

Detection of stereotyped hand flapping movements in Autistic Children using the Kinect Sensor: a Case Study

Nuno Gonçalves, Sandra Costa, José Rodrigues
R&D Centre Algoritmi
University of Minho
Campus de Azurém, 4800-058 Guimarães, Portugal

Filomena Soares
R&D Centre Algoritmi
University of Minho
Campus de Azurém, 4800-058 Guimarães, Portugal
filomena.soares@algoritmi.uminho.pt

Abstract— This paper presents a case study on the use of the Microsoft Kinect sensor and the gesture recognition algorithm, Dynamic Time Warping (DTW) to automatically detect stereotyped motor behaviours in children with Autism Spectrum Disorder (ASD). The proposed system was initially tested in a laboratory environment, and after the encouraging results obtained, the system was tested during the sessions with children with ASD in two schools with Special Needs Units. Video analysis was used to validate the results obtained using the Kinect application. This study provides a valuable tool to monitor stereotypes in order to understand and to cope with this problematic. In the end, it facilitates the identification of relevant behavioural patterns when studying interaction skills in children with ASD.

Keywords— *Stereotypical Motor Movements, Kinect Sensor, Gesture Recognition, Autism Spectrum Disorders.*

I. INTRODUCTION

Autism is a developmental disability that typically appears during the first three years of life, being defined as a global development disorder [1]-[4]. The symptoms and characteristics of autism spectrum disorders (ASD) are a result of a neurological disorder that affects the normal development of the brain in the areas of social interaction, communication skills, and emotion recognition. Children and adults with ASD have difficulties in verbal and non-verbal communication in social interactions [5], [6].

Typically, children with ASD manifest stereotyped behaviours, especially when there is a change in their daily routine or when the environment is full of new inputs and as a result, they tend to block the excessive feelings, manifesting stereotyped behaviours.

Given that children with ASD have difficulties in communicating, establishing relationships and responding appropriately to the environmental stimulus, failing to develop their full potential, the Robótica-Autismo Project studies whether a robotic mechanism, previously programmed is able to help to promote the referred skills in these children.

The work presented in this paper is part of a collaborative project [7]-[10] between the University of Minho, APPACDM

(a Portuguese association for mentally disabled people) of Braga and a group of schools, where the main goal is to improve the social life of children with ASD, in particular to promote their social interaction, by using a robot as a mediator or reward.

The intervention sessions between the child and the robot were recorded on video and pre-defined behaviour indicators were quantified. One of these indicators was the manifestation of stereotyped movements. Therapists measured this behaviour using traditional methods such as paper-pencil rating scales, direct observation or video-based methods that normally are time consuming.

Thus, the main objective of this work is to understand if the Kinect sensor can be used as a time-efficient tool to automatically detect in real-time stereotyped behaviours in children with ASD. This will allow characterizing children's behaviour during the sessions, when they are exposed to new situations, and in the future to use this input as a form to adapt the behaviour of the robot to the state of mind of the child.

This paper is divided in six sections: Introduction, where it is presented the paper focus and motivation. Section II is dedicated to the related work and the overall methodology is presented in Section III. Section IV presents the stereotyped behaviour detection based on the Kinect sensor, as well as the experimental approach, and the experimental methodology is explained in Section V. The results obtained in class environment with Autistic children are presented in Section VI, and the conclusions and further work are detailed in the last section.

II. RELATED WORK

Recently, systems have been developed to automatically detect stereotyped behaviours using accelerometers and pattern recognition algorithms with promising results.

The work presented by [11] and [12] uses a three-axis accelerometer to acquired data. The accelerometer is placed on each wrist and chest of the child to detect hand flapping and body rocking movement associated to these problematic. The pattern recognition algorithm used is based on the acquisition

of the characteristics at each component of the acceleration (O_x , O_y , O_z axis), where the distance between the average for each component of acceleration is analysed to know the orientation of the sensor. The variance of the different directions movements, the correlation coefficient to capture simultaneous movements, and the entropy to identify the types of stereotyped movements are also considered. These characteristics were analysed using Weka [13], which carries out a cross-validation with the input signals with the analysed signals.

The results are promising, but it is necessary to wear the sensors on the wrists and torso using elastic armbands, a situation that may not be well accepted by the children with ASD.

The Kinect sensor has been used in pattern recognition such as hand gestures [14], Human pose estimation [15] or gait analysis, but there is no related work to automatically detect stereotyped movements in children with ASD.

