A Self-Tuned Architecture for Human Activity Recognition Based on a Dynamical Recurrence Analysis of Wearable Sensor Data

M.-A. Zervou[1,2], G. Tzagkarakis[1], A. Panousopoulou[1], and P. Tsakalides[1,2]

[1]Institute of Computer Science, Foundation for Research and Technology-Hellas, Heraklion, Greece
[2]Department of Computer Science, University of Crete, Heraklion, Greece
E-mails: zervou@ics.forth.gr, gtzag@ics.forth.gr, n.panousopoulou@gmail.com, tsakalid@ics.forth.gr

Abstract—Human activity recognition (HAR) is encountered in a plethora of applications, such as pervasive health care systems and smart homes. The majority of existing HAR techniques employs features extracted from symbolic or frequency-domain representations of the associated data, whilst ignoring completely the behavior of the underlying data generating dynamical system. To address this problem, this work proposes a novel self-tuned architecture for feature extraction and activity recognition by modeling directly the inherent dynamics of wearable sensor data in higher-dimensional phase spaces, which encode state recurrences for each individual activity. Experimental evaluation on real data of leisure activities demonstrates an improved recognition accuracy of our method when compared against a state-of-the-art motif-based approach using symbolic representations.

Index Terms—Human activity recognition, recurrence quantification analysis, nonlinear data analysis, motif discovery, wearable sensors

I. INTRODUCTION

Human activity recognition (HAR) and classification using wearable sensor data is gaining an ever increasing interest, mainly due to its utility in a plethora of applications, such as in pervasive health care systems [1], smart homes [2], and surveillance systems for indoor and outdoor activities [3]. Nevertheless, the latest trends in HAR primarily focus on addressing the challenge of moving towards data processing architectures governed by the need to analyze and decipher complex activities while in data capture [4], [5]. To this end, motif discovery attempts to extract new, meaningful and unknown knowledge from data, as well as to monitor and track structural similarities in streaming data generated by a network of sensors. More specifically, time series motifs, which are approximately repeated subsequences in a longer time window [6], are exploited for representing the meaningful information content of the original data. Numerous studies have shown the potential of using motifs for detecting and classifying activities and events [7]–[9], by addressing the problem of motif discovery in unidimensional data.

Among the several existing motif discovery techniques, the symbolic aggregate approximation (SAX) method [10] has a prominent role, due to its conceptual simplicity and computational efficiency. Recently, a modified grammar induction algorithm, namely, the modified Sequitur algorithm [8], [11], [12] has been introduced, to improve the performance of SAX. Although such methods can lead to high precision results in the unidimensional case for relatively smooth data, their performance often degrades in more general cases. Furthermore, their enhanced performance comes at the cost of an increased sensitivity to the values of a set of parameters, whose accurate tuning is a demanding task.

To overcome these limitations, this work proposes an alternative approach for accurate human activity recognition, which exploits the temporal variability of the underlying dynamical system that generates the data associated with a specific activity. To this end, *recurrence quantification analysis* (RQA) [13] will be exploited to perform a sophisticated nonlinear analysis of sensor streams, while being also able to treat *nonstationary* and *short* data series. RQA comprises of a set of appropriate quantitative measures for the quantification of recurrent, typically small-scale, structures, and the detection of critical transitions in the system's dynamics (e.g. deterministic, stochastic, random).

RQA has been used recently in the field of HAR, demonstrating a promising performance. In [14], the complexity and recurrence properties of daily non-complex human activities measured by wearable sensors are investigated. Along these lines, [15] combines RQA and multiscale entropy to characterize trunk postural control and motor complexity during childhood, while [16] employs recurrence plots to quantify the local dynamic stability of human walking kinematics. On the other hand, the methods proposed in [17]–[19] focus on offline experimentation with benchmark datasets and various classifiers applied on time- or frequency-domain features.

