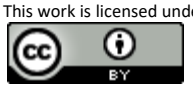
Available online at: https://ijact.in

Date of Submission	13/07/2020
Date of Acceptance	12/08/2020
Date of Publication	01/09/2020
Page numbers	3791-3797 (7 Pages)



ISSN:2320-0790

An International Journal of Advanced Computer Technology

GENERATING 3D DATASET OF GAIT AND FULL BODY MOVEMENT OF CHILDREN WITH AUTISM SPECTRUM DISORDERS COLLECTED BY KINECT V2 CAMERA

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

[1]College of Information Technology, University of Babylon, Babil, Iraq

Abstract: Until now, a three-dimensional dataset combines gait and body movement analysis of children with Autism Spectrum Disorders (ASD) in controlled environments has not been published. ASD mean disorders of neurodevelopment that last a lifetime which occurs in early childhood and usually associated with unusual movement and gait disturbances. Three-dimensional gait features captured by Kinect v2 can assist clinicians in diagnosing, clinical decision-making, and treatment planning of ASD. In this paper, Kinect v2 uses to build a 3D-skeleton-based gait dataset, which includes joints positions, the corresponding skeleton movement video, joints trajectories video captured by Kinect v2, and color videos captured by Samsung Note 9 rear camera. Besides building dataset, this paper classifies children with ASD from normal children by proposed system comprises four main stages: 1) Augmentation the database by using seven transformations to solve the problem of the small number of ASD cases, 2) Extracting features that we think play an important role in classification, 3) Reducing data dimensions using Principal Component Analysis (PCA) and 4) Using Multilayer Perceptron (MLP) to classify data. Classification accuracy when using eleven features result from PCA and MLP is 95% with 0.7 seconds to build the model.

Keywords: Autism spectrum disorders, Kinect v2, Gait analysis, body movement analysis, 3D-skeleton-based gait dataset, Dataset augmentation.

I. INTRODUCTION

Understanding the causes of Autism spectrum disorders has always been the dream of doctors and researchers. Autism spectrum disorders (ASD) are a pervasive neuro developmental disorder diagnose by the age of two years or sometimes even earlier and characterize by a triad of impairments: social communication problems, difficulties with reciprocal social interactions, and unusual patterns of repetitive behavior [1]. Unfortunately, no medications can cure Autism spectrum disorders or treat its core symptoms but can help some people affected feel better, while the early professional care can make a big difference in preparing these autistic kids for safety life.

Possible signs that could identify ASD are unusual behavior patterns [4], frequent body movements, and purposeless motor signs which referred to as stimming or stereotypy [2]. Stimming of ASD children has been gaining greater attention in recent years. However, little studied why these children exhibit differences in their movement and what is the relationship between these movements, which may lead to understand the underlying etiology of the disorders. Stimming can be studied from videos captured in two environment types. In uncontrolled environments, when the children are performing their daily activities, some self-stimulatory behaviors can be studied by automatically analyzing the captured videos. On the other hand, controlled environments can also use where a therapist exercises a defined protocol of play action with the child to elicit higher-level behaviors. Children's behaviors in both scenarios analyzing are equally important for early intervention and diagnosis.

In the same context, the second version of the Microsoft Kinect is one of the 3D cameras that could be used in clinical applications and captures ASD children's behavior in both environments. It uses 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 from the object. It contains a color camera with resolution (1920 * 1080), a depth camera with resolution (512 * 424), and a microphone multi-vector [3], in addition to 30 frames per second for recorded color video and tracked skeleton. Recently, it is widely used because of its cheap price, high accuracy, small size, high speed, ease of capturing motion based on a depth camera without the need for wearable devices.

