Semi-automatic behavioral change-point detection: a case study analyzing children interactions with a social agent

Vito Monteleone, Liliana Lo Presti, Marco La Cascia,
Soizic Gauthier, Jean Xavier, Mohamed Zaoui, Alain Berthoz,
David Cohen, Mohamed Chetouani and Salvatore Anzalone

***Abstract*—The study of human behaviors in cognitive sciences provides clues to understand and describe people's personal and interpersonal functioning. In particular, the temporal analysis of behavioral dynamics can be a powerful tool to reveal events, correlations and causalities but also to discover abnormal behaviors. However, the annotation of these dynamics can be expensive in terms of temporal and human resources. To tackle this challenge, this paper proposes a methodology to semi-automatically annotate behavioral data. Behavioral dynamics can be expressed as sequences of simple dynamical processes: transitions between such processes are generally known as change-points. This paper describes the necessary steps to detect and classify change-points in behavioral data by using a dataset collected in a real use-case scenario. This dataset includes motor observations from children with typical development and with neuro-developmental disorders. Abnormal movements which are present in such disorders are useful to validate the system in conditions that are challenging even for experienced annotators. Results show that the system: can be effective in the semi-automated annotation task; can be efficient in presence of abnormal behaviors; may achieve good performance when trained with limited manually annotated data.**

***Index Terms*—Change-point; Human Behavior; Semi-automated annotation**

I. Introduction

In their most general definition, behaviors are defined as "the internally coordinated responses to internal and/or external stimuli" [1]. In humans, this translates to the individuals' responses related to internal or perceived environmental stimuli, mediated by psychological states.

The study of such responses is carried out by psychology, psychiatry, neurosciences, and, more in general, by cognitive sciences, with the goal of providing clues about the inner mechanisms of the human brain that underlie perceptual and decisional processes [2]. While this knowledge can facilitate the study of personal and interpersonal people's functioning [3], it becomes particularly useful in the treatment of complex psychopathologies where such mechanisms are damaged [4]. Notably, such insights on human behaviors can also be applied to the design of products, systems and devices we use every day by making them easier, more comfortable and less frustrating [5].

Computer science contributed to such studies not only as a metaphor for innovative computational models [6] able to describe cognitive processes but also with useful methods and tools to capture and automatically or semi-automatically characterize humans' behaviors [7], [8]. However, automatic/semi-automatic behavioral analysis is not straightforward since it is unclear how observable behaviors should be measured and characterized. The problem is made particularly difficult due to the great behavioral heterogeneity that can characterize humans' responses, in particular during social interaction in which individuals' behaviors arise from interpersonal exchanges. Moreover, factors such as age, motor capabilities or the presence of cognitive impairments can make the problem even harder.

A. Human Behavior Characterization

Studies about observable human behaviors take advantage of measurements of multi-modal characteristics [9] such as gait, body poses, body movements, eye gaze, facial expressions, speech, or turn-taking. All these measurements are derived from signals acquired by a wide variety of sensors (microphones, cameras, wearable sensors such as accelerometers or gyroscopes). Thus, human behavioral responses can be represented in terms of one or more synchronized, multi-modal signals [10]. In this sense, human behaviors emerge from a mixture of several (often hidden) factors and through many communication channels [8]. For instance, the control of continuous motor behaviors involves simple coordinated movements of distinct joints as presented in [11]. Coordinating such movements allows individuals to both complete complex movements and skillfully adjust their balance to maintain control (e.g., adjusting posture) [12].

In literature, the problem of characterizing humans' behaviors has been approached mainly through three methodologies:

Vito Monteleone, Liliana Lo Presti and Marco La Cascia are with the Department of Engineering, University of Palermo, Palermo, Italy

David Cohen is head of the department of Child and Adolescent Psychiatry at La Salpêtrière Medical Hospital in Paris, France

Soizic Gauthier is with CRPMS Laboratory - EA 3522, Université Paris Diderot, Sorbonne Paris Cité in Paris, France

Jean Xavier is professor at Centre Hospitalier Laborit and at the Autism Resources Center Poitou-Charentes in Poitiers, France

Mohamed Zaoui is with the Collège de France in Paris, France

Alain Berthoz is professor emeritus at the Collège de France in Paris, France

Mohamed Chetouani is with the Faculté des Sciences et Ingénierie, Sorbonne Université in Paris, France

Salvatore Anzalone (corresponding author) is with the CHArt Laboratory - EA 4004, Université Paris 8 in Saint Denis, France. email: sanzalone@univ-paris8.fr

Manuscript received October 23, 2019; revised XXXX

- **Manual annotation**: trained human annotators carefully label the observations according to a previously agreed set of labeling criteria. Consistency between annotators is verified through inter-rater reliability measures like the Cohen's kappa or the Intra-class Correlation Coefficient (ICC) [13];
- **Automated annotation**: observed measurements are automatically processed according to a set of rules that can be explicitly programmed or implicitly inferred from data [14], [15];
- **Semi-automated annotation**: a subset of manually annotated measurements is employed as sample pattern into an automated system that characterizes the remaining observations accordingly [16], [17], [18].

Manual annotation fully relies on the human effort to characterise and annotate observations. While it can achieve an elevated degree of precision, it is extremely expensive in terms of temporal and human resources demanded to the human annotators. Recently, crowd-sourcing systems, such as the Amazon's Mechanical Turk, have been used to reduce such costs. However, it is unclear if the annotation collected with such systems would have the minimum quality required for fine-grained human behavior analysis from a psychological point of view. At the same time, ethical issues arise in the case of analysis of sensitive data [19].

Fully automated annotation systems can be effective when explicit rules are an adequate tool to achieve the requested degree of performance; as an alternative, rules can be inferred through clustering techniques able to reveal patterns and recurrences in data. However, also in this case, a post-processing revision by an expert is needed in order to interpret and validate the discovered clusters.

In recent years, impressive performances have been achieved through the use of deep neural networks [20], [21]. However, such performances are associated with the exploitation of very large annotated training sets that, in the field of behavioral analysis, may not be available. Furthermore, such models are very complex and still difficult to interpret. On the contrary, semi-automated annotation tries to combine the benefits from both the automated and manual annotation methodologies while minimizing their drawbacks. First, this methodology tries to limit the human annotators' effort in labeling the data. Secondly, it attempts to develop models that are easy to check and interpret for experts that are not necessarily computer scientists. Although a manually annotated set of observations is still required, reliable performance can be obtained by using a set of data that is smaller than the one required to train fully automated annotation systems.

In automated and semi-automated systems, complex behaviors are often modeled in terms of dynamical systems [22], [23]. In particular, behavior dynamics can be expressed as sequences of simple, stationary or quasi-stationary dynamical processes, each one characterised by its own set of parameters [24], [25], [26]. In this sense, complex behaviors emerge as the switching over time among simple behavioral classes [27]. A simplified model would consist of a finite state machine in which states represent simple behaviors, while transitions from state to state represent changes from a behavioral class to another [28], [29]. In correspondence with such state transitions, the parameters of the dynamical system change in accord to the behavioral class involved in the transition itself. For this reason, such transitions are generally known as change-points [30], [31], [32], [33], [34].

According to the level of details used to characterize the observed behavior, three main techniques can be adopted:

- **Global approaches**: transitions among behaviors are ignored. Observed measures are treated as part of a single dynamical process characterized through comprehensive statistical descriptors [35], [7];
- **Slicing**: as in global approaches, transitions are ignored but behaviors are described by statistical descriptors of a finite sequence of temporal slices of the observed measures [36], [24], [37], [38];
- **Local characterization**: change-points are explicitly detected and exploited to segment and characterize the observed measures [39].

Global approaches achieve a rough characterization of humans' behaviors, ignoring the details of their temporal evolution. Consequently, they can be very effective in revealing and describing features that are stationary over time. In contrast to the global approaches, local techniques may result in a finer and detailed characterization of humans' behaviors [40], [41]. Nonetheless, due to the inner stochastic nature of the behavioral measurements, such characterization is more sensitive to small fluctuations ascribable to noise and to the performance of the sensors used to observe the behavior [42]. In this sense, by considering time-slices of the observed behaviors, slicing techniques may help to reduce or partially filter out these fluctuations. As a consequence, a statistical description of the behavior would reveal quasi-stationary features among different slices. Such methodology can be particularly useful to describe how observed behaviors evolve in time in a more refined way. Notably, this technique can be used to study first impressions and early events [37], to analyze long-term scenarios [43] or to evaluate before-and-after effects [44] of particular events [30].

On the other hand, local analysis of the observed measurements can highlight breakpoints and transitions from one behavior to another. Such change-points are a powerful tool to segment the observations and reveal events, correlations, causalities and synchronous phenomena, but also to discover abnormal behaviors or delays in the expected behavioral changes.

B. Scope of the Paper

The aim of this paper is the local characterization of behavioral observations related to humans' movements. In particular, we focus on the detection and recognition of motor behavioral transitions defined as changes in the temporal evolution of motor behaviors. We will propose a semi-automated annotation system, tackling the change-point detection problem through a general-purpose methodology that is suitable in presence of heterogeneous behaviors for which no a-priori model is available.