Although these previous studies also try to exploit the inherent time-varying dynamics of the recorded data, however, either they rely on the time variations of statistical or recurrence measures, or they employ complex classifiers, in order to study qualitatively the differences between distinct activities. To address the above limitations, the contributions of our work are the following: (i) the underlying data generating

processes are modeled directly in a higher-dimensional phase space identifying more accurately the time-evolving dynamics of sensor streams; (ii) an efficient feature extraction scheme is designed for the discovery of information-rich patterns that best capture the underlying data dynamics; and (iii) a totally self-tuned architecture is designed for unsupervised HAR.

The rest of the paper is organized as follows: Section II introduces briefly a well-established motif discovery method, the so-called SAX, which has been extensively used for HAR. Section III analyzes in detail our proposed HAR architecture, based on RQA features and a linear-kernel support vector machine for activity recognition. Section IV evaluates the performance of our method on a real dataset of complex leisure activities, and compares its accuracy with a SAX-based approach. Finally, Section V summarizes the main outcomes of this work and gives directions for future extensions.

II. SAX-BASED MOTIF DISCOVERY FOR HAR

This section describes briefly the core architecture of a state-of-the-art online motif discovery method, shown in Fig. 1, which will be used as a benchmark to compare against our proposed RQA-based framework. The rationale for choosing motif discovery stems from the fact that the occurrence frequency of the motifs is used as a classification feature, instead of the symbolic representation itself. This resembles the RQA approach, which is based on the number of state recurrences in the phase space.

Discovering similarities directly in streaming data is typically a demanding task, in terms of computational and memory complexity. In order to improve the efficiency of data mining from time series, the symbolic aggregate approximation (SAX) is a well-established technique. SAX transforms the sensor data streams into strings of discrete symbols. The main benefit of this algorithm is the effective dimensionality reduction, while satisfying a lower bounding property, which guarantees that a distance measure applied on two symbolic strings lower bounds the true distance between the original time series [10]. Specifically, a sliding window runs over the time series, and a vertical segmentation is applied first, which divides the current window into equally sized non-overlapping segments. Then, for each segment the average value is computed followed by a horizontal segmentation, which associates the average to a symbol selected from a predetermined alphabet. This alphabet is constructed by dividing the whole range of values of a given time series into 2^Q intervals, where $Q \in \mathbb{N}$ is defined by the user according to the required granularity. Then, each symbol from the alphabet is assigned to an interval determined by a pair of breakpoints. In our implementation, the set of breakpoints is defined by the real numbers that divide the area under the standard Gaussian distribution into 2^Q equal regions.

To extract re-occurring patterns (a.k.a. motifs) from a SAX-based symbolic representation of the current window, the modified Sequitur algorithm is applied next. This recursive algorithm detects repetitions of bigrams, that is, of two consecutive symbols, and extracts them from the symbolic string by defining rules constituting a grammar. In order
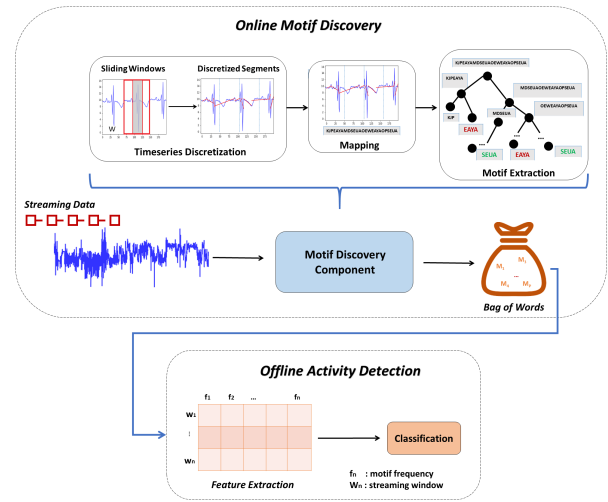

Fig. 1. SAX-based motif discovery and HAR architecture.

to obtain motifs of variable length, each rule corresponds to a set of bigrams, instead of matching a single bigram, as in the original Sequitur algorithm [11]. Having applied these rules, a bag of motifs is extracted for each streaming window, followed by a feature extraction process. In particular, a supervised learning approach is employed, where bags of motifs representing streams that correspond to the same target activity are merged, forming a new bag of unique core activity motifs. Then, for each motif in the core activity bag, the frequency of occurrence among the set of already extracted bags is measured. Finally, the frequencies of all motifs over all the available time windows constitute the feature matrix, to be further used for activity classification.