This research aims to provide a new publicly available three-dimensional dataset, Self-Stimulatory Behavior Dataset (SSBD) for autistic children to study the behaviors based on the skeleton data extracted by the Kinect v2 at the time that children walk towards the camera (controlled environment). This research proposes 1) Three-dimensional dataset with 25 body joints positions and 16 angles between joints captured by Kinect v2 in a controlled environment, 2) Dataset contains color videos, skeleton tracking videos and trajectory tracking videos for 50 cases with ASD, 3) Augmentation approach to increase the instances and improve classification accuracy, 4) Extracting about 1259 features for body movement which feed to Principal Component Analysis (PCA) to dimension reduction and 5) Using one of the most used classifiers for analyzing body movement and walking which is Artificial neural network.

The rest of this paper is organized as the following: Section 2 presents materials and methods which include data collection procedures and environment, data preprocessing, dataset augmentation approach, features extraction, dimension reduction, and classification approach. Section 3 presents the structure of the of dataset.

II. MATERIALS AND METHODS

To generate 3d dataset and classified data, body joints positions have been collected during walking of normal children and children with ASD in the following proposed system.

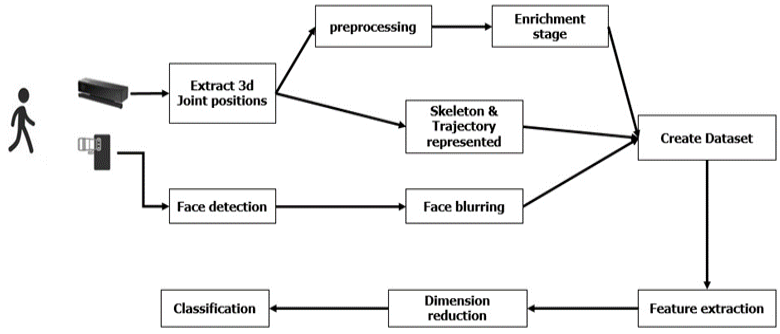


Figure1. The proposed system

A. Participants

Sixty-eight children with Autism spectrum disorders have been collected from seven ASD childcare centers in three different cities and fifty normal children from two kindergarten centers in Iraq. The degree of ASD was severe for some of them with a lack of response and great dispersion. In this case, we excluded some children and simulated the movement of others to end with fifty children that can be transferred to the following stages. Note that even color videos of excluded children have been included for scientific benefit. All participants were free of any lower-extremity injury before the data were collected, and all were free of any neurologic disorders or diseases that could interfere with their body movement patterns except Autism spectrum disorders. Before the collection of data, all parents of the participating children in this research have signed the informed consent of the world health organization. Demographic data of the two groups are shown in Table 1:

Table1: Mean height, weight & age value for each group

Parameter	ASD	Normal
Age (years)	4-12 years	6-11 years
Height (cm)	90-130 cm	100 -143 cm
Weight (kg)	20-58 kg	23-55 kg

B. Environmental setup and Procedure:

To ensure the best possible achievement, the temperature was measured periodically using mercury thermometer and it ranges from 20ºC to 22ºC. The ventilation was also good, which prevented the overheating of the camera. The camera placed away from direct sunlight and to ensure good lighting, the brightness measured frequently using the Lux Light Meter application on Samsung galaxy note 9 and it in range from 76 Lux to 87 Lux. The Kinect camera placed at a height of 0.75m and the recording was started 30 minutes after the camera was turned on. Children were asked to walk along a line, at normal speed, towards Kinect camera. The cameras recorded color video and skeleton tracking videos for 10 times then we choose one suitable gait cycle. Each time the participant walks about two gait cycles in the range of 1.5 to 4 meters in front of the camera. Then we extract one gait cycle to use in the following stages. The height of camera and distances between children and camera were chosen according to recommendations for getting the best data quality as shown in Figure 2.