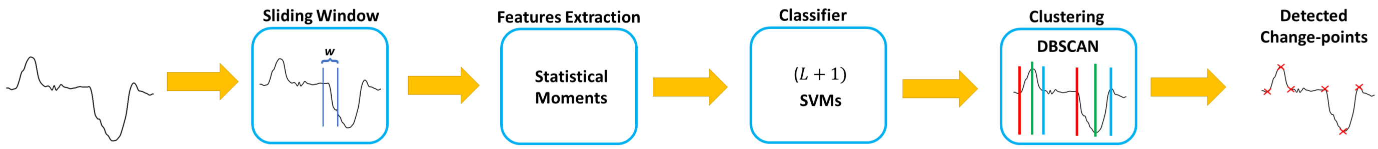

Fig. 1. The behavioral signal is analyzed by a sliding window approach. For each temporal window $\left(b_{i-\left[\frac{w}{2}\right]}, b_{i+\left[\frac{w}{2}\right]}\right)$, its description in terms of statistical moments $(\mu_i, \Sigma_i, S_i, K_i, \Delta_i,)$ is considered. This descriptor is fed as input to the SVM classifier that predicts the change-point class. Predictions for all possible temporal windows are aggregated through the DBSCAN clustering method. The time frame related to the cluster centers corresponds to the estimated change-points.

The main contributions of this paper are the characterization of the behavioral signal through statistical moments-based signatures, and the classification of change-points by Support Vector Machines (SVMs) [45]. Similarly to the methods in [46], [47], [48], we adopt a sliding window approach to extract feature descriptors from the data, and we detect and recognize change-points without any a-priori knowledge of the behavioral models. In contrast to the methods mentioned above, our feature descriptor embeds *contextual information* about the transitions from one behavior to another. In this way, our descriptor is not only able to fully characterize the statistical property of the change-point inside the sliding window, but also to take into account differences arising in the signal before and after the change-point, capturing the dynamics of the signal.

The methodology has been validated and tested in a real use-case scenario: a manually annotated dataset of humans' movements collected during an imitation task of a virtual agent [49], [50] acting as a tightrope walker [51]. The dataset involves behavioral observations from children with typical development (TD) but also children with neuro-developmental disorders (NDD): Autism Spectrum Disorder (ASD) and Developmental Coordination Disorder (DCD). The presence of abnormal movements [52], [53], [54], [55] from children with NDD gives the possibility of validating the proposed system in conditions that are challenging even for experienced annotators.

The behaviors annotated in the tightrope walker dataset are simple and quite stereotyped; however, their effective automated or semi-automated annotation can be generalized and transferred to a larger set of humans' behaviors. Achieving such effective annotation capabilities would result in fewer human resources committed to the manual analysis of the data, as well as on less time spent on this task.

In this paper, we want to demonstrate:

– The **feasibility** of the proposed approach in practical applications and its **effectiveness** in detecting and classifying change-points in humans' movements;
– The efficacy and the limits of the system while **tackling abnormal behaviors**, for instance when analyzing behaviors of children affected by NDD;
– The impact of the **size of the annotated training set** on the system performance to assess how many samples the training procedure requires to still achieve valuable results.

II. Change-points Detection

The assumption that the course of specific phenomena, represented in terms of time-series, follows the same fixed stationary process may not be realistic in several domains such as economics, business, engineering, medicine and social sciences. The discovery of specific time points in which the properties of the time-series change is an attracting and well-studied problem [30], [56], [57], [48], [34], [31], [32], [33]. A change-point is defined as the temporal boundary that separates two sequences of observations originating from two different statistical distributions. In this sense, change-point detection can be achieved by analyzing the parameters of such distributions.

Changes in the statistical moments of the distribution of the observations can be interpreted as possible change-points. Such changes can be particularly evident in the distribution of characteristics related to the observed measurements such as derivatives or energy spectra. Detection of change points is useful in the modelling and prediction of time series, and is found in application areas such as medical condition monitoring, climate change detection, speech and image analysis, and human activity analysis [30]. Detecting transitions based on the statistical moment changes is useful for segmenting motor activities, as for example the movement of an arm while performing the action of greeting someone. This movement can be broken down into atomic submovements [58], each of which can be identified by means of change-points.

A simple but effective change-point detection method is the CUSUM (cumulative sum) test [59], that involves the calculation of a cumulative sum of the observed characteristics: when the value of the cumulative sum exceeds a certain threshold value, a change-point has been detected.

The change-point detection problem can be stated in terms of a hypothesis test [60], [61] in which the null hypothesis H_0 assumes the absence of a change-point at a specified time in the time-series, while the alternative hypothesis H_1 assumes that there is one. Assuming that each observation in a time-series originated from some distribution fully described by a set of parameters θ_t with $t \in [0, N)$ indicating time, the hypotheses H_0 and H_1 can be formulated as follows:

$$H_0 : \theta_0 = \cdots = \theta_{p-1} = \theta_p = \theta_{p+1} = \cdots = \theta_{N-1} \quad (1)$$

$$H_1 : \theta_0 = \cdots = \theta_{p-1} = \theta_p \neq \theta_{p+1} = \cdots = \theta_{N-1} \quad (2)$$

where p represents the time when a change-point event occurs. The key factor in the formulation of H_1 is the inequality $\theta_p \neq \theta_{p+1}$: at some point in the time-series, and precisely between $t = p$ and $t = p + 1$, the underlying distribution changes. Consequently, the hypothesis H_1 assumes the presence of two different distributions characterized by the parameters θ_A and θ_B respectively. When $t \leq p$, $\theta_t = \theta_A$; whenever $t > p$, $\theta_t = \theta_B$.

According to this formulation, it is possible to define λ as the ratio between the likelihoods associated with the two hypotheses H_0 and H_1:

$$\lambda(t) = \frac{\mathcal{L}(H_0|t)}{\mathcal{L}(H_1|t)}. \tag{3}$$

This likelihood-ratio can provide a general decision test to classify observations in a sequence:

$$\begin{cases} \text{if } \lambda(t) > c & \text{do not reject } H_0 \\ \text{if } \lambda(t) \leq c & \text{reject } H_0 \end{cases} \tag{4}$$

where c is a predefined threshold selected to obtain a specified significance level. Without going further into the details, likelihood-ratio based classification systems, such as the Generalized likelihood ratio (GLR) [62] or the marginalized likelihood ratio (MLR) [63], exploit this model by finding suitable parameters that maximize the system capabilities of detecting change-points in the observed measurements.

Non-parametric tests have also been developed for applications in which no prior knowledge about the observed process distribution is available [64], [65]. Other statistical methods use Bayesian prior distributions to incorporate time-dependent information [30], [66].

Without being exhaustive, in recent years many machine learning algorithms have been designed or adapted to the change-point detection problem. Such methods usually employ sliding windows of subsequent observations [46], [47], [48] to reinforce their noise resistance or to capture smoothed transitions that otherwise could be difficult to detect. Under this point of view, supervised approaches have been proposed [46], [47], [67], [68], [69] to model the change-point detection problem as a binary classification problem. In this case, the goal is to discriminate between state transition sequences (i.e., observation sequences including the change-point) and within-state sequences. Multi-class classifiers have been adopted to solve the change-point estimation problem [30]. In that case, the goal is the recognition of the type of detected change-point, namely the kind of state transition that arises. Alternatively, unsupervised learning algorithms have been proposed to discover patterns in unlabeled data. Subspace modelling [57], [70] and graph-based technique [71], [72], [73], in particular, have been used as clustering methods to detect change-points.

III. Methodology

We assume that humans' behaviors are represented through a multi-dimensional time-series $B = \{b_1, b_2, \cdots, b_T\}$ where each sample b_t is a set of observations acquired at time $t \in [1, T]$. We adopt a sliding-window approach, such that W subsequent observations $\{b_i, b_{i+1}, \cdots, b_{i+W-1}\}$ centered in $b_{\lfloor \frac{W+i}{2} \rfloor}$ are used to calculate a behavioral feature $f_{\lfloor \frac{W+i}{2} \rfloor}$. By applying this technique, we transform the time-series B into the time-series $F = \{f_{\lfloor \frac{W}{2} \rfloor}, f_{\lfloor \frac{W}{2} \rfloor+1}, ... f_{T-\lfloor \frac{W}{2} \rfloor+1}\}$ where each element f_j is a feature vector extracted from the window centered in the sample b_j.

Provided with this feature representation, we apply a supervised classification technique to detect and recognize change-points from the feature time-series F.

We assume the availability of a dataset annotated by an expert in terms of change-points' timestamp and class, indicating with the latter the type of behavioral transition to be found. Consequently, our goal will not be the recognition of behaviors but the identification of transitions from one behavior to another.

In particular, we consider the problem of jointly recognizing among L different classes of change-points and detecting the time when such change-points arise. We aim at solving this problem without any a priori knowledge of the involved behavior classes. This problem is difficult because: behavior duration and type may largely change, there may be intra- and inter-subjects variations in behaviors and, in general, the transitions from one behavior to another are not abrupt.

We adopt a classifier that takes as input the feature descriptor of a temporal window f_t and provides as output the predicted change-point class (a label between 1 and L) or 0 if no change-point has been detected. Since the behavior duration varies, more subsequent windows may be recognized as belonging to the same change-point class. To refine the prediction step and estimate more precisely the instant when the transition arises, we apply a clustering technique to aggregate the predictions. The centroid of each of such clusters is used as estimated change-point.

