Automatic detection of stereotyped movements in autistic children using the Kinect sensor

Article *in* International Journal of Biomedical Engineering and Technology · January 2019

DOI: 10.1504/IJBET.2019.097621

5 authors, including:

Maha Jazouli
Sidi Mohamed Ben Abdellah University

Aicha Majda
Faculty of Science & Technology, Fès, Morocco

Djamel Djamal Merad
Aix-Marseille Université

Arsalane ZARGHILI
Sidi Mohamed Ben Abdellah University

Automatic detection of stereotyped movements in autistic children using the Kinect sensor

Maha Jazouli* and Aicha Majda

Laboratory of Intelligent Systems & Applications,
Faculty of Science and Technology,
Sidi Mohamed Benabdellah University,
Fez, Morocco
Email: maha.jazouli@usmba.ac.ma
Email: aicha_majda@yahoo.fr
*Corresponding author

Djamal Merad

Laboratory of Information Science and Systems,
Aix-Marseille University,
Marseille, France
Email: djamal.merad@univ-amu.fr

Rachid Aalouane

Laboratory of Clinical Neurosciences,
CHU HASSAN II Hospital,
Faculty of Medicine and Pharmacy,
Sidi Mohamed Benabdellah University,
Fez, Morocco
Email: aalouaner@yahoo.fr

Arsalane Zarghili

Laboratory of Intelligent Systems & Applications,
Faculty of Science and Technology,
Sidi Mohamed Benabdellah University,
Fez, Morocco
Email: zarghili@yahoo.fr

Abstract: Autism Spectrum Disorders (ASD) is a developmental disorder that affects communications, social skills or behaviours that can occur in some people. Children or adults with ASD often have repetitive motor movements or unusual behaviours. The objective of this work is to automatically detect stereotypical motor movements in real time using Kinect sensor. The approach is based on the $P Point-Cloud Recogniser to identify multi-stroke gestures as point clouds. This paper presents new methodology to automatically detect five stereotypical motor movements: body rocking, hand flapping, fingers flapping, hand on the face and hands behind back. With many ASD-children, our proposed system gives us satisfactory results. This can help to implement a smart video surveillance system and then helps clinicians in the diagnosing ASD.

Keywords: ASD; autism; stereotyped movement; repetitive motor movements; repetitive behaviours; gesture detection; Kinect sensor; point cloud; nearest-neighbour classifier; gesture recognition.

Reference to this paper should be made as follows: Jazouli, M., Majda, A., Merad, D., Aalouane, R. and Zarghili, A. (2019) 'Automatic detection of stereotyped movements in autistic children using the Kinect sensor', *Int. J. Biomedical Engineering and Technology*, Vol. 29, No. 3, pp.201–220.

Biographical notes: 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 is currently a PhD 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. She is a member of the IEEE Morocco Section.

Aicha Majda holds a PhD in engineering sciences from the University Mohamed V in Rabat in Morocco, and an Accreditation to Supervise Research "HDR" in Computer Sciences from the University Sidi Mohammed Ben Abdellah. Currently, she is teacher of Computer Science, in the Faculty of Science and Technics, and a research director in "Intelligent Systems and Applications" Laboratory, in the University Sidi Mohammed Ben Abdellah since 2002. She is also a coordinator of SIA Laboratory in biomedical research & health pole of Sidi Mohammed Ben Abdellah University since 2015, and vice president of the Moroccan Association of Engineering and Innovation in Health since 2016. She has many works published in computer vision.

Djamal Merad received an initial training of engineer, and supplemented by a DEA in Image and Vision at the University of Nice, he concluded his academic formation by a doctorate in Robotics at the University of Evry. In September 2008, he integrates the team Images and Models of LSIS lab, as a university lecturer. Since, his research activity is focused on the pattern recognition field and he has participated in several research projects.

Rachid Aalouane is a Doctor of Neurosciences at Sidi Mohamed Ben Abdellah University Fez, Morocco. He is a psychiatrist specialising in autism field.

Arsalane Zarghili is a Doctor of Science from Sidi Mohamed Ben Abdellah University (Fez, Morocco). He received his PhD 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. 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 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.

1 Introduction

Autism is a complex neurodevelopmental disability that lasts throughout life; it is a part of the Pervasive Developmental Disorders. Experts believe that most autism cases appear during the first three years of life. People with autism have a neurological disorder that has an effect on normal brain function. It can affect the development of the person's communication and social interaction skills. ASD is defined by a set of behaviours that affects individuals differently with varying degrees. As with all diseases, symptoms and their severity vary from one child to another, according to their age. Autism is characterised by difficulties with socialisation, communication and the presence of repetitive behaviours and restricted interests (Filipek et al., 1999). Many tools exist to diagnose autism. Among the most commonly used tools (Schopler et al., 1980), we find CARS (Childhood Autism Rating Scale), a degree of Autism Rating Scale for children based on their behaviour. It can identify children with autism and distinguish them from children with other developmental disorders. The development is assessed in its entirety with 15 categories of behaviour, among them, we find the stereotyped behaviours. Autistic individuals show many forms of stereotyped behaviours: the Repetitive Behaviour Scale-Revised (RBS-R) (Coronato et al., 2014; Lam and Aman, 2007) classifies them in several categories, such as body rocking, hand flapping, etc. Child psychologists call these stereotyped movements: Repetitive and Restrictive Behaviours (RRBs), which are considered as the main feature of Autism Spectrum Disorders (ASD). Research proves they are almost always present in very young children with ASD and persist over time as the child grows (Kim and Lord, 2010). Analysis and evaluation of the ASD is conventionally done by watching video recordings of the child's interaction, and manually evaluating his behaviour episodes. This data collection procedure is often very long, and several experimenters must control analysis to ensure reliable results (Noris, 2011). To overcome this problem, in the literature, researchers suggest automatic detection systems of stereotyped movements among autistic children. Most of automatic detection systems use the wireless three-axis accelerometers and pattern recognition algorithms to automatically detect stereotyped behaviours among ASD children (Albinali et al., 2009; Goodwin et al., 2011). On the other hand, the use of sensors on the wrists presents a disadvantage; it is not accepted by some autistic children. This creates an unusual environment for the child, which could sometimes lead to a wrong diagnosis. Recently, researchers are more and more interested in using the Kinect sensor in this field as a way to solve this problem (Gonçalves et al., 2012a; Sivalingam et al., 2012). Moreover, the recent studies suggest gesture recognition systems based on DTW (Dynamic Time Warping). In a recognition system using DTW, data clustering methods are used to reduce the number of gesture templates in the database, and thus reduce the computational cost; however, the recognition rate is degraded. In the following section, an overview of these methods will be presented.