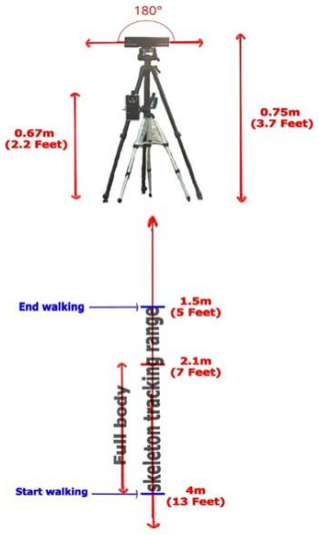


Figure2. Environmental setup of Kinect v2

C. Data Records

Two cameras have been used:

1- Kinect v2 which recorded position of each joint as coordinates (x, y, z): Each of these joints is records in specific range measured in meter; x ranging from -6 (max distance to right) to +6 (max distance to left), y ranging from -5 (max distance to bottom) to +5 (max distance to top), and z ranging from 0 (on surface of camera) to 8 (max depth from camera). We used C# 5.0 in visual studio 2012 connected with Kinect camera in order to extract body Joints and angles between joints which save as .csv file and draw body joints (skeleton) and trajectory of each joint as appearing in Figure 3. Table 2, 3 include joints and angles which extracted by Kinect v2:

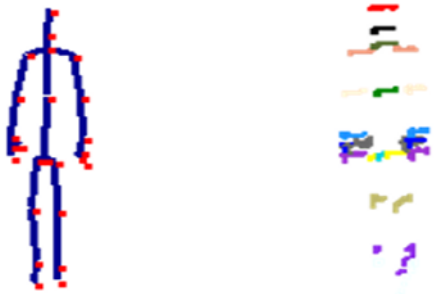


Figure3. Skeleton and trajectory of each joint

2- Samsung note 9 rear camera with digital 0.45X professional wide angle lens (58 MM): This camera recorded Full HD video with 60 fps and resolution 1920*1080. (12MP, f/1.5), in addition to wide angle lens which positioned on rear note 9 camera to get a wide view.

Kinect v2 captured twenty-five three-dimensional joints position, sixteen angles captured and two more features using to extract one gait cycle and stop the recording of Kinect v2 as shown in Figure 4.

Joints

no.	joint	no.	joint	no.	joint	no.	joint
1	Head	8	Wrist Left	15	Hand Tip Right	22	Ankle Left
2	Neck	9	Wrist Right	16	Spine Mid	23	Ankle Right
3	Spine Shoulder	10	Thumb Left	17	Spine Base	24	Foot Left
4	Shoulder Left	11	Thumb Right	18	Hip Left	25	Foot Right
5	Shoulder Right	12	Hand Left	19	Hip Right		
6	Elbow Left	13	Hand Right	20	Knee Left		
7	Elbow Right	14	Hand Tip Left	21	Knee Right		

Angles

no.	Abbreviation	Description
1	HESHL	Left Angle of Spine Shoulder between (Head, Left Shoulder)
2	HESHR	Right Angle of Spine Shoulder between (Head, Right Shoulder)
3	SPELL	Angle of Shoulder Left between (Spine Shoulder, Left Elbow)
4	SPELR	Angle of Shoulder Right between (Spine Shoulder, Right Elbow)
5	SHWRL	Angle of Left Elbow between (Left Shoulder, Left Wrist)
6	SHWRR	Angle of Right Elbow between (Right Shoulder, Right Wrist)
7	ELHAL	Angle of Left Wrist between (Left Elbow, Left Hand)
8	ELHAR	Angle of Right Wrist between (Right Elbow, Right Hand)
9	THHAL	Angle of Left Wrist between (Left Thumb, Left Hand)
10	THHAR	Angle of Right Wrist between (Right Thumb, Right Hand)
11	SPKNL	Angle of Left Hip between (Spine Base, Left Knee)
12	SPKNR	Angle of Right Hip between (Spine Base, Right Knee)
13	HIANL	Angle of Left Knee between (Left Hip, Left Ankle)
14	HIANR	Angle of Right Knee between (Right Hip, Right Ankle)
15	KNFOL	Angle of Left Ankle between (Left Knee, Left Foot)
16	KNFOR	Angle of Right Ankle between (Right Knee, Right Foot)

