Available online at: https://ijact.in

Date of Submission	13/10/2019
Date of Acceptance	16/01/2020
Date of Publication	31/01/2020
Page numbers	3566-3577 (12 Pages)

Cite This Paper: Ahmed AA Jubouri, Isra H Ali. A survey on movement analysis (hand, eye, body) and facial expressions-based diagnosis autism disorders using microsoft kinect v2, 9(1), COMPUSOFT, An International Journal of Advanced Computer Technology. PP. 3566-3577.

ISSN:2320-0790

COMPUSOFT
An International Journal of Advanced Computer Technology

A SURVEY ON MOVEMENT ANALYSIS (HAND, EYE, BODY) AND FACIAL EXPRESSIONS-BASED DIAGNOSIS AUTISM DISORDERS USING MICROSOFT KINECT V2

Ahmed A. Al-Jubouri[1], Israa Hadi Ali[2]

[1]College of IT, University of Babylon, Hillah, Iraq,
stud.ahmed.rahman@uobabylon.edu.iq
[2]College of IT, University of Babylon, Hillah, Iraq,
israa_hadi@itnet.uobabylon.edu.iq

Abstract: Kinect v2 may enhance the clinical practice of autism spectrum disorders (ASD). ASD means disorders of neurodevelopment that lasts a lifetime, which occurs in early childhood and usually associated with unusual movement and gait disturbances. The earlier diagnosis of ASD helps of providing well known of these disorders. The methods which are adopted by experts in diagnosis are expensive, time-consuming, and difficult to replicate, as it is based on manual observation and standard questionnaires to look for certain signs of behavior. This paper, to the best of our knowledge, is a first attempt to collect the previous researches of the Kinect v2 in the disorder's diagnosis. Relevant papers are divided into four groups which are: (1) papers suggest a system based on the analysis of facial expressions, (2) papers suggest a system based on the analysis of hand movement, (3) papers suggest a system based on analysis of eye movement, and (4) papers suggest a system based on analysis of body movement.

Keywords: Autism Spectrum Disorders, Autism, Kinect v2, facial expressions, hand movement analysis, eye movement analysis, body movement analysis

I. INTRODUCTION

Although autism has a considerable genetic component, it is currently diagnosed through behavior. Autism spectrum disorders (ASD) defined as lifelong neuro developmental disorders which could be identified in the early years of childhood, seen in children of ASD with difficulties in their social communication, repetitive patterns of behavior, and different degrees of impairment [1]. More important, early intervention reduces the risk of long-term disability for children, as they are more likely to become troubled adults. Referring to clinicians and researchers worldwide, there is an increase in prevalence rates of ASD cases in many parts of the world [2], meanwhile US occupied the first place with the number of autistic 168 per 10,000 children in 2018 [3]. In the analysis of autism's sex ratio, fifty-four research show that 4.2 boys are with autism for every girl [4], whereas Bloom J. (2017) mentioned that by 2025, 50% of the children will be infected around the world [5]. Despite more than half a century of research in difference research fields, none of the researchers have identified clear and confirmed causes that can be attributed to autism disease [6].

Children's behavior has been analyzed for the purpose of diagnosing autism spectrum disorders. The identification of behavior abnormalities could be beneficial for early detection and better treatment planning for children with

ASD. Current behavior assessment methods are often time-consuming and highly dependent on the clinician's judgment, which leads to subjective interpretations, while the advances in behavior analysis, instrumentation, and human-computer interaction system provides new insights in understanding all aspects of movement patterns and support an automated diagnosis of pathological disorders.
Human-computer interaction system is used to recognize innate and instinctive human expressions and return an expected and inspiring feedback. Popular culture has long viewed computer vision along with recognition of visual gestures as one of the "future" ways we will interact with our computers. HCI researchers focused on improving the usability of desktop computers (i.e., practitioners concentrated on how easy computers are to learn and use) and it will continue to evolve towards more natural forms of input [39].
In this regard, computer vision researches have taken its way towards the diagnosis and therapy of disorders [7]. One of an interesting point for these researches is a gray-scale intensity images captured with RGB cameras which matches human vision system, but the latter's sensitivity to light affect the accuracy of the resulting images. Whereas depth images of 3D camera considered more durable option, since it less sensitivity to light and contains distance information of with surfaces of scene objects, but the main disadvantage of these cameras is the high cost.

Kinect v2 is one of the 3D cameras which has RGB color camera with a resolution 1920*1080, depth sensor with a resolution 512* 424, and infrared sensor a resolution 512* 424.Its frame rate range from 15 to 30 fps. It used time-of-flight (ToF) to estimate the distance to an object surface using active light pulses from a single camera based on time and speed that light has taken to reflect devices, ToF cameras are cheap, small, high speed and dense depth maps cover every pixel [8]. The recognition processes from the object. Compared with other laser 3D scanning are considered inexpensive, robust, and lighter compared with the RGB camera, while occlusion and limitation of sensor viewpoint are still common challenges for depth sensor. Besides being a gesture-controlled console for gaming, Kinect offers depth sensing for a wide variety of emerging applications in computer vision, human-computer interactions, robotics, and augmented reality [9].
This paper organized: Section II briefly introduces methods and the results of use of Kinect v2 in diagnosis; Section III includes the discussion of methods; Finally, Section IV describes the conclusion.