The framework we implemented is presented in Figure 1. In the following, we provide more details about each of the steps required to apply the above-described methodology.

A. Feature Extraction

A change-point represents a switch from a dynamical system to another. Hence, in a change-point, the statistical properties of the signal within a temporal window change.

In particular, given a set of subsequent observations $\{b_{t-\lfloor \frac{W}{2} \rfloor}, \cdots, b_t, \cdots, b_{t+\lfloor \frac{W}{2} \rfloor}\}$ in a window of length W, we characterize the signal within the window by its statistical moments. The corresponding feature descriptor is defined as $f_t = \{\mu_t, \Sigma_t, S_t, K_t\ \Delta_t\}$ where: μ_t represents the mean value (1st order moment), Σ_t is the covariance matrix (2nd order moment), S_t and K_t represents the skewness and kurtosis of the multivariate distribution as defined by [74] (3rd and 4th order moments respectively). Moreover, the feature descriptor includes the value Δ_t, which is the difference between the maximal and minimal values of the signal components and, hence, describes the signal extension along with the various components. It is also possible to include the median of the signal values to make the descriptor more robust to outliers.

The feature descriptor is extended by including also other statistical properties of the signal such as the statistical moments of the velocity, acceleration and jerk of the signal (namely, the 1-st, 2-nd and 3-rd order derivatives of the signal) to capture local information about the way the signal changes over time [75], [76], [77].

Besides, the statistical moments of the signal curvature are included too in the descriptor. The curvature locally measures how fast a curve is changing direction at a given point. The formal definition of curvature is $C = |\frac{d\vec{T}}{d\vec{s}}|$ where $\vec{T}$ is the unit tangent of the curve and $\vec{s}$ is the arc length. In particular, for a three-dimensional signal $b(t)$, where t indicates time, the curvature measures the local magnitude of the acceleration of a particle moving with unit speed along the curve and is defined as $C = \frac{||b'(t) \times b''(t)||}{||b'(t)||^3}$, where $b'(t)$ and $b''(t)$ represent 1-st and 2-nd order of time derivatives respectively. For a one-dimensional signal $b(t)$, the curvature is simply defined as $C = \frac{b'(t) - b''(t)}{[1+(b'(t))^2]^{\frac{3}{2}}}$.

Contextual information, namely the properties of the signal before time $t - \lfloor \frac{W}{2} \rfloor$ and after time $t + \lfloor \frac{W}{2} \rfloor$, can help to detect and recognize a change-point at time t, and hence a transition from one behavior to another. As shown in Figure 2, it is possible to embed such contextual information into the change-point descriptor f_t by including the statistical moments of the signal in $\lfloor \frac{W}{2} \rfloor$ observations before and after the t-th window. By doing in this way, for a k-dimensional signal, the final descriptor f_t has a size equals to $7 \cdot (2k^2 + 3k) + 3k$. In particular, for $k = 1$, the descriptor has a size equals to 38.

The size of the above-described feature vector is independent of the number of considered observations W, namely the size of the temporal window. However, the value of W can have an impact on the descriptiveness of the feature vector, and should be chosen by considering the sampling frequency and the nature of the analyzed signal. Depending on the sampling frequency, a small value of W can be too local and result in a poor descriptor for the change-point detection problem. On the contrary, a too-large value of W increases the risk of using temporal windows that may contain more than one change-point [48]. Our proposed methodology aims at considering the case when no information is available about the duration of each behavior. The selection of the best value for W can be done empirically when an annotated dataset is available by selecting the best classification model among the ones trained by varying the value of W.

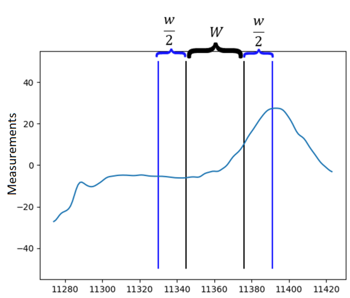

Fig. 2. Contextual information can be included in the descriptor extracted from a temporal window of length W by computing the statistical moments of the signal of the $\frac{W}{2}$ observations on the left and on the right side of the temporal window itself.

B. Descriptor Classification

We need to recognize if a descriptor f_j of the j-th window, computed as described in Sec. III-A, is not a change-point or is one of the L classes of possible behavior transitions. We model the problem as a classification one.

We assume that a dataset manually annotated by experts in the psychology field is available, and that annotators have provided the time and the class of the change-points in the behavioral time series.

We use the behavioral time series to extract a set of feature descriptors. We associate to each feature descriptor a label that depends on the available annotations.

When considering temporal windows of length W, there are no more than W subsequent temporal windows including an annotated time instant t. Intuitively, the descriptors associated to some of such temporal windows might differ very little (for example, subsequent temporal windows differ for only two observations). Hence, there is the risk to associate two similar descriptors with different labels, which may compromise the classifier training.

The most straightforward strategy to avoid this issue would be that of computing a feature descriptor for each window centered in the annotated change-point while avoiding to extract descriptors for overlapping temporal windows. However, this greatly limits the size of the training data, especially for rare change-point classes, with an impact on the accuracy of the classification model. Furthermore, this strategy strongly trusts in the experts' capabilities to precisely annotate change-points. In real experimental scenarios, different experts may not fully agree on the exact location of the change-point in the time series. In practice, during the data annotation process, different experts annotate independently the data and then discuss and resolve any divergence in the annotation results. Inter-rater reliability is in general measured by the Cohen's Kappa. Such kind of validation increases the cost for annotating data.

This suggests that a safer strategy is that of labelling in a consistent way all feature descriptors extracted from windows centered in observations that are close to the annotated change-point. To this purpose, we set a threshold τ on the absolute difference between the time of the central observation and the annotated change-point. Consequently, the size of τ will determine how many frames around the central observation are labeled as change-points.

Another important issue to consider when dealing with change-point classification is the nature of the resulting training set. Indeed, it is not unusual to deal with an unbalanced dataset, namely dataset in which change-point classes are not equally represented. This is especially true for the class representing the within-state sequences [78], [79]. To account for such issue it is important to apply a balancing procedure such that the number of samples in each change-point class becomes comparable. This has been achieved by performing a random under-sampling of the within-state class.

Many algorithms, as Random Forests, Support Vector Machines or Neural Networks, are available to solve the presented classification problem [80]. In this study, we focused on the use of SVM based on Radial Basis Function due to their flexibility and generalization abilities [81]. However, the choice of the best classification algorithm depends on the application and nature of dataset to be analyzed.

C. Change-Point Detection

As detailed in Sec. III-B, our classifier takes as input the feature descriptor of a temporal window and provides a label indicating if the center of the temporal window is a change-point and, in this case, the class of the detected change-point. During test, we apply a sliding window approach, meaning that we classify the feature descriptors of subsequent temporal windows. Two neighbor windows will then differ for two observations and their feature descriptor may not greatly vary. Hence, the classifiers will likely output the same predictions for close temporal windows. We adopt DBSCAN [82] to aggregate the time instants of such predictions in compact, separated clusters, without any overlap. Other clustering algorithms need to know the number of clusters they should compute, namely the number of change-points arising in the time series. In contrast, DBSCAN is an algorithm entirely based on the distances among samples (in our case, time instants); clusters are computed from the neighborhood relationships among the samples. DBSCAN is independently executed on each set of time instants classified as belonging to a specific change-point class. Then, the average value of the time instants in each resulting cluster is calculated. Such time instant corresponds to the predicted time when a change-point occurred in the signal.

D. Baseline Method

To assess the system capabilities, we adopted the baseline method introduced in the work of [7], which was devised for one-dimensional signals. The baseline method computes the second order derivative of the behavioral signal to measure the concavity of the signal itself. The second order derivative is then filtered by discarding all values below a threshold σ. The threshold σ is set to the standard deviation of the second derivative in the training data. The time instants of the unsuppressed values are clustered by using the K-means algorithm. The value K is set to the number of annotated change-points in the analyzed sequence. The centers of the found clusters are considered as predicted change-points. By using a-priori knowledge of the order in which the behavioral classes appear in the sequence, each detected change-point is associated with a category. We note here that the baseline method is simpler than the method we propose but is not fully automatic because it requires a-priori knowledge of the number of expected change-points and of the order in which behaviors emerge over time.

IV. Case Study

Motor behaviors can be defined by means of complex schemas or, as presented in [58], atomic sub-movements. In our methodology, we deal with the detection and recognition of the times in which a sub-movement ends and/or starts. These transitions can be defined as change-points.

To demonstrate the methodology introduced in this paper, we propose its application to the analysis of children's motor behaviors data collected by [50] and [49] during an imitation task where children had to reproduce the movements of a virtual agent acting as a tightrope walker (TW) [51].

In [50], authors focused particularly on how participants were able to imitate the TW and, eventually, perform a perspective change to the point of view of the virtual agent. As we will detail later, children's behaviors were described through time-series derived by their hand movements during the imitation task. In [50], authors investigated the ability of children in performing behavioral own-body-transformations by exploiting manually annotated change-points. Experts annotated change-points to analyze and compare the behaviors of TD children and children affected by NDD, specifically ASD and DCD. Change-points were annotated through visual analysis and filtered through shared consensus among the annotators.