In this paper, we propose a new automatic detection system for the assessment of stereotyped behaviours among children with ASD, while taking into account issues of intrusiveness, robustness, reliability, and their impact on the behaviour of the child. The stereotyped behaviours have been recognised as a core symptom of autism (Filipek et al., 1999). For this reason, the main motivation behind this work is to help the doctors in autism diagnosis. The main contributions of this work are the detection and recognition of five stereotyped gestures of autistic children in real time using the Kinect sensor, unlike previous researches, which were interested in detecting just one or two stereotyped gestures. The choice of five gestures was made after a deep study of autism specifically in CARS (Schopler et al., 1980) and an observation was conducted under normal conditions in an association for autistic children, in cooperation with a psychiatrist and researcher in the field of autism in the hospital CHU (Centre Hospitalier Universitaire Hassan II in Fez). In our study, we decided that these five gestures are the most popular and common among children with ASD. We propose a new automatic gesture classification method using $P point-Cloud Recogniser 3D for identifying multi-stroke gestures as point clouds (Vatavu et al., 2012). It is an instance-based nearest-neighbour classifier with a Euclidean scoring function. The $P recogniser avoids the storage complexity by representing gestures as clouds of Points and thus ignoring variable user behaviour in terms of stroke order and direction.

This paper is divided into five sections. They are structured as follows: Section 2 is dedicated to related work in automatic detection and recognition of stereotyped behaviours. In Section 3, we present the materials and methods used for recognising stereotyped movements and we describe the experimental protocol for our developed system. The results and the discussions are given in Section 4. The paper ends with the conclusion and a brief presentation of the future work.

2 Related works

Computer vision can help to automatically analyse image sequences, quantify the behaviour of the child and give a summary for specialists, as it can help reduce the processing time for diagnosis and provide better access to distance to clinicians. In literature, we find few works that address this issue (automatic detection of stereotyped movements in autistic children). Most recent studies in this field of autism focus on the recognition of gestures and postures of ASD children to determine their emotions (Alabbasi et al., 2015a; Alabbasi et al., 2015b; Larsson et al., 2015; Postawka and Śliwiński, 2015, Kang et al., 2016; Upadhyay and Dewangan, 2016).

Westeyn et al. (2005) present a system to recognise and monitor the behaviour of self-stimulation observed among autistic children. Their approach is based on Hidden Markov Models (HMM) using the three-axis accelerometers to distinguish between the different types of self-stimulatory behaviour and to determine when self-stimulation fails.

In Albinali et al. (2009) and Noris (2011) the authors develop a stereotyped movement recognition system for autistic people by using the acceleration data and a decision tree classifier type. This system allows the identification of two types of stereotyped movements: hand flapping and rocking of the body. The results are obtained with satisfactory accuracy in two different environments (class and laboratory). An accelerometer is placed on each wrist and the child's chest to detect hand flapping and body rocking associated with these problems. They are ultimately interested in creating a real-time recognition tool, and the decision trees have a desirable combination of properties.

Min and Tewfik (2010a, 2010b) describe novel methods to assist management by automatically detecting stereotypical behavioural patterns using accelerometer data. They use the Iterative Subspace Identification (ISI) algorithm to learn subspaces in which the sensor data lives. It extracts orthogonal subspaces which are used to generate dictionaries for clustering and for signal representation. It also applies to detect segments from acoustic data. The data were analysed using time-frequency and observing frequency band powers. Also Linear Predictive Coding (LPC) method is used to classify of the stereotypic and self-injurious behaviours.

Plötz et al. (2012) present a technique for using body accelerometers to assist in automated classification of problem behaviour during such direct observation. Using simulated data of episodes of severe behaviour acted out by trained specialists, they demonstrate how machine learning techniques can be used to segment relevant behavioural episodes from a continuous sensor stream and to classify them into distinct categories of severe behaviour (aggression, disruption, and self-injury).

Gonçalves et al. (2012a) consider two different approaches. The first approach uses gesture recognition algorithms on data from a Microsoft Kinect. The second approach uses a trademark device of Texas Instruments with built-in accelerometers and statistical methods to recognise stereotyped movements. 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 for automatic real-time detection of stereotypic behaviours in children with ASD. The Kinect sensor detects 51% of the stereotyped movements, while the eZ430-Chronos watch detects 76%. This result could be explained by the fact that one of the children tried to take off the eZ430-Chronos bracelet. It can also be observed that eZ430-Chronos watch detects false-positives in 24% of the time.

Mahmoud et al. (2013) propose a cyclic gestures descriptor that can detect and localise rhythmic body movements by tracking multidimensional tracklets and taking advantage of both colour and depth modalities. They demonstrate the importance of fusing depth and intensity features in a classification system to localise the rhythmic gesture as: hands fidgeting, legs fidgeting or rocking. Classification results are significantly higher (59%) than the majority of the classification baseline (35%).

Sivalingam et al. (2012) propose a novel system of a non-invasive multi-sensor configuration for the automatic monitoring of the behaviour of children in their natural environment using the Kinect platform. They present a computer vision system using cameras and depth sensors to track children across multiple views and for extended periods of time. In a first step, they mainly discuss the object tracking problem in a multi-sensor perception system in a non-structured classroom environment. In the second step, they work on objects tracking through the sensors. They propose the use of covariance descriptors based on 3D point clouds in combination with a dictionary learning framework for modelling the appearances of the objects in the scene and for fast matching. They use a Kalman filter to tackle noise in the depth sensors for robust tracking.

Sundar and Goecke (2014) propose an algorithm for detecting three types of self-stimulatory behaviours from publicly available unconstrained videos. The child's body is tracked in the video by a careful selection of poselet bounding box predictions using a nearest-neighbour algorithm. The Histogram of Dominant Motions (HDM) descriptor around the detected body regions is used to train a discriminatory model for classifying self-stimulatory behaviours.

In an overall view, most of the proposed systems in literature about the gesture recognition have an unsatisfactory recognition rate that leads to many false-positive results. On the other hand, the use of sensors on wrists presents a disadvantage of not being accepted by some autistic children. Also, most of the automatic detection systems of stereotyped behaviours among children with ASD are able to detect one or two stereotyped movements at most.

3 Proposed approach

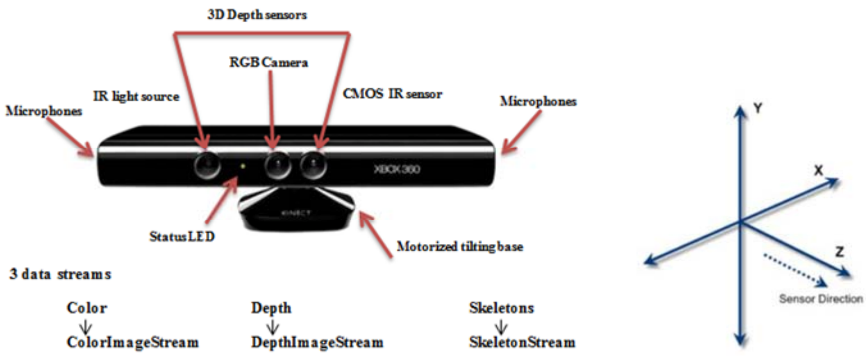
Stereotypical behaviours are one of the most common and least understood behaviours occurring among people with ASD. The presence of stereotypical movements has been a key feature of the syndrome of autism (Kanner, 1943). For this reason, we were interested in the detection and the recognition of five categories of stereotyped movements (or stereotyped gestures) that we have selected from CARS in real time (Schopler et al., 1980): body rocking, hand flapping, fingers flapping, hand on the face and hands behind back. These five stereotyped gestures are the most popular and common among children with ASD. In this paper, we propose a new automatic detection system for the assessment of stereotyped gestures in children with ASD. We used the Kinect sensor for data acquisition and $P Point-Cloud Recogniser for stereotyped gestures classification. The most recent studies seek to provide an automatic system to help the clinicians and doctors in the diagnosis task, such as Kumar and Moni (2011) and Bhattacharjee et al. (2015). So, also our system provides a novel clinical tool which facilitates for doctors the diagnosis of ASD.

