaked in practice as per the needs of the subject and specific instructions, if any, from a health care provider. The counter is used to keep a count of how long the subject has looked in the general direction of the webcam. Based on this counter’s value, audio based prodding suggestions are given by the system.

3.2. Base Model
This is the first level of Detection and Classification where the attention state of the subject is detected from tracking the face and detecting if the face is pointed directly at the webcam. Along with OpenCV, the MOSSE (Minimum Output Sum of Squared Error) [23] tracker takes the selected region of interest, which here is the frontal face of the subject and tracks it using adaptive correlation technique. The Viola Jones Face Detection Algorithm [24] is used to detect the frontal face. The algorithm works by detecting Haar-like features of the face and data of those features put together help to detect if the front of a face is present or not.

A bounding box is formed around the face and the counter is incremented once a face is detected. Whenever the subject is pointing the head away from the webcam, the algorithm does not track or detect the face - an obvious indication that eye contact is not being made with the primary, intended stimulus provider.

3.3. Higher Level Implementation: Eye Segmentation
This is the next level of Detection and Segmentation where eye segmentation is performed on the image of the face of the test subject. This is done in an attempt to ensure that when the face of the subject is pointed at the webcam, the eyes are also looking towards the same direction, and not slightly towards left or right. This is done by extracting eyes from the subjects’ faces in OpenCV with the help of scikit image module - “io”, which helps in reading and writing images.

During training, images from Columbia Gaze dataset [25] are read and the region of interest - the eyes of the subject - are extracted and stored in a folder with the help of the io module. The extracted images of the eyes are further used in the next step- classification if the subject is able to maintain eye contact or not.

3.3.1. Eye Region Extraction Model for Dataset Augmentation
For the model to determine whether a subject has made eye contact, given the eye region data, the model would be required to learn about the different gaze poses available. Thus for this purpose the dataset images were modified such that only the eye region was extracted with each extracted eye region being of size 130x50 pixels, which is of the same size as that of the images which the final project model extracted from the real live image data. Figure 3 represents the flowchart for the eye region extraction module.
3.4. Classification
As seen in the previous sub-section, we extract the eye region within the subjects’ facial features and then determine whether the subject is making eye contact or not. For the classification model used in this project, we made use of a CNN (Convolutional Neural Network) trained on the images present in the Columbia Gaze dataset [25]. This is a publicly available dataset consisting of 5,880 images of 56 people (32 men, 24 women) with different gaze and head poses. The resolution of each image was of 5,184 x 3,456 pixels, with the subjects between 18 and 36 years old and belonging to different ethnic groups. 21 of them wore glasses. Pictures were taken of each subject comprising of 5 horizontal head positions (0°, ±15°, ±30°), 7 horizontal gaze directions (0°, ±5°, ±10°, ±15°) and 3 vertical gaze directions (0°, ±10°) putting the number of gaze locking images per subject at a value of 5.

The trained model would be presented with the extracted eye region of each test subject using the image segmentation algorithm on the real-time feed, which is obtained using the webcam. The output would be binary indicating whether eye contact was made or not. If the result was in the affirmative, the counter would increase by one, otherwise it would retain the same count.
3.4.1. Convolutional Neural Network
Convolutional neural networks (CNN) are a variant of deep neural networks typically employed for the analysis of visual images. It finds application in many fields such as object recognition in images and videos, natural language processing, etc.

For the present dataset, the convolutional neural network used initially had 3 convolution layers with max pooling and batch normalisation done at each step. This was followed by a dense neural network with two output neurons. Dropout and flattening were also done. The activation function chosen for all the convolution layers was ReLu with the final activation function at the output being sigmoid.

The loss metric defined was categorical cross entropy which is generally used for multiclass classification. The Adam optimizer was chosen.

After several adjustments to the network done in pursuit of better accuracies, the current model consists of 5 convolution layers, with max pooling and batch normalisation at each step along with the other parameters such as Dropout, Flattening, activation functions and number of output neurons remaining the same.

The architecture for the final CNN model used in the final project model is shown in figure 4.

3.5. Final Project Model
The final step was to incorporate all the elements described above into the final project model. Initially, the name of the participant for whom the session is activated would be asked. This enables the model to provide more personalised responses based on the session progress.