In [49], authors explored children's behavioral imitation abilities in terms of interpersonal synchronization and motor coordination by exploiting the dataset in [50] through an automated characterization of temporal slices of the interaction with the virtual agent.

In this paper, we use the annotated data presented in [50] to apply our methodology and automatically discover change-points in the recorded behavioral signals. We use the data collected with TD children to train and test our system. Then, we use the data collected with children affected by ASD or DCD disorders to further investigate the potentiality and the limits of the proposed methodology. In the following, we provide more details about the considered scenario, the data used to asses our method, and the implementation details.

A. Experimental protocol

Authors of [51] used a colour movie of a computer-generated female tightrope walker (TW) to investigate whether individuals, under spontaneous conditions and without explicit instruction, embody another person's behavior. They designed a motor paradigm focusing on elementary mimicry to investigate, from the body posture, how individuals act together, focusing in particular on the achievement of own-body transformations, from an embodied, ego-centered viewpoint to a disembodied, hetero-centered viewpoint.

The studies in [49] and [50] extended the paradigm of the TW to adapt it to the study of children's behaviors by adjusting the size of the virtual character, giving it a cartoon-like aspect of a child. Fig. 3 shows an image of the TW in the front-facing orientation: the artificial TW projected on the wall stands on a rope and keeps its balancing by carrying a bar that, over time, it tilts laterally to its right or left. The TW's bar tilts have a maximum amplitude of 51° and a maximum duration of $3.2s$ (mean duration: $2.7s$). Tilts on the right or left are performed in random order 7 times. Children were provided with a wooden bar and invited to mimic the TW's behaviors by rotating their bar accordingly. To elicit perspective taking, two kinds of sequences were alternated 7 times: in the first type of sequence, the TW walks after the participant, by back-facing her/him; in the second type, the TW walks towards the participant, by front-facing her/him. In the first case, the participant can simply imitate the TW; in the second case, an effective imitation from the participant implies a mental rotation from the point of view of the TW.

Authors of [49] and [50] found behavioral differences in these tasks between TD children and ASD, DCD children. Children with such deficits, in particular, can experience difficulties with fine motor control abilities, producing gross, clumsy or uncoordinated movements [83]. Hence, we expect that the change-point detection and recognition in behavior

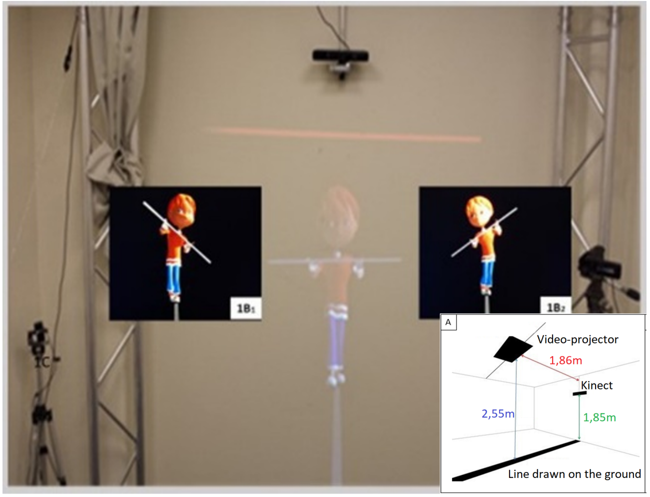

Fig. 3. The image shows the real-life projected Tightrope walker used in [49]. Images on the left and on the right show the TW while tilting its bar to the left and to the right.

time series of ASD and DCD subjects will be harder than in TD subjects considering the behavioral differences between these subject groups [84], [85]. In particular, we suppose the system will have more difficulties in dealing with ASD than DCD subjects.

Although for each frame participant's bar and the tightrope walker's bar angles are recorded, as well as the timing of the frame, and both the participant's and the TW's head inclinations, in this paper we have only dealt with the detection and the recognition of the change-points related to the movement angles of the participant's bar. The data of [50] are not simulated and they were produced in a real setup defined by psychologists. As mentioned above and due to the diversity of the developmental levels of the children participating in the study, we consider these data quite challenging and worthy of study.

B. Dataset

During each experimental session, videos of the participants were automatically and continuously recorded for offline analysis by a RGB-D sensor located in front of them, a Microsoft Kinect[1]. This sensor was used to estimate 3D poses of the participant (in terms of 3D skeletal data) at a frame-rate of about 25 fps. The angle of the child's bar tilt is calculated using the inclination of the 3D line passing through the two 3D points representing the child's hands.

The dataset includes experiments from 85 children imitating the TW. According to the experimental protocol, each experiment is composed by 7 sequences; however, only 70 children had completed all the planned sessions. Of these 70 children, 30 are TD, 14 DCD and 26 ASD.

Overall, the dataset adopted in this paper includes 490 bar tilts sequences of which 210 are of TD children, 98 of DCD children and 182 of ASD children.

Fig. 4 shows the signals measured during an experimental session. In particular, it shows the measured tilt angles in degrees of the bar of one of the participants (the blue line) and of the TW (the orange line) over time. As shown in the figure, the child's bar is initially in a horizontal position where the tilt angle is equal to $0°$. Then, for 7 times, the bar is moved and the tilt angle increases (decreases) till reaching a peak value. Later on, the tilt angle decreases (increases) till reaching again the value 0 (bar in horizontal position).

Change-points in [50] were manually identified by expert annotators by analyzing children's movements (see supplementary materials in [50]) and considering: the starting time of the bar movement; the time when the tilt angle reaches its peak value; the time of the end of the bar movement. Hence, three different change-point classes have been annotated. Although each sequence was composed of 7 tilt movements, experts found that, due to an ambiguous stimulus, the 6th tilt was consistently not reliable enough to be used for the analysis of the children's imitation capabilities. Therefore, we kept only 6 tilts as meaningful and each sequence of angles was annotated with 18 change-points (6 for each change-point class).

Annotating change-points in the signals acquired with ASD and DCD children is a complex activity even for psychiatrists [49]. Figs. 5 and 6 show examples of the tilt angle signal measured with a DCD and a ASD child respectively. As shown in the figures, the signals are less smoothed than the ones acquired with TD children. This is probably due to the difficulty of these children in replicating/imitating the TW's movements.

Overall, for the 3 different classes of change-points, the data include 8820 annotated change-points in the tilt angle sequences collected with all the 70 children. In particular, 3780 of these change-points are annotated in sequences collected with TD children.

C. Implementation details

Before extracting the feature descriptors, we applied a smoothing filter on each sequence and, in particular, we used the Savitzky-Golay filter. Then, z-score normalization was applied to the data. Finally, our system has been trained considering the 3780 change-points manually annotated in the TD children data.

As discussed in Sec. III-B, to train our model it is necessary to associate each feature vector with a label indicating whether a change-point occurs in the considered window and, eventually, the type of change-point. This labeling task can be a challenging activity that can be performed through different strategies.

It is possible to label as change-point only the window in which the central frame has been annotated as change-point; otherwise, it is possible to annotate as change-point all the windows such that the distance between the annotated change-point and the central observation is below a threshold τ. Two different values of τ have been tested: $\tau = \{0, 2\}$ The maximum threshold $\tau = 2$ was empirically chosen according to the frame rate and the scale of the event that we want to detect, while the window size was chosen among $W = \{15, 31, 61\}$.

[1]Microsoft Kinect website: https://developer.microsoft.com/en-us/windows/kinect

We trained SVMs with a 1-vs-1 strategy for the 3 change-point classes and the additional within-state, background class by using Radial Basis Function (RBF) kernel[86].

Then we aggregated the positive predictions by applying the DBSCAN algorithm. The minimum number of samples in each cluster was set to 3. This value varies according to the application and the scale of the event to be detected.

During test, we adopted a sliding window approach and classified all windows with a stride equal to 1. Sliding window stride value defines the number of frames skipped in between adjacent sliding window scans. All experiments were performed in cross-subject validation.

V. Experimental results

To validate our system, we conducted a series of experiments by using the previously described tightrope walker database. In particular, as already disclosed, we focused on:

– **The evaluation of the performance of the system** in detecting and recognizing change-points in behaviors of TD children. Performance metrics computed in this case can be seen as measures from a *best-case scenario* since the signals representing children's behaviors are smooth and readable (see Fig. 4). The method was compared against a baseline method presented in [7].
– **The efficacy and the limits of the system** in presence of *NDD*. In this particular case, the system deals with abnormal signals that are not easy to be labeled even for expert annotators (see Figs. 5 and 6);
– **The trade-off** between training set size and achieved system performances.

To assess the system capability of precisely detecting change-points, we measured the F1-Score, the Mean Absolute Error (MAE), the Missing Rate (MR) and the Precision (P) achieved by our system.

