

Preliminary Study on determining stereotypical motor movements

Nuno Gonçalves, José L. Rodrigues, Sandra Costa, Filomena Soares, *Member, IEEE*

Abstract— Stereotypical motor movements are one of the most common and least understood behaviors occurring in individuals with Autism Spectrum Disorder (ASD). To overcome problems with traditional methods for measuring stereotypical motor movements in persons with ASD, the Kinect sensor from Microsoft and gesture recognition algorithms were used to automatically detect hand flapping. This study provides a valuable tool to monitor stereotypes in order to understand and to cope with this problem. At the end it facilitates to identify behavioral patterns especially relevant when studying interaction skills in children with ASD.

Keywords - Stereotypical motor movements, Kinect Sensor, gesture recognition, ASD

I. INTRODUCTION

Autism Spectrum Disorders (ASD) affect as many as 1 in 150 children[1] and typically manifests itself during the first three years of life, being defined as a global development disorder [2-5]. Stereotyped behaviors, typically made by people with ASD, are defined as repetitive interests and/or motor or vocal sequences that appear to the observer to be invariant in form and without any obvious eliciting stimulus or adaptive function [6].

Typically, autistic children have stereotypical behavior in particular when there is a change in their daily routine or when the environment is full of new perceptions; as a consequence, they tend to block the excess of sensations, manifesting repetitive movements. Therapists measure this behavior using traditional methods such as paper-pencil rating scales, direct observation or video-based methods that can be inaccurate and time consuming.

Recently, systems have been developed to automatically detect stereotypical behaviors using accelerometers and pattern recognition algorithms [7-9]. The results have been promising, but it is necessary to worn the sensors on the wrists and torso using elastic armbands, a situation that may not be well accepted by the child. The Kinect sensor could

suppress the disadvantages referred, as the sensor is based on the concept of Human-Device interaction through the user's senses, especially hearing and vision. The sensor in Figure 1 contains a RGB color camera, a depth sensor and multi microphone array which allows capturing body motion in 3D, gesture and face recognition, obtaining the users joints coordinates. The depth sensor is an infrared laser in combination with a CMOS monochrome sensor which enables the Kinect device processing 3D scenarios in any ambient lighting.

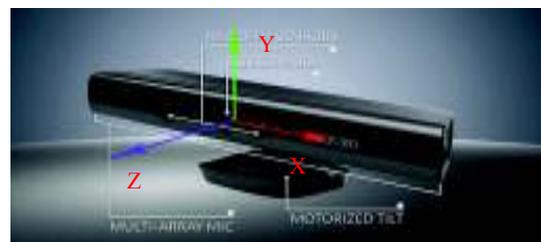


Figure 1 – Constitution of the Kinect [10]

The work presented in this paper is part of a collaborative project [11-13] between the University of Minho, APPACDM (an association for mentally disabled people) of Braga and a group of schools with Special Education units, where the main objective is to improve the social life of children with autism spectrum disorders, in particular to promoting their social interaction and communication, by using a robotic platform as mediator. The experimental sessions are recorded on video for further analysis in terms of pre-defined indicators. By using the Kinect sensor based methodology in real time, during the intervention sessions may enable to analyze the amount/intensity of stereotypies performed, which can be an indicator of the child's emotional state.

Thus the main objective of this work is to understand if the Kinect sensor from Microsoft with gesture recognition algorithms can be used as a time-efficient tool to automatic detect in real-time the stereotypical behavior of children with ASD. This will allow characterizing children behavior during the interventions sessions, when they are exposed to new situations.

The paper is divided in four sections: Introduction, where it is presented the paper focus and motivation; Section II is dedicated to the dynamic time warping algorithm; Section III presents the stereotyped behavior detection based on the Kinect sensor, as well as the experimental approach; Results obtained in laboratory environment are presented in Section IV. Finally, the conclusions and further work are detailed in the last section.

* This work was supported in part by the Portuguese Foundation for Science and Technology under the Project Contract RIPD/ADA/109407/2009.