During the model operation, based on the participant responses, suitable model responses are generated. The responses are made based on the amount of time the participant’s eye contact is recorded by the model. The CNN model is required to determine if eye contact was made based on which a count variable is recorded. This variable is thus used to determine the total time eye contact is made.

For this, the webcam is used to acquire images of the participants, which is then passed into the eye region extraction module developed primarily to be used on the real time image data acquired. In addition, there is also a countdown timer which is used to maintain strict control over each session duration.

Figure 5 is used to represent the final model flowchart.

4. Results and Discussion
From the initial dataset [25], the eyes region of each image was obtained after resizing each image from their original size of 5184x3456 pixels to 600x500 pixels. This made it easier to extract the eyes region from each image. Represented in figure 6, is a comparison of the original image available in the dataset to their eye region extracted images.

The dataset consisted of 56 participant images, each participant having 105 images representing different eye and head positions. However, for training our model we selected a condensed version of the dataset due to the fact that the original dataset presented a heavy skew in the ratio of the number of images of eye contact to the number of images with no eye contact. The subset consisted of 280 images of each class and was therefore had a total of 560 images.

To improve the model accuracy, several measures were tried and tested. More layers were added and the loss metric was changed to binary cross entropy. However, training the network over 600 epochs seemed to cause overfitting and a resulting drop in accuracy to 68%.
Figure 4. Final CNN model architecture
Following this, a call back called early stopping was used. Early stopping is a method used to avoid overfitting. This is done by having the user specify a measure of performance that the call back will monitor and when a certain pre-specified condition is reached, the training process is stopped. After this, the model reported an accuracy of 70%.

In order to improve the performance, the loss metric was changed back to categorical cross entropy with early stopping. However, since the accuracy then fell to 68% after 27 epochs, it was evident that some changes were still needed.

The final model did not include the early stopping call back. Instead, the network was trained for 384 epochs - a number greater than any that the call backs allowed but less than the initial number of 600 epochs. 384 was chosen due to the fact that the network showed minimum loss at this stage while being trained for 600 epochs before rising. The resulting accuracy reached a value of 76%. This model was deployed for the final project. The confusion matrix is as seen in figure 7. Furthermore, a ten-fold cross-validation was also carried out, with the average accuracy coming to 74.46%.

Once the model was trained, tested and deployed in the final project, there seemed to be a decrease in accuracy. This decrease was due to the inaccuracies involved in the eye extraction.
Figure 6. Examples of the extracted eye region from images taken from the dataset [25]

code for real time data wherein the region containing both the eyes of the subject were not successfully extracted at every time instance. Figure 8 shows in the form of table, some of the images extracted when running the final project model on the dataset images. Finally, when the entire project was tested it was concluded that the project has much room for improvement in terms of the necessary accuracy required. This was due to the relatively low classification model accuracy coupled with the low real time eye extraction model accuracy.

Although the final project model was able to extract the correct region of interest within the real time image frame and classify it correctly, it did not do so on a consistent basis. However, it did show that it is possible to develop a model which can be used for improving and reinforcing eye contact based on eye contact detection and classification along with ABA principles.

Thus in order to develop a more robust and accurate model, an improved classification model as well as a more precise image processing algorithm for eye region extraction would need to be developed.
Figure 7. Confusion matrix of the final model: 0 - eye contact not detected; 1 - eye contact detected

Figure 8. The extracted eye region from dataset [25] images using the final project eye extraction module
5. Future advancements in the field of Image Processing (Quantum Image Processing)

There are certain options that one may pursue while aiming to improve on accuracy and overall performance. It could be implementing more advanced neural networks or artificially increasing the size of the dataset by augmenting the dataset. Furthermore, datasets and test subjects consisting of children with ASD could help with a more realistic understanding of the performance of the proposed methodology and the right changes to be made, if any. Another area holding the promise of novelty and advancement is Quantum Visual Tracking and Image Processing.

