Utilizing SURF Features and KLT Tracking Algorithm in Augmented Reality (AR), Using Kinect V.2 with the Aim of Autism Therapy

Conference Paper · February 2020

Seyed Muhammad Hossein Mousavi
Independent Scientist

S.Younes Mirinezhad
Buali Sina University

Utilizing SURF Features and KLT Tracking Algorithm in Augmented Reality (AR), Using Kinect V.2 with the Aim of Autism Therapy

Seyed Muhammad Hossein Mousavi
Department of Computer Engineering
Bu Ali Sina University
Hamadan, Iran
mosavi.a.i.buali@gmail.com

S.Younes Mirinezhad
Department of Computer Engineering
Bu Ali Sina University
Hamadan, Iran
y.mirinezhad93@basu.ac.ir

Atiye Ilanloo
Faculty of Humanities-Psychology
Islamic Azad University-Rasht
Gilan, Iran
elanlooatiye@gmail.com

Narges Aghsaghloo
Faculty of Humanities-Psychology
Islamic Azad University-Saveh
Markazi, Iran
n.aghsaghloo@gmail.com

***Abstract*— Using Augmented Reality (AR) played important role in learning, education, therapy and especially autism therapy in last decade. Rehabilitating people with Autistic Spectrum Disorder (ASD) using AR was so mentionable in the therapy process. This paper presents a new method for rehabilitation and therapy for people with ASD using Microsoft Kinect V.2. System starts with human detection using depth sensor. In the second step system detects the subject's face using Viola and Jones algorithm and recognize subject's face using Convolutional Neural Network (CNN) using color sensor. Third step is consisting of feature extraction process using Bag of Features (BOF) for Content Based Image Retrieval (CBIR) purpose and tracing these points using Kanade–Lucas–Tomasi (KLT) algorithm for finalizing AR process. In the meanwhile, morphing subject's face to a new avatar takes place (randomly), and also emotion recognition using CNN helps the human therapist. Experiments is done on 4 children with ASD. Therapy estimation after a weak showed experiment's improvement during 7 days or meetings.**

Keywords*— Augmented Reality (AR), Autistic Spectrum Disorder (ASD), Microsoft Kinect V.2, Depth sensor, Content Based Image Retrieval (CBIR), Kanade–Lucas–Tomasi (KLT) tracking.

I. INTRODUCTION (*HEADING 1*)

People with Autistic Spectrum Disorder (ASD), need different ways to learn things. One of them is game therapy [1]. Game therapy systems could be made by image processing techniques and had good effect on autistic children in recent researches. Using range sensors beside of color sensors could increase the effect and efficiency of the system. Interacting with computer is so fun for children, especially autistic children. Due to that, a human-computer interaction system is made for this kind of rehabilitation. Proposed system is a kind of game therapy one, but with a little change could be turn to a learning system. People with ASD are part of the community and they effect the community. They could be rehabilitated and act like normal people if they be treated well. As this research aids this purpose, it rises the importance of the research. As people with ASD effect the community less than normal people and there is no absolute cure for them, there is no high number of researches about them in comparison with normal people, which is the limitation of the research that should be addressed. we made a novel and interactive system based on color and depth images using Augmented Reality (AR) + Kinect senor to rehabilitate children with ASD. As the system interacts with the subject and shows colorful and morphable faces to the subject, it is very efficient approach for autism rehabilitation in children. It is a rehabilitation system, not educational (but could be used in educational purposes with some changes and advances). Paper consist of VI sections, which starts with introduction on the subject's definition in section I. Section II demonstrates the materials. Section III pays to some of the important and new researches on the subject by other researchers. Section IV demonstrates all the algorithms and using them in the proposed Augmented Reality (AR) game therapy system. Results on subjects are included in section V, and conclusion, suggestions and discussion on the subject is placed in section VI.

II. MATERIALS

This section is about materials and methods which are used in the research. All definitions from basic needs to final therapy estimation is defined in this section. The section is consisting of 6 parts which are related to Augmented Reality (AR), employed sensor type, autism therapy and final recognition structure.

A. *Image Retrieval and Content Based Image Retrieval (CBIR)*

Browsing, searching and retrieving images from a large database of images is called image retrieval. Most methods of image retrieval utilize some method of adding metadata such as captioning, keywords, or descriptions to the images so that retrieval can be performed over the notation words. These types of image retrievals are called concept- based image retrieval, which is so traditional and time consuming. In new approach which is opposed of traditional one, system analyzes the contents of the image rather than the metadata such as keywords, tags, or descriptions associated with the image. This new type is called content-based image retrieval, which uses colors, shapes, textures, or any other information that can be derived from the image itself [2]. Figure 1 shows the general structure of image retrieval process.

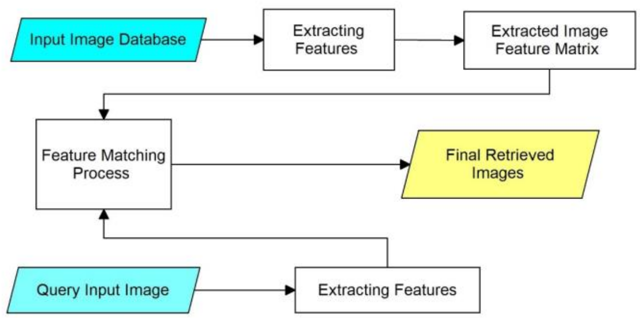


Fig. 1. General structure of image retrieval process

B. *Augmented Reality (AR)*

Augmented reality (AR) is a live direct or indirect view of a physical, real-world environment whose elements are "augmented" by computer-generated perceptual information, ideally across multiple sensory modalities, including visual, auditory, haptic, somatosensory, and olfactory [3]. The overlaid sensory information can be constructive (i.e. additive to the natural environment) or destructive (i.e. masking of the natural environment) and is spatial registered with the physical world such that it is perceived as an immersive aspect of the real environment [4]. In this way, Augmented reality alters one's current perception of a real-world environment, whereas virtual reality replaces the real-world environment with a simulated one [5]. The primary value of augmented reality is that it brings components of the digital world into a person's perception of the real world, and does so not as a simple display of data, but through the integration of immersive sensations that are perceived as natural parts of an environment. Hardware components for augmented reality are: processor, display, sensors and input devices. Modern mobile computing devices like smartphones and tablet computers contain these elements which often include a camera and MEMS sensors such as accelerometer, GPS, and solid-state compass, making them suitable AR platforms [6]. A key measure of AR systems is how realistically they integrate augmentations with the real world. The software must derive real world coordinates, independent from the camera, from camera images. That process is called image registration, and uses different methods of computer vision, mostly related to video tracking [7]. Many computer vision methods of augmented reality are inherited from visual odometry. Usually those methods consist of two parts. The first stage is to detect interest points, constant markers or optical flow in the camera images. This step can use feature detection methods like corner detection, blob detection, edge detection or thresholding, and other image processing methods [8]. The second stage restores a real-world coordinate system from the data obtained in the first stage. AR has application in Literature, Archaeology [9], Architecture [10], Visual art [11], Commerce [12], Education [13], Emergency management/search and rescue [14], Video games [15], Industrial design, Medical, Military, Navigation, Broadcast and live events, Tourism and sightseeing, Music and more.