Nuno Gonçalves is a Researcher in University of Minho. He is with the R&D Centre Algoritmi, Campus de Azurém, 4800-058 Guimarães, Portugal (email: nuno.goncalves@algoritmi.uminho.pt)

José L. Rodrigues is a Researcher in University of Minho. He is with the R&D Centre Algoritmi, Campus de Azurém, 4800-058 Guimarães, Portugal (email: Jose.rodrigues@algoritmi.uminho.pt)

Sandra Costa is a PhD Student in University of Minho (SFRH/BD/71600/2010). She is with the R&D Centre Algoritmi, Campus de Azurém, 4800-058 Guimarães, Portugal (e-mail: sandra.costa@algoritmi.uminho.pt).

Filomena Soares is with University of Minho. She is with the R&D Centre Algoritmi, Campus de Azurém, 4800-058 Guimarães, Portugal (e-mail: filomena.soares@algoritmi.uminho.pt)

II. DYNAMIC TIME WARPING ALGORITHM

In this work it was used the Dynamic Time warping (DTW) algorithm [14-17]. This algorithm is based on a dynamic programming that measures the similarity between patterns which may vary with time of different durations. The main concept of the algorithm is to compare characteristics of a particular pattern to be checked against a template reference previously registered.

The algorithm has been used in voice recognition patterns, signature verification or gesture recognition [17, 18].

The final objective of the DTW algorithm is to compare a pattern standard with an input pattern that we want to be recognized. The algorithm has to yield a match result between the two patterns. At the end a measure of overall distance/similarity between the two patterns must be generated. The smaller the distance, more similar the patterns will be.

Assuming that we have two time sequences, X and Y, of length n and m respectively, where $X = \{x_1, \dots, x_n\}$ and $Y = \{y_1, \dots, y_m\}$ and the objective is to compute the matching cost matrix: $DTW(X, Y)$.

The two sequences can be arranged on the sides of a matrix, with one sequence placed on the top and the other is located up to the left hand side, Figure 2; both sequences start at the bottom left side of the matrix.

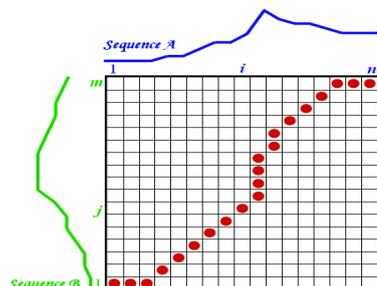


Figure 2 – Example of the matching cost matrix [14]

To find the best match or alignment in the matching cost matrix one needs to find a path, also known as the warping path, through the grid which minimizes the total distance. The procedure involves finding all possible routes through the matrix.

The major optimizations to the DTW algorithm arise from the necessity of a good path through the grid. These are outlined in Sakoe and Chiba [15] and can be summarized as follows:

1. Boundary condition: the path has to start at the bottom left and ends at the top right;
2. Monotonic condition: the path cannot turn back on itself; both i and j indexes either stay the same or increase, they never decrease;
3. Continuity condition: the path advances one step at time. Both i and j can only increase by at most 1 on each step along the path;

4. Slope constraint condition: the path should not be too steep or too shallow. This prevents very short sequences matching very long ones.

Figure 3 shows an example an admissible warping path satisfying the conditions.

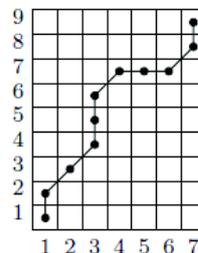


Figure 3 – Example of the warping path [16]

The matching cost is computed, based on dynamic programming presented, as:

$$D(i, j) = d(x_i, y_j) + \min \begin{cases} D(i, j - 1) \\ D(i - 1, j) \\ D(i - 1, j - 1) \end{cases} \quad (1)$$

where the distance function $d(x_i, y_j)$, also known as the Euclidian Distance, is defined as

$$d(x_i, y_j) = \sqrt{(x_i - y_j)^2} \quad (2)$$

where x_i and y_j are the coordinates of point i

After computing the matching cost matrix the DTW distance is calculated by dividing the sum of the warping path by the length of the standard pattern.

