# 3D-AD: 3D-Autism Dataset for repetitive behaviours with Kinect sensor

Omar RIHAWI, Djamal MERAD, Jean-luc DAMOISEAUX
Aix-Marseille University, LSIS-UMR CNRS 7296,
163 Avenue of Luminy, Cedex 9, Marseille 13288, France
First.Last AT univ-amu.fr

## Abstract

*Autism spectrum disorders (ASD) is a disorder that affects communication, social skills or behaviours of some people. Children or adults with ASD often have some common repetitive behaviours or self-stimulatory behaviours. These behaviours usually refer to specific behaviours such as flapping, rocking, spinning, etc. This work investigates these behaviours and provides a benchmark dataset for researchers. In our knowledge, this dataset is the first 3D-dataset available online[1] in the area of 3D recognitions of complex and repetitive behaviours of autistic people. The 3D-Autism Dataset (3D-AD) is captured with Kinect sensor. We explore different categories of autistic repetitive behaviours: static and dynamic ones, simple and complex ones. Experiments have been done using dynamic time warping to detect these behaviours.*

## 1. Introduction

Autism spectrum disorder (ASD) is a complex disorder that affects some people in their behaviours and their communication with others. This disease can be detected by the age of 2 or in some cases even earlier. However, the exact causes of autism are unknown until now. Any child may show any of the early warning signs of autism should be evaluated by a specialised physician to help him. The early professional care can make a large difference in preparing these autistic kids for safety life. In many experiences, a good early intervention program has made some differences in their behaviours. ASD people are often restricted, inflexible, and even obsessive in their behaviours, activities, and interests. Different symptoms can be shown between different people. They could have different tolerance levels for different behaviours [3].

In older ages, ASD behaviours can be recognized by: 1) Many repetitive behaviours such as: rocking, flapping, finger flicking, balling fists, manipulating objects. 2) Child is behind in physical/cognitive tasks. 3) Language delay (verbal and non-verbal) or inappropriate use of language (echolalia). 4) No eye contact or smile. 5) Inability to hold a conversation. 6) Inability to play with toys, and does not play creatively/imaginatively. 7) Irrational fears. 8) Seems out of control emotionally tantrums, screaming, unable to be comforted. 9) Refusing to respond. 10) Running away/isolating self, could have overstimulated, unsure, unconfident (feeling unsafe).

However, ASD people have some common repetitive and restricted behaviours [1], which are in our study the most important behaviours such as: rocking back and forth, stimming (abbreviated from self-stimulation, autistic people stimulate themselves) [11], spinning (or walking) in a circle, finger flicking, hand flapping, head banging, repetitive hand movements, toe walking, staring at lights, moving fingers in front of the eyes, snapping fingers, tapping ears, scratching, lining up toys, spinning objects, wheel spinning, watching moving objects, flicking light switches on and off, repeating words or repeating noises.

In this work, we explore some of these repetitive behaviours. Especially: *hands on the face* , *hands back* , *tapping ears* , *head banging* (or rocking back and forth) , *flicking* , *hands stimming* (repetitive hand movements) , *hand moving front of the face* , *toe walking* , *walking in circles* , *play with a toy* from/to different positions repetitively.

In the second section, we explore some of related works in the area of detection and recognition of human activities with 3D-depth images. The third section introduces the different categories of ASD behaviours, the feature extraction and the algorithm which has been used to detect ASD behaviours. The fourth and last section details experimentations on the benchmark dataset (3D-AD) and shows some results of the autistic behaviours recognition with dynamic time warping (DTW) method.

## 2. Related work

Computer utilities for autism have been recognized early on 70s [15]. In that time, researches on autism was not so widespread, because personal computers were not became

[1] See http://www.lsis.org/3d-ad/3D-AD.zip

978-1-5386-2939-0/17/$31.00 ©2017 European Union 

widely available to private individuals and researchers. Furthermore, in very recent years there have been quantum leaps in the development of computer-based audio-visual (multimedia) technology. According to [15] and until about 1981, the number of publications per year was rarely (1 per year). Then a discernible jump occurred to reach 57 publications per year. The number of computer applications and researches of autism on the Internet nowadays is continuing to be increased rapidly.

In [12], a research has been done on autistic vision domains and especially on gaze tracking and object detection. Other works, as [4] and [5], try to detect stereotyped behaviours (hand flapping) with Kinect sensor and a specific watch (eZ430-Chronos). Their approach uses *dynamic time wrapping* algorithm (DTW [17]) to recognize a hand flapping gesture. With their approach, the Kinect sensor can detect only $51\%$ number of times of these stereotypical movements. They use only hand coordination to identify the movements. That can make the result inefficient for other repetitive movements. Moreover, in [5] they test their platform with non-autistic people only.