Other

Abbreviation	Description
DFRToFL	The distance between the feet
MinDBFAC	The shortest distance between the feet and the camera

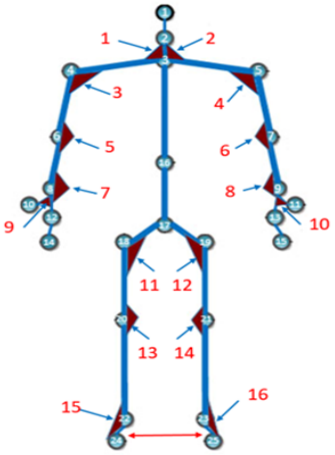


Figure4. Joints and angles recorded by Kinect v2

In the same context, we encountered three limitations with Kinect v2:

1- Self-occlusion in capturing lateral body movement: This occurred when part of a body hidden by another, as shown in Figure 5. This limitation has been overridden by adopting the front walk toward the camera.

2- The limited range of skeletons tracking: Kinect v2 tracking skeleton in the range [1.5m -4m] from a camera which in some cases prevented the completion

of two gait cycles, this limitation has been overcome by extracted one gait cycle.

3- Frame rate of skeleton data reduces when recording color frames to 15 fps rather than 30 fps. To avoid this problem, Kinect camera was used to tracking skeleton with 30 fps and used Samsung galaxy note 9 to record color video with 60 fps.

(a) **(b)**

Figure5. a) Self-occlusion in lateral body movement captured by Kinect v2, b) Image captured by Samsung Note 9

D. Data pre-processing

The gait cycle has been processed as the following:

a) Extract one gait cycle

Each child walked in front of cameras for 2.5m for ten times where Kinect v2 recorded joints movement features. One of these features is the distance between feet which used to extract features of one gait cycle. This feature is rising and falling two times in a single gait cycle as shown in Figure 6.

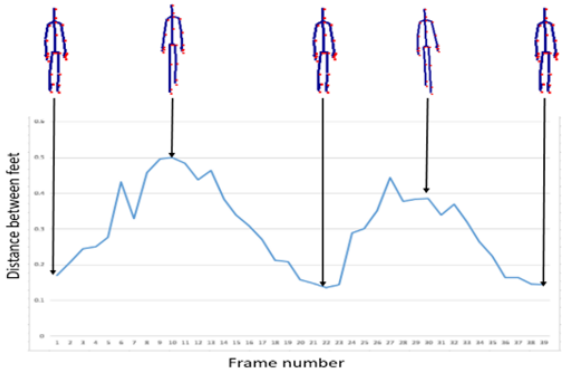


Figure6. Extract one gait cycle using distance between feet feature

b) Replacing of missing data

Choosing the optimal method to deal with missing values is always based on trial and error. In general, there are three methods: eliminate missing data, ignore the missing value during analysis, and replace missing value by another value. In eliminate data and ignore the missing value during analysis approaches, the sample size of data is reduced. In this paper, since we have data of one gait cycle then these approaches are excluded. On the other hand, replace missing value could be a good approach (except for replacing by mean which sometimes affected by outliers). This research-based on mice function in R language version 3.2.6 to replace missing values based on the impulse of many times. At first, the mice function detects the variables which have missing values, then missing values are replaced by Predictive Mean Matching (PMM) [5, 6]. In this method, a small set of candidates (usually 3, 5, or 10 candidates) has been formed for each missing entry where all candidates in the set have predicted values closest to the predicted value for the missing entry. Then one of the candidates is randomly selected and replaces the missing value. The assumption is that missing data are missing at random, which means that the probability that a value is missing depends only on observed value and can be predicted using them. Since it based on values observed elsewhere so it is realistic, evading problems with meaningless imputations and reduce the bias introduced in a dataset through imputation. Also, the model is implicit, so it is less vulnerable to model misspecification since there is no need to define an explicit model for the distribution of the missing values.

c) Face detection and blurring

Faces have been detected by two methods: Haar Cascades and Multi-Task Cascaded Convolutional Neural Network. The first method quickly detected the face, but in some cases, it failed so the second method has been used. Since MTCNN slower than the Haar method, image dimensions are reduced to 30 percent to increase detection speed. Then, Gaussian blur filer has been used to blurring detected face.