- The F1-score is computed for the (binary) detection problem. In this way, all feature vectors that are classified as belonging to one of the change-point classes are considered as positive cases. The vectors that are classified as within-state sequences are considered as negative cases. The F1-score is defined as:
$$F_1 = 2 \cdot \frac{P \cdot R}{P + R} \tag{5}$$
where P indicates the precision value and R stands for recall, and measures the proportion of true positives with respect to the number of annotated positive samples.
- The Mean Absolute Error (MAE) measures the average absolute difference between the time instant when a change-point has been detected and the time instant when the change-point has been annotated. Time is expressed as frame number (data have been collected with a frame rate equals to 25). MAE is computed only for the valid detection. A detected change-point is considered valid if its distance to the annotated change-point in terms of number of frames is lower than 100.
- The Missing Rate (MR) is defined as the percentage of annotated change-points that are not detected by our system.
- The precision (P) is defined as the proportion of true positives with respect to all positive predictions made by our system.

Metric:	F1-score	MR %	MAE	P
S-SW(15, 0)	0.880 ± 0.07	21.0 ± 2.7	4.44 ± 1.16	0.800 ± 0.11
S-SW(15, 2)	0.900 ± 0.07	14.0 ± 2.4	4.00 ± 1.13	0.850 ± 0.10
C-SW(15, 0)	0.940 ± 0.05	13.8 ± 4.2	4.41 ± 1.34	0.910 ± 0.08
C-SW(15, 2)	0.945 ± 0.03	14.0 ± 4.0	4.50 ± 1.30	0.910 ± 0.10
S-SW(31, 0)	0.940 ± 0.05	4.0 ± 6.7	4.39 ± 1.55	0.930 ± 0.07
S-SW(31, 2)	0.942 ± 0.08	4.2 ± 6.4	4.25 ± 1.60	0.932 ± 0.06
C-SW(31, 0)	0.946 ± 0.05	4.3 ± 6.5	4.40 ± 1.79	**0.940** ± 0.06
C-SW(31, 2)	**0.950** ± 0.05	**3.4** ± 5.0	**4.07** ± 1.64	0.938 ± 0.07
S-SW(61, 0)	0.948 ± 0.04	4.1 ± 4.8	4.65 ± 1.71	0.940 ± 0.06
S-SW(61, 2)	0.947 ± 0.07	4.2 ± 4.4	4.63 ± 1.80	**0.943** ± 0.06
C-SW(61, 0)	0.950 ± 0.08	4.1 ± 4.5	4.45 ± 1.20	0.942 ± 0.04
C-SW(61, 2)	0.950 ± 0.05	3.9 ± 5.1	4.58 ± 1.67	0.940 ± 0.05
Baseline M.	0.71	14.0	9.38	0.62

TABLE I
THE TABLE REPORTS THE PERFORMANCE OF OUR SYSTEM (C-SW) WITH $\tau = 2$ AND $\tau = 0$ ON SEQUENCES COLLECTED FROM TD CHILDREN. THE FIRST NUMBER INDICATES THE VALUE OF THE VARIABLE W.

A. Evaluation of the performances of the system

A leave-one-subject-out protocol has been employed to validate the presented system using sequences collected from TD children: for each subject, we trained a model with the samples from all the other subjects; samples from the excluded subject were used as test set. The protocol has been repeated using three different window lengths ($W = 15$, $W = 31$ and $W = 61$) and two different descriptors: a simple one (*S-SW*) resuming the statistical properties of the signal in each window, and an extended one (*C-SW*) able to include contextual information considering $\lfloor \frac{W}{2} \rfloor$ observations before and after each temporal window, as discussed in III-A. Average performances were compared between them and against a baseline method.

Table I shows the results that we obtained by varying W and τ. For each metric, computed as average over the per-subject predictions, we report the average value and the standard deviation. As shown in the table, the performance obtained by setting $\tau = 0$ are slightly inferior to those obtained by setting $\tau = 2$.

The table also shows that all variants of the proposed method (C-SW and S-SW) outperform the baseline method in all the selected performance metrics. As for the impact of the window size W on the system performances, while the F1-Score and the precision (P) do not change much, the amount of missing rate (MR) and the mean average error (MAE) slightly increase when increasing the window size suggesting that the value of W can affect the general reliability of the system in terms of accuracy in localizing the change-points. Consequently, a meaningful choice of the window length would be a compromise of 31 frames ($W = 31$). Since the system captures data at 25 fps, a value of W equals to 31 turns out to process temporal windows slightly larger than 1 second. Finally, the table shows that embedding contextual information into the descriptor (*C-SW*) tends to improve the system performances.

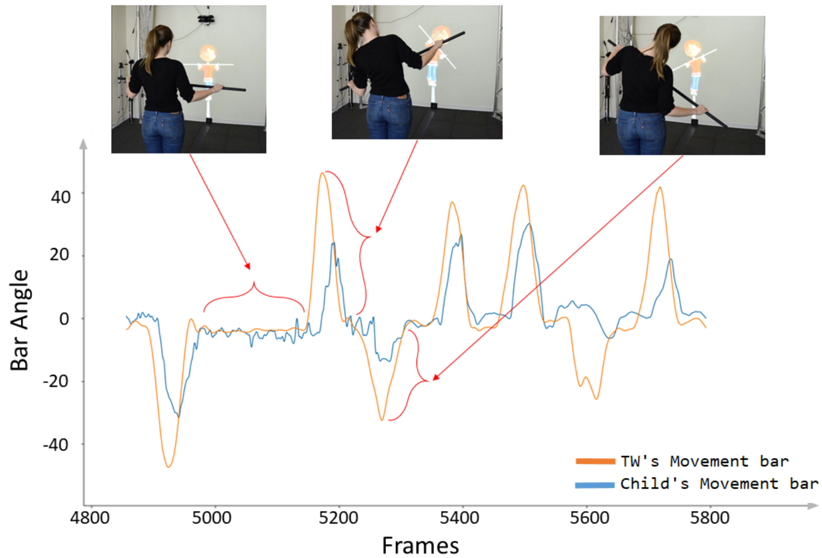

Fig. 4. The blue line represents the tilt angle of a TD child's bar. The orange line represents the reference tilt angle of the TW's bar. The red arrows indicate the movement phases that the child performs during the emulation of the TW. The transition from one phase to another is a change point.

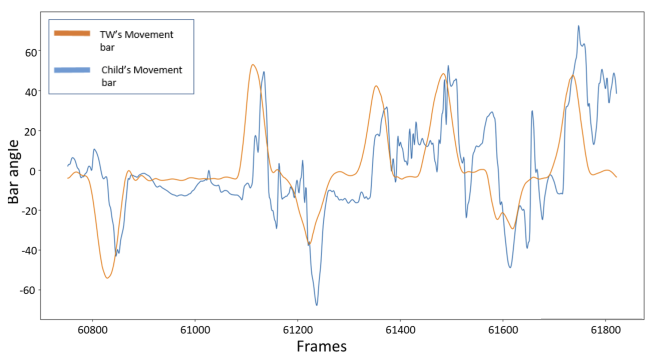

Fig. 5. The blue line represents the tilt angle of a DCD child's bar. The process of the change-points detection is not a simple activity: although in the plot is possible to identify some patterns of child's bar movements, it is still challenging to precisely detect and recognize the time when a movement startsends.

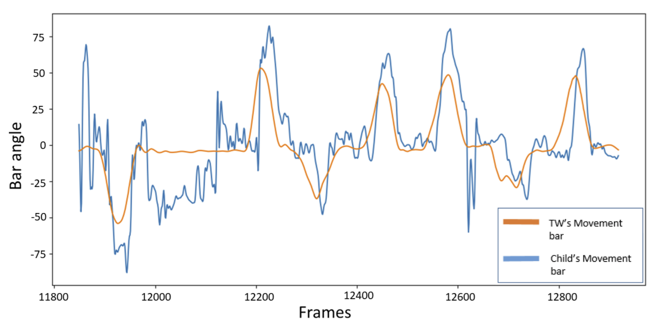

Fig. 6. The blue line represents the tilt angle of a ASD child's bar movements. The precise detection and recognition of change-points in movements of ASD children is an even more complex activity than the one present in the DCD plot.

Finally, we analysed the change-point class recognition capabilities of the system. Table II reports the confusion matrix for the change-point classes "Start", "Peak", "End" and "Background". In particular, the Background class represents the within-state class. The confusion matrix highlights that the system is able to correctly detect and recognize change-point classes with a good degree of reliability.

T vs. P (TD)	Start	Peak	End	Background
Start	94.96%	0.08%	0.16%	4.78%
Peak	0.25%	93.30%	1.09%	5.36%
End	0.16%	0.13%	92.20%	7.51%
Background	0.02%	0.03%	0.06%	99.88%

TABLE II
THE TABLE SHOWS THE CONFUSION MATRIX FOR THE CHANGE-POINT CLASSIFICATION PROBLEM ON BEHAVIORAL DATA FROM CHILDREN IN TYPICAL DEVELOPMENT.

B. *Efficacy and limits in presence of abnormal behaviors*

We investigated how the system deals with data coming from abnormal behaviors. Therefore, from the tightrope walker database, we selected the data of children with NDD. As already pointed out in Sec. IV, annotating such kind of behavioral signals is challenging even for experts, hence, any system, automated or semi-automated, able to help and speed-up this process is especially important. In any case, due to such difficulties, we expect a drop in the performance of our system.

Average performances of the system were obtained by training it with TD data only, using the most effective parameter set previously found. Then, the obtained model has been tested with data from children with NDD. As a consequence, we asked to the system an important ability of generalization and abstraction from the training sample set.