In [8], authors propose a system that is able to detect five categories of gestures among autistic children: *body rocking*, *hand flapping*, *fingers flapping*, *hand on the face* and *hands behind back*. For these gestures they use only five types of joints: head, shoulder, elbow, wrist and hand in kinect's skeleton. $\$P$ Point-Cloud Recognizer method was used to recognize gestures. However, they do not have a benchmark dataset.

There are many researches on non-autistic actions recognition, and that provides many techniques and methods to be tested in autistic behaviours area later. Kinect has provided robust human skeleton tracking. With such these depth sensors, body part tracking has become high-level recognition tasks. Several approaches have built on it and more complex human actions can be detected. Algorithms such as Dynamic Time Warpping (DTW [17]) can be applied for tracking human actions. Many algorithms can be used with the skeleton data.

For example, in features extraction areas we can find many methods as a *bag of 3D points* in [9]. Authors propose a bag of 3D points from the depth maps, then clustering the points in order to obtain vocabulary postures. These vocabulary postures is used as a reference to all other postures, and then any movement could be represented by these reference postures. In [23], different *spatiotemporal interest points* (STIP)-based features are evaluated for human action recognition.

Different from joint locations features, there are various types of histogram based descriptors are proposed for action recognition by many researches. For example, histograms of 3D joint locations (HOJ3D) [21], histograms of oriented displacements (HOD) [6], histograms of oriented 4D normal (HON4D) [14].

A different features is proposed in [13], they use a sequence of the most informative joints (called SMIJ). Other work in [10], Pyramidal Motion Features (PMF) are used for human action recognition, sequence of 3D skeletal joints are computed as motion features. In order to be invariant, the descriptor use the relative position of joints. Another method, *Eigen-Joints* was proposed in [22], they use the properties of posture, motion, and offset of each frame based on 3D silhouettes to extract the features. Then, they employ the Naive-Bayes-Nearest-Neighbor classifier for multi-class action classification. In [7], the covariance matrix of skeletal joint locations over time is utilized as a discriminative descriptor for a sequence. [20] propose *Actionlet* ensemble model to represent the interaction of a subset of skeletal joints. Different from the methods which utilize skeletal features or statistical information of skeletal features for action recognition they use Fourier pyramid feature of each join [19].

In [2], authors use a weighted dynamic time warping on some joints' features to improve the results. This is a logical solution because some joints can move more than others in many human actions. Others like [16] represent a human model using a feature vector defined by 15 joints on a 3D human skeleton model. They use Dynamic Time Warping (DTW) with automatic feature weighing on each joint to achieve real-time action recognition. Similarly, [18] also used DTW. The features based on relative angles among body parts.

## 3. ASD repetitive behaviours

As we mention before, children with ASD are often restricted, inflexible, and even obsessive in their behaviours, activities, and interests. Different self-stimulatory behaviours can be shown by different children. However, there are some of common behaviours could clearly be seeing such as: flapping, rocking, spinning, etc.

We plan to detect some of these repetitive behaviours: *hands on the face* , *hands back* , *tapping ears* , *head banging* (or rocking back and forth) , *flicking* , *hands stimming* (repetitive hand movements) , *hand moving front of the face* , *toe walking* , *walking in circles* , *play with a toy* from/to different positions repetitively.

We divide these repetitive behaviours into two main categories according to their movements:

1. Static behaviours: *hands on the face*, *hands back* and *tapping ears*. Actually, in this type of behaviours there are no real movement in the middle of the action or we call it static behaviours.

2. Dynamic behaviours: where they can be done with some movements during a period of time, and these behaviours can be divided again into two categories:

- Simple behaviours: rocking back and forth (or *head banging*), *flicking*, repetitive hand movements (or *hands stimming*), *hand moving front of the face*.
- Complex behaviours: *toe walking*, *walking in circles*, *play with a toy* from/to different positions repetitively.

### 3.1. Features extraction

Human actions normally come from body movements, or changes in joints' positions and angles. Different children have different sizes, and that can make same actions hard to be detected with different scales and rotations. For that, it is necessary to normalize the body size and rotate all body parts to the same reference direction. Then the detection should be invariant to scale and rotation.