E. Augmentation Approach

The augmentation approach is shown below:

a) Dataset Augmentation

Dataset Augmentation [12] is done by applying a set of transformations to the original dataset to increase the diversity of data available for training models; enhance size and quality of dataset and avoid overfitting with taking into account preserving the label of data, these transformations have described below:

1-Jittering: Jittering is implemented as a way to simulate additive sensor noise. Each sensor has a different type of mechanical noise. Simulating random sensor noise increases the robustness of the training data against various types of sensors and their multiplicative and additive noises. Gaussian noise is used in this thesis to add jittering to raw training data.

2-Scaling: Scaling is another technique adopted in data augmentation which changes the magnitude of the raw data but preserves the labels. This variation observes in situations where in the dimensions of the implement to which the sensor is attached changes, such as a change in length of excavator boom, etc.

3-Translation: Shifting images left, right can be a very useful transformation to avoid positional bias in the data.

4-Flipping: Horizontal axis flipping is much more common than flipping the vertical axis.

5-Slicing: It is a subsampling method to randomly extract continuous slices from the original time series. The length of the slices is a tunable parameter. For the classification problem, the labels of sliced samples are the same as the original time series.

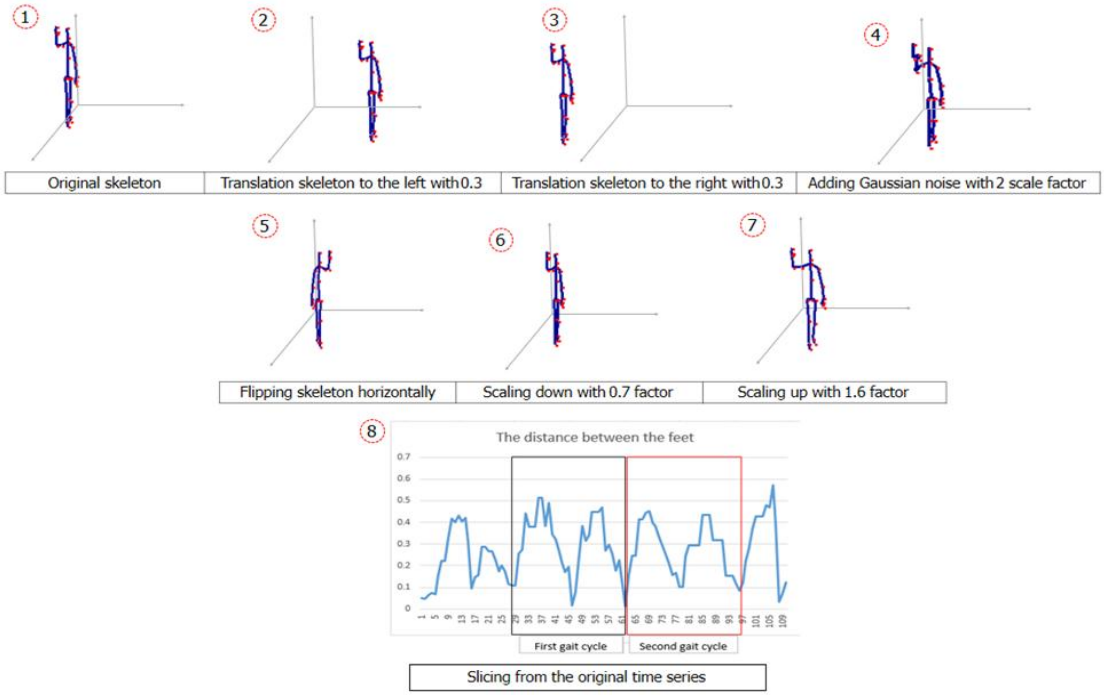


Figure7. Results of augmentation dataset

Before classifying data, the augmentation dataset is shuffled and divided into training (seventy percent) and testing set (thirty percent). Augmentation dataset has shuffled block by block where each block contains instance and its augmentation to ensure no leakage of augmentation instances to the testing set after dividing dataset. On the other hand, augmentation instances have been used only in training set while testing set contains the original instance since the goal of data augmentation is increased the diversity of data available for training models