3.1 Database acquisition using the Kinect sensor

Gesture or movement detection is intended to decide which parts of an image sequence correspond to a movement. In general, automatic movement detection in an image sequence consists of three main steps: identifying gesture potential locations, extracting the necessary features from the image sequence, and classifying each sub-window automatically to determine if it contains the searched gesture or not. However, in the literature, there are many gesture detection methods. We will give a brief overview of the most commonly used methods. They can be classified as follows: (1) interest point detection, which is a recent terminology in computer vision that refers to the detection of interest points for subsequent processing (Lowe, 2004). Generally, interest points detection algorithms focus on specific points of the selected contour, according to specific criteria (Verbeke, 2007). (2) Background subtraction, or foreground detection, this method is used to build a representation of the scene 'background model', and then, it can detect some changes from this model for each new image in the sequence (Penne, 2011). (3) Image segmentation, which is the process of partitioning the image into multiple similar segments in order to locate objects and boundaries (lines, curves, etc.) in images. The most common methods are: based on the mean shift algorithm, graph-cut algorithm, and active-contour-based algorithm. (4) Supervised classification, which consists of detecting an object by automatically learning different views of the object from a set of examples by a supervised learning mechanism (Penne, 2011).

In our case, we believe that the best method to detect stereotyped gestures in autistic children can be the method of interest point detection using the Kinect sensor v1 by Microsoft. It is the most suitable due to its ability to monitor the joints. Kinect is a device for the Microsoft hardware, is a motion-sensing camera to manipulate games without a controller (Figure 1). The major components of a Kinect are its colour sensor, IR depth sensors, IR emitter, microphone arrays, and a stepper motor that can be tilted to change the Kinect camera angles. The Kinect for Windows is a toolkit for developing applications for Kinect devices. It provides a new opportunity for the developer to build many applications in different domains. We can build different applications with Kinect sensor for healthcare, such as exercise measurement, monitoring patients, their body movements, and so on (Rahimi and Golzarian, 2015). This Toolkit is an additional installer that comes with a set of extended components, such as Skeleton tracking SDK, which helps to track human body. The Kinect for Windows SDK provides us with a set of APIs that allow easy access to the skeleton joints. The SDK supports the tracking of up to 20 joint points (Figure 2). Each and every joint position is identified by its name (head, shoulders, elbows, wrists, hands, spine, hips, knees, ankles, and foots) (Figure 2). All the joints are represented as three dimensions (x, y, z), where X and Y define the position of the joint and Z represents the distance from the sensor. To get the proper coordinates, the sensor calculates the three views of the same image: front view, left view, and top view, by which the sensor defines the 3D body proposal. Microsoft Kinect is a rapidly developing, an inexpensive and portable monitoring platform. It allows basic tracking of skeleton points in 3D space; it can track 20 body skeleton points at 30 frames per second. The coordinate system for the skeleton data is a full 3D system with values in meters (Witten and Frank, 1999) (Figure 1).

Figure 1 Kinect Sensor and Kinect sensor coordinate system



We mentioned at the beginning that our proposed system enables the detection and recognition of five categories of gestures among autistic children: body rocking, hand flapping, fingers flapping, hand on the face and hands behind back. For data acquisition, we used five different types of joints to characterise these gestures: head, shoulder, elbow, wrist and hand (Figure 3). So, in our system, we tracked just these five types of joints.

3.2 $P Point-Cloud Recogniser method

After the data acquisition and construction of a database, we move on to the classification and recognition phase. For this purpose, we adopt the $P Point-Cloud Recogniser algorithm, it is the latest in the dollar family of recognisers ($1, $N). $P is fast, simple and accurate gesture recognition approach. The $P Point-Cloud Recogniser is a 2D gesture recogniser designed for rapid prototyping of gesture-based user interfaces. It is an instance based on the nearest-neighbour classifier with a Euclidean distance function. The technique is not limited only to the gesture recognition with unistroke, but it can also handle multi-strokes recognition. Compared with other algorithms, $P has a great advantage to be simple to implement; it may be encoded into hundred lines to both define and recognise the gesture while giving high recognition rate.

Figure 2 Points tracked by the Kinect sensor (see online version for colours)

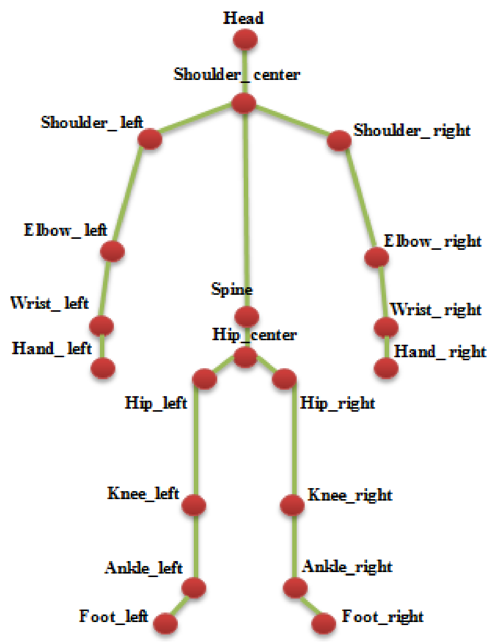
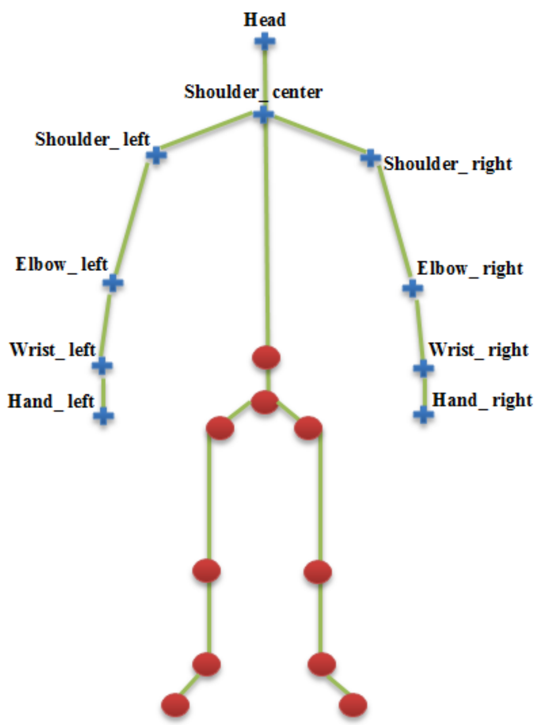


Figure 3 Points tracked for data acquisition (see online version for colours)