Each joint in the Kinect's skeleton has three coordinates (x,y,z) refer to Kinect position (0,0,0). In order to be invariant, we can use angles only between relative joints (joints and their parent-joints). That means, there is no scale information and it should be invariant between different children.

As we mention before, we have two categories of autistic behaviours. Static behaviours such as *hands on the face*, *hands back*, etc. Dynamic behaviours such as rocking, flicking etc. In static behaviours, we can easily detect these behaviours by computing the distance between some joints. For example in *hands on the face*, the distance between hands and head can be the same even with different scales and rotations. But, such of these methods could not work well with dynamic behaviours. Where there are a lot of movements in body parts, and the body could be in different positions from Kinect device. However the behaviours are static or dynamic, we plan to detect both kinds of behaviours with a robust method to any transformation or in other words invariant to scale and rotation.

Features are built form all skeleton joints. Each joint provides two features $\theta_1$ & $\theta_2$ which are the euclidean angles in 3D space between YZ and ZX for each frame (see figure 1.A):

$$\theta_1 = arctan(\frac{Y_{joint} - Y_{ParentJoint}}{Z_{joint} - Z_{ParentJoint}}) \quad (1)$$

$$\theta_2 = arctan(\frac{Z_{joint} - Z_{ParentJoint}}{X_{joint} - X_{ParentJoint}}) \quad (2)$$

All joints are already oriented to Torso direction to get rid of the rotation. If children are not standing in front of the Kinect, his torso should make an angle with the Z-axis (Z from Kinect device to the children area). Then we can apply a rotation around Y-axis to get rid of torso angle which is:

$$\theta_{Torso} = arctan(\frac{Z_{RightShoulder} - Z_{LeftShoulder}}{X_{LeftShoulder} - X_{RightShoulder}}) \quad (3)$$

Then we can rotate the torso along Y-axis (Y from Kinect device to up) by $180 - \theta_{Torso}$ because the Z also is inverted (from Kinect to the object). These features are invariant to the scale (children size) and the camera position as well, that because we chose only two features $\theta_1$ & $\theta_2$ (angles between joints), that means there is no scale information in each feature. This method should be invariant between different children. Then, an action could be a set of frames or a set of features from all joints.

### 3.2. Action detection

We use *dynamic time warping* (DTW) method [17] to detect a repetitive action. We consider that a repetitive action is a re-occurrence of the same signal of $\theta_1$ & $\theta_2$ for each skeleton joint (see figure 2).

To detect a match between a query and a template signals we need to compute a threshold of the accepted value (accepted error) in DTW-distance. Both signals (query and template) are sets of $\theta$s values, each value between 0 and 360 degrees. Each one comes from a state of body skeleton joints. If we consider that the $Sim_i$ (the similarity factor of $\theta_i$) is the accepted error value for $\theta_i$ of each joint. Then, the skeletons sequence of $W$ size (window size of the query) should have an accepted error value of $W \times Sim_i$ for $\theta_i$. Then, the global threshold of accepted DTW-distance can be calculated from this:

$$threshold = \sum_{i=1}^{n} W \times Sim_i \quad (4)$$

where $n$ is the number of $\theta$s values. In our case $n = 2\times$ *Number of joints*, as we have two angles $\theta_1$ & $\theta_2$ for each joint.

Now, if we accept a difference in one degree only ($\pm1$) of each joint as an accepted error value. Then, the $Sim_i = 1$ for each state, and $\sum_{i=1}^{n} W \times 1 = n \times W$ should be the global accepted threshold.

### 3.3. Algorithm

A general framework of our system can be with these steps:

- First, get skeleton joints from kinect for each frame.
- Extract all joints' features from skeleton joints and add them to actual sequence history.
- If there are no changes in features' values from $\epsilon$ time ($\epsilon$ can be defined by user), we try to detect static behaviours.
- Else try to detect a repetitive action by compare DTW-distances from the actual time to the passed time in the history. A repetitive means the same sequence of