II. METHODS

One of important using of the camera is to diagnose certain diseases such as autism spectrum disorders. This research collects researches of this field into 4 groups, as shown in Figure 1.

2.1 Facial Expressions Analysis

In 1952, Aseh proved that facial behaviors associated with an emotion where Floyd Allport (1924), Tomkins (1962) and Darwin (1872) agreed with him [7]. Since then, there has been a lot of researches in this area, some researches tried to improve human-computer interaction, while other research used facial behaviors psychologically to improve social communication for children with some disorders such as ASD. Facial expression information can extract from the static image in a process called Localizing or from sequence images in a process called tracking. Also, the distinction should be made between the features used to represent the face which called face model features and the prominent features of the face (eyes, nose, mouth, and chin) which called facial features [10]. The face can represent as a set of features (analytic representation), as a whole unit (holistic representation), or as a combination of these (hybrid approach) [9]. One of the modern devices used for this purpose is Kinect v2 which has face tracking engine tracks at the speed of 4-8 ms per frame depending on CPU only (does not use GPU, since it may be needed to render graphics). Also, its software development kit (SDK) provided 121 3D points representing the face. Table 1 shows the studies that used Kinect v2 to diagnose autism spectrum disorders based on facial expressions.

However, the automatic analysis of the facial expressions has certain limitations because their short duration (they occur as fast as 1/15–1/25 of a second) and their low movement amplitude. It is necessary to point out here that the shutter speed of Kinect v2 is relatively slower compared with polikovsky et al. (2009), Yan W-J et al. (2013) with a shutter speed reached to 200 fps, and Zhao G et al. (2014) with a shutter speed reached to 100 fps.

2.2 Hand Movement Analysis

Sign language and survey research show various gesture communications modalities [11]. Today, Human-Computer Interaction (HCI) is the demand of the world where advances in computer vision and gesture recognition considered being more workable than traditional input devices such as (mouse, keyboard, touch screen, and joystick) [12].
Hand gestures are easy, spontaneous, and powerful communication modalities for interaction. They are also an important marker of diagnose some disorders such as autism spectrum disorders, which are common in most children with these disorders, as much research of diagnosis has focused on it.

The latest version of Kinect provides major enhancements in hand tracking and finger poses, also it has demonstrated a significant performance of light sensitivity in gesture recognition, however, there are still unsolved issues of gesture finding, and segmenting with depth information needs to be further treated. Recently, researchers have focused on the recognition problem instead of the problem of gesture segmentation in the applications of hand gesture recognition. The gesture segmentation methods were usually utilized by the direct distance interval setting [4, 5, 29] or the hand being the front most object [30]. The simplified methods demonstrated to be quick and effective;

however, the distance between the hand and Kinect sensor is restricted so that the hand gestures can only be recognized by moving the hand to a specific position and keeping the distance during the whole process. Table 2 shows the studies that used Kinect v2 to diagnose autism spectrum disorders based on hand gestures.

2.3 Eye Movement Analysis

Visual behavior has long been discovered by eye movement analysis. Eyes movement is also a rich source of information for recognizing emotion and activity. Independently of what we are looking at, our eye movement patterns have the potential to reveal much about the activities themselves. This includes information on physical activities, such as driving a car, information on visual tasks, such as reading [16], information on cognitive processes of visual perception, such as attention or saliency determination [32], but it cannot reveal thought contents [33]. Researches in the eye-movement field have adopted many new disciplines since the early 2000s, many of which are full of researchers with experience in statistics and experimental design. These researches have led to advances in an understanding of how the brain processes tasks, and the role that the visual system plays in this. All studies used known eye movement characteristics to model visual behavior during specific tasks. This development of researches in the eye movement field has driven by technological development, such as modern video-based eye trackers [6]. 3D eye-tracking method using the HD face model of Kinect v2 is one of the modern techniques used in tracking eye movement, because the proposed method uses accurate 3D ocular feature positions and a 3D human eye scheme, which can track an eye gaze position more accurately and promptly than previous methods. Table 3 shows the studies that used Kinect v2 to diagnose autism spectrum disorders based on eye movement analysis.

2.4 Body Movement Analysis

The occurrence of revolution in many tasks will be inevitable if the machines can automatically interpret the daily activities of people. For this reason, human motion analysis has been an active research area in computer vision, whose goal is to segment, capture and recognize human motion in real-time, perhaps predict ongoing human activities, and diagnoses some disease [13]. Tracking movements of a human body have been done in much researches and applications by placing active or passive markers at the body joints important to extract low-level features. On the other hand, construction 3D or 2D dynamics models of the human body from these markers have been generated by virtual wire-frame connections [14]. Sometimes, the place of markers may negatively affect the state of the people. After the release of Microsoft Kinect, all these obstacles were overcome, as this camera extracts the positions of the joints without needed to markers or wearable devices. Kinect v2 can track up to six skeletons at one time. Each of these skeletons has 25 joints.