C. *Microsoft Kinect V.2 and Depth Data*

Kinect is one of the most useful depth sensors to have. It is very cheaper than other depth sensors and efficient. It can be used on Microsoft Xbox 360 (Kinect V.1) or Xbox one (Kinect V.2) consoles or be used as a developer device. Kinect 2.0 was released with Xbox One on November 22, 2013. Because of the lower price and high power to use, a lot of developers and researcher use it as a main depth device. It could record RGB and Depth video frames with 1920*1080 resolution for RGB images and 512*424 for Depth images on 30 fps. Also, it is capable to work between 0.8-5.0-meter ranges. An RGB-D image is simply a combination of an RGB image and its corresponding depth image. A depth image is an image channel in which each pixel relates to a distance between the image plane and the corresponding object in the RGB image. It is also termed as 2.5D or Range image [16]. Figure 2 (a) shows Kinect V.1 VS V.2 specifications.

D. *Autism and Autism Therapy*

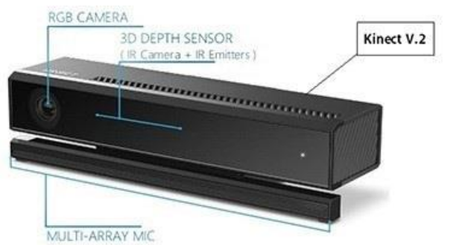
Autism is a developmental disorder characterized by troubles with social interaction and communication [17]. Often there is also restricted and repetitive behavior [17]. Parents usually notice signs in the first two or three years of their child's life [17], [18]. These signs often develop step by step, though some children with autism reach their developmental milestones at a normal pace and then worsen [19]. Autism is caused by a combination of genetic and environmental factors [20]. Autism therapies are interventions that attempt to lessen the deficits and problem behaviors associated with autism spectrum disorder (ASD) in order to increase the quality of life and functional independence of autistic individuals. Treatment is typically catered to person's needs. Treatments fall into two major categories: educational interventions and medical management. Training and support are also given to families of those with ASD [21]. Studies of interventions have some methodological problems that prevent definitive conclusions about efficacy [22].

E. *Human, Face Detection and Face Recognition*

The process of distinguishing an object from its background is called object detection. Now if this object be a human or face, then it is called human or face detection. It is mentionable that Human detection process will be done using depth sensor in the paper (more details in section IV). Now imagine face detection problem with five individual and it is needed to select a specific individual with specific face shape and attributes. Distinguishing and selecting that specific and desire individual from the other individuals is called face recognition [23]. Object or face recognition usually take advantage of supervised learning techniques, which we are going to use one of those techniques called Convolutional Neural Network (CNN), which is based on learning the labeled extracted features [24].Due to the small data (4 subjects), CNN works fast here. Figure 2 (b) shows shallow vs deep NN structure.

F. Gesture Recognition

Gesture recognition is a human-computer interaction process, which aids image processing techniques to perform. Image data could be color or range types, which Kinect sensor supports both of them with different resolutions as it mentioned above. Gesture is a mention or hint to the computer which performs by human and computer response to it. These gestures perform by body parts (usually hands), which they need visual sensors to receive them. So, gesture is not used for acoustic processes. For example, with moving up the right hand, computer says something or vice versa. Mentioned gesture could be done by color images, but in-depth image, it is possible to use another dimension which is range or distance between object and sensor. For example, moving back the left hand, which just depth sensor could understand it [25].



	Kinect for Windows v1	Kinect for Windows v2
Color	640 × 480 @ 30fps	1920 × 1080 @ 30fps
Depth	320 × 240 @ 30fps	512 × 424 @ 30fps
Sensor	Structured Light (PrimeSense Light Coding)	Time of Flight (ToF)
Range	0.8~4.0 m	0.5~4.5 m
Angle of View Horizontal / Vertical	57 / 43 degree	70 / 60 degree
Microphone Array	◉	◉

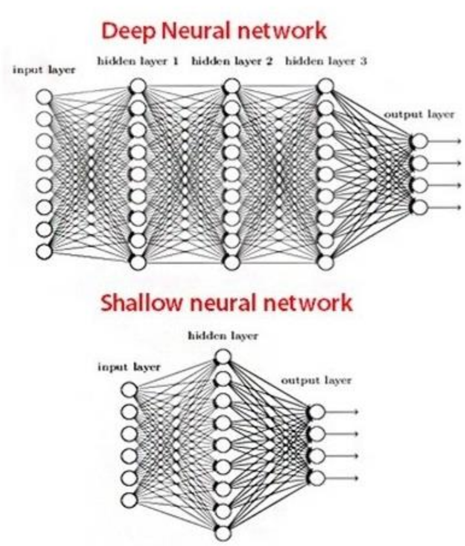


Fig. 2. (a). Kinect V1 VS V2 Spec and (b). Shallow neural network VS Deep neural network structure

III. Prior Related Researches

In 2007, Richard, Emmanuelle, et al succeeded to make a system for pairing and provides visual, olfactory or auditory cues to help children in decision making. 93 children from a French elementary school (including 11 cognitive disabled ones) participated in a preliminary study. Autistic children were able to express some positive emotions when confronted to the application [26]. Bai, Zhen et al, made an AR system to examine the potential of using AR technologies to promote pretend play behaviors in children with autism in 2012 [27]. Also, in 2013 they enhanced the system for children between the ages of 4 and 7 who have been diagnosed with ASC [28]. In 2015 Chen, Chien-Hsu, I-Jui Lee et al, researched on assessing the possibility of enabling three adolescents with ASD to become aware of facial expressions observed in situations in a school setting simulated using augmented reality (AR) technology in 2015 [29]. Okay we demonstrate some of the related works on AR for Autism therapy using color sensors, and now some of the related works using color and depth sensors, especially Kinect will be discussed. In 2012 Casas, Xavier, et al, made an augmented reality system for teaching key developmental abilities for individuals with ASD. They used Kinect V.1 and mirror, as the interaction tools for teaching skills to the subjects [30]. In 2013, Bartoli, Laura, et al, described a field study that extends the current body of empirical evidence of the potential benefits of touch less motion-based gaming for autistic children using Kinect V.1. They experiment on 5 autistic children over 20 hours in 5 meeting with assistant of a therapist [31]. Also, in 2015, Fu, Ying, et al, made a game for rehabilitation, based on Kinect V.1 which applied to assist the rehabilitation of Mental Retardation (MR) children [32]. Another mentionable research is Roglić, Miloš, et al work on a Kinect-based game for autistic children in Serbia which, is a fine-tuning of motor skills, and performance improvement of basic cognitive tasks through practicing five game categories. At the end therapists assess subject's results [33]. Research has been done using Kinect V.2 in 2016. Finally, in 2017, Persefoni Karamanoli and et al, present a review examines studies that have used AR for direct intervention to people with ASD [34]. Also [35] and [36] are mentionable researches in this area.