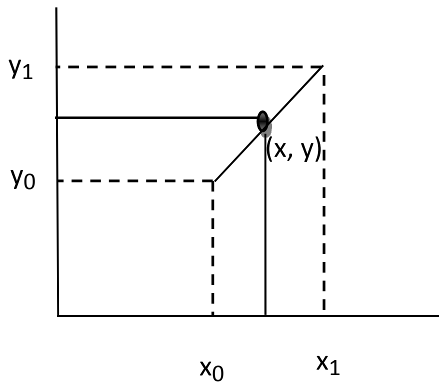


A gesture by a person is defined by a set of points. These points are then compared with another previously stored set of points (templates) using a measure of Euclidean distance. The recognised gesture is for which this distance is minimal (Figure 4). The number of points to define a gesture will depend on the speed of execution and the sampling frequency of the device. A gesture can be performed at different positions, different orientations and different scales. This is why the $P-recogniser was built in such a way that it is insensitive to all these types of variations. Moreover, learning a gesture is done once, i.e. requires only a one passage to create a template unlike HMM and neural networks. The $P-recogniser consists of three steps (Figure 4): resampled, scaled and translated, and comparison using Euclidean distance. Each of these steps is explained in more detail below.

- *Resample the Point Path*

  To compare the gesture made with pre-existing templates we must first resample the created set of points (stroke). Firstly, we calculated the total points length created by the gesture. Secondly, we divided this total by *N*-1 where *N* should be not too small which would cause a significant loss of accuracy or too big that would cause too many comparisons while consuming time (formula (1)). We look down the path of the original gesture, and then we determine the points by using the linear interpolation (formula (2)). It allows us calculate the distance between the points 2 by 2 of gestures. At the end of this first step and in the current time all gestures must have the same size *N*.

  $$I = length \div (N-1) \tag{1}$$

  $$\begin{aligned} \frac{y-y_0}{x-x_0} &= \frac{y_1-y_0}{x_1-x_0} \\ y &= y_0 + (x-x_0)\frac{y_1-y_0}{x_1-x_0} \end{aligned} \tag{2}$$

- *Scale and Translate*

  The gesture is scaled relative to the reference rectangle (reference square or bounding box). This is calculated by taking the minimum and maximum of x, y and z points (min_x, max_x, min_y, max_y, min_z, max_z) (formula (3)). After scaling, the gesture is translated to a reference point. To simplify computation, we choose (0, 0, 0) as centroid to translate the gesture.

$$Scale = Max\left(x_{max} - x_{min}, y_{max} - y_{min}, z_{max} - z_{min}\right) \tag{3}$$

- *Comparison using Euclidean distance*

  The list of obtained points was then compared to each of the existing templates. The template corresponds to a reference gesture, which is associated to a list of points and a name. Each comparison, which is based on a Euclidean distance, is used to calculate a score that reflects the degree of similarity between the template and the emitted gesture (formula (4)). The template that gets the highest score is considered the gesture executed (Vatavu et al., 2012). Owing to its straightforward internal representation of gestures as point clouds, we thought to use the $P in the context of recognition of stereotyped gestures of children with ASD. In this work, our data, which is acquired by the Kinect sensor, is in a space of three dimensions (x, y, z). That is why, we emerged the $P Point-Cloud Recogniser 2D to 3D. We chose this algorithm because it can handle both single and multi-stroke gestures efficiently, which allows to recognise easily complex gestures. On the other hand, we have been interested in gesture recognition independently of children. $P achieved 99.3% for user-dependent and 96.6% for user-independent testing to Euclidean's distance performance 99.5% and 97.1% (Vatavu et al., 2012). The algorithm is based on Euclidean distances (formula (4)); it computes the sum of distances between matched points (i.e. the distance between the two gestures (C and T)). The task of the recogniser is to match the point cloud of the candidate gesture (C) to the point cloud of each template (T) in the training set and compute a matching distance. In the tradition of the nearest-neighbour approach, the template located at the smallest distance from C delivers the classification result.

$$\sum_{i=1}^{n} \left\| C_i - T_j \right\| = \sum_{i=1}^{n} \sqrt{\left(C_{i.x} - T_{j.x}\right)^2 + \left(C_{i.y} - T_{j.y}\right)^2 + \left(C_{i.z} - T_{j.z}\right)^2} \tag{4}$$

3.3 Experimental protocol

In this section, we explain our application which detects stereotypical movements among autistic individuals using the Kinect sensor and the $P Point-Cloud Recogniser algorithm. We have been interested in five stereotyped gestures of autistic children: body rocking, hand flapping, fingers flapping, hand in the face and hand behind back (Figure 5). The choice of these gestures was made after a deep study of autism specifically in CARS (Schopler et al., 1980), in cooperation with a psychiatrist and researcher in the field of autism in the hospital CHU (Centre Hospitalier Universitaire Hassan II in Fez). In our study, we decide that these five gestures are the most popular and common among children with ASD. Moreover, in the literature, most of the work is based on one or two stereotyped gestures (Albinali et al., 2009; Goodwin et al., 2011). In our case, we have used five stereotyped movements, which will provide more information to the clinicians and doctors.

The first functionality of our system is to capture the data. For this purpose, we decided to use the Microsoft Kinect sensor to detect posture. We chose it because it does not require any contact with the body, unlike other systems that require people to wear clothing or attach sensors directly to their bodies, which puts the child in an abnormal environment (Gonçalves et al., 2012a; Min and Tewfik, 2010a; Min and Tewfik, 2010b; Plötz et al., 2012). Additionally, it is an inexpensive commodity hardware and readily available. As we described in Section 3, the Kinect sensor provides around 20 joints positions skeleton tracker. We focused on five types of joints for data acquisition: head, shoulder, elbow, wrist and hand. The joints choice was made because we were interested in movements that are in connection with these joints. The following diagram shows the steps of the general system (Figure 6).

In order to acquire gestures data, we had difficulties in capturing data from autistic children, whether on administrative or parental level. In order to avoid blocking at this stage, we decided to imitate these gestures by non-autistic people and test our system before moving to autistic children. Normal individuals (non-autistic people) have interpreted and imitated the gestures of autistic children by watching real video sequences. Each gesture is characterised by a sequence of joints movements. The selection joints are done according to these five gestures to be detected. The gestures are identified by five types of joints: head, shoulder (right, left, and centre), elbow (right, left), wrist (right, left) and hand (right, left) (Figure 3). These joints are extracted using Kinect sensor. The Kinect skeletal tracking allows the application to choose the skeleton to track among the six individuals recognised in the field of view. By default, skeletal tracking will select the first two recognised persons in the field of view. In our system, we tracked only one skeleton. To detect skeleton, the person needs to be in front of the sensor, making sure the sensor can see their head and upper body (about 2 metres); no specific pose or calibration action needs to be taken for a person to be tracked. Whenever a new skeleton frame is ready, we extract information of the five joints on space (x, y, z) during a specific time period, which gives us a gesture represented by a vector in space x, y, z. This vector should be normalised so that the gesture reaches a level determined in advance. The data were stored independent of the person and the gesture speed.

We evaluated our system with our own database, containing ten normal non-autistic people, each person was asked to make each gesture ten times. This gave us a database of 500 gestures. Thereafter, we did the test on five real patients (autistic children) where each child has one or more stereotyped gestures. We obtained a test data of ten stereotyped gestures of autistic children.