5.1. Quantum Visual Tracking and Quantum Image Processing

When the task is of locating a moving object within a video clip, it is known as Visual Tracking (VT) with a lot of attention being given to detection tracking, which detects objects using a variety of discriminatory machine learning classifiers. With this approach, each frame pair is run through in a chronologically contiguous manner [26]. The basic idea of quantum mathematics is about the development of states which do not accept 0 and 1, but rather values represented as follows:

\[ |0\rangle + \beta|1\rangle \]

where \( \alpha \) and \( \beta \) are the normalized probability amplitudes. If we consider the classical \( N \) bits, we can apply the superposition principle to encode the classical information of \( N \) bits in \( \log_2 N \) quantum bits.

Quantum behaviour can be simulated on a classic computer, but the computing costs increase exponentially with the number of qubits (quantum counterparts of binary bits). Therefore, even though several quantum image models as well as quantum analogs of image processing techniques such as Quantum Fourier transform (QFT) and quantum wavelet transform (QWT) have been proposed that are more efficient than their classical counterparts [27], they cannot currently be implemented because hardware support for quantum image processing and quantum visual tracking has not yet been developed. Because of the limitations we have implemented QFT and tried to analyze its result. QFT being the most crucial step of QVT, the implementation helps us to analyze the overall process and reach close to the desired aim.

5.2. Quantum Fourier Transform

Quantum Fourier Transform (short for QFT) invented by Don Coppersmith is a linear transform on a quantum bit, part of numerous quantum algorithms (which decomposes and computes discrete logarithms), quantum phase estimation algorithms, etc. The circuit above in figure 9 represents a Quantum Fourier Transform circuit. The circuit used 3 qubits as the inputs which spanned over a Hamiltonian Space of 8, thus giving the overall transform for a total of 8 vector spaces. The circuit is a combination of Unitary Rotation Gates and Hadamard Gates. The Hadamard gates are used to have the qubits in the superposition states and the Unitary Rotation gates are used to apply the required rotation to the states of qubits at each step [28].

The above QFT Circuit is an integral part of the Quantum Image Processing field and
Quantum Visual Tracking Circuit. It also forms the building block for various other circuits which have a revolutionary application in each field of computation such as Shor’s Factoring Algorithm thus providing a possible direction to new research in the field of image processing and object tracking. As we see the above graph in figure 10, when the Quantum state of $|5>$ was passed through the circuit and the circuit was allowed to be compiled over the IBM Quantum Computers in Stockholm, the results of the readout predicted a probability of 0.688 to find the 010 state which corresponded in accordance with the theoretical results.

The other probabilities that we see while the graph is plotted is because of the existence of errors present in the qubits and the gates used. As the total depth of the circuits used increases, the error increases per qubit and thus the inaccuracies in the result increases. To visualize the effect of QFT on the state of the qubits, a study of the Bloch-Sphere of the qubits is taken to get the knowledge of the changes in the state and the transformations following the 3 qubit input that has been provided [29].

6. Conclusion

The observations stated thus far show that there is great scope for improved technological intervention and innovation in the treatments associated with Autism Spectrum Disorder particularly those pertaining to artificial intelligence, machine learning and robotics. With the added advantage of personalizing based on individual requirements, it makes room for the fact that there is widespread heterogeneity and that each person will have different needs and requirements as mandated by an associated health care expert.

From the viewpoint of possible future improvements, the focus would be on improving the CNN model accuracy as well as the eye extraction accuracy in the final project model. Future work may also focus on developing models which aim at other exercises present in the artificial intelligence integrated robot assisted treatment protocol.

Keeping in mind the current research progress being made in the field of Quantum Visual Tracking and Quantum Image Processing, algorithms developed from such domains may be implemented further using real world Quantum Computers when the adequate hardware becomes available, specifically for the use case discussed in the project and thus, pave the way for a comparison of the results of the Classical CNN with the aforementioned Quantum Visual Tracking.
References