IV. Proposed AR Therapy System

This section going to demonstrate all the algorithms and methods, which will be used in the proposed method. Due to lack of enough space and much algorithms, demonstration on each algorithm is summarized and referred to the main reference paper, but the main process is intact and with full details. Morphing faces processed and recorded in Abrosoft FantaMorph software; Flowcharts drawn in Diagram Designer software; experiment environment created using Edrw Max software; coding and GUI Implementation is done using Matlab R2018b software. For using Kinect V.2, USB 3.0 controller dedicated to the Kinect for Windows v2 sensor*, 64-bit (x64) processor, 4 GB Memory (or more), Physical dual-core 3.1 GHz (2 logical cores per physical) or faster processor, DX11 capable graphics adapter is require, which we used a system with 16 GB of ram, Quad-Core 4.0 GHz processor, Nvidia GTX 1050 (2 GB RAM), USB 3 port supported motherboard and 256 GB SSD Hard drive for faster computation for whole process. It is mentionable that Kinect V.2 needs a lot of power source, so power supply is essential for having 30 fps video for both color and depth data.

A. Human Detection Using Depth Data

In this section, a simple segmentation process accrues after background removal pre-processing which extract human body from the other staff in the scene. Due to prepared recording condition and no more objects except the subject, this process distinguishes closer pixels (human or subject) from the farthest.

B. Viola & Jones Face Detection Algorithm

Viola & jones algorithm is one of the best face detection algorithms which is using for a long period of time in face image processing. This algorithm is so fast and robust and it could be used for depth images too with lower accuracy. This algorithm is an object detection algorithm and could be used for any learned object but mainly uses for face. The algorithm has four stages:

- Haar Feature Selection
- Creating an Integral Image
- Adaboost Training
- Cascading Classifiers

After detecting face by this algorithm, simply it can be extracted by cropping technique. For more information and details about this algorithm could refer to [37]. This algorithm constructs a “strong” classifier as a linear combination of weighted simple “weak” classifiers.

$$h\left(X\right)=sgn\left(\sum_{j=1}^{M}a_j h_j\left(X\right)\right) \tag{1}$$

Which h_j is weak classifier for each a subject. Each weak classifier is a threshold function based on the feature f_j.

$$h_j\left(X\right)=\begin{cases}-s_j \; if \; f_j<\theta_j \\ s_j \; otherwise\end{cases} \tag{2}$$

The threshold value θ_j and the polarity $s_j \in \pm$ are determined in the training, as well as the coefficients α_j.

C. Face Recognition Using Convolutional Neural Network (CNN)

For more information about NN, Shallow NN and Deep NN (deep learning) please refer to [38]. In this part one of the deep learning’s algorithms called Convolutional Neural Network (CNN) is used to recognize subject’s faces. Due to lack of number of subjects (4 subjects), system performs fast enough even using deep learning. A convolutional neural network, is an architecture commonly used for deep learning. CNNs are often used to recognize objects and scenes, and perform object detection and segmentation. They learn directly from image data, eliminating the need for manual feature extraction. A convolutional neural network can have tens or hundreds of layers that each learn to detect different features of an image. Filters are applied to each training image at different resolutions, and the output of each convolved image is used as the input to the next layer. The filters can start as very simple features, such as brightness and edges, and increase in complexity to features that uniquely define the object as the layers progress. Figure 3 shows an example of using CNN algorithm in image processing, which it is an example of a network with many convolutional layers. Filters are applied to each training image at different resolutions, and the output of each convolved image is used as the input to the next layer. Also, for emotion recognition, the same method is employed.

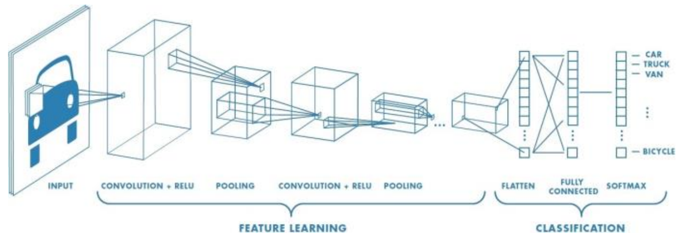


Fig. 3. An example of using CNN algorithm in image processing [39]

D. Speeded Up Robust Features (SURF) Feature

SURF is a feature detection and descriptor algorithm. It is the enhanced version of the Scale-Invariant Feature Transform SIFT feature [40] and even 3-5 times faster than Sift, because of the integral image calculations. In this method, feature points not change during rotation of the image. In the other hand final matrix is 64*64 but SIFT is 128*128 dimensions, but there is no accuracy loss in the SURF. SURF feature steps are as follow:

1. Calculating integral image for faster computations
2. Calculating feature points using Hessian matrix
3. Making scale space
4. Determining Extremum points
5. Constructing feature vector

SURF changes images during iterations with different Gaussian low pass filter in different octaves (consist of n image with same size but different blurring using sigma factor) and subtracting each octave from next lowered image version of itself. Here is SURF paper for more information [41]. SURF also uses the determinant of the Hessian matrix for selecting the scale. Given a point p=(x, y) in an image I, the Hessian matrix H (p, σ) at point p and scale σ, is as follow:

$$H\left(p,\sigma\right)=\begin{pmatrix} L_{xx}\left(p,\sigma\right) & L_{xy}\left(p,\sigma\right) \\ L_{xy}\left(p,\sigma\right) & L_{xy}\left(p,\sigma\right)\end{pmatrix} \tag{3}$$

Where L_{xx} (p,σ) etc. is the convolution of the second-order derivative of Gaussian with the image I(x,y) at point x. More details on mentioned five steps is based on [41].

E. Bag of Features (BOF) or Bag of visual Words (BOW)

A common technique used to implement a Content-Based Image Retrieval (CBIR) system is Bag of visual Words (BOW), also known as Bag of Features (BOF) [42]. This technique is also known as Vector Space Model (VSM) [43].

