Stereotypical Motor Movement Recognition Using Microsoft Kinect with Artificial Neural Network

M. Jazouli, S. Elhoufi, A. Majda, A. Zarghili, R. Aalouane

***Abstract*—**Autism spectrum disorder is a complex developmental disability. It is defined by a certain set of behaviors. Persons with Autism Spectrum Disorders (ASD) frequently engage in stereotyped and repetitive motor movements. The objective of this article is to propose a method to automatically detect this unusual behavior. Our study provides a clinical tool which facilitates for doctors the diagnosis of ASD. We focus on automatic identification of five repetitive gestures among autistic children in real time: body rocking, hand flapping, fingers flapping, hand on the face and hands behind back. In this paper, we present a gesture recognition system for children with autism, which consists of three modules: model-based movement tracking, feature extraction, and gesture recognition using artificial neural network (ANN). The first one uses the Microsoft Kinect sensor, the second one chooses points of interest from the 3D skeleton to characterize the gestures, and the last one proposes a neural connectionist model to perform the supervised classification of data. The experimental results show that our system can achieve above 93.3% recognition rate.

***Keywords*—**ASD, stereotypical motor movements, repetitive gesture, kinect, artificial neural network.

I. Introduction

AUTISM or more generally ASD is a group of complex disorders of brain development. Symptoms are present in the early developmental period, usually detected by the parents in the first two years of the child's life. Autism is characterized by impaired verbal and nonverbal communication (emotional, facial expression and gesture) deviant social interactions, abnormalities of sensory motor activity [1]. The main areas of difficulty are in social communication, social interaction, and restricted or repetitive behaviors and interests. Analysis of the natural behavior of the child is the key to early detection of ASDs. The identification and diagnosis process for ASD varies among clinicians and communities. Three elements are often considered essential: surveillance, prevention, and diagnosis. To analyze and measure ASD, this is conventionally done by watching video recordings of the child's interaction with his environment, and manually assessing the child's behavior episodes. This data collection procedure is often very long, and analysis must be controlled by several experimenters to ensure reliable results [2]. The objective of our research is to help clinicians and doctors to perform the task of early detection and automatic measurement of ASD, focusing on computer vision tools to measure and identify behavioral markers associated with ASD. These behaviors often manifest in an intense and focused interest in a particular subject matter; stereotyped body movements like hand flapping and spinning; and an unusual and heightened sensitivity to everyday sounds or textures. We propose in our research to work on automated systems to measure and identify behavioral markers associated with ASD while considering issues of intrusiveness, robustness, reliability, and their impact on child behavior.

In this paper, we present a novel automatic detection system for non-invasive assessment of stereotyped behaviors among children with ASD. Stereotyped gestures are a set of gestures that are understudied in the field of automatic detection of nonverbal visual cues. We are interested in the detection and recognition of the five categories of gestures that we have selected from CARS (Childhood Autism Rating Scale) in real-time: body rocking, hand flapping, fingers flapping, hand on the face and hands behind back. To achieve our goal, we use the Microsoft Kinect sensor for feature extraction and artificial networks of neurons for the classification and recognition of stereotyped movements. ANN functions are powerful computational tools for nonlinear prediction problems in various applications, including image coding, pattern recognition, Speech recognition, fault detection, and medical imaging [3]-[5]. The main motivation behind the use of an ANN in this paper is based on good adjustment to a set of data. An ANN generates a better output function than the classical approximation methods. In the classical approximation methods, a set of data is adjusted to a n-degree polynomial function [6].

Methodically, the paper is shackled into five sections. Firstly, we will review the previous work on automatic recognition of gestures. While proposed solutions for implementing our approach appear in Section III, and experiments results aimed at evaluating the efficiency and the robustness of our application will be the focus of Section IV. We conclude the paper in Section V.