III. EXPERIMENTAL SETUP DEVELOPMENT

The Kinect is a sensor introduced by Microsoft in order to control video game console (Xbox 360) without using any controller, based on the concept of Natural Interaction (NI). This concept is supported in the Human-device interaction through human senses, especially hearing and vision. The device contains an RGB colour camera, a depth sensor and a microphone multi vector (Figure 1), which allows the capture of body motion, gesture recognition, recognition of voice commands and face. The depth sensor consists of an infrared laser combined with a monochrome CMOS sensor allows processing scenes in any ambient lighting.



Figure 1 – Kinect Sensor

Using the depth map of the obtained data it is possible to determine human forms (Figure 2), known as skeleton tracking, to establish the positions of various joints of the human skeleton. For example, tracking the skeleton determines where the user's head, hands and centre of mass are located. The skeleton tracking provides the coordinates O_x , O_y , O_z and their values for each joint detected.

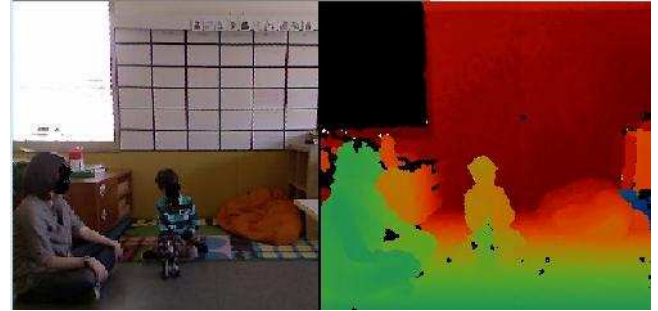


Figure 2 – Colour and depth image obtained by the Kinect Sensor

To recognize stereotyped motor movement with the Kinect sensor it was used the Dynamic Time Warping (DTW) algorithm [16], [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.

This algorithm has been used in voice recognition patterns, signature verification, and gesture recognition [16], [18]

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

Some preliminary tests were performed using the Kinect sensor and the DTW algorithm [19].

IV. STEREOTYPED BEHAVIOUR DETECTION

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

A. Kinect referential

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, i.e. 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 3.

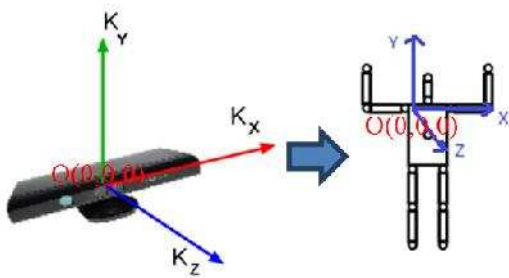


Figure 3 – Illustration of the converted referential

B. Application description

In order to use the Kinect sensor, an application based on WPF (Windows Presentation Form) and C# language was developed. After the initialization of the application it is necessary to wait until the Kinect detects a user (represented by the red colour 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 (Figure 5-e). By pressing the hand flapping button, the gesture is recorded for twenty frames. This value of twenty frames was defined in experimental pre-tests, being considered a good value for a template sequence. When the recording ends, the application goes to the reading mode, ready to identify the stereotypes (Figure 4).

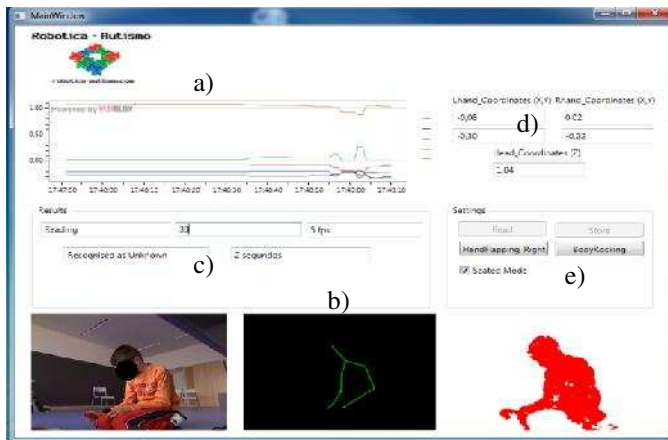


Figure 4 –Application developed for the Kinect sensor

The interface is segmented in four different sectors: a) presents a graph that keeps track of the head position (only Z coordinate) characterized by the orange line. The remaining lines represent the hands position (X and Y coordinates; sector b) shows the RGB, depth video and the skeleton of the child obtained by the Kinect sensor; sector c) presents the display area of the results; and sector d) display the hands positions (X, Y, Z coordinates) at the current state.