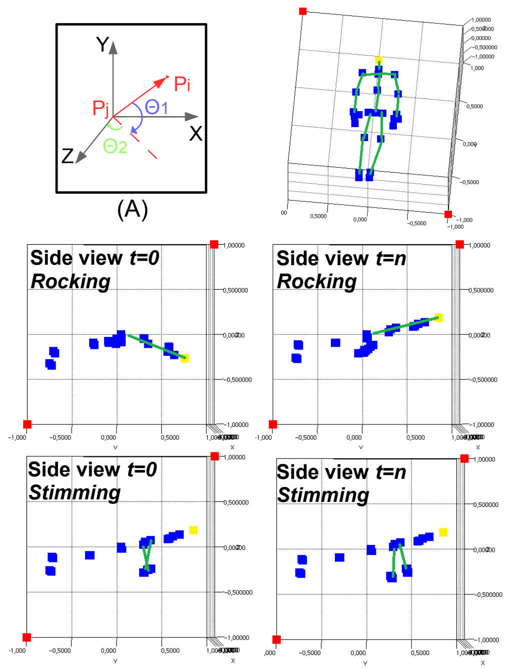


Figure 1. ASD behaviours: rocking and stimming. (A) $\theta_1$ & $\theta_2$ features for each skeleton joint Pi to its parent Pj.

features more than 2 times. Which should be computed by searching the best match between two time-slices (see figure 2 for an example):

1. Each $t_0$ time, we compute DTW-distance from two time-slices A and B. Time-slices A begin from $t_0$ to $t_n$ and B from $t_n$ to $t_m$.
2. Change the length of time slices A & B (change n and m), and recompute DTW-distance.

- Once we have a best match, we should now determine the total number and the global period of this behaviour through: 1) Wait until detect one more repetitive in the future, or wait to fill a new time-slice C (C between $t_0$ and $t_p$). $t_p$ from the future and should be between n and m. 2) Each new detection of the same repetitive should add 1 to the total number of this behaviour and reset $t_0 = t_p$. 3) Still try to detect a new C time-slice until we reach a new $t_p$ that is $> max(n, m)$ and DTW-distance always is $> threshold$. Then the system should end the period of this behaviour, which is from $t_m$ to last $t_p$ ( last $t_0$).

- Finish and return to first step.

## 4. Experiments

First step in any experiment is collecting the dataset. We plan to create a first collection of 3D autistic repetitive be-

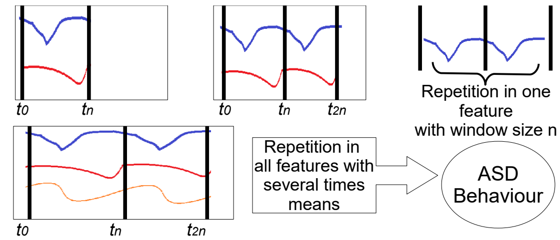


Figure 2. Repetitive behaviours detector: two windows with the same time-slices (size = N)

haviours (3D-AD) benchmark dataset.

### 4.1. 3D-AD: 3D autism dataset

Since there is no 3D benchmark dataset until now with depth videos templates of autistic behaviours that can be available online, we collect our dataset and make it available online for the first time in the literature ( http://www.lsis.org/3d-ad/3D-AD.zip ).

Our dataset contains these actions (from simple to complex): *hands on the face*, *hands back*, *tapping ears*, *head banging* (or rocking back and forth), *flicking*, *hands stimming*, *hand moving front of the face*, *toe walking*, *walking in circles* and *play with a toy* from/to different positions repetitively.

Each action has been repeated for at least 10 times with non-autistic people. The depth maps have been captured at rate 33 frames per second with Kinect-v2 camera.

### 4.2. Detection algorithm

There are many methods in literature for 3D human actions detection. Many researchers have used dynamic time warping to compare a similarity between different actions. We use Kinectv2 device to create a sequence of skeleton joints features. Then, we use dynamic time warping distance to detect a similarity between two actions with a accepted threshold. First step is extracting all joints' features from skeleton joints for all autistic behaviours. Then, if there are no changes in features' values for $\epsilon$ time (*e.g.* $\epsilon$=1 sec.), we should detect static behaviours (*hands on the face*, *hands back* and *tapping ears*). If there are any change in features' values, then we should detect any dynamic behaviour (*head banging*, *flicking*, *stimming*, *hand moving front of the face*, *toe walking*, *walking in circles* or *play with a toy*).

### 4.3. Discussion

Repetitive behaviours are divided into two categories according to their relations to the movement: Static and dynamic behaviours. We compare the results of each category with different queries (Qs) and templates (Ts). These Qs

| Static Behaviours | Hands back T | Hands on face T | Tapping ears |
|---|---|---|---|
| Hands back Q | $< 0.002$ | $> 0.012$ | $> 0.013$ |
| Hands on face Q | $> 0.012$ | $< 0.004$ | $> 0.008$ |
| Tapping ears Q | $> 0.013$ | $> 0.008$ | $< 0.003$ |

Table 1. DTW-distance between Q (query) and T (template) of static behaviours

and Ts have been randomly chosen from each behaviour's set.