[1] http://www.rehabcouncil.nic.in/writereaddata/autism.pdf
[2] https://www.who.int/news-room/fact-sheets/detail/autism-spectrum-disorders?text=Epidemiology,figures
[3] https://www.nimh.nih.gov/health/topics/autism-spectrum-disorders-asd/index.shtml
[4] [5] Neely L, Gerow S, Rispoli M, Lang R and Pullen N. Treatment of echolalia in individuals with autism spectrum disorder: a systematic review. 2016. Rev J Autism Dev Disord 3, 82–91.
[6] Campisi L, Imran N, Nazeer A, Skokauskas N, Azeem M W, Autism spectrum disorder. 2018. Br. Med. Bull., 127(1), 91–100.
[7] Frith U and Happé F. Autism spectrum disorder. 2005. Curr Biol. 15(19), 786-90.
[8] Christensen J, Grønborg T K, Sørensen M J, Schendel D, Parner E T, Pedersen L H and Vestergaard M. Prenatal valproate exposure and risk of autism spectrum disorders and childhood autism. 2013. JAMA. 309(16) 1696-703.
[9] Dietert R R, Dietert J M and Dewitt C. Environmental risk factors for autism. 2011. Emerg. Health Threats J, 4(1), 7111.
[10] Santh N, AnDrrews J C, McSheeters M L and Warren Z E. Nutritional and dietary interventions for autism spectrum disorder: a systematic review. 2017. Pediatr. 139(6):20170346.
[11] Giuseppe P, Carolis B D and Esposito F. Artificial intelligence for robot-assisted treatment of autism. 2017. WAIH@API*IA.
[12] https://www.psychologytoday.com/us/therapy-types/applied-behavior-analysis.
[13] https://www.healthline.com/health/aba-therapy.
[14] Skinner, B. F. (1938). The behavior of organisms: an experimental analysis. Appleton-Century.
[15] Roane H S, Fisher W W and Carr J E. Applied behavior analysis as treatment for autism spectrum disorder. 2016. J. Pediatr.
[16] Salvador M, Marsh A S, Gutierrez A and Mahoor M H. Development of an ABA Autism Intervention Delivered by a Humanoid Robot. 2016. Lecture Notes in Computer Science. 9979.
[17] https://familytreeaba.com/overview
[18] Palestra G, Varni G, Chetouani M and Esposito F: A multimodal and multilevel system for robotics treatment of autism in children. 2016. Proceedings of the International Workshop on Social Learning and Multimodal Interaction for Designing Artificial Agents. 3. ACM (2016).
[19] Chevalier P, Martin J C, Isableu B, Bazile C, Jacob D O and Tapus A. Joint attention using human-robot interaction: impact of sensory preferences of children with autism. 2016. Robot and Human Interactive Communication (RO-MAN), 2016 25th IEEE International Symposium. 849–854. IEEE (2016)
[20] Zheng, Z, Young E M, Swanson A R, Weitlauf A S, Warren Z E and Sarkar N. Robot-mediated imitation skill training for children with autism. 2016. IEEE Trans. Neural Syst. Rehabilitation Eng. 24(6), 682–691 (2016)
[21] Ismail L I, Shamshudin S, Yussof H, Hanapiah F A, Zahari N I. Estimation of concentration by eye contact measurement in robot-based intervention program with autistic children. 2012. International Symp. on Robotics and Intelligent Sensors 2012 (IRIS 2012).
[22] Boccanfuso L and O’Kane J M. An adaptive robot design with hand and face tracking for use in autism therapy. 2011. Int J of Soc Robotics. 3, 337–347.
[23] https://learnopencv.com/object-tracking-using-opencv-cpp-python/
[24] Viola P and Jones M. Rapid object detection using a boosted cascade of simple features. 2001. Proc. of the 2001 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition.
[25] Smith B A, Yin Q, Feiner S K and Nayar S K. Gaze locking: Passive eye contact detection for human-object interaction.2013. ACM Symp. on User Interface Software and Technology (UIST).271-80.
[26] Yang H, Shao L, Zheng F, Wang L, and Z. Song, Recent advances and trends in visual tracking: A review. 2011. Neurocomput., 74, 3823 (2011).
[27] Smeulders A, Chu D, Cucchiara R, Calderara S, Dehghan A, and Shah M. Visual tracking: An experimental survey. 2014. IEEE Trans. Pattern Anal. Mach. Intell., 36, 1442-68.
[28] Montanaro A, Quantum algorithms: an overview. 2016. npj Quantum Information, 2, 15023.
[29] Grover L. K. Quantum mechanics helps in searching for a needle in a haystack, 1997. Phys. Rev. Lett. 79, 325-8.