Instead of using actual words as in document retrieval, bag of features uses image features as the visual words that describe an image. Image features are an important part of CBIR systems. These image features are used to gauge similarity between images and can include global image features such as color, texture, and shape. Image features can also be local image features such as Speeded-Up Robust Features (SURF), Histogram of Gradients (HOG), or Local Binary Patterns (LBP). In this paper, we use SURF feature. Figure 4 shows a simple example of BOW procedure.

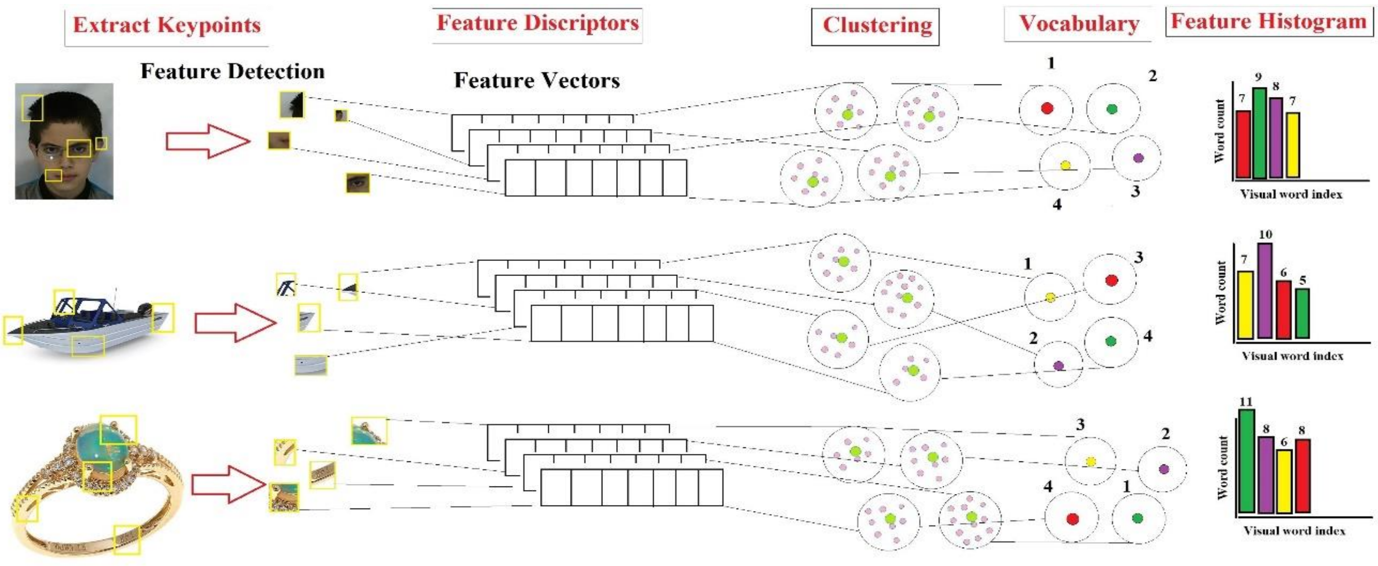


Fig. 4. A simple example of BOW procedure

F. *Kanade-Lucas-Tomasi (KLT) Tracking Algorithm*

Kanade-Lucas-Tomasi (KLT) is an implementation, in the C programming language, of a feature tracker for the computer vision community. The source code is in the public domain, available for both commercial and non-commercial use which is done by Tomasi, Carlo, and Takeo Kanade in 1991 [44]. Jianbo Shi and Carlo Tomasi enhanced algorithm for better feature tracking purposes in 1994 [45]. In short, good features are located by examining the minimum eigenvalue of each 2 by 2 gradient matrix, and features are tracked using a Newton-Raphson method of minimizing the difference between the two windows. Multi-resolution tracking allows for relatively large displacements between images. KLT tracking algorithm in the simple form is as follow:

1. Detect Harris corners in the first frame.
2. For each Harris corner compute motion (translation or affine) between consecutive frames.
3. Link motion vectors in successive frames to get a track for each Harris point.
4. Introduce new Harris points by applying Harris detector at every m (10 or 15) frames.
5. Track new and old Harris points using steps 1-3.

Figure 5 represents the Graphical User Interface (GUI) of proposed system. As it is clear in this figure, subject's emotion or expression is calculated using CNN algorithm and presented as gauge, which helps human therapist for better therapy estimation. Figure 6 represents whole the process of proposed AR therapy system in details, but summarized form of the system will demonstrate. System starts with color and depth data acquisition using Kinect V.2. Then human detection using simple segmentation takes place with depth sensor. After positioning subject in 2.5 meter from the sensor, face detection using Viola and Jones algorithm accrues with color sensor. In the next step and after background removal pre-processing, face recognition phase takes place to identify the subject (using color sensor and CNN algorithm). Next step is consisting of recognizing proper gesture which is moving hand backward, and if gesture performed correctly, then Proposed AR process starts. Full process of proposed AR therapy system is demonstrated in Figure 7, but in a general view, feature extraction happens using SURF feature and Tracking is done using KLT algorithm in each frame. Next, subject's face avatar changes to a new morphed face and therapy estimation will achieve and record with the aid of therapist. Proposed AR structure flowchart's is presented in Figure 7.