III. STEREOTYPED BEHAVIOR DETECTION

In this section it is explained how the application detects a stereotyped behavior using the Kinect sensor and the DTW algorithm

A. Kinect calibration

The first step to be carried out with the Kinect is the extraction of the coordinates of the selected joints, the joints of the hands, wrists and head. Once the coordinates are referenced to the sensor, ie, the Kinect sensor is $O(0,0,0)$ and the gesture recognition algorithm does not depend on the user's height or where he/ she is sitting, it is necessary to convert the user as the new reference. For convenience it was chosen the shoulders of the user, as the new reference coordinate $O(0,0,0)$, Figure 4.

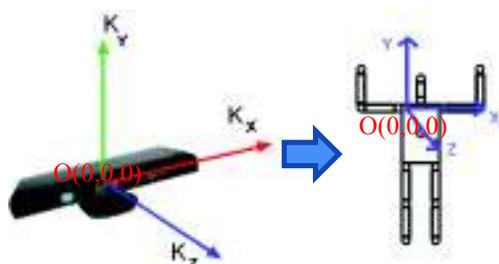


Figure 4 – Illustration of the converted referential

B. Application description

Figure 5 shows the interface application developed.

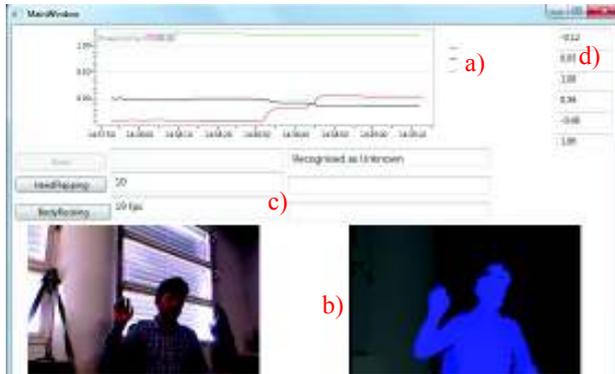


Figure 5 – Interface developed

The interface is segmented into four different sectors: a) presents a graph that keeps track, in this case, of the left hand position (X and Y coordinates), characterized by the black and red lines and represented by the green line is the Z coordinate of the head; sector b) shows the RGB and Depth video obtained by the Kinect sensor; sector c) presents the display area of the results; sector d) display the hands positions (X,Y,Z coordinates) at the current state.

After the application is initialized it is necessary to wait until the Kinect detects a user (represented by the blue color in the depth image, Figure 5). When the user is detected, it has to record the hand flapping movement to serve as a template sequence to the DTW algorithm. By pressing the *Hand flapping* bottom, the gesture is recorded for twenty frames, Figure 6. The value of twenty frames was chosen because it demonstrated as a sufficient value for a good template sequence. When the recording ends the application goes to the reading mode, ready to identify the stereotypes.

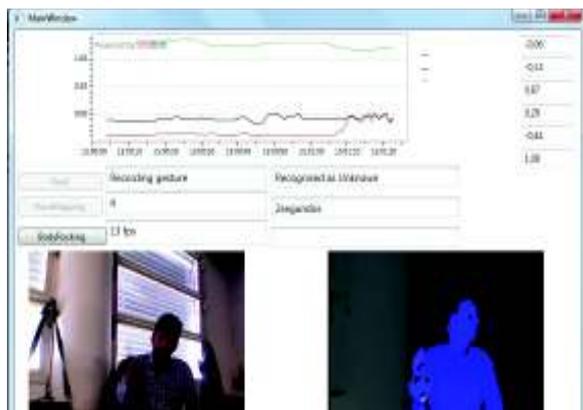


Figure 6 – Recording the hand flapping

The sampling period (sample sequence) chosen for the application in the reading mode was five frames. This value was chosen by trial and error. The application at every five frames stores the hand coordinates (only the Y coordinate was considered relevant) in an array and through the DTW algorithm it is verified if there is a similarity with the template sequence of the hand flapping. If a match is found it is considered a stereotypical movement and starts to count

the time duration of the hand flapping. To avoid computational burden, the calculated distance between the coordinates of the template sequence with the coordinates of the sample sequence must be smaller than an established threshold.

Figure 7 shows an example stereotypical movement detected by the application.