III. PROPOSED RQA-BASED FEATURE EXTRACTION FOR HAR

In contrast to a SAX-based approach, our proposed method capitalizes on the efficiency of RQA to extract the underlying dynamics of a recorded data stream by mapping the time series in a higher-dimensional phase space of trajectories. A major advantage of our feature extraction approach, when compared with symbolic representations for motif discovery, is the *fully self-tuned nature*, in the sense that no prior parameter fine-tuning is required in a manual fashion, as is the case with SAX-based techniques.

More specifically, a recurrence plot (RP) is derived first, which depicts those times at which a state of a dynamical system recurs, thus revealing all the times when the *phase space trajectory* of the dynamical system visits roughly the same area in the phase space. To this end, RPs enable the investigation of an m-dimensional phase space trajectory through a two-dimensional representation of its recurrences. Such recurrence of a state occurring at time i, at a different time j is represented within a two-dimensional square matrix with ones (recurrence) and zeros (non-recurrence), where both axes are time axes.

Given a time series of length N, $\{r_i\}_{i=1}^N$, a phase space trajectory can be reconstructed via time-delay embedding,

$$\mathbf{x}_i = [r_i, r_{i+\tau}, \ldots, r_{i+(m-1)\tau}] , \quad i = 1, \ldots, N_s , \quad (1)$$

where m is the embedding dimension, τ is the delay, and $N_s = N - (m-1)\tau$ is the number of states. Having constructed a phase space representation, an RP is defined as follows,

$$\mathbf{R}_{i,j} = \Theta \left(\varepsilon - \|\mathbf{x}_i - \mathbf{x}_j\|_p \right) , \quad i, j = 1, \ldots, N_s , \quad (2)$$

where $\mathbf{x}_i$, $\mathbf{x}_j \in \mathbb{R}^m$ are the states, ε is a threshold, $\|\cdot\|_p$ denotes a general ℓ_p norm, and $\Theta(\cdot)$ is the Heaviside step function, whose discrete form is defined by

$$\Theta(n) = \begin{cases} 1, & \text{if } n \geq 0 \\ 0, & \text{if } n < 0 \end{cases} , \quad n \in \mathbb{R} . \quad (3)$$

The resulting matrix $\mathbf{R}$ exhibits the main diagonal, $\mathbf{R}_{i,i} = 1$, $i = 1, \ldots, N_s$, also known as the *line of identity* (LOI). Typically, several linear (and/or curvilinear) structures appear in RPs, which give hints about the time evolution of the high-dimensional phase space trajectories. Besides, a major advantage of RPs is that they can also be applied to rather *short* and even *nonstationary* data. Fig. 2 shows the RPs for two distinct time windows corresponding to core and non-core activities, respectively. Clearly, in the former case the RP is able to identify state recurrences, which are visualized in the form of shorter or longer diagonal and vertical line segments, whereas in the later case, the absence of structure in the associated time window is expressed in the form of isolated points in the RP.
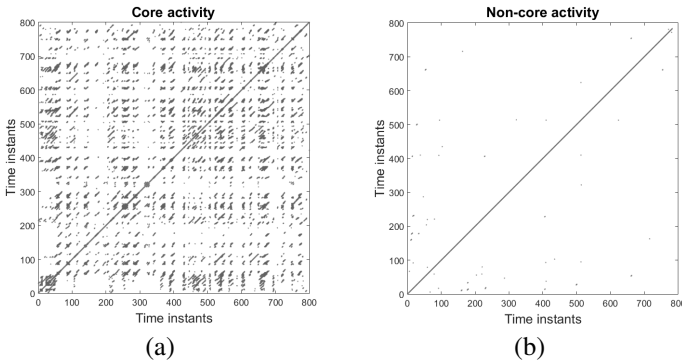

Fig. 2. RPs of (a) core and (b) non-core activity segments.