II. Related Work

Recently, the most of systems have been developed to automatically detect stereotypical motor using accelerometers and pattern recognition algorithms. In [7], [8] the authors are interested in creating a real-time recognition tool. They developed a stereotyped movement recognition system for autistic people by using the acceleration data and a decision tree classifier type. They have tested their system on two types

M. J., S. E., A. M. and A. Z. are with the Laboratory of Intelligent Systems & Applications, Faculty of Science and Technology, Sidi Mohamed Benabdellah University, Fez, Morocco, (e-mail: maha.jazouli@usmba.ac.ma, safae.elhoufi@usmba.ac.ma, aicha_majda@yahoo.fr, zarghili@yahoo.fr).

R. A. is with the Laboratory of Psychiatry, Chu Hassan II Hospital, Faculty of Medicine and Pharmacy, Sidi Mohamed Benabdellah University, Fez, Morocco, (e-mail: aalouaner@yahoo.fr).

of stereotyped movements: hand flapping and rocking of the body. Measures by wireless accelerometer sensing technology and machine learning techniques provide an automatic, time efficient, and accurate measure of stereotypical motor movements [9]-[13]. Still, the use of sensors on wrists presents a disadvantage of not being accepted by some autistic children, which affects the results. To overcome this problem, in the literature, researchers are more and more interested in using the Kinect sensor. Many researchers have been working with it using pattern recognition such as hand gestures, human pose estimation or gait analysis, but there are few of research studied gesture recognition associated with the field of autism by analyzing images in video sequences using the Kinect sensor. The work presented by [10] recognition system based on the Dynamic Time Warping (DTW) algorithm. The DTW has been used in voice recognition patterns, signature verification, and gesture recognition [14], [15]. The main objective of this work is to understand whether the Kinect sensor and the eZ430-Chronos with accelerometers can be used as an effective tool.

After preliminary studies of the academic literature, the most of the automatic detection system of stereotyped behaviors in children with ASD is interested to detect one or two stereotyped movements on maximum. That is why we were interested to automatically detect five stereotypical motor movements (body rocking, hand flapping, fingers flapping, hand in the face and hands behind back) in real time by proposing a method that gives good results. On the other hand, our system proposes a system with the Kinect sensor because the use of sensors on wrists presents a disadvantage of not being accepted by some autistic children.

III. System Developed

Stereotypical motor movements are one of the most common and least understood behaviors occurring in persons with ASD. The main of this study is to automatically detect stereotyped motion disorders of patients with ASD in real time. Our work is based on the Kinect sensor with an ANN approach. We propose an automatic detection system for non-invasive assessment of stereotyped behaviors in children with ASD. In this paper, we focus on five stereotyped motion disorders: body rocking, hand flapping, fingers flapping, hand on the face and hands behind back.