Metric:	F1-score	MR %	MAE	P
TD	0.95 $\pm$ 0.050	3.4 $\pm$ 5.0	4.07 $\pm$ 1.64	0.94 $\pm$ 0.069
DCD	0.91 $\pm$ 0.001	4.2 $\pm$ 0.3	4.38 $\pm$ 0.15	0.88 $\pm$ 0.002
ASD	0.82 $\pm$ 0.002	6.5 $\pm$ 3.0	5.20 $\pm$ 0.14	0.73 $\pm$ 0.030

TABLE III
THE TABLE REPORTS THE PERFORMANCE OF OUR SYSTEM $(C - SW(W = 31))$ ON SEQUENCES COLLECTED FROM THREE GROUPS OF CHILDREN: TD, ASD AND DCD.

Table III compares the performance achieved by our system (*C-SW* $(W = 31)$) for the three groups of children: TD, DCD and ASD. We conducted an analysis of the recognition performances of the system in presence of NDDs. Table IV and V report the confusion matrices on the change-point classes predicted by the classifier on behavioral data from children with DCD and ASD, respectively. Such confusion matrices highlight that the system is still able of distinguishing the change-points from the background class, whilst the recognition performances of the change-point classes deteriorate, especially when analysing ASD children data.

Overall, as highlighted also by [49], the system performances decay on NDD children data, highlighting the difficulties the system has in analysing such data. The performances deterioration show a correspondence with the examined pathological groups: since DCD is a deficit on fine motor control skills, observations from DCD children are less precise than the ones collected from TD children because anomalous movements are introduced; in ASD children performances got worst as effect of motor control and social interaction deficits, that manifest themselves in more abnormal and erratic movements.

C. Trade-off between training set size and system performances

Finally, we have investigated how the size of the training set affects the performance of our model. Using a cross-subject validation protocol, we have trained our classifier with a variable number of subjects varying within $\{5, 10, \ldots, 95\}\%$ of the total number of TD children. As shown in Fig. 7, by varying the percentage of subjects in the training set, the MAE value decreases. This result clearly shows that a wider training set allows for a more precise localization of the change-points. In a similar way, the MR value also decreases. Finally, while the F1-score keeps growing while increasing the size of the training set, the precision value has an inflection and starts to grow again after the 80% of the subjects are used for training purposes. Overall, with small dataset size, the system exhibits high recall but is not precise yet. By increasing the size of the training set, the precision value increases as well.

The experimental results, however, highlight how the system needs just the 30% of the total number of subjects to achieve a 90% of the F1-score.

T vs. P (DCD)	Start	Peak	End	Background
Start	92.50%	0.21%	0.21%	7.00%
Peak	0.00%	92.50%	0.63%	6.77%
End	0.21%	0.21%	86%	13.34%
Background	0.03%	0.03%	0.06%	99.86%

TABLE IV
THE TABLE SHOWS THE CONFUSION MATRIX ON THE PREDICTED CHANGE-POINT CLASSES WHEN ANALYSING BEHAVIORAL DATA FROM CHILDREN WITH DEVELOPMENTAL COORDINATION DISORDER (DCD).

T vs. P (ASD)	Start	Peak	End	Background
Start	88.00%	0.38%	0.25%	10.77%
Peak	0.76%	86.00%	0.12%	12.54%
End	0.88%	0.63%	83.00%	14.00%
Background	0.04%	0.05%	0.06%	99.00%

TABLE V
THE TABLE SHOWS THE CONFUSION MATRIX ON THE PREDICTED CHANGE-POINT CLASSES WHEN ANALYSING BEHAVIORAL DATA FROM CHILDREN WITH AUTISM SPECTRUM DISORDER (ASD).

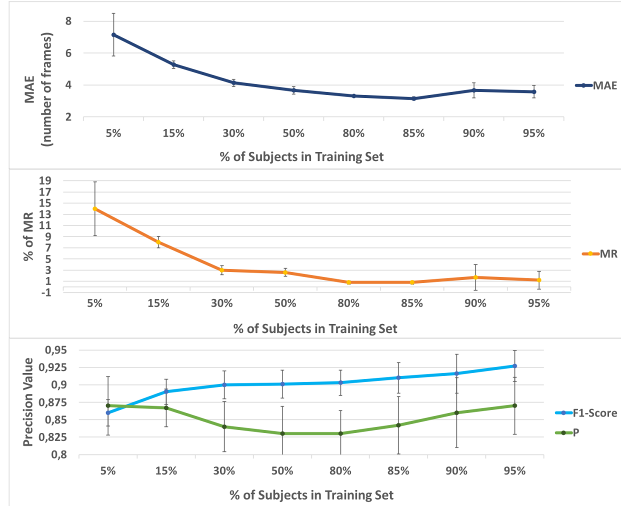

Fig. 7. The plots are obtained by varying the percentage of TD subjects in the training set. The plots show trend and standard deviation of: Mean Absolute Error (MAE), Missing Rate (MR), F1-score and Precision values.

VI. Discussion

Experiments on data from children with TD show the effectiveness of the system in detecting and identifying change-points ($F1 \approx 0.95$) against the background and among three classes. Behavioral data from children with DCD and ASD have been exploited to verify the reliability of the system facing abnormal movements. Despite an expected decrease in performances, the system is still able to obtain an acceptable recognition rate in accord to the particular NDD analysed (DCD, $F1 \approx 0.91$ and ASD, $F1 \approx 0.82$).

Table I reports the performance of our semi-automated annotation system on sequences collected from TD Children. Such results show that the use of a sliding window approach to extract features from data is successful in detection and recognition of change-points without a-priori knowledge of the behavioral models. Experiments testing different window lengths show the need of choosing a window size W that takes into account the signal sampling rate and the dynamics of the behaviors that should be modeled.

The adopted signal characterization includes a set of statistical features of the signal within a temporal window, and demonstrated the capability of describing change-points and, hence, the transition from a motor behavior to another. We presented, in particular, two different descriptors: a simple one called *S-SW* that resumes the statistical properties of the signal in each window and an extended one, *C-SW*, able to include contextual information before and after each temporal window.

Results in Table I focus on experiments that employ such descriptors, varying the values of W and τ. The best resulting configuration in terms of F1-score, MR, and MAE, exploiting the proposed dataset, is the one that takes into account contextual information, namely the descriptor *C-SW* with $\tau = 2$. While showing the efficacy of the proposed approach for modeling change-points, such results also underline the more general importance of including contextual information: the descriptor should not only include a statistical characterization of the point and its window, but should also include information describing its closest surrounding area.

The feasibility and the effectiveness of the proposed approach in detecting and classifying change-points are also confirmed by the plot in Fig. 7. Indeed, through an analysis of the impact of the training set size on the performances of the system, we found that it is possible to achieve an F1-score of 90% by supplying a 30% of the number of subjects in the sample set. Consequently, it is possible to conceive in future works focusing on behavior analysis and understanding, the use of a semi-automated annotation methodology based on the proposed system: in datasets similar to the one exploited in this work, it is possible to hypothesize a reduction of the 70% of the effort spent on the annotation of the whole dataset.

The proposed system, in fact, has not the ambition of operating in real-time or as a fully automated change-point annotation system, but as a convenient, semi-automated offline tool for psychologists, psychiatrists, computer scientists, cognitive scientists and other practitioners working on behavior understanding, affective computing, social robotics or, more in general, human-machine interaction, that need an efficient tool for annotating behavioral observations. After the manual annotation of a small, randomized set of behavioral data, the presented system can be employed to complete the annotation of the whole dataset. However, according to the degree of reliability requested by the behavioral analysis, and due to the not negligible presence of errors, especially in case of behavioral anomalies or NDD, the detected change-points should always be reviewed by expert annotators. Despite these limits, the use of the developed system would result in a fast and efficient workflow for the annotation of behavioral observations.

VII. Conclusions and Future Works

In this paper, we presented a system aimed at the detection and recognition of change-points in behavioral data related to humans' movements. In particular, the proposed methodology has been used to implement a semi-automated annotation system able of achieving a fine, local characterization of behavioral observations. While general purpose, the system has been tested using a database of behavioral data collected during an imitation experiment involving interactions between children and a virtual avatar acting as a tightrope walker. Such behavioral data have been exploited to evaluate the general performances of the system and its reliability.

While the developed system has been imagined as a general purpose tool, it has been tested on a single human behavior dataset that focuses on children movements. Consequently, more experiments with other humans' movements datasets are needed to confirm the overall extent of such performances. Future works will focus also on testing the system while exploiting other modalities, such as gazing or eye direction, able to describe the human engagement's evolution. At the same time, the presented tool will be extended to support different, synchronized modalities. Such extension will also account for the different sampling rates of the multi-modal sensors used to capture human behaviors, such as skeleton data from RGB-D cameras and gaze information from eye trackers (usually faster than RGB-D). The study of more informative features remain in any case an interesting topic for future investigations.

We will also further focus on domain-adaptation techniques to transfer the system ability of classifying TD data onto NDD data. Finally, we have the ambition of enlarging the testing domains of the presented framework beyond the analysis of human activities by considering, for instance, the analysis of physical, astronomical or financial observations.