Table 1: Overview of studies used Kinect v2 in facial expressions analysis of children with autism spectrum disorders

Author	**Year**	**Population**	**Significant Findings**
Youssef et al.[8]	2013	Auto-Optimized Multimodal Expression Recognition Framework Using 3D Kinect Data for ASD Therapeutic Aid	Aim: Children recognize emotions Methods: Two different frameworks to recognize 6 basic expressions using 3-dimensional data sequences captured in real-time using the Kinect sensor. At first used the SVM classifier on the whole set of facial key-points, while the second applied SVM to upper and lower key-points separately. Result: Best accuracy for individuals who's not taking part in the training of classifiers was 38.8% (SVM) and 34.0% (k-NN) while the best accuracy for others was 78.6% (SVM) and 81.8% (k-NN).
Li & Elmaghraby [15]	2014	A Framework for using Games for Behavioral Analysis of Autistic Children	Aim: Enable children to play serious games, evaluate treatment sessions based on these games, and possible directions for further treatment Method: Analysis of emotions of children after recording their interactions using Kinect camera Results: There are two approaches to making the results contribute to children's evaluation: The results are either generated in real-time and sent to control games or processed and stored separately
Tang et al. [7]	2016	Helping Neuro-typical Individuals to "Read" the Emotion of Children with Autism Spectrum Disorder: An Internet of- Things Approach	Aim: helping neuro-typical individuals to "read" the emotions of children with disorders Method: internet of things environment comprises several embedded sensors and objects such as IP camera wearable watch and Kinect v2 which produced emotion labels in the play environment by capture real-time facial and behavioral data. Data can be captured in real-time through four modules: Uno, Yun shield, base shield and the sensor which create Arduino unit the data can be stored in the cloud and retrieved for further analysis. Result: it is the first attempt to create an IoT-based gaming environment to "read" the emotions of children with ASD and open up a channel for social interactions with other children.
Zhao et al. [9]	2018	Facial Expression Detection Employing a Brain-Computer Interface	Aim: Using brain-computer interfaces to detect happy, sad, angry, and surprise facial expressions. Method: The components of the facial expression tracking system are the Microsoft Kinect V2 for facial tracking, automated pre-recorded animations display software developed with the Unity 3D game engine and asks users to mimic shown facial expressions and Muse BCI headband to provide raw voltages associated to brain activity from its sensors on the left ear, left forehead, right forehead, and right ear. Results: It showed that brain signals can monitor face expression changes under certain conditions, although special care has to be considered when processing the data given a large amount of noise caused by muscle movement and multimodal stimuli.

			More accurate brain activity is provided in case of the lack of facial muscular activity. A 72% classification accuracy using a random forest classifier was obtained.
Ramirez-Duque et al. [16]	2018	Robot-Assisted Autism Spectrum Disorder Diagnostic Based on Artificial Reasoning	Aim: In naturalistic behavioral observation, the early detection of ASD signs may be improved through child-robot interaction CRI and technological-based tools for an automated behavior assessment. Method: data of each RGBD sensor in the multi-camera system are processed for two nodes: driver level node to transform the streaming data of the RGBD sensor into the ROS messages format and processing node to execute the face analysis algorithm. Results: This study showed the feasibility of identifying and quantifies differences in the patterns of behavior of TD children and ASD elicited by the CRI intervention. Through the proof of a concept, it is evidenced here the system able to improve the traditional tools used in ASD diagnosis.

Table 2: Overview of studies used Kinect v2 in hand movement Analysis of children with autism spectrum disorders

Author	Year	Population	Significant Findings
Gomez-Donoso et al. [17]	2016	Automatic Schaeffer's Gestures Recognition System	Aim: It aims to recognize the whole of Schaeffer's sign language and create a continuous real-time classification system by determining the starts and ends of gestures. Method: When a person makes a gesture in front of the camera, the motion is captured by the Kinect v2. The gesture object is sent to the Gesture Class Pre Selection (GCPS) module, which selects a subset of classes for the gesture, discarding others to improve performance. Then, the candidate's classes and the gesture itself are sent to the classifier. The classifier compares the unknown gesture with every gesture present in the model using Dynamic Time Warping (DTW), and it uses the Nearest Neighbor (NN) algorithm to select the one with the shortest distance. Result: The system can recognize a subset of 11 gestures of Schaeffer's sign language online. Drawback: It is necessary to achieve a real-time response, but it leads to decreases the accuracy of the system.
Oprea et al. [18]	2017	A Recurrent Neural Network based Schaeffer Gesture Recognition System	Aim: It uses the Schaeffer language recognition system to teach children how to communicate using gestures, speed up the learning process, and increase a child's interest using a technological approach. Method: The sliding-window system was implantation to classify Schaeffer sign language. RNN-based models are fed with variable-length sequences of feature vectors where each vector represents a human body pose at a given time step. Besides vanilla RNNs, different models, such as LSTMs and GRUs together with skeletal data, provided by a single Kinect

			v2 camera, have been implemented and tested. Result: A subset of 25 gestures has been successfully recognized. A Long Short-Term Memory (LSTM) model has been implemented for this purpose reporting a 93.13% classification success rate over a subset of 25 Schaeffer gestures.
Marinoiu et al. [19]	2018	3D Human Sensing, Action and Emotion Recognition in Robot Assisted Therapy of Children with Autism	Aim: It records non-staged videos for therapy sessions based on the robot and introduces new tasks to recognize emotion defined. Method: The research considers RGB and depth videos recorded by Kinect v2 camera (at 30 FPS) placed right above the robot head towards the child. They experiment with several skeleton-based action recognition models and perform ablation studies with different 2D and 3D human body reconstructions. Also, use a cross-validation setting on the upper-body joints of the human skeleton. Result: The results show that properly adapted, the current 2D and 3D reconstruction methods from RGB data compete with industrial-grade RGB-D Kinect systems. With action recognition baselines in the 40-50% performance range, the large-scale data introduced by them represents a challenge in modeling behavior, with an impact in both computer vision and child-robot interaction with applications to autism.