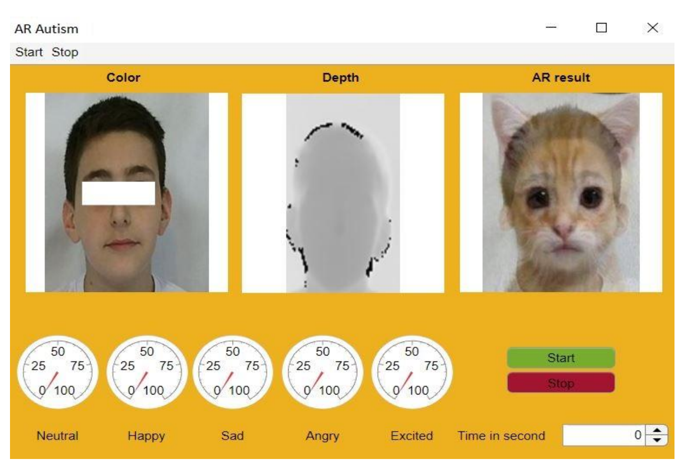


Fig. 5. Proposed system GUI

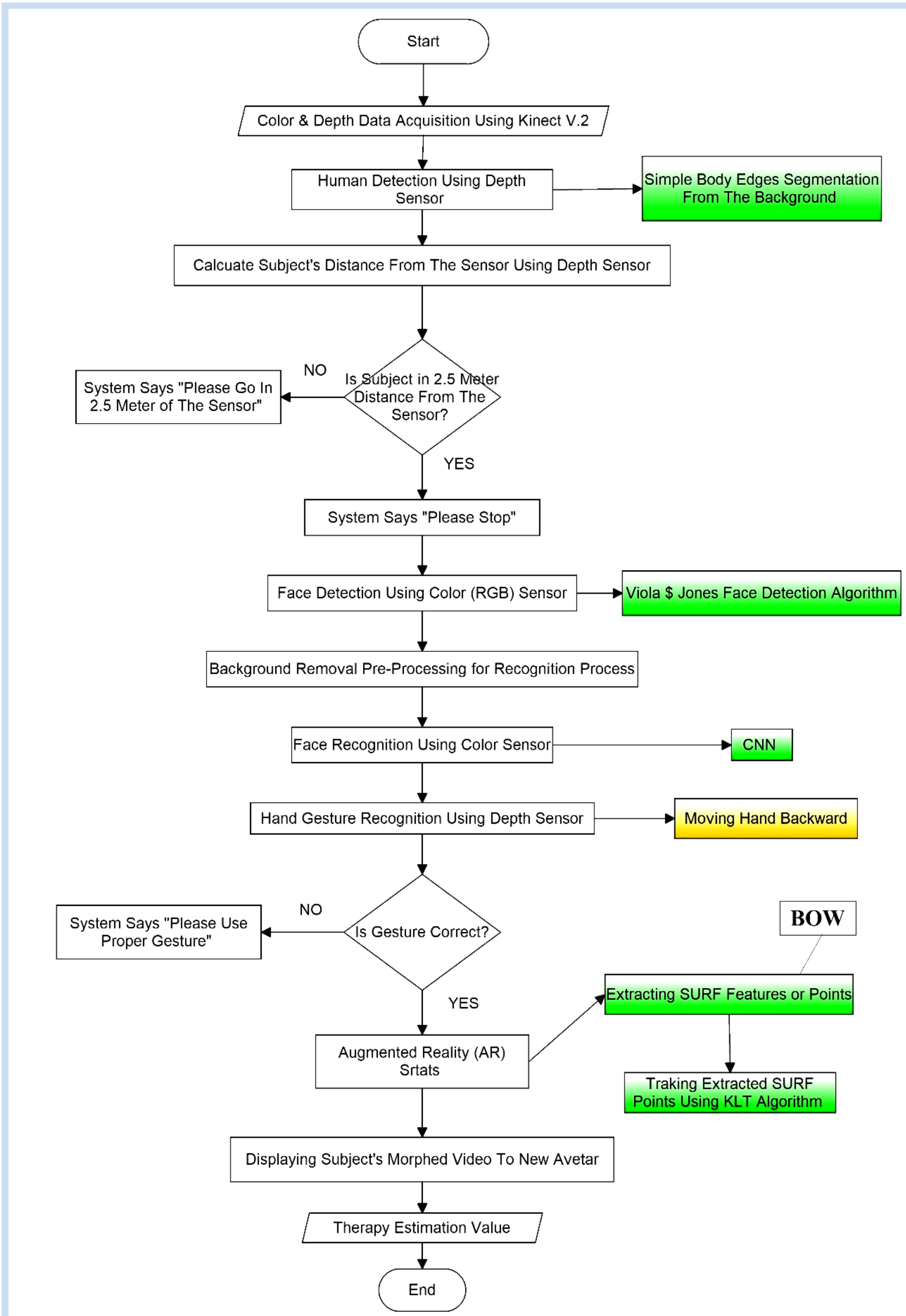


Fig. 6. Full process of proposed AR therapy system in details

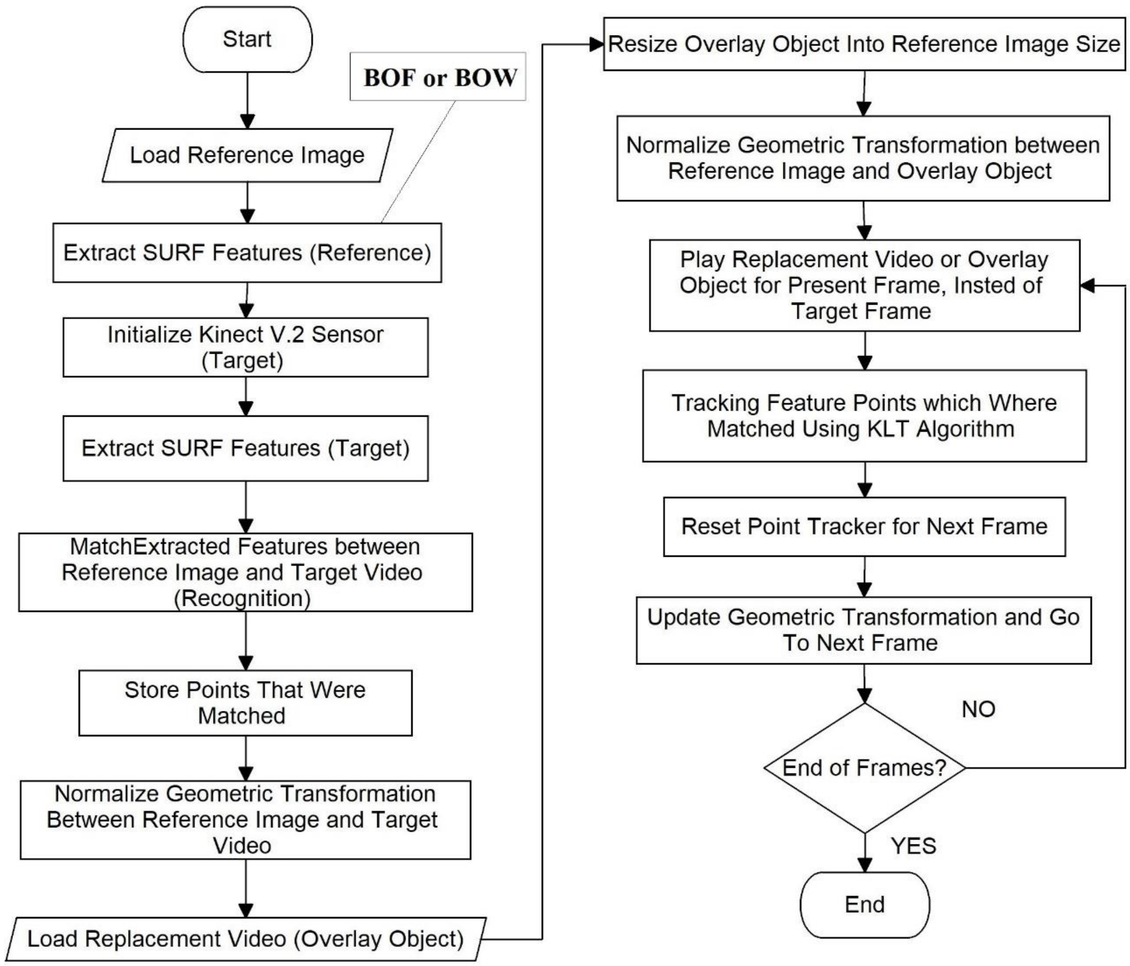


Fig. 7. Proposed AR structure's flowchart

V. RESULTS

Experiment had been done using Kinect V.2 as main sensor. In this experiment, 4 autistic children (subjects) and 1 therapist was involved. Experiment took place in 7 days or meetings. Figure 8 and 9 represent experiments environment schema in graphical and real form. Subjects average association in each meeting is presented in Figure 10. Also, Subjects average distress rate in each meeting is presented in Figure 11. As it is clear in Figures 10 and 11, as meeting days increases, ASD subjects tend to associate more in next meeting. In the other hand, they showed less distress as meeting days increases. Figure 12 shows two example of the morphing action on two subjects in the experiment. Due to the identity protection, their eyes are protected in first and third frames. Different types of pet or domestic animal avatars are used in the morphing process. In each meeting animal avatar changes for each subject. All subjects had good reaction to three specific pets which are kitten, owl and bunny. Table I presents profile of each subject. Table II represents positive, negative and average estimation result gained by subjects and therapist respectively. As it is clear in Table II subject 3 is so friendly and in the other hand subject 2 is not. At the end and in the seventh day, all the 4 subjects were friendly and results was better. In the experiment subjects, especially subject3 was so excited about touch less gesture and more likely liked the morphing faces. Changing face during time (frames) and even in moving head situation was funny for subject 3, annoying