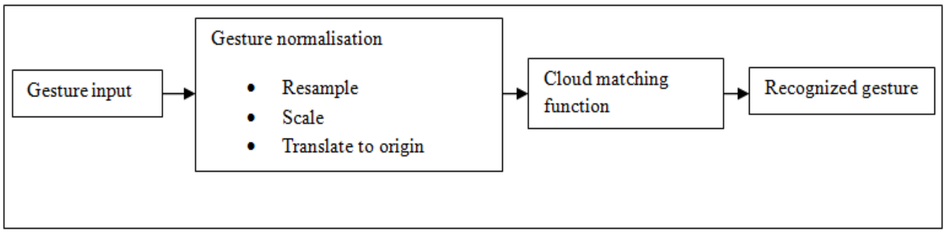
After data acquisition, we used 50 gestures in the learning (ten training for each gesture type) (formula (5)). To recognise a new gesture, the $P Point-Cloud algorithm takes as input a list of points that corresponds to this gesture (vector). The gesture vector goes through preprocessing phases: resampling, scaling, translation (see Section 3.2).

Afterwards, $P recogniser is used to match the point cloud of the candidate gesture (new gesture) to the point cloud of each template in the training set and compute a matching distance.

$$learning_{data} = \{g_1, g_2, \ldots, g_{10}\} \text{ where } g \text{ represents the gesture} \tag{5}$$

The application was based on WPF (Windows Presentation Form) and C# language. After having initialised the application, it was necessary to wait until the Kinect detects a person. The interface was segmented into different areas (Figure 7). Area (a) allows adding a new gesture to our database. So far, we have treated just five types of gestures (area c). Area (b) presents gesture recognition in real time. In area (d), we store the list of all the detected gestures in the observation. Finally, area (e) shows the RGB, the skeleton of the child obtained by the Kinect sensor.

Figure 4 $P Point-Cloud gesture recogniser



Note: Cloud matching function: match two clouds (points and template) by performing repeated alignments between their points (each new alignment starts with a different starting point) (Vatavu et al., 2012).

Figure 5 Stereotyped gestures of autistic children

Figure 6 General system of automatic detection and recognition of stereotyped movements (see online version for colours)

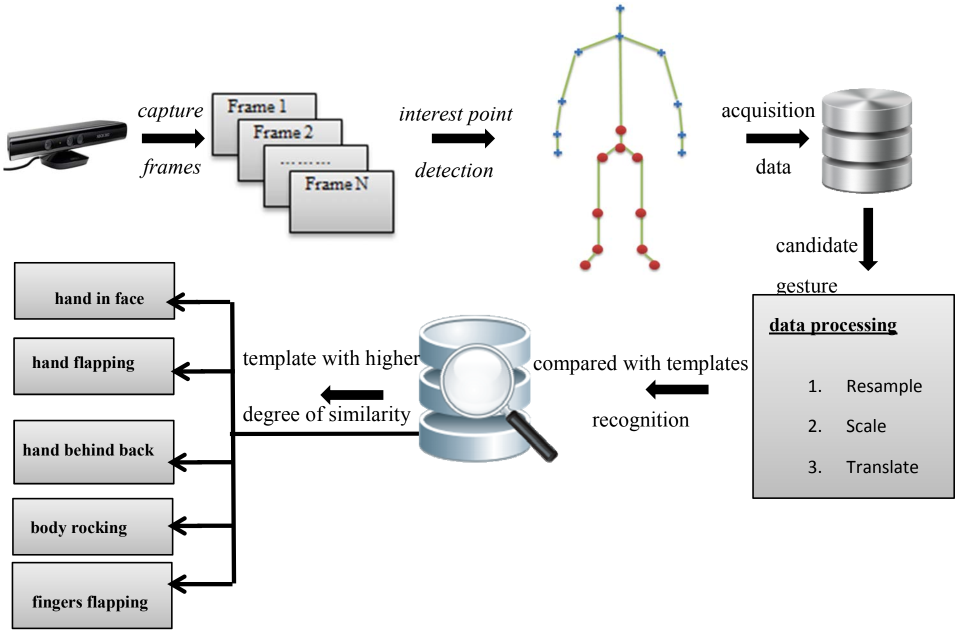
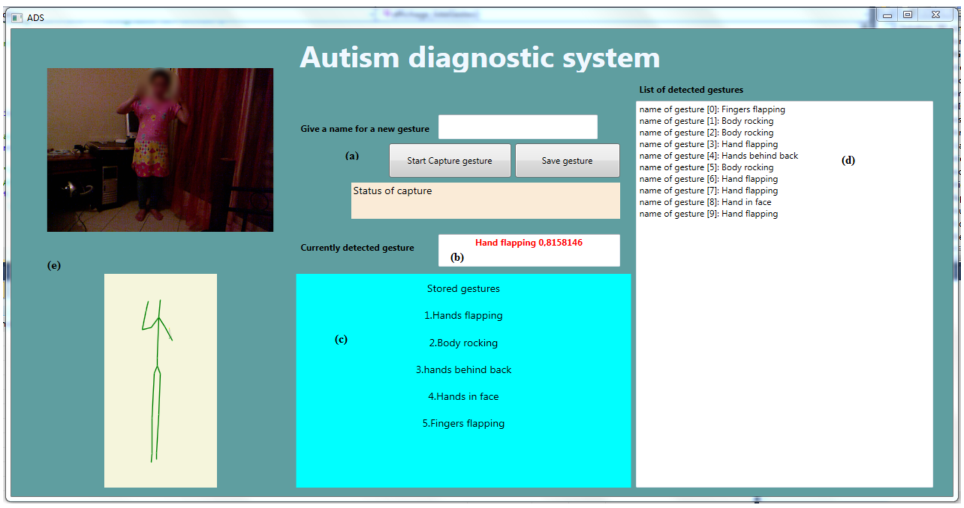


Figure 7 Interface of our application



4 Results and discussions

In this section, we show and discuss the interesting results obtained by our system. Our recognition method can be effectively used for detecting stereotyped gestures or movements among autistic children. The results were obtained through these models of repetitions of the gestures made by these non-autistic people (Section 4). Table 1 shows the recognition of the stereotyped movements. Our system can recognise five stereotyped gestures: body rocking, hand flapping, fingers flapping, hand in the face and hand behind back. So, when the gesture does not belong to any class of the five categories, it is classified as unknown gesture.

$$test_{data} = \{p_1, p_2, \ldots, p_{10}\} \tag{6}$$

We compared the results of our approach based on the $P algorithm with a DTW algorithm. This algorithm is based on a dynamic programming for measuring similarity between two temporal sequences which may vary in time of different durations. The core notion of the algorithm is to compare the characteristics of a particular pattern which will be checked against a template reference previously registered (Gonçalves et al., 2012a; Gonçalves et al., 2012b; Gonçalves et al., 2012c; Gonçalves et al., 2014; Gonçalves et al., 2014). The DTW has been used in voice recognition patterns, signature verification, and gesture recognition (Myers et al., 1980; Sakoe and Chiba, 1978).