Table 1 shows the confusion-matrix of different static behaviours. These behaviours are easy to be detected because they are static. DTW-distance between same action are less than $0.004$, and we can easily distinguish between different actions at the threshold= $0.008$ which is the smaller difference between *hands on face* and *tapping ears* behaviours. This similarity comes from that these two actions have many joints with similar positions and then similar features (*e.g*. shoulder and elbow). To improve the detection in the future, a weighted-DTW method should show more clear differences between these actions.

For dynamic behaviours, we distinguish between simple behaviours in table 2 and complex behaviours in table 3. In table 2, *Hands stimming* gives us the smallest DTW-distance by 0.001, then *head banging* with 0.002, *Hands front of face* 0.0025 and *flicking* with worst distance by 0.004. *Flicking* also gives a similarity with other actions (*stimming*, *front of face*). In fact, autistic people can make flicking with different positions of hands. That could cause the similarity with other actions.

Table 3 shows differences of DTW-distance between more complex autistic behaviours: *Toe walking*, *Walking in circles* and *play with a toy*. *Toe walking* gives best detection. However, the DTW-distance between these actions shows that they are very close to each other (DTW-distance less than 0.002). These actions are hard to be detected because it is hard to be captured with one Kinect device. For example, in *walking in circles* many parts of the skeleton body could be behind the body and as a consequence hard to be detected. The solution should be in these cases by adding two/three Kinect devices to keep capture all bodies parts.

In table 4, we compare between the number of repetitive non-static behaviours that have been detected and the real number of these behaviours. Best results for *stimming* (26/27) and *hand move front of the face* (19/18).

| Non-Static | Nb of RD / Real Nb |
|---|---|
| Head Banging | 33 / 49 |
| Flicking | 21 / 30 |
| Hands stimming | 27 / 26 |
| Hands front face | 19 / 18 |
| Toe walking | 45 / 63 |
| Walking in circles | 44 / 49 |
| Play with a toy | 24 / 17 |

Table 4. The number of detected repetitive behaviours / the total number of real repetitive behaviours.

## 5. Conclusions

A system, that is able to detect autistic behaviours from early ages, can help many autistic people. It can aim these people to be early evaluated by a professional, and then they can quickly get the necessary aids to be held from some of their behaviours. Examples of these autistic behaviours are rocking, flapping, spinning, etc. Most of these behaviours are repetitive and common between most of autistic children. This paper investigates these behaviours and provides a benchmark dataset available online 3D-AD[2]. In our knowledge, this dataset is the first benchmark dataset available in the area of 3d detections and recognitions of complex repetitive behaviours with Kinect sensor for ASD people. In 3D-AD, We divide repetitive behaviours into two categories according to their body movements: Static and dynamic behaviours. Experiments using dynamic time warping show a good recognition of static behaviours and some of dynamic ones. In the future, we plan to investigate different recognition methods on our dataset to improve the detection process.

## References


[1] I. Carcani-Rathwell, S. Rabe-Hasketh, and P. J. Santosh. Repetitive and stereotyped behaviours in pervasive developmental disorders. *Journal of Child Psychology and Psychiatry*, 47(6):573–581, 2006.

[2] S. Celebi, A. S. Aydin, T. T. Temiz, and T. Arici. Gesture Recognition Using Skeleton Data with Weighted Dynamic Time Warping. In *Computer Vision Theory and Applications*. Visapp, 2013.

[3] J. Clements and E. Zarkowska. *Behavioural Concerns and Autistic Spectrum Disorders*. Jessica Kingsley, 2000.

[4] N. Gonalves, J. L. Rodrigues, S. Costa, and F. Soares. Automatic detection of stereotyped hand flapping movements: Two different approaches. In *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*, pages 392–397, Sept 2012.

[5] N. Gonalves, J. L. Rodrigues, S. Costa, and F. Soares. Preliminary study on determining stereotypical motor move-


[2] See http://www.lsis.org/3d-ad/3D-AD.zip

| Simple Behaviours | head banging T | Flicking T | Hands stimming T | Hands front of face T |
|---|---|---|---|---|
| head banging Q | $< 0.002$ | $> 0.013$ | $> 0.01$ | $> 0.01$ |
| Flicking Q | $> 0.013$ | $< 0.004$ | $> 0.0061$ | $> 0.006$ |
| Hands stimming Q | $> 0.01$ | $> 0.0061$ | $< 0.001$ | $> 0.006$ |
| Hands front of face Q | $> 0.01$ | $> 0.006$ | $> 0.006$ | $< 0.0025$ |