for subject 2 and sometimes autistic actions showed from subjects 1 and 4 during experiment (maybe due to confusion). It is clear that we select funny and sometimes not so funny avatars to morph for therapy, because using annoying avatars lead to sickness. We used different resolutions and frame rates in the experiment but mostly with 800*600 resolution and 15 fps due to the faster process. Despite of using deep learning technique, system was so optimized and fast. As rehabilitation systems are based on estimation and each rehabilitation system uses different approach, so comparison with other researches is meaningless. And that's due to lack of precise percentage. Research on this subject, in proposed form is done for the first time in this research. Also, figure 13 represents final estimation of therapy (system vs human expert).

In order to demonstrate subject's levels, three famous metrics are presented.

The Social Responsiveness Scale (SRS) [46] measures social ability of children from 4 years to 18 years old. It is used primarily with individuals with Autism Spectrum Disorder (ASD), family members of individuals with ASD, and others who have social impairments. The Social Communication Questionnaire (SCQ) [47] is one tool clinicians use when screening an individual for Autism Spectrum Disorder (ASD). It is a measure for caregivers to complete. Caregivers need to be familiar with the individual's developmental history and current behavior. The Autism Diagnostic Observation Schedule-Generic (ADOS-G) [48] is a semi structured, standardized assessment of social interaction, communication, play, and imaginative use of materials for individuals suspected of having autism spectrum disorders.

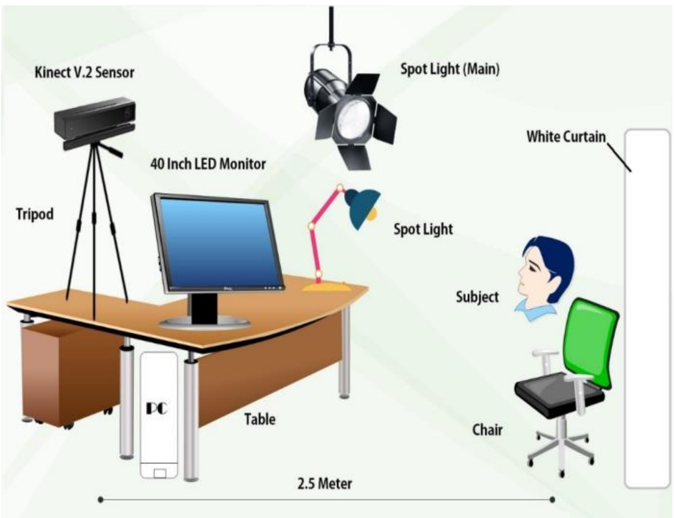


Fig. 8. Experiment environment (graphical)

Fig. 9. Experiment environment (real)