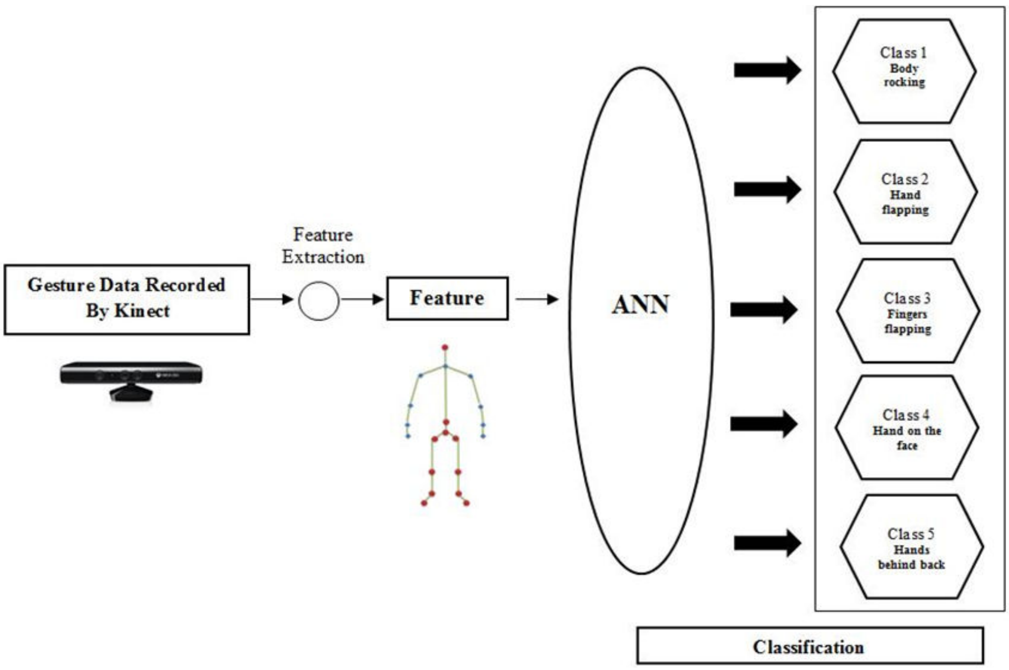


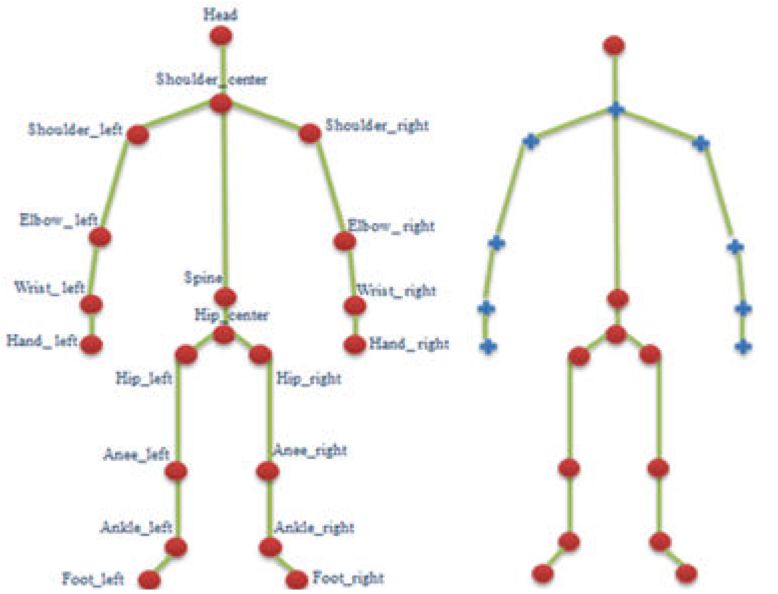
Fig. 1 Global architecture of our system for stereotypical motor movement recognition

A. Extraction Features

In order to acquire motion data, in this research, we use 3D skeleton model information generated from Microsoft's Kinect sensor. The Microsoft Kinect is a set of sensors developed as a peripheral device for use with the Xbox 360 gaming console Fig. 1. Using image, audio and depth sensors, it detects movements, identifies faces, and recognizes speech of players, allowing them to play games using only their own bodies as controls. Microsoft Kinect provides an inexpensive and portable monitoring platform, which along with providing depth information also provides good quality color image data. 3D skeleton enables developing applications capable of tracking human posture in real time. The skeleton tracking engine determines 3D positions of semantic skeleton feature points. One frame data contain the three-dimensional positions of 20 joints over time. The 20 joints include head, shoulders, elbows, wrists, hands, spine (shoulder, mid and base), hips, knees, ankles and feet as shown in Fig. 3 (a).

Fig. 2 Kinect Sensor and coordinate system

In our system, we are interested to detect and recognize five stereotypical movements: body rocking, hand flapping, fingers flapping, hand on the face and hands behind back. These gestures depend on some of joints skeleton. For this data acquisition, we focused on four types of joints: shoulder, elbow, wrist and hand Fig. 3 (b).



(a) Points tracked by the Kinect sensor (b) Points tracked by our system.

Fig. 3 3D skeleton model

B. Classification and Recognition Using ANN

ANN is a biologically motivated learning machine inspired by the biological neuron and nervous system processes [16], [17].

For an artificial network of neurons, each neuron is interconnected with other neurons and form layers in order to solve a specific problem on the data input in the network [18]. ANN contains three layers of nodes from left to right [19]. The input layer of nodes is only a fan-out layer. The middle layer is called the hidden layer, since the output of their neurons is not accessible from the outside. The output layer is a continuous function, which is obtained after the training. The network architecture is shown in Fig. 4.