The values in Tables 1 and 2 (Figures 8–9) are the results of the maximum probability assignment rule, which assigns gestures to the class for which they have the highest posterior probability of template gestures. In this method, the task of the recogniser remains to match the point cloud of the candidate gesture (*C*) to the point cloud of each template (*T*) in the training set and compute a matching distance. The distance is used to depict the greatest similarity between gestures by calculating the minimum distance between them (formula (1)).

After testing the application with non-autistic people, and obtaining a recognition rate that exceeds 90%. It is necessary to test the system in the field of autism. Table 2 shows the results obtained with autistic children from 'Association Miroir Pour L'enfant Autiste'. The system was tested on five ASD children aged between 5 and 10 years. We obtained similar results to previous with $P recogniser and non-autistic children, which proves the great reliability and robustness of the proposed system. The result in Table 2 (Figure 9) was obtained using the training data of normal people (non-autistic) and the testing data were captured from autistic children.

In order to quantify the recognition rate, Table 3 shows the confusion matrix for our method using $P recogniser. We achieve a 94% recognition rate with an error rate of 6% on this set of data.

To evaluate the performance of our stereotyped gesture recognition with our approach we used the recognition rate defined as follow:

$$accuracy = \frac{K}{N} \tag{7}$$

where *K is* the number of test samples correctly recognised and *N* is the total number of test samples. The number $N = 10$ indicates the number of test cases for each gesture.

Tables 1 and 2 show a satisfaction at the level of the $P algorithm by comparing it with the DTW algorithm. The $P recogniser and DTW algorithms work by creating a template time series for each gesture that needs to be recognised, and then warping the real-time signals to each of the templates to find the best match independently of who made the gesture and how they did it. $P recogniser reached a recognition rate of over 90% with only ten training per gesture. By cons, by using the DTW algorithm, we had a percentage of 44% false detection using the same training data (Figure 10). The advantage of having a reduced learning base is the reduction the calculation time, especially when using the Kinect sensor in real time. $P recogniser also gives good results, because, among its advantages, we found it sensitive to any type of variation: person, speed of movement, equipment, orientation. We chose comparing it with DTW because it is a powerful classifier that works very well for recognising temporal gestures. Our test results proved that $P recogniser is more efficient than DTW, $P recogniser allows reducing the number of gesture templates in the database, and thus reducing the computational cost and it gave a very good recognition rate.

To summarise, our study provides a novel clinical tool which facilitates for doctors the diagnosis of ASD. Traditional methods for measuring stereotyped movements rely primarily on paper-and-pencil rating scales (Lewis and Bodfish, 1998; Rojahn et al., 2000), direct behavioural observation, video-based methods (Matson and Nebel-Schwalm, 2007), and kinematic analyses (Sprague and Newell, 1996; Newell et al., 1999; Berkson et al., 2001), all of which are potentially limiting. To overcome these problems, the early studies in this field have proposed to use the accelerometer to detect automatically the stereotyped movement, but the use of it on wrists presents a disadvantage of not being accepted by some autistic children. For all of these reasons, the aim of the current work was to explore whether the Microsoft Kinect sensor combined to gesture recognition algorithms can provide an automated measure of stereotyped movements in real time. And that may be more objective, detailed, and precise than rating scales and direct behavioural observation, and more time-efficient and mobile than video-based methods and kinematic analyses. We evaluated our method, the total rate of correct recognitions over all gestures was 94% related with the confusion matrix given in Table 3 and the most of gestures are perfectly classified with a very good classification rate, which are satisfactory results compared to literature. For the next steps, we consider that is necessary to build a bigger database for autistic children, using people with different ages and ethnicities and also take into account most of the existing stereotyped movements to make the recognition more general.

Table 1 The obtained results with non-autistic people

Result	*Recogniser*	*P1*	*P2*	*P3*	*P4*	*P5*	*P6*	*P7*	*P8*	*P9*	*P10*	*Rate*
Hand flapping	$P	90%	90%	91%	92%	88%	95%	91%	96%	92%	90%	91.5%
	DTW	84%	83%	74%	77%	76%	81%	71%	72%	73%	80%	77.1%
Fingers flapping	$P	93%	91%	92%	89%	91%	92%	94%	91%	90%	91%	91.4%
	DTW	71%	78%	82%	67%	70%	71%	75%	87%	71%	88%	76%
Hand in the face	$P	95%	90%	90%	92%	90%	89%	90%	92%	91%	87%	90.6%
	DTW	77%	76%	85%	72%	75%	77%	70%	77%	87%	79%	77.5%
Hands behind back	$P	91%	95%	92%	91%	94%	92%	85%	95%	95%	91%	92.1%
	DTW	84%	89%	77%	70%	83%	72%	78%	81%	71%	73%	77.8%
Body rocking	$P	93%	90%	91%	94%	88%	95%	95%	91%	94%	91%	92.2%
	DTW	73%	88%	80%	72%	74%	77%	70%	79%	70%	81%	76.4%

Note: *P* represents the test data for each person (formula (6)).

Table 2 The obtained results with autistic children

Result	*Hand flapping*		*Fingers flapping*		*Hand in the face*		*Hands behind back*		*Body rocking*	
Recogniser	*$P*	*DTW*	*$P*	*DTW*	*$P*	*DTW*	*$P*	*DTW*	*$P*	*DTW*
Child 1	–	–	91%	68%	92%	70%	–	–	90%	71%
Child 2	–	–	81%	67%	–	–	–	–	90%	73%
Child 3	92%	71%	–	–	–	–	–	–	–	–
Child 4	–	–	–	–	–	–	94%	75%	91%	70%
Child 5	91%	67%	–	–	91%	71%	–	–	–	–

Table 3 Confusion matrix for the stereotyped gesture recognition using $P recogniser

Class	*Hand flapping*	*Fingers flapping*	*Hand in the face*	*Hands behind back*	*Body rocking*	*Rate error*
Hand flapping	**9** **18.0%**	**1** **2.0%**	0 0.0%	0 0.0%	0 0.0%	90% 10.0%
Fingers flapping	**2** **4.0%**	**8** **16.0%**	0 0.0%	0 0.0%	0 0.0%	80% 20.0%
Hand in the face	0 0.0%	0 0.0%	**10** **20.0%**	0 0.0%	0 0.0%	100% 0.0%
Hands behind back	0 0.0%	0 0.0%	0 0.0%	**10** **20.0%**	0 0.0%	100% 0.0%
Body rocking	0 0.0%	0 0.0%	0 0.0%	0 0.0%	**10** **20.0%**	100% 0.0%
Rate error	81.82% 18.18%	88.89% 11.11%	100% 0.0%	100% 0.0%	100% 0.0%	**94%** **6.0%**

Figure 8 Classification rate for five stereotyped movements of $P vs. DTW with non-autistic people