The sampling period chosen for the application in the reading mode was five frames. This value was defined in experimental pre-tests through a trial and error methodology. The application at every five frames stores the hand

coordinates in an array. Since the hand flapping movement is generally performed along the Oy axis, only the Oy coordinates were considered relevant. Through the DTW algorithm it is verified if there is a similarity between the template sequence and the array with the hand coordinates. If a match is found, it is considered a stereotyped movement and starts to count the time duration of the hand flapping movement.

V. EXPERIMENTAL TEST METHODOLOGY

Initially the system was tested in a laboratory environment with seven adults, without any associated pathologies.

Since the results were encouraging, the system was tested in three special education units with children with ASD. The system was tested during sessions where the robot was used in a triadic relationship (researcher, child and robot) as a mediator or a reward. The sessions lasted 10 minutes and were recorded on video by three independent cameras; the Kinect system was also installed in the classroom.

In the first unit, APACI (Association of Parents and Friends of Children Unsuitable) of Barcelos, Portugal, the system was tested with two three-years-old children. The final tests were performed with five children with ASD (mean age, M = 9, standard deviation, SD = 3.31) in six sessions in two schools with Special Education Units in Leça da Palmeira, Portugal.

VI. RESULTS

The results obtained in APACI were not quantified due to relevant factors that were not previously considered. The height of children and the table where the experiment was performed (object between the child and the researcher) did not allow the system to correctly detect the children. These difficulties were overcome by conducting the experiments on the floor, i.e., without any object, except the robot, between the child and the researcher and some improvements in the pattern recognition algorithm were performed (Figure 5).



Figure 5 – Intervention session with the Kinect held on the ground

Table 1 shows the results obtained with the five children from the Special Education Units in Leça da Palmeira. The quantification of the stereotyped motor movements performed

in the sessions and detected by the application and by video analysis are presented. Video analysis was used to validate the results obtained using the Kinect application. The results correspond to the average values of the motor manifestation in the six sessions.

TABLE I. HAND FLAPPING VALUES DETECTED BY THE KINECT APPLICATION AND OBTAINED BY VIDEO ANALYSIS (MEAN VALUES IN SIX SESSIONS)

Child	Kinect Application	Video Analysis
1	7.7	5.5
2	3.7	13
3	0	0
4	0.2	0.3
5	0.5	0.5

It can be observed in Table 1, the average of stereotyped movements detect by the Kinect sensor in Child 1 were higher than those really performed by the child. This result indicates that the Kinect sensor detected false-positive values. These values may be justified by the influence of the researcher in the Kinect detection area.

In the sessions performed with Child 2 and Child 4, the Kinect sensor detected less stereotyped movements than the stereotypes held by the child. This situation can be explained by the fact that the children may have left the Kinect detection area or due to noise signals. The researcher may have entered the detection area and the Kinect sensor cannot distinguish between the child and the researcher; the application cannot distinguish whose movements it has to track.

Child 3 did not manifest any kind of stereotyped movement, correctly determined by the application.

It is worth mention that Child 5 participated in only two sessions. The stereotyped movements performed were correctly determined by the Kinect application.

VII. CONCLUSION AND FUTURE WORK

This paper presents an application based on the Kinect sensor and on the Dynamic Time Warping algorithm. These gesture recognition algorithms automatically detect in real time stereotyped behaviour, such as hand flapping. These motor movements are often shown by children with ASD. The objective of the application was to automatically detect hand flapping during the intervention sessions while using a robot.

The system was tested in a laboratory environment with subjects without any pathology associated. It was also tested in a classroom environment with children with ASD. The results were promising, but the system still needs some improvements. The image obtained by the Kinect sensor has to be enhanced in order to eliminate the background noise of the image, thereby decreasing the probability of detecting false positives. It is necessary to improve the accuracy of the application when the child is far from the Kinect or too close to it. Due to the fact that the Kinect sensor detects more than one person, it is essential to implement the algorithm that makes only the trace of the person of interest, which in this case is the child.

As future work, it is intended to improve the gesture recognition algorithms; the use of the Hidden Markov Models (HMM) could be a possibility. Body rocking gesture determination is also to be included in the application.