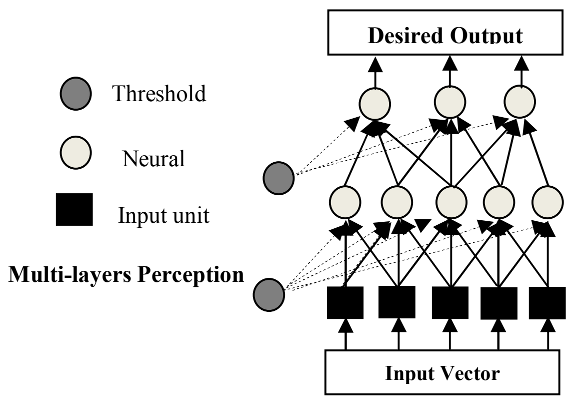


Fig. 4 The architecture of the multi-layer perception

IV. Experimental Results

We used a system with three layers, driven by the backpropagation algorithm. The backpropagation algorithm requires a random initialization of synaptic weights. The network is fully connected between two layers and is thresholded. The input layer is supplied with the feature vectors. Once the learning is completed, the network can process the new data. It is the neuron whose output is the highest that determines the class. Five classes were defined, and network output layer contains five neurons (one per class). The following table indicates the desired output vector for each of the five classes:

TABLE I
The Desired Output Vector for Each of the 5 Classes

Class	Denomination	Desired outputs
1	Body rocking	(1, 0, 0, 0, 0)
2	Hand flapping	(0, 1, 0, 0, 0)
3	Fingers flapping	(0, 0, 1, 0, 0)
4	Hand on the face	(0, 0, 0, 1, 0)
5	Hands behind back	(0, 0, 0, 0, 1)

In this part, we are going to present the results from the training and testing procedure of the algorithm and also the performance of the neural network that was designed for gestures recognition. The database consists of ten autistic children, taking into consideration the five stereotypical gestures (see Fig. 5). This gives us a base of 50 gestures. We used 70% in the learning phase and 30% in the test.

A. Training the Network

During the training sessions, the weights and threshold of the network are iteratively adjusted to minimize the network performance error. We give below the network architecture and the confusion matrix corresponding to the learning phase. We obtain a recognition rate of 97.1% (see Table II).

B. Testing the Network

After training, the performance of the network has to be tested with a test set (set of similar data unused during training). A first indication is given by the percentage of correct classifications of the training set records. Nevertheless, the performance of the network with a test set is more relevant. In the test step, the input data are fed into the network and the desired values are compared to the network's output values [19]. The agreement or disagreement of the results thus gives an indication of the performance of the trained network. We achieve a 93.3% recognition rate with an error rate of 6.7% on this set of data (see Table III).

V. Conclusion

In this paper, the ability to recognize the gestures among children with autism using the Kinect sensor has been achieved by the classifier ANN in a very fast manner. The system recognizes five stereotypical motor movements in real time: body rocking, hand flapping, fingers flapping, hand in the face and hands behind back.

Fig. 5 Stereotyped gestures of autistic children

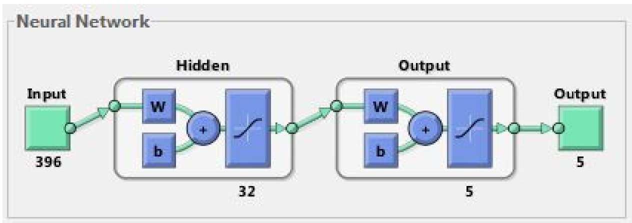


Fig. 6 The network architecture

The method exploits the learning capabilities and generalization of a multilayer perceptron. The idea consists on getting such a network understands a number of examples of each possible gesture. To reduce the amount of data to be processed, the autistic child is characterized by points of interest extracted from the 3D skeleton. At the end of the learning, the generalization capabilities of the multilayer perceptron allow him the recognition on non-learned gestures. The results obtained are satisfactory, proving the robustness of the method in front of a deficient pretreatment or acquisition. The proposed system is easy to implement and use, furthermore the recognition is very fast. The application detected 93.3% of the stereotyped movements, which are satisfactory results compared to literature.