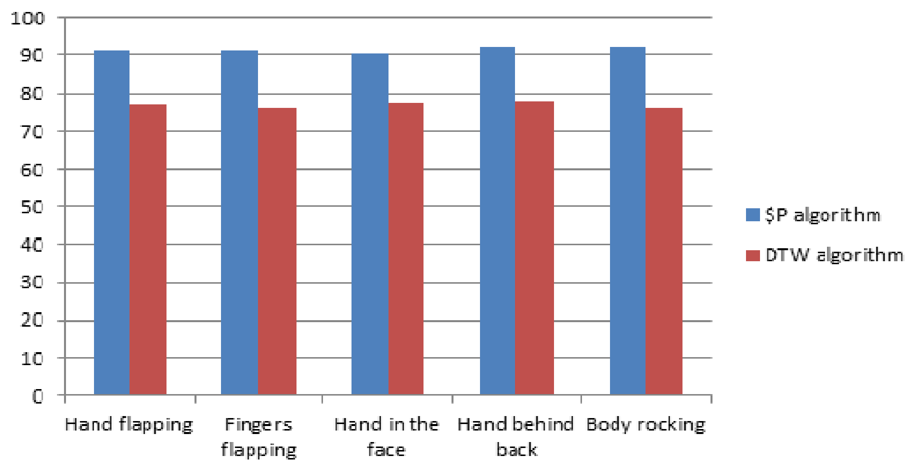


Figure 9 Classification rate for five stereotyped movements of \$P vs. DTW with autistic children

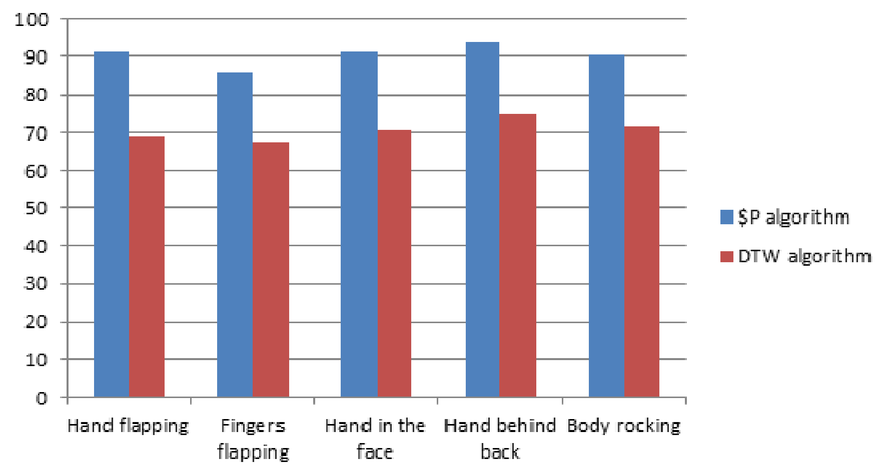
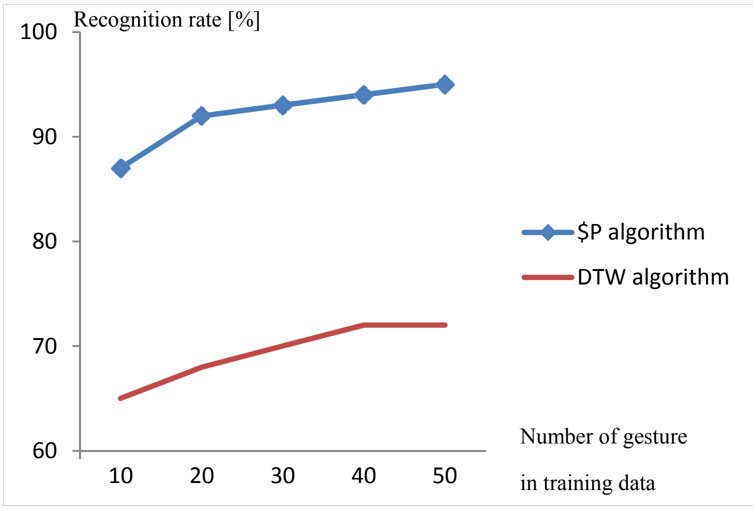


Figure 10 Recognition performance of \$P vs. DTW



5 Conclusion

In this paper, we create a new system based on the Kinect sensor to automatically detect stereotyped movements, in particular five of the gestures: body rocking, hand flapping, fingers flapping, hand in the face and hand behind back. We have applied the \$P recogniser to identify the stereotyped movements. In this work, we have demonstrated that the \$P method, which we converted in three-dimensional space, can be useful in reducing the time and complexity of the gesture recogniser system. The system has been tested on children with ASD. In the evaluation process, the results are measurable and quantifiable. The application detects 94% of the stereotyped gestures, which are satisfactory results compared to literature. For future work, it is necessary to improve the gesture recognition algorithms with the Kinect sensor, due to the false-positives detected and build a real database for autistic children. Subsequently, it is necessary to complete the system to help the doctors make decisions during diagnosis of autism.

References

Alabbasi, H.A., Moldoveanu, P. and Moldoveanu, A. (2015a) 'Real time facial emotion recognition using Kinect V2 sensor', *IOSR Journal of Computer Engineering*, Vol. 17, No. 3, pp.61–68.

Alabbasi, H.A., Moldoveanu, P., Moldoveanu, A. and Zaid, S. (2015b) 'Facial emotion expressions recognition with brain activities using Kinect sensor V2', *International Research Journal of Engineering and Technology*, Vol. 2, No. 2, pp.421–428.

Albinali, F., Goodwin, M.S. and Intille, S.S. (2009) 'Recognizing stereotypical motor movements in the laboratory and classroom : a case study with children on the autism spectrum', *The Proceedings of the 11th International Conference on Ubiquitous Computing*, ACM, Orlando, FL, USA, pp.71–80.

Berkson, G., Andriacchi, T. and Sherman, L. (2001) 'More information on the nature of stereotyped body-rocking', *American Journal on Mental Retardation*, Vol. 106, pp.205–208.

Bhattacharjee, S., Das, Z. and Neogi, B. (2015) 'Design and simulation aspect towards modelling of automatic cardiovascular disorder diagnosis system', *International Journal of Biomedical Engineering and Technology*, Vol. 19, No. 4, pp.303–315.

Coronato, A., De Pietro, G. and Paragliola, G. (2014) 'A situation-aware system for the detection of motion disorders of patients with Autism Spectrum Disorders', *Journal Expert Systems with Applications*, Vol. 41, No. 17, pp.7868–7877.

Filipek, P.A., Accardo, P.J., Baranek, G.T., Cook Jr., E.H., Dawson, G., Gordon, B., et al. (1999) 'The screening and diagnosis of autistic spectrum disorders', *Journal of Autism and Developmental Disorders*, Vol. 29, No. 6, pp.439–484.

Goodwin, M.S., Intille, S.S., Albinali, F. and Velicer, W.F. (2011) 'Automated detection of stereotypical motor movements', *Journal of Autism and Developmental Disorders*, Vol. 41, No. 6, pp.770–782.

Gonçalves, N., Rodrigues, J.L., Costa, S. and Soares, F. (2012a) 'Automatic detection of stereotyped hand flapping movements: two different approaches', *The 21st IEEE International Symposium on Robot and Human Interactive Communication*, IEEE, Paris, France.

Gonçalves, N., Rodrigues, J.L., Costa, S. and Soares, F. (2012b) 'Automatic detection of stereotypical motor movements', *The 26th European Conference on Solid-State Transducers, EUROSENSOR*, Kraków, Poland.