Table 3: Overview of studies used Kinect v2 in Eye Movement Analysis of children with autism spectrum disorders

Author	Year	Population	Significant Findings
Tapus et al. [3]	2012	Children with Autism Social Engagement in Interaction with Nao, an Imitative Robot - A Series of Single Case Experiments	Aim: It aims to determine if the interaction with the robot helps 4 single-subjects with ASD to show more social participation. Method: Real-time arm movements are imitated by the robot. Besides the robot, they used a Kinect sensor. The skeleton data (i.e. joints and rotation positions) is transmitted to a robot which is mirroring the user upper-body position. Different behavioral criteria (i.e. smile/laughter, gaze shifting, and eye gaze) were analyzed based on video interaction. Result: High variability in reactions of children with the robot. The second and third children showed not affected by the robot, while the first child shows best-shared attention.
Jaiswal et al. [4]	2017	Automatic Detection of ADHD and ASD from Expressive Behavior in RGBD Data	Aim: It presents a novel method to aid diagnostic predictions about the presence/absence of ADHD and ASD by automatic visual analysis of a person's behavior. Method: Training the classifiers involves computing high-level feature descriptors corresponding to facial expressions (facial AUs), head pose and motion, etc. each video is first divided into 12 segments corresponding to the 12 stories that the participants have to read while they were recorded. For each video segment, histogram-based feature descriptors are computed separately using pre-trained classifiers/ regressors

			that detect individual behavioral cues. The combined set of feature descriptors from all segments in a recording are used for training the ADHD/ASD classification models. Result: Using state-of-the-art facial expression analysis based on Dynamic Deep Learning and 3D analysis of behavior, they attain classification rates of 96% for Controls Vs Condition (ADHD/ASD) group and 94% for Comorbid (ADHD+ASD) vs ASD only group. It shows that the system is a potentially useful time-saving contribution to the diagnostic field of ADHD and ASD.

Table 4: Overview of studies used Kinect v2 in Body Movement Analysis of children with autism spectrum disorders

Author	Year	Population	Significant Findings
Magrini et al. **[20]**	2016	An Interactive Multimedia System for Treating Autism Spectrum Disorder	Aim: It presents a real-time gesture tracking system uses in active well-being self-assessment activities and apply to medical coaching and music-therapy. Method: The system was developed in two different steps; the first version is based on the video capture and processing, where all the body recognition routines are software-based. This system needed a controlled environment and a technician during each session. The second version relies on the use of Kinect v2 which allows the tracking of full-body movements in 3D space and has the peculiarity of being installed in the user's home, and provides an intuitive interface, to be easily used by the children's families. Result: Children developed skills in establishing joint attention, imitation of caregivers, communicating with gestures and symbols. This promises to transfer the behavior shown in the setting to the external environment, increasing communication and interaction in the real world.
Bi Ge [21]	2016	Detecting Engagement Levels For Autism Intervention Therapy Using RGB-D Camera	Aim: The motivation is to develop an autonomous robot system able to perform autism intervention therapy. This work focuses on detecting engagement /disengagement levels of a child in a therapy session as a first step in designing a therapy robot. Method: Using Kinect and kinematic movements extracted from Kinect as features towards modeling the engagement level of a child during an autism intervention therapy session. Autism intervention therapy usually comprises a therapist, a student (child) and a task. Some tasks include a piece of paper, a book, an interactive toy, and a touch-screen tablet. While the student is working on the task, the therapist will issue prompts whenever the child appears to be distracted and stops concentrating on the task. Result: Using AdaBoost with Mean Joint Coordinates achieved the best single-feature performance at 96% accuracy while AdaBoost with all the features gives an accuracy of 97% appears to be the best performing combination.

KANG et al. [22]	2016	Automated Tracking and Quantification of Autistic Behavioral Symptoms Using Microsoft Kinect	Aim: Novel approach to detecting autistic symptoms using the Microsoft Kinect v.2 to objectively and automatically quantify autistic body movements. Method: Kinect V2 recordings were obtained while twelve actors performed nine separate repetitions of hand flapping, body rocking, and spinning. Visual Gesture Builder (VGB) is an **interactive tool for building models of body gestures using the Random** Forest Regression (RFR) machine learning classifier. VGB uses two detection technologies: AdaBoostTrigger and RFRProgress. Each of the videos was analyzed with Visual Gesture Builder and provided a graph that compares frame number with gesture completeness. Also, MatLab's solution trained the videos, the output showing the frequency and duration of the motor movements. Once the MatLab code was developed for each motor movement, test videos were run through the code and the outputs were obtained as text files. Result: Overall, the VGB approach outperformed the MatLab approach on each of the three stereotypical motor movements investigated.
Barmaki [23]	**2016**	Gesture Assessment of Teachers in an Immersive Rehearsal Environment	Aim: It uses machine-learning algorithms that semi-automate and automates event tagging in training systems developed to improve human interaction. It provided real-time feedback to people participating in classroom rehearsal in a virtual environment. Method: It used the RGB-D sensor (Microsoft Kinect V2) for human full-body tracking and the gesture recognition process. There were five targeted closed gestures to train in this research. Result: Besides technical contributions, this thesis presents a road map for future research, ranging from feedback modalities to automated event tagging. These studies provided invaluable clues to understand the necessity and demand for social skill training in the teaching context.
Zaraki et al. [24]	2017	Toward Autonomous Child-Robot Interaction: Development of an Interactive Architecture for the Humanoid Kaspar Robot	Aim: Teach and improve the social skills and capabilities of the children through interacting and playing with robots Method: The first step towards creating an autonomous CRI relies on Kaspar's competency to recognize and track different toys and children's high-level communicative signals. Interactive architecture (IA) has three main interconnected layers, which deal with Kaspar's Sense or perception, Think or decision making and action planning processes, and Act or low-level motor control and action performing capabilities. The first function of IA is to perceive the environment, using various sensors and devices, and the second is to control the robot's behavior appropriately to advance the game and close the interaction loop with children. Result: Preliminary results indicate that the architecture of the system for autonomous CRI studies is promising.
Budman et	2019	Quantifying the social symptoms	Aim: This study shows the utility of several physical measures of interactions between child and doctor for quantifying social