Table 2. DTW-distance between Q (query) and T (template) of simple stimming behaviours

| Complex behaviours | Toe walking T | Walking in circles T | Play with a toy T |
|---|---|---|---|
| Toe walking Q | $< 0.0009$ | $> 0.004$ | $> 0.002$ |
| Walking in circles Q | $> 0.004$ | $< 0.0014$ | $> 0.002$ |
| Play with a toy Q | $> 0.002$ | $> 0.002$ | $< 0.0015$ |

Table 3. DTW-distance between Q (query) and T (template) of of complex behaviours


ments. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1598–1601, Aug 2012.

[6] M. A. Gowayyed, M. Torki, M. E. Hussein, and M. El-Saban. Histogram of oriented displacements (hod): Describing trajectories of human joints for action recognition. In *Proceedings of the 13 Int. Conf. on Artificial Intelligence*, IJCAI '13, pages 1351–1357. AAAI Press, 2013.

[7] M. E. Hussein, M. Torki, M. A. Gowayyed, and M. El-Saban. Human action recognition using a temporal hierarchy of covariance descriptors on 3d joint locations. In *Proceedings of the 13International Joint Conference on Artificial Intelligence*, IJCAI '13, pages 2466–2472. AAAI Press, 2013.

[8] M. Jazouli, A. Majda, D. Merad, R. Aalouane, and A. Zarghili. Automatic detection of stereotyped movements in autistic children using the kinect sensor. *Biomedical Engineering and Technology*, apr 2017.

[9] W. Li, Z. Zhang, and Z. Liu. Action recognition based on a bag of 3d points. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, pages 9–14, June 2010.

[10] G. LU, Y. ZHOU, X. LI, and C. LV. *Action Recognition by Extracting Pyramidal Motion Features from Skeleton Sequences*, pages 251–258. Springer Berlin Heidelberg, Berlin, Heidelberg, 2015.

[11] W. Napier and R. Bergemann. *An Asperger Leader's Guide to Living and Leading Change*. Asperger's Employment Skills Guides. Jessica Kingsley Publishers, 2013.

[12] B. Noris. Machine vision-based analysis of gaze and visual context: an application to visual behavior of children with autism spectrum disorders. PhD Thesis, 2011.

[13] F. Ofli, R. Chaudhry, G. Kurillo, R. Vidal, and R. Bajcsy. Sequence of the most informative joints (smij): A new representation for human skeletal action recognition. *Journal of Visual Communication and Image Representation*, 25(1):24 – 38, 2014. Visual Understanding and Applications with RGB-D Cameras.

[14] O. Oreifej and Z. Liu. Hon4d: Histogram of oriented 4d normals for activity recognition from depth sequences. In *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pages 716–723, June 2013.

[15] B. O. Ploog, A. Scharf, D. Nelson, and P. J. Brooks. Use of computer-assisted technologies (cat) to enhance social, communicative, and language development in children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 43(2):301–322, 2013.

[16] M. Reyes, G. Domnguez, and S. Escalera. Feature weighting in dynamic time warping for gesture recognition in depth data. In *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pages 1182–1188, Nov 2011.

[17] H. Sakoe and S. Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26(1):43–49, Feb 1978.

[18] S. Sempena, N. U. Maulidevi, and P. R. Aryan. Human action recognition using dynamic time warping. In *Proc. of 11Int. Conf. on Electrical Engineering and Informatics*, pages 1–5, July 2011.

[19] R. Vemulapalli, F. Arrate, and R. Chellappa. R3dg features: Relative 3d geometry-based skeletal representations for human action recognition. *Computer Vision and Image Understanding*, 152:155 – 166, 2016.

[20] J. Wang, Z. Liu, Y. Wu, and J. Yuan. Learning actionlet ensemble for 3d human action recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(5):914–927, May 2014.

[21] L. Xia, C. C. Chen, and J. K. Aggarwal. View invariant human action recognition using histograms of 3d joints. In *IEEE Conf. on Computer Vision and Pattern Recognition Workshops*, pages 20–27, June 2012.

[22] X. Yang and Y. Tian. Effective 3d action recognition using eigenjoints. *J. Vis. Comun. Image Represent.*, 25(1):2–11, Jan. 2014.

[23] Y. Zhu, W. Chen, and G. Guo. Evaluating spatiotemporal interest point features for depth-based action recognition. *Image and Vision Computing*, 32(8):453 – 464, 2014.