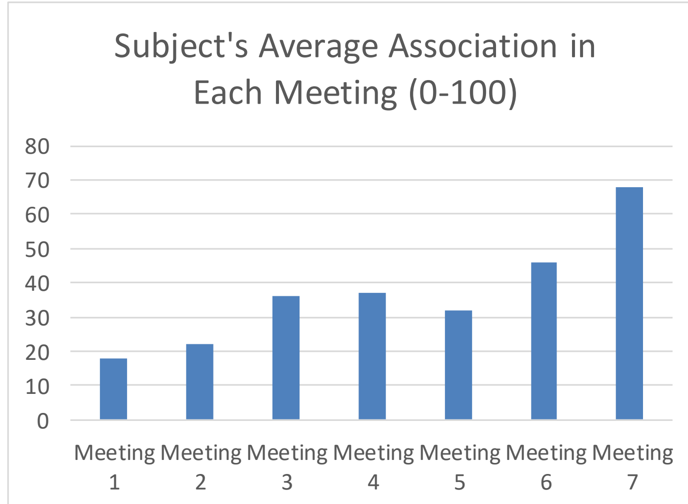


Fig. 10. Subjects average association in each meeting

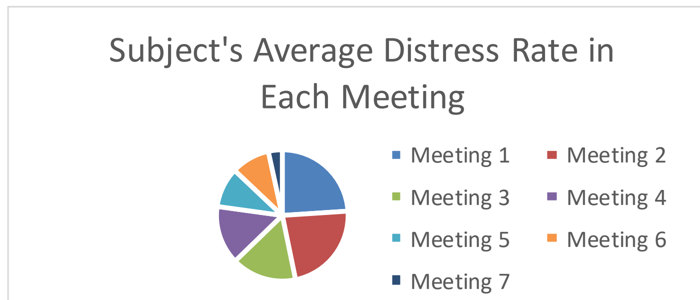


Fig. 11. Subjects average distress rate in each meeting

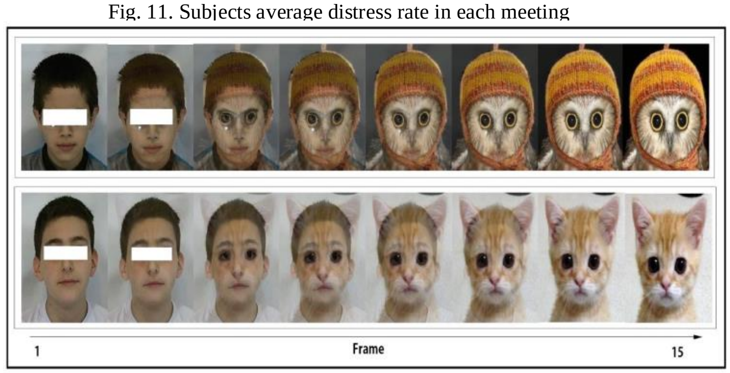


Fig. 12. Two examples of the morphing action on two subjects in the experiment

TABLE I. PROFILE OF THE PARTICIPANTS WITH ASD

Participant (Gender)	**Age**	**SRS (cutoff=60)**	**SCQ (Cutoff=15)**	**ADOS-G (cutoff=7)**
ASD1 (m)	7.36	52	6	11
ASD1 (m)	5.12	57	7	12
ASD1 (f)	6.39	65	11	14
ASD1 (m)	8.47	80	17	18

TABLE II. POSITIVE, NEGATIVE AND AVERAGE ESTIMATION RESULT GAINED BY SUBJECTS AND THERAPIST

Meeting	Positive Description	Negative Description	Average Therapy Estimation
1	One of the subjects behaved nice	Lack of trust in first meeting	5%
2	Subject 3 liked all her morphs	Subject 1 did not participate	4%
3	Subjects 1 and 3 liked their morphs	Subject 2 and 4 had neutral expression	7%
4	Subject 3 request for more AR	Subject 1 was sick	6%
5	All the subjects participate	-	11%
6	All the subjects participate	Subject 2 hated one of his morphs	7%
7	All the subjects participate	-	12%

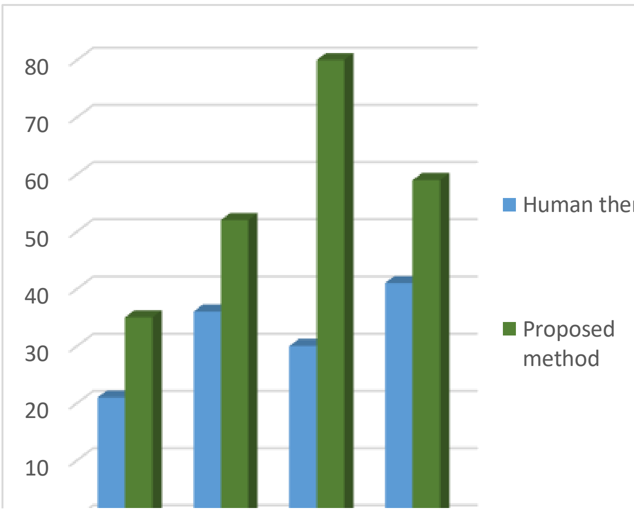


Fig. 13. Percentage of therapy (system vs human expert)

VI. CONCLUSION, DISCUSSION AND SUGGESTIONS

Combining depth images to color for AR applications, especially AR therapy for children with ASD will increase the efficiency and accuracy of the therapy. As it is clear in section V, experiment improved day after day using such a depth, color AR system. We used Kinect V.2 for this purpose and with aid of some of the fastest algorithms in feature extraction and tracking, could achieve proper results. With changing recognition process with algorithms such as K-Nearest Neighborhood (KNN) or Support Vector Machine (SVM) runtime speed increases significantly, which it is one of the suggestions. Also, for even faster computation for real-time processes, it is possible to use Ellipse fitting technique for face detection part. Our GUI was good but simple. So, using colorful GUI, and even toys around the experiment environment could optimize the experiment, which in this field our experiment was a little bit poor bot not so bad. Due to prepared experiment environment, using algorithms such as SIFT [40] for feature extraction or Kalman tracking algorithm could be useful. As results was promising and satisfactory, using such an AR system with a little change could be used in other medical subjects for learning process or even education and entertainment in other similar kind of disease such as Intellectual disability, Schizophrenia, Obsessive-compulsive disorder (OCD), Avoidant personality disorder and more.

REFERENCES

[1] Wilkinson, Nathan, Rebecca P. Ang, and Dion H. Goh. "Online video game therapy for mental health concerns: a review." International journal of social psychiatry 54.4 (2008): 370-382.

[2] Datta, Ritendra, et al. "Image retrieval: Ideas, influences, and trends of the new age." ACM Computing Surveys (Csur) 40.2 (2008): 5.

[3] Schuettel, Patrick (2017). The Concise Fintech Compendium. Fribourg: School of Management Fribourg/Switzerland.

[4] Rosenberg, L.B. (1992). "The Use of Virtual Fixtures As Perceptual Overlays to Enhance Operator Performance in Remote Environments". Technical Report AL-TR-0089, USAF Armstrong Laboratory, Wright-Patterson AFB OH, 1992.

[5] Steuer, Jonathan. Defining Virtual Reality: Dimensions Determining Telepresence Archived 24 May 2016 at the Wayback Machine., Department of Communication, Stanford University. 15 October 1993.

[6] Metz, Rachel. Augmented Reality Is Finally Getting Real Technology Review, 2 August 2012.

[7] Azuma, Ronald; Balliot, Yohan; Behringer, Reinhold; Feiner, Steven; Julier, Simon; MacIntyre, Blair. Recent Advances in Augmented Reality Computers & Graphics, November 2001.

[8] State, Andrei; Hirota, Gentaro; Chen,David T; Garrett, William; Livingston, Mark. Superior Augmented Reality Registration by Integrating Landmark Tracking and Magnetic Tracking, Department of Computer ScienceUniversity of North Carolina at Chapel Hill.

[9] Dähne, Patrick; Karigiannis, John N. "Archeoguide: System Architecture of a Mobile Outdoor Augmented Reality System". Retrieved 2010-01-06.

[10] Lee, Gun A., et al. "CityViewAR: A mobile outdoor AR application for city visualization." Mixed and Augmented Reality (ISMAR-AMH), 2012 IEEE International Symposium on. IEEE, 2012.

[11] Slenske, Michael. "Amir Baradaran Gives Tourists a Reason to Photograph Famous Art". Art In America Magazine. Retrieved 2017-02-22.

[12] Katts, Rima. Elizabeth Arden brings new fragrance to life with augmented reality Mobile Marketer, 19 September 2012.

[13] Santos, M. E. C.; Chen, A.; Taketomi, T.; Yamamoto, G.; Miyazaki, J.; Kato, H. (January 2014). "Augmented Reality Learning Experiences: Survey of Prototype Design and Evaluation". IEEE Transactions on Learning Technologies. 7 (1): 38–56. doi:10.1109/tlt.2013.37. ISSN 1939-1382.