al. [25]		of autism using motion capture	ASD symptoms. These measures include the distance between them, the percentage of time the child approaches or moves away from the doctor and the direction of movement Method: Four Cameras (Kinect V2) was installed at a height of 1.5m in corners of the room (5.8 by 3m), and each camera records infrared depth images. Children were recorded during a clinical assessment. Result: These measures show that the variance in ADOS scores is 30% when using only 5m segments of "free play" from the recorded ADOS assessments. The study shows the great benefit of capturing movement to assess social symptoms in autism spectrum disorders for doctors and researchers. Continued development of this technique, motion capture measures and the use of advanced devices are likely to yield valuable information that helps determine the severity of symptoms and changes in these disorders.
Piana et al. [26]	2019	Effects of Computerized Emotional Training on Children with High Functioning Autism	Aim: An evaluation study of a serious game and a system for the automatic emotion recognition designed for helping autistic children to learn to recognize and express emotions usingtheir full-body movement is presented. Method: Three-dimensional motion data of full-body movements are obtained from RGB-D sensors and used to recognize emotions by linear SVMs. Result: Increase of task accuracy from the beginning to the end of training sessions in the trained group. In particular, while the increase of recognition accuracy was concentrated in the first sessions of the game, the increase in expression accuracy was more gradual throughout all sessions. The training seems to produce a transfer effect on facial expression recognition.

III. DISCUSSIONS

Recent researches in computer vision have established the importance of gesture recognition systems for human-computer interaction. Gestures are defined as the movements of a body part to convey meaning or information. Besides the role of gesture recognition in tracking body movement, which are used to interpreting those movements, translate them into a meaningful form, express feelings and communicate ideas specifically in some disorders characterized by weak social communication [12]. Most technical researches in the diagnosis of ASD over the years focused on four main areas: facial expressions analysis, gesture analysis, eye movement analysis, and body movement analysis as shown in figure 1.

Face expressions reflect people's emotional state, intentions, and important events in the environment [27]. To conduct interpersonal communication and normal social interaction, it is necessary to understand facial expressions and try to extract socially relevant information. Recognizing facial expressions is an early social skill that emerges in children early in life [28]. Walker- Andrews (1998) found that children could recognize happiness, sadness anger, surprise, and fear at 4 months when these expressions were presented in a familiar context and the use of emotional expressions of the social signal appears for infants at 8 or 10 months. Children's social development is greatly influenced by failing to recognize early emotions, as the child cannot recognize the feelings and responses of others. Despite many studies that have lasted for more than half a century, there is still no consensus on whether identifying basic emotions is a fundamental difficulty for individuals with autism spectrum disorders. If this weakness exists, it is not clear whether reading some emotions is affected and avoided or weakened, or all emotions are affected equally [27].One study has shown that children with autism have no difficulty distinguishing feelings of happiness and sadness while they cannot perceive surprise. Another study found defects in the identification of disgust and anger [29]. Other studies show that there are more difficulties in recognizing fear and anger than other basic expressions. Harms et al. (2010) believed that "behavioral studies are only slightly more likely to find facial emotion recognition deficits in autism than not" [30]