The visual interpretation of RPs, which is often difficult and subjective, is enhanced by means of several numerical measures for the quantification of the structure and complexity of RPs [20]. These quantification measures provide a global picture of the underlying dynamical behavior during the entire period covered by the HAR data. The temporal evolution of RQA measures and the subsequent detection of transient dynamics are enabled for each recorded data stream by employing a windowed version of RQA. Doing so, the corresponding quantification measures are computed in small windows, which are then merged to form our feature
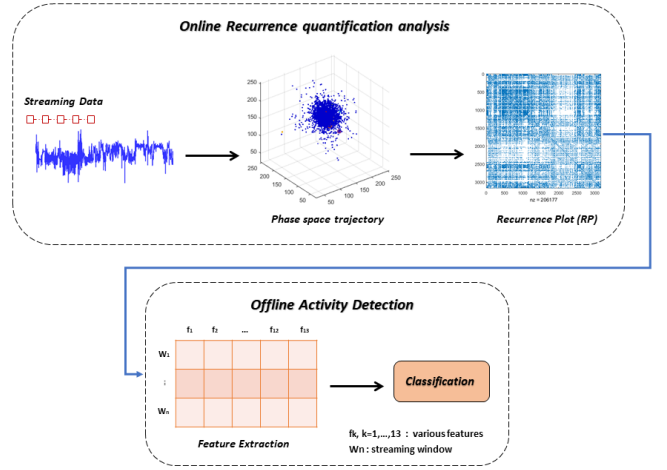

Fig. 3. Proposed RQA-based feature extraction and HAR architecture.

matrix. Furthermore, it is noted that the length of the sliding window yields a compromise between resolving small-scale local fluctuations and detecting recurrence structures located farther away from the LOI. The following ten RQA measures are utilized in order to form our feature matrix (ref. [21] for the definitions): recurrence rate, determinism, average diagonal length, length of longest diagonal line, entropy of diagonal length, laminarity, trapping time, length of longest vertical line, clustering coefficient, and transitivity. Finally, a linear-kernel support vector machine (SVM) is applied on the feature matrix for activity recognition. Fig. 3 shows the overall architecture of our proposed RQA-based HAR system.

Estimation of embedding parameters: In our implementation, the optimal time delay τ is estimated as the first minimum of the average mutual information (AMI) function [22]. Concerning the embedding dimension m, a minimal sufficient value is estimated using the method of false nearest neighbours (FNN) [23]. Furthermore, the maximum norm is used as our selected distance metric for the construction of the RP, which is defined as $\|\mathbf{x}\|_{max} = \max_{i=1,\ldots,N} |x_i|$, while a rule-of-thumb is currently used to set the threshold $\varepsilon = 0.2\sqrt{m}$.

IV. PERFORMANCE EVALUATION

In order to evaluate the performance of our proposed RQA-based HAR architecture, and compare against the motif-based counterpart, we employ a publicly available dataset [7], which includes leisure activities data for a group of six volunteers. Each volunteer performs a single activity once a day over a period of five days. A remarkable aspect of this dataset is that the sampling rate varies, being higher during the target activity. We particularly focus on three leisure activities, namely, "cycling", "playing the guitar" and "dancing flamenco". These activities are selected such that both periodic and intense (in "cycling"), as well as sharp and delicate movements (in "dancing flamenco" and "playing the guitar"), are examined.