For future work, on the one hand it is intended to compare and to improve the gesture recognition algorithms, in order to reduce error rates. On the other hand, we need an exhaustive analysis of stereotypical motor movements.

TABLE II
Confusion Matrix for Training Set

Classes	1	2	3	4	5	Rate Error
1	**7**	0	0	0	0	100%
	20.0%	0.0%	0.0%	0.0%	0.0%	0.0%
2	0	**7**	**1**	0	0	87.5%
	0.0%	20.0%	2.9%	0.0%	0.0%	12.5%
3	0	0	**6**	0	0	100%
	0.0%	0.0%	17.1%	0.0%	0.0%	0.0%
4	0	0	0	**7**	0	100%
	0.0%	0.0%	0.0%	20.0%	0.0%	0.0%
5	0	0	0	0	**7**	100%
	0.0%	0.0%	0.0%	0.0%	20.0%	0.0%
Rate Error	100%	100%	87.5%	100%	100%	**97.1%**
	0.0%	0.0%	14.3%	0.0%	0.0%	2.9%

TABLE III
CONFUSION MATRIX FOR TESTING SET

Classes	1	2	3	4	5	Rate Error
1	**3**	0	0	0	0	100%
	20.0%	0.0%	0.0%	0.0%	0.0%	0.0%
2	0	**3**	**0**	0	0	100.0%
	0.0%	20.0%	0.0%	0.0%	0.0%	0.0%
3	0	0	**3**	0	1	75.0%
	0.0%	0.0%	20.0%	0.0%	6.7%	25.0%
4	0	0	0	**3**	0	100%
	0.0%	0.0%	0.0%	20.0%	0.0%	0.0%
5	0	0	0	0	**2**	100%
	0.0%	0.0%	0.0%	0.0%	13.3%	0.0%
Rate Error	100%	100%	100%	100%	66.7%	**93.3%**
	0.0%	0.0%	0.0%	0.0%	33.3%	6.7%

REFERENCES

[1] P. A. Filipek, P. J. Accardo, G. T.Baranek, E. H. Cook, G.Dawson,B. Gordon, et al., "The screening and diagnosis of autistic spectrum disorders", Journal of Autism and Developmental Disorders, 29 (6), 439–484, 1999.

[2] B. NORIS, "Machine Vision-Based Analysis of Gaze and Visual Context: an Application to Visual Behavior of Children with Autism Spectrum Disorders",Lausanne, écolepolytechnique fédérale,2011.

[3] J. Stark, "A neural network to compute the Hutchinson metric in fractal image processing", IEEE Transactions on Neural Networks, 2 (1), 156–158,1991.

[4] J. Jiang, P. Trundle, J. Ren, "Medical image analysis with artificial neural networks." Computerized Medical Imaging and Graphics, 34 (8), 617–631, 2010.

[5] K. Saravanan, S. Sasithra, "Review on classification based on artificial neural networks", International Journal of Ambient Systems and Applications (IJASA) Vol.2, No.4, December 2014.

[6] J. A. Munoz-Rodrıguez, A. Asundi, R. Rodrıguez-Vera, "Shape detection of moving objects based on a neural network of a light line", Optics Communications, 221, 73–86, 2003.

[7] F. Albinali, M. S. Goodwin, and S. S. Intille, "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, 2009.

[8] M. S. Goodwin, S. S. Intille, F. Albinali, and W. F. Velicer, "Automated detection of stereotypical motor movements.", Journal of autism and developmental disorders, vol. 41, no. 6, pp. 770–82, 2011.

[9] F. Albinali, M. S. Goodwin, and S. S. Intille, "Detecting stereotypical motor movementsin the classroom using accelerometry and pattern recognition algorithms.", Pervasiveand Mobile Computing, 8(1), 103–114, 2012.

[10] N. Gonçalves, J. L. Rodrigues, S. Costa, and F. Soares, "Automatic detection of stereotyped hand flapping movements: two different approaches.", The 21st IEEE International Symposium on Robot and Human Interactive Communication, 2012.

[11] C.H. Min, A.H. Tewfik,"Automatic characterization and detection of behavioralpatterns using linear predictive coding of accelerometer sensor data.", In: Engineering in Medicine and Biology Society (EMBC), pp. 220–223, 2010.