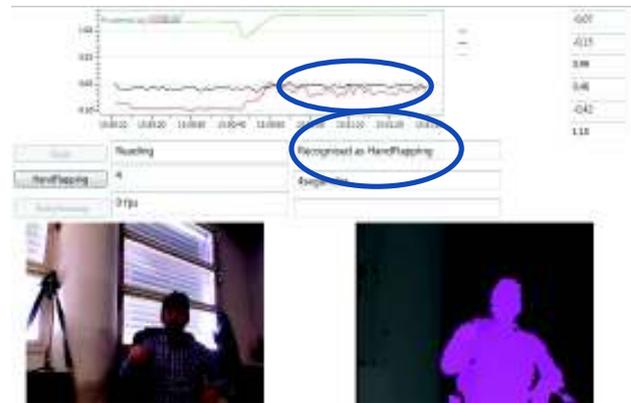


Figure 7 – Hand flapping detection

IV. RESULTS

For this preliminary study it was only considered the hand flapping as a stereotyped movement in a laboratory environment with two persons, subject A and subject B, without any pathology associated.

It is worth mention that only the subject A recorded the gesture that served as the template sequence for these trials. It was recorded the number of occurrences of the stereotyped movement and the duration of the stereotyped movement. Both subjects simulated four times, during 5 minutes in total, a stereotypy behavior with different durations.

Tables 1 and 2 show the results obtained for subject A and B, respectively, comparing the real movements performed by each subject to the corresponding movement detected by the developed application. The comparison is performed in terms of:

- number of times that the subjects performed stereotyped movement (n_r);
- time duration of the performed stereotyped movement (r_t);
- number of times that the application detected stereotyped behavior (n_d);
- time duration of the detected stereotyped behavior (t_d);
- false-positives (f-p)
- time duration of the false-positives (t_{f-p}).

As it can be observed in Tables 1 and 2, for both subjects, every time the stereotypes were simulated the application successfully detected them. It is worth mention that the time duration of the simulated stereotyped behavior and the corresponding stereotypy detected by the application were similar. With subject B, it was detected a stereotypy lasting 15 seconds when in fact it only lasted eight seconds. This happened because the application previously detected a false-positive. In some cases, stereotypes which lasted more

than ten seconds the application considers as more than one stereotyped behavior.

Table 1 – Results obtained with subject A

	n_r	r_t (seconds)	n_d	t_d (seconds)	f-p	t_f-p (seconds)
Subject A	1	6	1	5	no	0
	1	8	1	7	no	0
	1	17	1	15	no	0
	0	0	1	7	yes	7
	0	0	1	1	yes	1
	0	0	1	1	yes	1
	1	28	2	20 (11+9)	no	
Total:	4	59	8	56	3	9

Table 2 – Results obtained with subject B

	n_r	r_t (seconds)	n_d	t_d (seconds)	f-p	t_f-p (seconds)
Subject B	0	0	1	7	yes	7
	1	8	1	15	no	0
	1	18	2	21 (7+14)	no	0
	1	20	1	25	no	0
	0	0	1	5	yes	5
	1	50	3	34 (12+14+8)	no	0
Total:	4	96	9	107	2	12

V. CONCLUSION AND FUTURE WORK

In this paper we presented a new system based on the Kinect sensor to automatic detect stereotyped behaviors, in particular the hand flapping. We applied the DTW algorithm to recognize the hand flapping gesture. The system was tested in a laboratory environment with non ASD persons. Although the results showed to be promising, some aspects need to be improved, namely the noise of the depth image that can lead to false-positives in the identification, and improve the accuracy of the application when the user sits at extremes distance, or too close to the Kinect sensor.

As future work, it is intended to add the body rocking gesture recognition, eliminating the image noise and finally the testing with ASD children during the intervention sessions with the robot as the mediator.

ACKNOWLEDGMENT

The authors are also grateful to the Portuguese Foundation for Science and Technology, FCT- in Portuguese *Fundação para a Ciência e a Tecnologia*, for funding through the R&D project reference RIPD/ADA/109407/2009. This work is supported by a QREN initiative, from UE/FEDER (in Portuguese *Fundo Europeu de Desenvolvimento Regional*) funds through the "Programa Operacional Factores de Competitividade – COMPETE".