b) Three-dimensional projection

In general, projection is the representation of points in the coordination system of dimension N into a coordination system with fewer dimensions [7]. Recently, computer graphics has been used a lot for 3D projection, which means maps point in three-dimensions onto a two-dimensional plane. There are two main graphical projection categories: parallel projection and perspective projection. Parallel projection is a projection of an object in three-dimensional space onto the projection plane where projection lines are parallel to each other, while perspective projection occurs when projector lines converge at the center of projection (Vanishing point), which results in many visual effects of an object. Perspective projection depends on the relative position of the eye and the view plane and considered more realistic than a parallel projection since it nearly resembles human vision and photography [7]. Kinect v2 software development kit provided build-in Map Camera Points To Color Space function which for the 3D projection of joints recorded directly from camera while skeleton tracking. In this paper, for the purpose of 3D projection of joints recorded at different times and resulting from augmentation stage, there was a need to build a special projection matrix [8,9] as shown in Equation 1:

1- Representing 3D joints in homogeneous coordinate [10]: This step aims to represent 3D joints in projective space by producing N+1 numbers form N-coordinates. In this paper, we add a variable w into existing coordinates to represent each joints position (x, y, z) as (x, y, z, w).
2- Determining affine matrix: This matrix used to correct geometric distortions that result from non-ideal angles of the camera.
3- Calculating the intrinsic matrix of Kinect v2: It calculated based on information captured by build-in GetDepthCamerIntrinsics function of Kinect v2:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} fx * sx & 0 & cx \\ 0 & fy * sy & cy \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r11 & r12 & r13 & t1 \\ r21 & r22 & r23 & t2 \\ r31 & r32 & r33 & t3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad (1)$$

(2d joints) (Intrinsic matrix) (Affine matrix) (3d joints)

Where fx and fy are focal length of the camera; sx and sy are scale factor; cx and cy are the principle point of camera and r 's are rotation factors. To get projection results closed as much as possible to Kinect v2 mapping, GetDepthCameraIntrinsics function provides focal length (both equal 366.8) and principle point (cx=260.3, cy =

208). We also used scale factors (sx, sy) equal to 2 and multiply principle points by 3 and 2.2, respectively. On the other hand, rotation factors r11, r33 equal -1 and r22 equal 1 to rotate joints with 180 degrees around y coordination. Figure 8 shows skeleton results from two approaches:

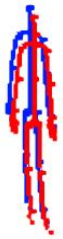


Figure8. Skeleton result from MapCameraPointsColorSpace (blue) and skeleton of our projection (red)

F. Feature Extraction

It is a method of creating an important point, distance, angles, range of movement, or any features or combination of features that may contribute to the distinction of the movement of persons. It is very complex since gait patterns categorized by time dependence, high dimensionality, and nonlinearities [11]. From the gathered data described above, we extracted difference features and gather them into four groups:

i. The distance between the joints: The distance has been measured using Euclidean distance between 3D joints positions where:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \quad (2)$$

ii. The distance between some joints and the ground: It gives important information in classification gait of children with Autism spectrum disorders and typical children gait, where:

- Head to ground distance distinguishes weak visual communication in children with ASD
- SpineBase to ground distance distinguishes jumping and bounce
- Hand tip to ground distance distinguishes flutter, put hands on-ear and fingertip in front of eyes
- Ankle to ground distance distinguishes toe walking and jump