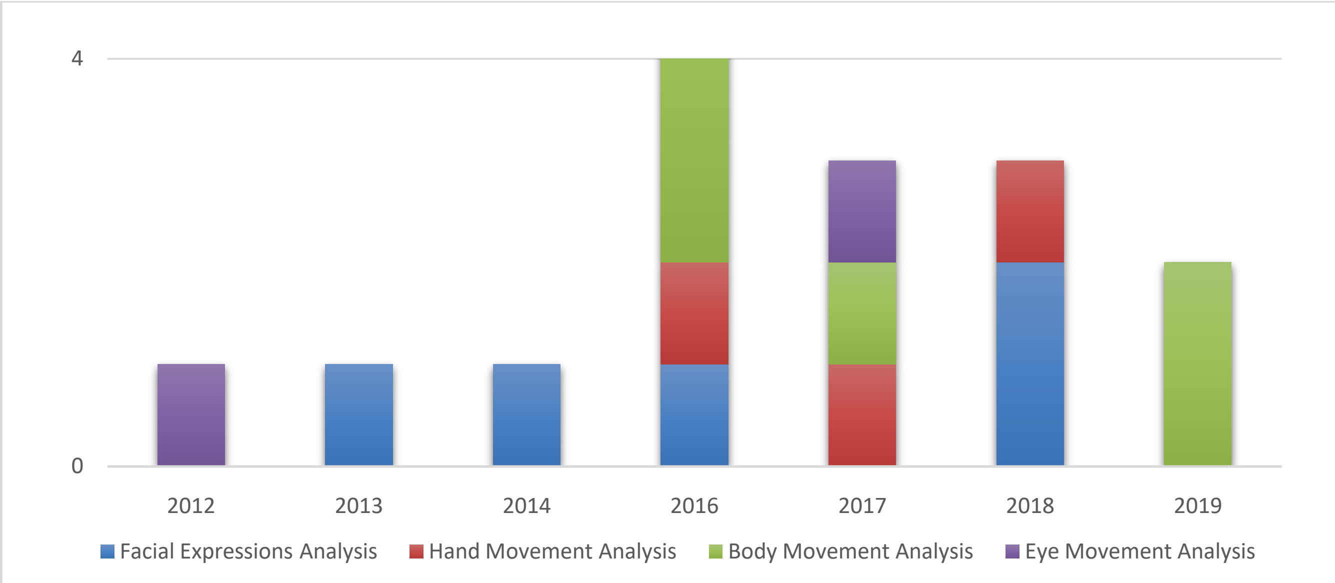

Figure 2: Number of studies over the years which used Kinect v2 in the diagnosis of autism spectrum disorders

There are two models in which the face-tracker operates with skeleton based information and without. In the 1st models, one can pass an array with two head points (neck position, head position) to StartTracking/ContinueTracking methods. These points allow the face tracker to find a face faster and easier, so this model is cheaper in terms of computer resources and sometimes more reliable at big head rotations. The 2nd model only requires color frame plus the depth frame to be passed with an optional region of interest parameter that tells the face tracker to search on RGB frame for a user's face. If the region of interest is not passed, then the face tracker will try to find a face on a full RGB frame which is the slowest model of operation of StartTracking method [36].

People use hand gestures to express their thoughts and feelings. So as to analyze these gestures, one of the two basic approaches is required; vision-based analysis or glove-based analysis. The former provides much better convenience for the user where non-contact-type devices (such as Kinect v2) are the preferable ones for man-machine-interfaces to adapt the computer to the human being and detect in real-time stereotyped behaviors [31]. The latter, such as DataGlove, can measure the position of the hand and the finger joint angles precisely. It has to be adjusted to the user and, due to the cables needed to transmit data to the computer; it is cumbersome to wear them. The better option to use in the case of people with ASD is a vision-based analysis. Kinect sensor is usually selected to acquire gestures to obtain extra depth information, and it has demonstrated a significant performance of light sensitivity in gesture recognition. Finally, there are two factors to consider when studying and analyzing gestures; First, the lack of training data makes most algorithms able to cover only a limited set of action types. Second, the data are captured in different environmental settings, with different lighting conditions, camera settings, and fixed or dynamic camera positions.

On the other hand, eyes movement is also a rich source of information for recognizing emotion and activity. There are three basic eye movement types: simultaneous movement of both eyes which calls saccade, the gaze which held upon a specific location in the visual scene which calls fixations, and regular opening and closing of the eyelids to spread fluid of eye across the corneal surface which calls blink. Developing sensors to record eye movements in daily life is still an active topic of research. However, these remain expensive, with demanding video processing tasks requiring bulky auxiliary equipment [32]. Studies tracking the eye movement of children with autism have found abnormally frequent gazes to the mouth and abnormally infrequent gazes to the eyes. According to Trepagnier et al. (2002), high functioning people with autism have fewer gazes onto faces than controls [34]. While Pelphrey et al. (2002), found decreased gazes to face features (e.g., eyes, nose, mouth) and increased gazes to nonfeatural elements of static faces compared to controls.

Tracking techniques are used for the analysis of human body movements. Human motion analysis has been widely used by researchers since the 1960s. Individual differences, emotions, environment, culture and many other factors determined human movement. Recognizing actions automatically are a major challenge because it is influenced by many factors; the great diversity in the shape, size of the human body. Besides the different actions taken by people to do the same work, even the same person may do the same work in different ways at different times. In this context, Aggarwal and Wang recommend tracking identification points (or low-level features).

In this technique, identification points are associated with points in a kinematic or dynamic model. These identification points are usually extracted from a time sequence of 2D images using digital processing methods such as pattern recognition and segmentation.

Finally, we need to clarify that most of the researches covered in this paper are clear, high accurate, and fast processing data, but some important aspects are often overlooked such as the research environment (brightness and sunlight, temperature, occlusions, and air draft), placement of Kinect camera (high, distance from an object, leveling) which affects the performance of Kinect and the accuracy of the results [35, 37, 38]. Also researches did not determine a type of ASD and T-score of person that system work with them.