REFERENCES

- [1] U.S. CDC, "Prevalence of Autism Spectrum Disorders - Autism and Developmental Disabilities.
- [2] Schopler, E., G. B. Mesibow (Eds), (1987). Neurobiological issues in autism, en Current issues in autism (series). New York: Plenum Press H. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1985, ch. 4.
- [3] Giellberg, C. & M. Coleman (Eds) (1992). The biology of the autistic syndromes, 2nd ed., Londres: Mackeith.
- [4] Happé, F. (1994), *Introducción al autismo*. Trad. María Núñez Bernardos (1998). Madrid: Alianza Editorial, S.A.
- [5] A. P. A., *Diagnostic and Statistical Manual on Mental Disorders*, 4th Ed., vol. IV-TR. Washington, DC: Amer. Psychiatric Publishing, 2000.
- [6] A. Baumeister and R. Forehand, "Stereotyped acts," in *Int'l. Rev. of Res. in Mental Retardation*: VI., Ed. New York, 1973, pp. 55-96.
- [7] Fahd Albinali, Mathew S. Goodwin, Stephen S.Intille, "Recognizing Stereotypical Motor Movements in the Laboratory and Classroom: A case Study with Children on the Autism Spectrum" ,in *UbiComp 2009*, September 2009, Orlando, Florida, USA.
- [8] Cheol-Hong, Ahmed H. Tewfik, "Automatic Characterization and Detection of Behavioral Patterns using Linear Predictive Coding of Accelerometer Sensor Data", in *IEEE EMBS 2010, Engineering in Medicine and Biology Society*, September 2010, Buenos Aires, Argentina.
- [9] Fahd Albinali, Mathew S. Goodwin, Stephen S.Intille, Wayne F. Velicer, "Automated Detection of Stereotypical Motor Movements", *Journal of Autism and Developmental Disorders*, Springer Ed, Netherlands, Vol. 41, Issue 6, pp. 770-782.
- [10] Valentino Frati, Domenico Prattichizzo – "Using Kinect for Hand Tracking and rendering in wearable haptics" – *IEEE World Haptics Conference 2011*, Istanbul, Turkey
- [11] Sandra Costa, Jorge Resende, Filomena Oliveira Soares, Manuel João Ferreira, Cristina P. Santos, Fátima Moreira, Applications of simple robots to encourage social receptiveness of adolescents with autism, in *EMBC 2009, Engineering in Medicine and Biology Conference*, September 2009, Minneapolis, Minnesota, EUA.
- [12] Sandra Costa, Cristina Santos, Filomena Soares, Manuel João Ferreira e Fátima Moreira, Promoting the interaction amongst autistic adolescents using robots, in *EMBC 2010, Engineering in Medicine and Biology Conference*, September 2010, Buenos Aires, Argentina.
- [13] Sandra Costa, Filomena Soares, Cristina Santos, Manuel J. Ferreira, Fátima Moreira, Ana Paula Pereira, Fernanda Cunha, An approach to promote social and communication behaviors in children with Autism Spectrum Disorders: Robot based intervention, in *Ro-Man 2011, 20th IEEE International Symposium on Robot and Human Interactive Communication*, August 2011, Atlanta, Georgia, EUA.
- [14] <http://www.psb.ugent.be/cbd/papers/gentxwarper/DTWalgorithm.htm> (accessed March 2012)
- [15] Müller, Meinard "Information Retrieval for Music and Motion", New York: Springer-Verlag, 1985, chapter 4
- [16] <http://kinectdtw.codeplex.com/> (accessed March 2012)
- [17] Sakoe and Chiba, "Dynamic Programming Algorithm Optimization for Spoken Word Recognition", *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. ASSP-26, NO. 1, February 1978.
- [18] Cory Myers, Lawrence R. Rabiner and Aaron E. Rosenberg, "Performance Tradeoffs in Dynamic Time Warping Algorithms for Isolated Word Recognition", *IEEE Transactions on Acoustics, Speech and Signal Processing*, Volume 28, Issue 6, 1980, pp: 623 – 635.