With Kinect v2, floorClipPlane property calculates joint to ground distance by getting 4 float-point value (A, B, C) for orientation of the plane in the 3D space and W for the distance between the plane and the origin of the coordinate system. Then distance can calculate by:

$$d = \frac{Ax + By + CZ + W}{\sqrt{A^2 + B^2 + C^2}} \quad (3)$$

iii. The range of motion for each joint coordination which determines how far joints can move in different directions and distinguishes gross gait coordination. range of motion of each joint calculates by:

ROM (joint.x) =max (joint.x) – min (joint.x) (4)

ROM (joint.y) =max (joint.y) – min (joint.y) (5)

iv. Other attributes:

- The body is divided into upper and lower half according to y coordinates of SpineBase and determines the position of hand tip based on it. This can determine any abnormal movement of the hand.
- Step length, step width, and the distance between feet and stride length, as bellow:

Step length = Joint.z (t (1)) – Joint.z (t (2)) (6)

Step width =|Joint.x(t(2))-Joint.x(t(1))|(7)

Stride length= Joint.z (t (1)) – Joint.z (t (n)) (8)

- Gait cycle time, stand time and swing time

After calculating statistical measures (mean, variance, standard deviation) of all features, we get about 1259 features which feed to dimension reduction technique. All extracted features have shown in Figure 9:

G. Dimension reduction and classification

In this paper, all features have been fed into PCA and as the result, we get thirty-one features, only the top eleven features have been selected based on its standard division and passed to a multilayer perception classifier in weka. These features represent eleven nodes in the input layer which followed by six hidden nodes and two output nodes, typical children with children with ASD. Classification accuracy with eleven features results from PCA and MLP is 95% with 0.7 seconds to build the model.

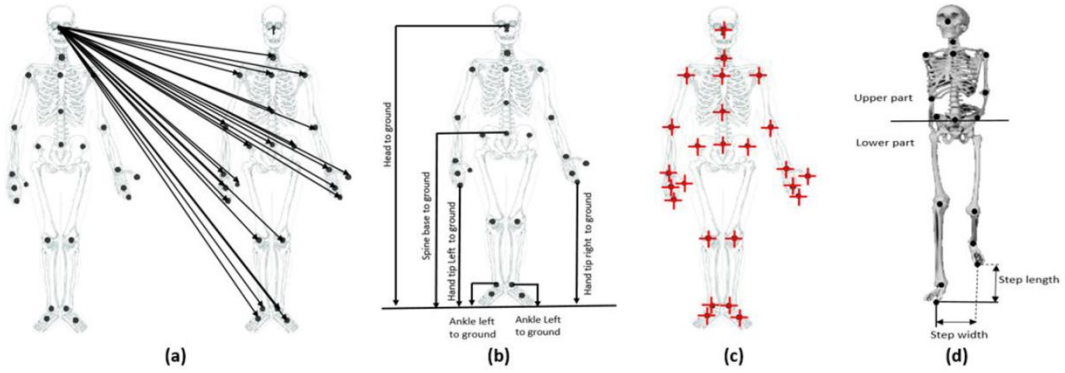


Figure9. Extracted features, (a) distance between joints, (b) Distance between head, spine base, Hand tip (left and right), Ankle (left and right), (c) Range of movement joint on x and y coordination, (d) step length, step width and parts of body

H. DATASET

Each of the 50 cases has a folder consisting of:

Folder	The contents
Color video	Color videos in (.avi format)
Skeleton video	Original video and videos after augmentation (.avi format)
Joints trajectories	videos of trajectories of 25 joint of each chid (.avi format)
3d Dataset folder (All in .xlsx format)	• File for full children gait in front of camera • File for the extracted one gait cycle • File after replace missing value • File for extracted features • seven files result from augmentation

Dataset folder also contain:

1- File for full dataset after calculating the statistical measures in (.csv) format.
2- File for the dataset after the normalization in (.csv) format
3- File for color video of excluded cases.