IV. Conclusion

Autism spectrum disorders are complex and lifelong neuro developmental conditions that occur in early childhood and are associated with unusual movement and gait disturbances. Earlier identification of these pervasive disorders could provide help in diagnosis and establish rapid quantitative clinical judgment. In this paper, various applications of the Kinect in the diagnosis of autism spectrum disorders have been examined. The research has classified studies of (1) facial expressions analysis, (2) hand movement analysis, (3) eye movement analysis, and (4) body movement analysis. The results of the review show that in contrast to eye movement analysis, body movement analysis is gaining better popularity in ASD diagnosis researches as show in figure 2. Most of covered researches overlooked several factors that affect the performance of Kinect and the accuracy of the results such as (researches environment, placement of camera greatly). Also, they need to determine a type of ASD and T-score of person. The direction for future work is to use Kinect v2 in the early diagnosis of autism spectrum disorders through gait analysis.

V. references

[1] American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders, 5th ed. Arlington: VA American Psychiatric Publishing, 2013.

[2] C. Rhind et al., “An examination of autism spectrum traits in adolescents with anorexia nervosa and their parents.” Molecular autism vol. 5, no.1, p. 56. 20 Dec. 2014. doi:10.1186/2040-2392-5-56.

[3] A. Tapus et al., “Children with Autism Social Engagement in Interaction with Nao , an Imitative Robot- A Series of Single Case Experiments,” Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems, vol. 13, Issue 3, p. 315 –347, Jan 2012. doi:10.1075/is.13.3.01tap.

[4] S. Jaiswal et al., “Automatic Detection of ADHD and ASD from Expressive Behaviour in RGBD Data,” 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), Washington, DC, USA, 30 May-3 June 2017. pp. 715-722. doi:10.1109/FG.2017.95.

[5] J. Bloom, American Council On Science And Health, June 7, 2017. [Online]. Available: https://www.acsh.org/news/2017/06/07/mit-researcher-glyphosate-will-cause-half-all-children-be-autistic-2025-yeah-sure-11337. [Last Accessed: July 2019]

[6] J. Orquin and K. Holmqvist, “Threats to the validity of eye-movement research in psychology” Behavior Research Methods, vol. 50, p.1645–1656, Aug 2018. [Online]. Available: doi:10.3758/s13428-017-0998-z.

[7] T. Y. Tang, “Helping Neuro-typical Individuals to "Read" the Emotion of Children with Autism Spectrum Disorder: an Internet-of-Things Approach,” IDC '16 Proceedings of the The 15th International Conference on Interaction Design and Children, Manchester, United Kingdom, June 21 - 24, 2016. pp. 666-671. doi: 10.1145/2930674.2936009.

[8] A. E. Youssef et al., "Auto-Optimized Multimodal Expression Recognition Framework Using 3D Kinect Data for ASD Therapeutic Aid," International Journal of Modeling and Optimization, vol. 3, no. 2, pp. 112-115, 2013. [Online]. Available: 10.7763/IJMO.2013.V3.247.

[9] D. Zhao et al., “Facial Expression Detection Employing a Brain Computer Interface,” 2018 9th International Conference on Information, Intelligence, Systems and Applications (IISA), Zakynthos, Greece, Greece, 23-25 July 2018. pp. 1-2. doi: 10.1109/IISA.2018.8633661.

[10] M. Pantic and L. Ü. M. Rothkrantz, “Automatic analysis of facial expressions: The state of the art,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, no. 12, pp. 1424–1445, Dec 2000. doi: 10.1109/34.895976.

[11] T. Baudel and M. Baudouin-Lafon, “Charade: Remote Control of Objects Using Free-Hand Gestures,” Comm. ACM, vol. 36, no. 7, pp. 28-35, Jul 1993. [Online]. Available: doi:10.1145/159544.159562.

[12] GRS Murthy and RS Jadon, "A review of vision based hand gestures recognition," International Journal of Information Technology and Knowledge Management, vol. 2, no.2, pp. 405-410. Jul 2009.

[13] L. Chen et al., “A survey of human motion analysis using depth imagery,” Pattern Recognition Letters, vol.34, Issue 15, pp.1995-2006, 1 November 2013. doi: https://doi.org/10.1016/j.patrec.2013.02.006.

[14] L. M. Pedro and G. Augusto, “Kinect evaluation for human body movement analysis,” 2012 4th IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)Z, Rome, Italy, 24-27 June 2012. doi: 10.1109/BioRob.2012.6290751.

[15] Y. Li and A. S. Elmaghraby, “A framework for using games for behavioral analysis of autistic children,” 2014 Computer Games: AI, Animation, Mobile, Multimedia, Educational and Serious Games (CGAMES), Louisville, KY, USA, 28-30 July 2014. pp. 130–133. doi: 10.1109/CGames.2014.6934157.

[16] A.Ramírez-Duque et al., “Robot-Assisted Autism Spectrum Disorder Diagnostic Based on Artificial Reasoning,” J Intell Robot Syst, vol. 96, pp.267–281, 29 March 2019. doi: 10.1007/s10846-018-00975-y.

[17] F. Gomez-donoso et al., “Automatic Schaeffer ’ s Gestures Recognition System,” Expert Systems, vol. 33, no.5, pp. 480–488, 13 July 2016. doi: https://doi.org/10.1111/exsy.12160.

[18] S. Oprea et al., “A recurrent neural network based Schaeffer gesture recognition system,” 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, 14-19 May 2017. pp. 425–431. July 2017. doi: 10.1109/IJCNN.2017.7965885.