Acknowledgements

Authors would like to thank Francois Villa and Berangere Thirioux for all the help and their feedbacks on the development of this project. Authors thank also Olivier Houde, Ania Aite, Gregoire Borst, and all the LaPsyDE laboratory from Paris-Descartes University and Caen University.

References

[1] D. A. Levitis, W. Z. Lidicker Jr, and G. Freund, "Behavioural biologists do not agree on what constitutes behaviour," *Animal behaviour*, vol. 78, no. 1, pp. 103–110, 2009.

[2] P. Thagard, *Mind: Introduction to cognitive science*. MIT press Cambridge, MA, 2005, vol. 17.

[3] P. J. Corr and G. Matthews, *The Cambridge handbook of personality psychology*. Cambridge University Press Cambridge, 2009.

[4] A. P. Association *et al.*, "Diagnostic and statistical manual of mental disorders," *BMC Med*, vol. 17, pp. 133–137, 2013.

[5] C. L. Breazeal, *Designing sociable robots*. MIT press, 2004.

[6] R. Sun, *The Cambridge handbook of computational psychology*. Cambridge University Press, 2008.

[7] S. M. Anzalone, S. Boucenna, S. Ivaldi, and M. Chetouani, "Evaluating the engagement with social robots," *International Journal of Social Robotics*, vol. 7, no. 4, pp. 465–478, 2015.

[8] A. Vinciarelli, M. Pantic, and H. Bourlard, "Social signal processing: Survey of an emerging domain," *Image and vision computing*, vol. 27, no. 12, pp. 1743–1759, 2009.

[9] A. Vinciarelli, M. Pantic, D. Heylen, C. Pelachaud, I. Poggi, F. D'Errico, and M. Schroeder, "Bridging the gap between social animal and unsocial machine: A survey of social signal processing," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 69–87, 2011.

[10] S. Ghidoni, S. M. Anzalone, M. Munaro, S. Michieletto, and E. Menegatti, "A distributed perception infrastructure for robot assisted living," *Robotics and Autonomous Systems*, vol. 62, no. 9, pp. 1316–1328, 2014.

[11] J. Massion, "Postural control system," *Current opinion in neurobiology*, vol. 4, no. 6, pp. 877–887, 1994.

[12] E. K. Bojanek, Z. Wang, S. P. White, and M. W. Mosconi, "Postural control processes during standing and step initiation in autism spectrum disorder," *Journal of Neurodevelopmental Disorders*, vol. 12, no. 1, p. 1, 2020.

[13] L. Ogden, "Collaborative tasks, collaborative children: An analysis of reciprocity during peer interaction at key stage 1," *British Educational Research Journal*, vol. 26, no. 2, pp. 211–226, 2000.

[14] X. P. Burgos-Artizzu, P. Dollár, D. Lin, D. J. Anderson, and P. Perona, "Social behavior recognition in continuous video," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2012, pp. 1322–1329.

[15] J. Rehg, G. Abowd, A. Rozga, M. Romero, M. Clements, S. Sclaroff, I. Essa, O. Ousley, Y. Li, C. Kim *et al.*, "Decoding children's social behavior," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 3414–3421.

[16] A. Mahdhaoui, M. Chetouani, R. S. Cassel, C. Saint-Georges, E. Parlato, M. C. Laznik, F. Apicella, F. Muratori, S. Maestro, and D. Cohen, "Computerized home video detection for motherese may help to study impaired interaction between infants who become autistic and their parents," *International Journal of Methods in Psychiatric Research*, vol. 20, no. 1, pp. e6–e18, 2011.

[17] K. A. Funes Mora, L. Nguyen, D. Gatica-Perez, and J.-M. Odobez, "A semi-automated system for accurate gaze coding in natural dyadic interactions," in *Proceedings of the 15th ACM on International conference on multimodal interaction*. ACM, 2013, pp. 87–90.

[18] M. Fung, Y. Jin, R. Zhao, and M. E. Hoque, "Roc speak: semi-automated personalized feedback on nonverbal behavior from recorded videos," in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2015, pp. 1167–1178.

[19] H. Salam, O. Celiktutan, I. Hupont, H. Gunes, and M. Chetouani, "Fully automatic analysis of engagement and its relationship to personality in human-robot interactions," *IEEE Access*, vol. 5, pp. 705–721, 2016.

[20] Z. Shou, J. Pan, J. Chan, K. Miyazawa, H. Mansour, A. Vetro, X. Giro-i Nieto, and S.-F. Chang, "Online detection of action start in untrimmed, streaming videos," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 534–551.

[21] B. Wang and M. Hoai, "Back to the beginning: Starting point detection for early recognition of ongoing human actions," *Computer Vision and Image Understanding*, vol. 175, pp. 24–31, 2018.

[22] H. Dindo, L. Lo Presti, M. La Cascia, A. Chella, and R. Dedić, "Hankelet-based action classification for motor intention recognition," *Robotics and Autonomous Systems*, vol. 94, pp. 120–133, 2017.

[23] L. Lo Presti and M. La Cascia, "A novel time series kernel for sequences generated by lti systems," in *Asian Conference on Computer Vision*. Springer, 2016, pp. 433–451.

[24] L. Lo Presti, M. La Cascia, S. Sclaroff, and O. Camps, "Hankelet-based dynamical systems modeling for 3d action recognition," *Image and Vision Computing*, vol. 44, pp. 29–43, 2015.

[25] B. Mesot and D. Barber, "Switching linear dynamical systems for noise robust speech recognition," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 6, pp. 1850–1858, 2007.

[26] N. Ozay, M. Sznaier, C. M. Lagoa, and O. I. Camps, "A sparsification approach to set membership identification of switched affine systems," *IEEE Transactions on Automatic Control*, vol. 57, no. 3, pp. 634–648, 2011.

[27] L. Lo Presti, M. La Cascia, S. Sclaroff, and O. Camps, "Gesture modeling by hanklet-based hidden markov model," in *Asian Conference on Computer Vision*. Springer, 2014, pp. 529–546.

[28] N. İkizler and D. A. Forsyth, "Searching for complex human activities with no visual examples," *International Journal of Computer Vision*, vol. 80, no. 3, pp. 337–357, 2008.

[29] M. Dong and D. He, "A segmental hidden semi-markov model (hsmm)-based diagnostics and prognostics framework and methodology," *Mechanical systems and signal processing*, vol. 21, no. 5, pp. 2248–2266, 2007.

[30] S. Aminikhanghahi and D. J. Cook, "A survey of methods for time series change point detection," *Knowledge and information systems*, vol. 51, no. 2, pp. 339–367, 2017.

[31] M. Boettcher, "Contrast and change mining," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 1, no. 3, pp. 215–230, 2011.

[32] S. Hido, T. Idé, H. Kashima, H. Kubo, and H. Matsuzawa, "Unsupervised change analysis using supervised learning," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, 2008, pp. 148–159.

[33] D. Kifer, S. Ben-David, and J. Gehrke, "Detecting change in data streams," in *Proceedings of the Thirtieth international conference on Very large data bases-Volume 30*. VLDB Endowment, 2004, pp. 180–191.

[34] R. Iniesta, D. Stahl, and P. McGuffin, "Machine learning, statistical learning and the future of biological research in psychiatry," *Psychological medicine*, vol. 46, no. 12, pp. 2455–2465, 2016.

[35] L. Lo Presti, S. Sclaroff, and A. Rozga, "Joint alignment and modeling of correlated behavior streams," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2013, pp. 730–737.

[36] C. Achard, X. Qu, A. Mokhber, and M. Milgram, "A novel approach for recognition of human actions with semi-global features," *Machine Vision and Applications*, vol. 19, no. 1, pp. 27–34, 2008.

[37] S. M. Anzalone, G. Varni, S. Ivaldi, and M. Chetouani, "Automated prediction of extraversion during human–humanoid interaction," *International Journal of Social Robotics*, vol. 9, no. 3, pp. 385–399, 2017.

[38] V. Monteleone, L. Lo Presti, and M. La Cascia, "Hop: Histogram of patterns for human action representation," in *International Conference on Image Analysis and Processing*. Springer, 2017, pp. 457–468.

[39] L. Senger, M. Schröer, J. H. Metzen, and E. A. Kirchner, "Velocity-based multiple change-point inference for unsupervised segmentation of human movement behavior," in *2014 22nd International Conference on Pattern Recognition*. IEEE, 2014, pp. 4564–4569.

[40] N. Ambady and R. Rosenthal, "Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis." *Psychological bulletin*, vol. 111, no. 2, p. 256, 1992.

[41] A. Pentland, "A computational model of social signalin," in *18th International Conference on Pattern Recognition (ICPR'06)*, vol. 1. IEEE, 2006, pp. 1080–1083.

[42] T. Liu and A. Kappas, "Predicting engagement breakdown in hri using thin-slices of facial expressions," in *Workshops at the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[43] T. Kanda, R. Sato, N. Saiwaki, and H. Ishiguro, "A two-month field trial in an elementary school for long-term human–robot interaction," *IEEE Transactions on robotics*, vol. 23, no. 5, pp. 962–971, 2007.

[44] S. Ivaldi, S. M. Anzalone, W. Rousseau, O. Sigaud, and M. Chetouani, "Robot initiative in a team learning task increases the rhythm of interaction but not the perceived engagement," *Frontiers in neurorobotics*, vol. 8, p. 5, 2014.