III. CONCLUSION

The primary goal of this paper is to create a 3D dataset for the gait of children with Autism spectrum disorders and classify them from normal children based on gait analysis and body movement analysis. After obtaining approval from the parents, the children were asked to walk toward the camera, which tracked the movement of the joints and captured the angles between the joints. We extracted a single gait cycle based on the distance between the feet and then extracted features that we think are important for classification. Given that the number of cases is relatively few, a dataset augmentation based on seven transformations saves the label of movement, increases the cases in the dataset, and improves classification model. A dimension reduction approach was used to make feature transformations and to produce fewer features with better classification accuracy. In the end, we used an artificial neural network for the classification task. Classification accuracy when using eleven features results from PCA and MLP is 95% with 0.7 seconds to build the model.

ACKNOWLEDGEMENTS

We would like to extend our thanks to the management of following autistic child care centers for help with the collection of our data: Baghdad government center for the care of children with Autism spectrum, Iraqi Association for Psychological Therapy, Al- Brotherhood of Love Institute for Autism Spectrum Care, Al-Shams Center for Autistic Children Care, Al-Noor Center for Autistic Children Care, Al-Safa Center for Autistic Children Care, makers of Hope Institute for Autism Spectrum Care.

IV. REFERENCE

[1] American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders (5th Edition), 5th ed. Zagreb: Naklada Slap, Jastrebarsko, Croatia, 2013

[2] L. Smith et al., What is stimming, MedicalNewsToday, Feb. 2018. [Online]. Available: https://www.medicalnewstoday.com/articles/319714. Last Accessed: 17 May 2020.

[3] M. Rahman, “Beginning Microsoft Kinect for Windows SDK 2.0: Motion and Depth Sensing for Natural User Interfaces”, 1st ed. Montreal, Quebec, Canada: Apress, 2017.

[4] A. Jubouri and I. Ali, “A survey on movement analysis (hand, eye, body) and facial expressions-based diagnosis Autism spectrum disorders using microsoftkinect v2,” COMPUSOFT, An International Journal of advanced Computer Technology, vol. 9, Issue.1, pp. 3566-3577, Jan 2020. [Online]. Available: DOI: http://dx.doi.org/10.6084/ijact.v9i1.1045

[5] D. Rubin, "Statistical Matching Using File Concatenation with adjusted Weights and Multiple Imputations". Taylor & Francis: Journal of Business & Economic Statistics, 1986.doi:10.2307/1391390.

[6] R. Little, "Missing-Data adjustments in Large Surveys, “Journal of Business & Economic Statistics, vol.6, no.3, pp. 287–296, June 1988. [Online]: doi:10.2307/1391878.

[7] J. Foley, “Introduction to Computer Graphics,” California: addison-Wesley, 1994.

[8] Gartia, “Project 3: Camera Calibration and Fundamental Matrix Estimation with RANSAC,” Georgia Tech, GTID: 903136557. Accessed on: 16 Mar. 2020, Available: https://www.cc.gatech.edu/classes/AY2016/cs4476_fall/results/proj3/html/agartia3/index.html

[9] T. Motta et al.,” Kinect Projection Mapping,” SBC Journal on Interactive Systems, vol. 5, no. 3, 2014

[10] F. Berchtold and J. Hausen, “Homogeneous coordinates for algebraic varieties,” Journal of Algebra, vol.266, no.2,pp. 636–670, 2003. [Online]. Available: doi:10.1016/S0021-8693(03)00285-0.

[11] M. Karg et al., “Recognition of Affect Based on Gait Patterns,” IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), Vol. 40, Issue.4, Mar. 2010, pp. 1050 – 1061. Aug. 2010. [Online]. Available: DOI: 10.1109/TSMCB.2010.2044040.

[12] Q. Wen et al., “Time Series Data Augmentation for Deep Learning: A Survey,” arXiv preprint arXiv:2002.12478. 2020