[14] "Augmented Reality--Emerging Technology for Emergency Management", Emergency Management Magazine, September 24, 2009.

[15] Hawkins, Mathew. Augmented Reality Used To Enhance Both Pool And Air Hockey Game Set WatchOctober 15, 2011.

[16] Ganguly, Suranjan, Debotosh Bhattacharjee, and Mita Nasipuri. "3D Image Acquisition and Analysis of Range Face Images for Registration and Recognition." Handbook of Research on Emerging Perspectives in Intelligent Pattern Recognition, Analysis, and Image Processing (2015): 185.

[17] Landa RJ (2008). "Diagnosis of autism spectrum disorders in the first 3 years of life". Nat Clin Pract Neurol. 4 (3): 138–47. doi:10.1038/ncpneuro0731 Freely accessible. PMID 18253102

[18] Autism Spectrum Disorder, 299.00 (F84.0). In: American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition. American Psychiatric Publishing; 2013.

[19] Stefanatos GA (2008). "Regression in autistic spectrum disorders". Neuropsychol Rev. 18 (4): 305–19. doi:10.1007/s11065-008-9073-y.

[20] Chaste P, Leboyer M (2012). "Autism risk factors: genes, environment, and gene-environment interactions". Dialogues in Clinical Neuroscience. 14: 281–92. PMC 3513682.

[21] Myers SM, Johnson CP, Council on Children with Disabilities. Management of children with autism spectrum disorders. Pediatrics. 2007;120(5):1162–82. doi:10.1542/peds.2007-2362.

[22] Ospina MB, Krebs Seida J, Clark B et al. Behavioural and developmental interventions for autism spectrum disorder: a clinical systematic review. PLoS ONE. 2008;3(11):e3755. doi:10.1371/journal.pone.0003755. PMID 19015734. PMC 2582449.

[23] Duda, Richard O., Peter E. Hart, and David G. Stork. Pattern classification. Vol. 2. New York: Wiley, 1973.

[24] Goodfellow, Ian, et al. Deep learning. Vol. 1. Cambridge: MIT press, 2016.

[25] Richard, Emmanuelle, et al. "Augmented reality for rehabilitation of cognitive disabled children: A preliminary study." Virtual Rehabilitation, 2007. IEEE, 2007.

[26] Richard, Emmanuelle, et al. "Augmented reality for rehabilitation of cognitive disabled children: A preliminary study." Virtual Rehabilitation, 2007. IEEE, 2007.

[27] Bai, Zhen, Alan F. Blackwell, and George Coulouris. "Making Pretense Visible and Graspable: An augmented reality approach to promote pretend play." Mixed and Augmented Reality (ISMAR), 2012 IEEE International Symposium on. IEEE, 2012.

[28] Bai, Zhen, Alan F. Blackwell, and George Coulouris. "Through the looking glass: Pretend play for children with autism." Mixed and augmented reality (ISMAR), 2013 IEEE International Symposium on. IEEE, 2013.

[29] Chen, Chien-Hsu, I-Jui Lee, and Ling-Yi Lin. "Augmented reality-based self-facial modeling to promote the emotional expression and social skills of adolescents with autism spectrum disorders." Research in developmental disabilities 36 (2015): 396-403.

[30] Casas, Xavier, et al. "A Kinect-based Augmented Reality System for Individuals with Autism Spectrum Disorders." GRAPP/IVAPP. 2012.

[31] Bartoli, Laura, et al. "Exploring motion-based touchless games for autistic children's learning." Proceedings of the 12th International Conference on Interaction Design and Children. ACM, 2013.

[32] Fu, Ying, et al. "Game System for Rehabilitation Based on Kinect is Effective for Mental Retardation." MATEC Web of Conferences. Vol. 22. EDP Sciences, 2015.

[33] Roglić, Miloš, et al. "Serious gaming based on Kinect technology for autistic children in Serbia." Neural Networks and Applications (NEUREL), 2016 13th Symposium on. IEEE, 2016.

[34] Karagiannidis, Charalampos. "The Application of Augmented Reality for Intervention to People with Autism Spectrum Disorders.", IOSR Journal of Mobile Computing & Application (IOSR-JMCA), 2017.

[35] Seyed Muhammad Hossein Mousavi. “Using Genetic Programming for Making a New Evolutionary Artwork, Based on Human-Computer Interactions for Autism Rehabilitation”, The third international conference on Intelligent Decision Science (IDS 2018), Tehran-Iran, 2018.

[36] Mousavi, Seyed Muhammad Hossein, 2018, Using Genetic Programming for Making a New Evolutionary Artwork, Based on Human-Computer Interactions for Autism Rehabilitation, The Third International Conference on Intelligent Decision Science, تهران, انجمن سيستم هاي فازي ايران - دانشگاه آزاداسلامي واحد علوم تحقيقات, https://www.civilica.com/Paper-IDS03-IDS03_107.html

[37] Viola, Paul, and Michael Jones. "Rapid object detection using a boosted cascade of simple features." Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on. Vol. 1. IEEE, 2001.

[38] Nielsen, Michael A. Neural networks and deep learning. Vol. 25. San Francisco, CA, USA:: Determination press, 2015.

[39] https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

[40] Lowe, David G. (1999). "Object recognition from local scale-invariant features" (PDF). Proceedings of the International Conference on Computer Vision. 2. pp. 1150–1157.

[41] Bay, Herbert, Tinne Tuytelaars, and Luc Van Gool. "Surf: Speeded up robust features." Computer vision–ECCV 2006 (2006): 404-417.

[42] Philbin, James, et al. "Object retrieval with large vocabularies and fast spatial matching." Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on. IEEE, 2007.

[43] Salton, Gerard, Anita Wong, and Chung-Shu Yang. "A vector space model for automatic indexing." Communications of the ACM 18.11 (1975): 613-620.

[44] Tomasi, Carlo, and Takeo Kanade. "Detection and tracking of point features." (1991).

[45] Shi, Jianbo. "Good features to track." Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on. IEEE, 1994.

[46] Constantino, John N., et al. "Validation of a brief quantitative measure of autistic traits: comparison of the social responsiveness scale with the autism diagnostic interview-revised." Journal of autism and developmental disorders 33.4 (2003): 427-433.

[47] Chandler, Susie, et al. "Validation of the social communication questionnaire in a population cohort of children with autism spectrum disorders." Journal of the American Academy of Child & Adolescent Psychiatry 46.10 (2007): 1324-1332.

[48] Lord, Catherine, et al. "The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism." Journal of autism and developmental disorders 30.3 (2000): 205-223.