ACKNOWLEDGMENT

This work has been supported by FCT – Fundação para a Ciência e Tecnologia in the scope of the project: PEst-OE/EEI/UI0319/2014.

The authors are also grateful to the Portuguese Foundation for Science and Technology, FCT- in Portuguese Fundação para a Ciência e a Tecnologia, by the R&D project reference RIPD/ADA/109407/2009 and the PhD scholarship SFRH/BD/71600/2010.

REFERENCES

- [1] Schopler, E., & Mesibov, G. B. (Eds.), 1987. Neurobiological issues in autism. Springer.
- [2] Gillberg, C., & Coleman, M., 1992. The biology of the autistic syndromes. Mac Keith Press.
- [3] Happé, F., & Bernardos, M. N., 2007. Introducción al autismo. Alianza Editorial.
- [4] American Psychiatric Association (Ed.), 2000. Diagnostic and statistical manual of mental disorders: DSM-IV-TR®. American Psychiatric Pub.
- [5] Sutinen, E., Virmajoki-Tyrväinen, M., & Virnes, M., 2004. Concretizing Technologies in Special Education for Developing Social Skills. In Proceedings of the 2nd Cambridge Workshop on Universal Access and Assistive Technology (CWUAAT), Cambridge, UK (pp. 147-156).
- [6] Baumeister, A. A., & Forehand, R., 1973. Stereotyped Acts., International Review of Research in Mental Retardation, 6, 55.
- [7] Costa, S., Resende, J., Soares, F. O., Ferreira, M. J., Santos, C. P., & Moreira, F., 2009. Applications of simple robots to encourage social receptiveness of adolescents with autism. In Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE (pp. 5072-5075). IEEE.
- [8] Costa, S., Soares, F., Santos, C., Ferreira, M. J., Moreira, F., Pereira, A. P., & Cunha, F., 2011. An approach to promote social and communication behaviours in children with Autism Spectrum Disorders: Robot based intervention. In RO-MAN, 2011 IEEE (pp. 101-106). IEEE.
- [9] Costa, S., Santos, C., Soares, F., Ferreira, M., & Moreira, F., 2010. Promoting interaction amongst autistic adolescents using robots. In Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE (pp. 3856-3859). IEEE.
- [10] Silva, S., Soares, F., Costa, S., Pereira, A. P., & Moreira, F., 2012. Development of skills in children with ASD using a robotic platform. In Bioengineering (ENBENG), 2012 IEEE 2nd Portuguese Meeting in (pp. 1-4). IEEE.
- [11] Albinali, F., Goodwin, M. S., & Intille, S. S., 2009. Recognizing stereotypical motor movements in the laboratory and classroom: a case study with children on the autism spectrum. In Proceedings of the 11th international conference on Ubiquitous computing (pp. 71-80). ACM.
- [12] Goodwin, M. S., Intille, S. S., Albinali, F., & Velicer, W. F., 2011. Automated detection of stereotypical motor movements. Journal of autism and developmental disorders, 41(6), 770-782.
- [13] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H., 2009. The WEKA data mining software: an update. ACM SIGKDD Explorations Newsletter, 11(1), 10-18.
- [14] Tang, M., 2011. Recognizing hand gestures with microsoft's kinect. Web Site: http://www.stanford.edu/class/ee368/Project_11/Reports/Tang_Hand_Gesture_Recognition.pdf.
- [15] Qiao, M., Cheng, J., & Zhao, W., 2012. Model-Based Human Pose Estimation with Hierarchical ICP from Single Depth Images.

**2014 IEEE International Conference on
Autonomous Robot Systems and Competitions (ICARSC)
May 14-15, Espinho, Portugal**

- In *Advances in Automation and Robotics*, Vol. 2 (pp. 27-35). Springer Berlin Heidelberg.
- [16] Sakoe, H., & Chiba, S., 1978. Dynamic programming algorithm optimization for spoken word recognition. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 26(1), 43-49.
- [17] Müller, M., 2007. *Information retrieval for music and motion*. Springer.
- [18] Myers, C., Rabiner, L., & Rosenberg, A., 1980. Performance tradeoffs in dynamic time warping algorithms for isolated word recognition. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 28(6), 623-635.
- [19] Gonçalves, N., Rodrigues, J. L., Costa, S., & Soares, F., 2012. Preliminary study on determining stereotypical motor movements. In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE* (pp. 1598-1601). IEEE.