The recorded data correspond to the acceleration measured across the x-axis direction. The length of the non-overlapping

Fig. 4. Confusion matrices for "cycling" (1st row), "dancing flamenco" (2nd row) and "playing the guitar" (3rd row) activities, using motif discovery (left column) and RQA (right column). Green boxes represent true positive (TPR) and true negative (FPR) rates for class 1 (activity) and 2 (non-activity). Red boxes correspond to false positive (FPR) and false negative (FNR) rates. Gray boxes represent the precision and recall percentage (in green) and the error rate (in red). Blue boxes contain the classification accuracy (in green) and expected error rate (in red).

sliding windows is set equal to 1 minute. Instead of performing calculations on the entire dataset, we select a total of 100 consecutive windows, including time intervals before, during, and after the target activity. The sensor data and metadata of each activity during the five days of execution are concatenated and then divided randomly into training and testing subsets containing 75% and 25% of the data, respectively. For the comparison with the state-of-the-art motif-based SAX-Sequitur scheme, the optimal parameters of the associated algorithms are tuned using a nested cross-validation process. Specifically, the word length is set equal to 4, the alphabet size equal to 5, and the sliding window length to 80 samples.

The two HAR architectures are implemented in MATLAB, on a desktop computer equipped with a CPU processor (Intel Core i5-4590) clocked at 3.30GHz, and a 8 GB RAM, while the CRP toolbox (http://tocsy.pik-potsdam.de/CRPtoolbox/) has been employed for implementing the RQA. Lastly, a non-linear classifier, specifically, a linear-kernel SVM, is applied on the generated feature matrix in order to perform activity recognition, which is addressed as a classification problem. The choice of this classifier is motivated by its fast execution, as well as by its high accuracy, especially in the case of a large number of available features. We emphasize, though, that the classification step is decoupled from the feature extraction step, thus the overall performance of a HAR architecture can be further improved by employing a better classifier.

The performance of our RQA-based HAR architecture is compared against the SAX-Sequitur approach, in terms of classification accuracy, F-score and memory complexity. As illustrated in Fig. 4, our method outperforms significantly the motif-based scheme for every activity. In particular, the results corresponding to the "cycling" and "playing the guitar" activity reveal that for periodic, yet complex activities, our proposed RQA-based feature matrix is better capable of detecting accurately the temporal variations of the underlying dynamical system that generates the corresponding data, yielding a high accuracy at the order of 92% and 93.6%, respectively.

TABLE I
F-SCORE FOR EACH TARGET ACTIVITY CLASS (CLASS 1)

	F-score for target activity class (%)		
	Cycling	Flamenco	Guitar
Motif-based	39.29	46.88	74.80
RQA-based	88.10	79.19	94.29

TABLE II
NUMBER OF FEATURES EXTRACTED FROM 100 STREAMS

	Number of features		
	Cycling	Flamenco	Guitar
Motif-based	2783	884	323
RQA-based	130	130	130

Moreover, Table I demonstrates that, although the motif-based method achieved a 72.8% accuracy in correctly classifying a window as a core or non-core activity, however, it yields a low F-score for class 1 (target activity). On the contrary, our RQA-based method achieves a high accuracy, while recognizing precisely the target activity, as demonstrated by the high F-scores for both classes and the zero false negative rate (FNR). This also verifies the capability of the employed classifier to discriminate accurately between the core activity and any other activity.