[12] J. L. Rodrigues, N. Gonc, S. Costa, and F. Soares, "Stereotyped movement recognition in children with ASD.",Sensors and Actuators A: Physical, vol. 202, pp. 162–169, 2013.

[13] T. Westeyn, K. Vadas, X. Bian, T. Starner, and G. D. Abowd, "Recognizing Mimicked Autistic Self–Stimulatory Behaviors Using HMMs.", In: Proceedings. Ninth IEEE International Symposium on pp. 164–167, 2005.

[14] H. Sakoe and S. Chiba,"Dynamic programming algorithm optimization for spoken word recognition", IEEE Trans. Acoustics, Speech, Signal Processing, 26(1), 43–49, 1978.

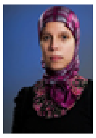
[15] C. Myers, Lawrence R. Rabiner and Aaron E. Rosenberg, "Performance Tradeoffs in Dynamic Time Warping Algorithms for Isolated Word Recognition", IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 28,pp: 623 – 635, 1980.

[16] E. Kolman, M. Margaliot, "Are artificial neural networks white boxes?", IEEE Transactions on Neural Networks, 16 (4), 844–852, 2005.

[17] S. Haykin, "Neural Networks and Learning Machines", 3rd ed. Prentice-Hall, Ontario, 2008.

[18] R. LEPAGE, B. SOLAIMAN, "Les réseaux de neurones artificiels et leurs applications en imagerie et en vision par ordinateur", Coop ÉTS, 2003.

[19] D. Rebya, S. Lek , I.Dimopoulos, J. Joachim, J. Lauga, S. Aulagnier, "Artificial neural networks as a classification method in the behavioural sciences", Behavioural Processes, 40, 35–43, 1997.

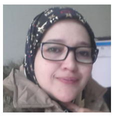


Maha Jazouli received her master's degree in intelligent systems and networks from the Faculty of Science and Technology, in University Sidi Mohamed Ben Abdellah, Fez, Morocco, in 2013. She's currently a Ph.D student at the Intelligent Systems and Applications Laboratory at the Faculty of Science and Technology, Fez.

Her research interests include medical video processing, automatic object detection and robust object tracking. Ms. Jazouli is a member of the IEEE Morocco Section.

Safae El Houfi received her master's degree in intelligent systems and networks from the Faculty of Science and Technology, in University Sidi Mohamed Ben Abdellah, Fez, Morocco, in 2013. She's currently a Ph.D student at the Intelligent Systems and Applications Laboratory at the Faculty of Science and Technology, Fez.

Her interest areas include pattern recognition ,3D image processing and 3D modeling.



Aicha Majda is a Doctor of Science from Mohamed V University (Rabat). She received her Ph.D. in 2OOO in Engineering Sciences and joined Sidi Mohamed Ben Abdellah University as professor at the Computer Science department of the Faculty of Science and Technology. She obtained her HDR in 2013. She is a member of Intelligent Systems and Applications Laboratory. Her current research interests include multimedia signal processing (e.g., image and video Coding, Analysis and recognition).

Arsalane Zarghili is a Doctor of Science from Sidi Mohamed Ben Abdellah University (Fez-Morocco). He received his Ph.D in 2001 and joined the same University in 2002 as assistant professor at the computer science department of the Faculty of Science and Technology of Fez (FST). In 2007, he was head of the computer sciences department and chair of the Master of Software Quality at FST-Fez. He lectures Programming, Distributed system, compilation and image processing, for both undergraduate and master levels. In 2008, he obtained his HDR in information processing, promoted to associate professor. In 2015, he is promoted to full professor at the same department. In 2011, he is the co-founder and the head of the Laboratory of Intelligent Systems and Applications at FST of Fez. He is a member of the steering committee of the department of computer sciences and was a member of the faculty board. In 2011, he is the chair of Master Intelligent Systems and Networks. He is also IEEE member since 2011. His main research is about pattern recognition, image indexing and retrieval systems in cultural heritage, biometric, and Healthcare. He also works on Arabic Natural Language Processing.