[19] E. Marinoiu et al., "3D Human Sensing, Action and Emotion Recognition in Robot Assisted Therapy of Children with Autism," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18-23 June 2018. pp. 2158–2167. December 2018. doi: 10.1109/CVPR.2018.00230.

[20] M. Magrini et al., "An Interactive Multimedia System for Treating Autism Spectrum Disorder," Springer, Cham, vol. 9914, pp. 331–342, November 2016. doi: https://doi.org/10.1007/978-3-319-48881-3_23.

[21] B. Ge, "Detecting engagement levels for autism intervention therapy using RGB-D camera," M.S. thesis, School of Electrical and Computer Engineering, Georgia Institute of Technology, May 2016. Accessed on: 25 June 2019. http://hdl.handle.net/1853/55043

[22] JY Kang et al.," Automated Tracking and Quantification of Autistic Behavioral Symptoms Using Microsoft Kinect," Stud Health Technol Inform, vol.220, pp. 167 – 170, 2016. doi: 10.3233/978-1-61499-625-5-167.

[23] R. Barmaki, "Gesture Assessment of Teachers in an Immersive Rehearsal Environment," Ph.D. dissertation, Computer Science, University of Central Florida, August 2016. [Online]. Available: http://purl.fcla.edu/fcla/etd/CFE0006260

[24] A. Zaraki et al., "Toward autonomous child-robot interaction: development of an interactive architecture for the humanoid kaspar robot," In 3rd Workshop on Child-Robot Interaction (CRI2017) in International Conference on Human Robot Interaction (ACM/IEEE HRI 2017), Vienna, Austria, Mar 2017, pp. 6-9.

[25] I.Budman et al., "Quantifying the social symptoms of autism using motion capture," Springer Nature, vol. 9,Issue 7712, pp. 1–8, May 2019. doi: 10.1038/s41598-019-44180-9.

[26] S. Piana et al., "Effects of Computerized Emotional Training on Children with High Functioning Autism," IEEE Trans. Affect. Comput., p. 1. May 2019. doi: 10.1109/TAFFC.2019.2916023.

[27] M. Uljarevic and A. Hamilton, "Recognition of Emotions in Autism: A Formal Meta-Analysis," J Autism Dev Disord, vol.43, pp.1517–1526. 2013. doi: 10.1007/s10803-012-1695-5

[28] K. A. Pelphreyet al., "Visual Scanning of Faces in Autism," J Autism Dev Disord, vol. 32, no. 4, pp. 249-261, Aug 2002. doi:10.1023/a:1016374617369.

[29] K. Humphreys et al., "A fine-grained analysis of facial expression processing in high-functioning adults with autism," Neuropsychologia, vol. 45. Issue 4, pp. 685–695, 2007. doi: 10.1016/j.neuropsychologia.2006.08.003.

[30] C. Cook et al., "Alexithymia, not autism, predicts poor recognition of emotional facial expressions," Psychol Sci, vol. 24, Issue 5, pp. 723-732, May 2013. doi: 10.1177/0956797612463582.

[31] C. Nolker and H. Ritter, "Detection of Fingertips in Human Hand Movement Sequences," Springer, Berlin, Heidelberg, volume 1371, pp.209-218. May 2006. doi: https://doi.org/10.1007/BFb0053001

[32] A. Bulling et al., "Eye Movement Analysis for Activity Recognition Using Electrooculography Andreas," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol: 33, Issue: 4, pp. 741 – 753, 2010. doi: 10.1109/TPAMI.2010.86.

[33] J. Hyönä, (2011). Documenting literature, "The usefulness and limitations of eye-tracking in the study of reading (and writing)". [Online]. Available: http://www.writingpro.eu/upload/presentations/Summerschool_EyeTracking_JukkaHyona.pdf

[34] Spezio ML et al., "Analysis of face gaze in autism using 'Bubbles'," vol. 45, pp. 144–151, 2007. DOI: 10.1016/j.neuropsychologia.2006.04.027.

[35] D. Fiedler and H. Muller, "Impact of Thermal and Environmental Conditions on the Kinect Sensor". Advances in Depth Image Analysis and Applications, vol 7854, 2013. https://doi.org/10.1007/978-3-642-40303-3_3.

[36] N. Smolyanskiy et al., "Real-time 3D face tracking based on active appearance model constrained by depth data". Image and Vision Computing, vol. 32, Issue 11, pp. 860–869, Nov 2014. [Online]. Available: doi: https://doi.org/10.1016/j.imavis.2014.08.005

[37] E. Lachat et al., "Assessment and Calibration of a RGB-D Camera (Kinect v2 Sensor) Towards a Potential Use for Close-Range 3D Modeling". Remote Sensing, vol.7, pp. 13070-13097, 2015. [Online]. Available: doi: DOI:10.3390/rs71013070

[38] R. Seggers, "People Tracking in Outdoor Environments Evaluating the Kinect 2 Performance in Different Lighting Conditions", Computer Science. June 26th, 2015.

[39] G.R.S. Murthy and R.S. Jadon, "Computer Vision Based Human Computer Interaction. Journal of Artificial Intelligence," Journal of Artificial Intelligence, vol. 4, pp. 245-256. December 08, 2011. Available doi:10.3923/jai.2011.245.256