Gonçalves, N., Rodrigues, J.L., Costa, S. and Soares, F. (2012c) 'Preliminary study on determining stereotypical motor movements', *The Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, IEEE, San Diego, CA, USA.

Gonçalves, N., Costa, S., Rodrigues, J. and Soares, F. (2014) 'Detection of stereotyped hand flapping movements in Autistic children using the Kinect sensor: a case study', *The IEEE International Conference on Autonomous Robot Systems and Competitions*, IEEE, Espinho, Portugal.

Kang, J.Y., Kim, R., Kim, H., Kang, Y., Hahn, S., Fu, Z., Khalid, M.I., Schenck, E. and Thesen, T. (2016) 'Automated tracking and quantification of autistic behavioral symptoms using Microsoft Kinect', *Studies in Health Technology and Informatics*, Vol. 220, pp.167–170.

Kanner, L. (1943) 'Autistic disturbances of affective contact', *Nervous Child*, Vol. 2, pp.217–250.

Kim, S.H. and Lord, C. (2010) 'Restricted and repetitive behaviors in toddlers and preschoolers with autism spectrum disorders based on the Autism Diagnostic Observation Schedule (ADOS)', *Autism Research*, Vol. 3, No. 4, pp.162–173.

Kumar, S.S. and Moni, R.S. (2011) 'Diagnosis of liver tumour from CT images using contourlet transform', *International Journal of Biomedical Engineering and Technology*, Vol. 7, No. 3, pp.276–290.

Lam, K.S. and Aman, M.G. (2007) 'The repetitive behavior scale-revised: independent validation in individuals with autism spectrum disorders', *Journal of Autism and Developmental Disorders*, Vol. 37, No. 5, pp.855–866.

Larsson, A., Bonet, I., Rázuri, J.G., Rahmani, R., Sundgren, D. and Moran, A. (2015) 'Recognition of emotions by the emotional feedback through behavioral human poses', *International Journal of Computer Science Issues*, Vol. 12, No. 1, pp.7–17.

Lewis, M.H. and Bodfish, J.W. (1998) 'Repetitive behavior disorders in autism', *Mental Retardation and Developmental Disabilities Research Reviews*, Vol. 4, pp.80–89.

Lowe, D.G. (2004) 'Distinctive image features from scale invariant keypoints', *International Journal of Computer Vision*, Vol. 60, pp.91–110.

Mahmoud, M., Morency, L-P. and Robinson, P. (2013) 'Automatic multimodal descriptors of rhythmic body movement', *The 15th ACM International Conference on Multimodal Interaction*, ACM, Sydney, Australia.

Matson, J.L. and Nebel-Schwalm, M. (2007) 'Assessing challenging behaviors in children with autism spectrum disorders: a review', *Research in Developmental Disabilities*, Vol. 28, pp.567–578.

Min, C-H. and Tewfik, A. (2010a) 'Automatic characterization and detection of behavioral patterns using linear predictive coding of accelerometer sensor data', *The Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Vol. 2010, pp.220–223.

Min, C-H. and Tewfik, A.H. (2010b) 'Novel pattern detection in children with Autism Spectrum Disorder using iterative subspace identification', *The IEEE International Conference Acoustics Speech and Signal Processing (ICASSP)*, IEEE, Dallas, TX, USA.

Myers, C., Rabiner, L. and Rosenberg, A. (1980) 'Performance tradeoffs in dynamic time warping algorithms for isolated word recognition', *The IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. 28, No. 6, pp.623–635.

Newell, K.M., Incledon, T., Bodfish, J.W. and Sprague, R.L. (1999) 'Variability of stereotypic body-rocking in adults with mental retardation', *American Journal on Mental Retardation*, Vol. 104, pp.279–288.

Noris, B. (2011) *Machine Vision-Based Analysis of Gaze and Visual Context : An Application to Visual Behavior of Children with Autism Spectrum Disorders*, École Polytechnique Fédérale, Lausanne.

Penne, T. (2011) *Développement d'un système de tracking vidéo sur caméra robotisée*, Université Blaise Pascal – Clermont-Ferrand 2, Aubière. [In French]

Plötz, T., Hammerla, N. and Rozga, A. (2012) 'Automatic assessment of problem behavior in individuals with developmental disabilities', *The 14th International Conference on Ubiquitous Computing (Ubicomp)*, ACM, Pittsburgh, PA, USA.

Postawka, A. and Sliwinski, P. (2015) 'Recognition and modeling of atypical children behavior', *The Artificial Intelligence and Soft Computing*, Vol. 9119, pp.757–767.

Rahimi, N. and Golzarian, J. (2015) 'Assessing the Kinect technology acceptance/diffusion into healthcare (the case of physical therapy)', *International Journal of Medical Engineering and Informatics*, Vol. 7, No. 3, pp.269–292.

Rojahn, J., Matlock, S.T. and Tasse, M.J. (2000) 'The stereotyped behavior scale: psychometric properties and norms', *Research in Developmental Disabilities*, Vol. 21, pp.437–454.

Sakoe, H. and Chiba, S. (1978) 'Dynamic programming algorithm optimization for spoken word recognition', *The IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. 26, No. 1, pp.43–49.

Schopler, E., Reichler, R. and DeVellis, R. (1980) 'Toward objective classification of childhood autism: childhood autism rating scale (CARS)', *Journal of Autism and Developmental Disorders*, Vol. 10, No. 1, pp.91–103.

Sivalingam, R., Cherian, A., Fasching, J., Walczak, N., Bird, N., Morellas, V. and Papanikolopoulos, N. (2012) 'A multi-sensor visual tracking system for behavior monitoring of at-risk children', *The Proceedings of the IEEE International Conference on Robotics and Automation*, IEEE, Saint Paul, MN, USA.

Sprague, R.L. and Newell, K.M. (Eds) (1996) *Stereotyped Movements: Brain and Behavior Relationships*, American Psychological Association, Washington, DC.

Sundar, R.S. and Goecke, R. (2014) 'Detecting self-stimulatory behaviours for autism diagnosis', *The IEEE International Conference on Image Processing*, IEEE, Paris, France.

Upadhyay, A. and Dewangan, A.P.A.K. (2016) 'Facial expression recognition: a review', *International Research Journal of Engineering and Technology*, Vol. 3, pp.1616–1620.

Vatavu, R., Anthony, L. and Wobbrock, J. (2012) 'Gestures as point clouds: a $P recognizer for user interface prototypes', *The 14th ACM International Conference on Multimodal Interaction*, ACM, Santa Monica, CA, USA.

Verbeke, N. (2007) *Suivi d'objets en mouvement dans une séquence vidéo*, Université Paris Descartes, Paris. [In French]

Westeyn, T., Vadas, K., Bian, X., Starner, T. and Abowd, G.D. (2005) 'Recognizing mimicked autistic self-stimulatory behaviors using HMMs', *The IEEE International Symposium on Wearable Computers*, IEEE Computer Society, Washington, DC.

Witten, I.H. and Frank, E. (1999) *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*, Morgan Kaufmann, San Francisco, CA.