[45] N. Cristianini, J. Shawe-Taylor *et al.*, *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press, 2000.

[46] S. Reddy, M. Mun, J. Burke, D. Estrin, M. Hansen, and M. Srivastava, "Using mobile phones to determine transportation modes," *ACM Transactions on Sensor Networks (TOSN)*, vol. 6, no. 2, p. 13, 2010.

[47] Y. Zheng, L. Liu, L. Wang, and X. Xie, "Learning transportation mode from raw gps data for geographic applications on the web," in *Proceedings of the 17th international conference on World Wide Web*. ACM, 2008, pp. 247–256.

[48] E. Ahmed, A. Clark, and G. Mohay, "A novel sliding window based change detection algorithm for asymmetric traffic," in *2008 IFIP International Conference on Network and Parallel Computing*. IEEE, 2008, pp. 168–175.

[49] J. Xavier, S. Gauthier, D. Cohen, M. Zahoui, M. Chetouani, F. Villa, A. Berthoz, and S. Anzalone, "Interpersonal synchronization, motor coordination, and control are impaired during a dynamic imitation task in children with autism spectrum disorder," *Frontiers in psychology*, vol. 9, p. 1467, 2018.

[50] S. Gauthier, S. M. Anzalone, D. Cohen, M. Zaoui, M. Chetouani, F. Villa, A. Berthoz, and J. Xavier, "Behavioral own-body-transformations in children and adolescents with typical development, autism spectrum disorder, and developmental coordination disorder," *Frontiers in psychology*, vol. 9, p. 676, 2018.

[51] B. Thirioux, G. Jorland, M. Bret, M.-H. Tramus, and A. Berthoz, "Walking on a line: A motor paradigm using rotation and reflection symmetry to study mental body transformations," *Brain and cognition*, vol. 70, no. 2, pp. 191–200, 2009.

[52] B. Ingersoll, "The social role of imitation in autism: Implications for the treatment of imitation deficits," *Infants & Young Children*, vol. 21, no. 2, pp. 107–119, 2008.

[53] H. Roeyers, P. Van Oost, and S. Bothuyne, "Immediate imitation and joint attention in young children with autism," *Development and psychopathology*, vol. 10, no. 3, pp. 441–450, 1998.

[54] A. Whiten and J. Brown, "Imitation and the reading of other minds: Perspectives from the study of autism, normal children and non-human primates," *Intersubjective communication and emotion in early ontogeny*, pp. 260–280, 1998.

[55] M. Mazeau, "Troubles neurovisuels et praxiques: un élément déterminant du pronostic à long terme," *Médecine thérapeutique/Pédiatrie*, vol. 3, no. 4, pp. 273–80, 2000.

[56] T. Idé and H. Kashima, "Eigenspace-based anomaly detection in computer systems," in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2004, pp. 440–449.

[57] Y. Kawahara, T. Yairi, and K. Machida, "Change-point detection in time-series data based on subspace identification," in *Seventh IEEE International Conference on Data Mining (ICDM 2007)*. IEEE, 2007, pp. 559–564.

[58] T. Flash and N. Hogan, "The coordination of arm movements: an experimentally confirmed mathematical model," *Journal of neuroscience*, vol. 5, no. 7, pp. 1688–1703, 1985.

[59] M. Basseville, I. V. Nikiforov *et al.*, *Detection of abrupt changes: theory and application*. Prentice Hall Englewood Cliffs, 1993, vol. 104.

[60] B. James, K. L. James, and D. Siegmund, "Tests for a change-point," *Biometrika*, vol. 74, no. 1, pp. 71–83, 1987.

[61] J. Chen and A. K. Gupta, "On change point detection and estimation," *Communications in statistics-simulation and computation*, vol. 30, no. 3, pp. 665–697, 2001.

[62] A. Willsky and H. Jones, “A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems,” *IEEE Transactions on Automatic control*, vol. 21, no. 1, pp. 108–112, 1976.

[63] F. Gustafsson, “The marginalized likelihood ratio test for detecting abrupt changes,” *IEEE Transactions on automatic control*, vol. 41, no. 1, pp. 66–78, 1996.

[64] I. T. Young, “Proof without prejudice: use of the kolmogorov-smirnov test for the analysis of histograms from flow systems and other sources.” *Journal of Histochemistry & Cytochemistry*, vol. 25, no. 7, pp. 935–941, 1977.

[65] E. Brodsky and B. S. Darkhovsky, *Nonparametric methods in change point problems*. Springer Science & Business Media, 2013, vol. 243.

[66] B. P. Carlin, A. E. Gelfand, and A. F. Smith, “Hierarchical bayesian analysis of changepoint problems,” *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 41, no. 2, pp. 389–405, 1992.

[67] I. Cleland, M. Han, C. Nugent, H. Lee, S. McClean, S. Zhang, and S. Lee, “Evaluation of prompted annotation of activity data recorded from a smart phone,” *Sensors*, vol. 14, no. 9, pp. 15 861–15 879, 2014.

[68] M. Han, Y.-K. Lee, S. Lee *et al.*, “Comprehensive context recognizer based on multimodal sensors in a smartphone,” *Sensors*, vol. 12, no. 9, pp. 12 588–12 605, 2012.

[69] L. Wei and E. Keogh, “Semi-supervised time series classification,” in *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2006, pp. 748–753.

[70] S. Liu, M. Yamada, N. Collier, and M. Sugiyama, “Change-point detection in time-series data by relative density-ratio estimation,” *Neural Networks*, vol. 43, pp. 72–83, 2013.

[71] J. Zhang and M. Small, “Complex network from pseudoperiodic time series: Topology versus dynamics,” *Physical review letters*, vol. 96, no. 23, p. 238701, 2006.

[72] L. Lacasa, B. Luque, F. Ballesteros, J. Luque, and J. C. Nuno, “From time series to complex networks: The visibility graph,” *Proceedings of the National Academy of Sciences*, vol. 105, no. 13, pp. 4972–4975, 2008.

[73] P. R. Rosenbaum, “An exact distribution-free test comparing two multivariate distributions based on adjacency,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 4, pp. 515–530, 2005.

[74] K. V. Mardia, “Measures of multivariate skewness and kurtosis with applications,” *Biometrika*, vol. 57, no. 3, pp. 519–530, 1970.

[75] D. Eager, A.-M. Pendrill, and N. Reistad, “Beyond velocity and acceleration: jerk, snap and higher derivatives,” *European Journal of Physics*, vol. 37, no. 6, p. 065008, 2016.

[76] N. Hogan and D. Sternad, “Sensitivity of smoothness measures to movement duration, amplitude, and arrests,” *Journal of motor behavior*, vol. 41, no. 6, pp. 529–534, 2009.

[77] S. Balasubramanian, A. Melendez-Calderon, and E. Burdet, “A robust and sensitive metric for quantifying movement smoothness,” *IEEE transactions on biomedical engineering*, vol. 59, no. 8, pp. 2126–2136, 2011.

[78] K. D. Feuz, D. J. Cook, C. Rosasco, K. Robertson, and M. Schmitter-Edgecombe, “Automated detection of activity transitions for prompting,” *IEEE transactions on human-machine systems*, vol. 45, no. 5, pp. 575–585, 2014.

[79] F. Desobry, M. Davy, and C. Doncarli, “An online kernel change detection algorithm,” *IEEE Trans. Signal Processing*, vol. 53, no. 8-2, pp. 2961–2974, 2005.

[80] R. Caruana and A. Niculescu-Mizil, “An empirical comparison of supervised learning algorithms,” in *Proceedings of the 23rd international conference on Machine learning*, 2006, pp. 161–168.

[81] S.-i. Amari and S. Wu, “Improving support vector machine classifiers by modifying kernel functions,” *Neural Networks*, vol. 12, no. 6, pp. 783–789, 1999.

[82] M. Ester, H.-P. Kriegel, J. Sander, X. Xu *et al.*, “A density-based algorithm for discovering clusters in large spatial databases with noise.” in *Kdd*, vol. 96, no. 34, 1996, pp. 226–231.

[83] A. P. Association *et al.*, *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub, 2013.

[84] S. M. Anzalone, E. Tilmont, S. Boucenna, J. Xavier, A.-L. Jouen, N. Bodeau, K. Maharatna, M. Chetouani, D. Cohen, M. S. Group *et al.*, “How children with autism spectrum disorder behave and explore the 4-dimensional (spatial 3d+ time) environment during a joint attention induction task with a robot,” *Research in Autism Spectrum Disorders*, vol. 8, no. 7, pp. 814–826, 2014.

[85] S. M. Anzalone, J. Xavier, S. Boucenna, L. Billeci, A. Narzisi, F. Muratori, D. Cohen, and M. Chetouani, “Quantifying patterns of joint attention during human-robot interactions: An application for autism spectrum disorder assessment,” *Pattern Recognition Letters*, vol. 118, pp. 42–50, 2019.

[86] B. Scholkopf, K.-K. Sung, C. J. Burges, F. Girosi, P. Niyogi, T. Poggio, and V. Vapnik, “Comparing support vector machines with gaussian kernels to radial basis function classifiers,” *IEEE transactions on Signal Processing*, vol. 45, no. 11, pp. 2758–2765, 1997.