Considering the "dancing flamenco" activity, the motif-based scheme performs rather poorly, achieving a total accuracy of 45.6% and an F-score equal to 46.88% for the core activity class. This result reveals a degraded performance of the motif discovery scheme, which is not capable of detecting accurately such a complex activity with varying sampling rate. On the contrary, our RQA-based method yields a significantly higher accuracy and F-score values, at the order of 75.2% and 79.19%, respectively. This also verifies the efficiency of
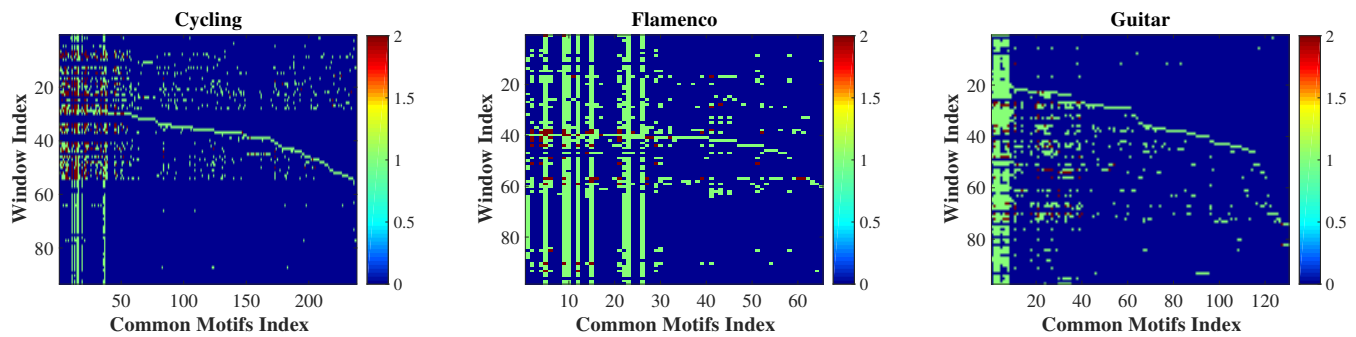
Fig. 5. Target activity motif representation in each streaming window for the three activities. Target activities appear in the windows 29-77, 30-89 and 13-85, respectively.

our RQA-based feature extraction scheme to generate features with an increased discriminating capability between the core and non-core activity, yielding an FNR at the order of 4%.

Finally, Fig. 5 indicates that for the motif-based scheme, the motifs referring to the target activity appear both in target and non-target activity windows. Specifically, for the "dancing flamenco" and "playing the guitar", the low performance is primarily due to the fact that these activities include various irregular and complex motions, which are quite difficult to capture via the extracted motifs. To this end, Table II indicates that the number of features extracted by the motif-based method not only differs significantly between the activities, but is also much larger compared with our RQA-based approach.

V. CONCLUSIONS AND FUTURE WORK

In this work, we designed and implemented a HAR architecture based on a representation of wearable sensor data in higher-dimensional phase spaces using the RQA method for capturing the underlying dynamics of the data. The experimental evaluation on real leisure data revealed the superiority of our RQA-based framework in extracting and exploiting the underlying temporal dynamics of the data generating processes, resulting in significantly higher classification accuracy for complex activities, when compared against a state-of-the-art online motif discovery scheme based on symbolic representations. An extension of this work will consider the use of multidimensional (multichannel) HAR data based on joint RPs and a multidimensional extension of RQA. We expect that the incorporation of potential correlations and joint dynamics between the multiple data streams will further increase the overall classification accuracy.

REFERENCES

[1] V. Osmani et al., "Human activity recognition in pervasive health-care: Supporting efficient remote collaboration," *J. Network and Comp. App.*, vol. 31, no. 4, pp. 628–655, 2008.

[2] L. Chen, C. D. Nugent, and H. Wang, "A knowledge-driven approach to activity recognition in smart homes," *IEEE Trans. Knowledge and Data Eng.*, vol. 24, no. 6, pp. 961–974, June 2012.

[3] P. Bharti et al., "HuMAn: Complex activity recognition with multi-modal multi-positional body sensing," *IEEE Trans. Mobile Comp.*, pp. 1–1, Early access 2018.

[4] K. Karagiannaki et al., "An online feature selection architecture for human activity recognition," in *IEEE Intl. Conf. on Acoust., Speech and Sig. Process. (ICASSP)*, 2017, pp. 2522–2526.

[5] T. Ya, W. Hongxing, and T. Jindong, "An adaptive-gain complementary filter for real-time human motion tracking with MARG sensors in free-living environments," *IEEE Trans. Neural Systems and Rehab. Eng.*, vol. 21, no. 2, pp. 254–264, 2013.

[6] B. Chiu, E. Keogh, and S. Lonardi, "Probabilistic discovery of time series motifs," in *Proc. 9th ACM SIGKDD*, 2003, pp. 493–498.

[7] E. Berlin and K. V. Laerhoven, "Detecting leisure activities with dense motif discovery," in *Proc. ACM Conf. on Ubiq. Comp.*, 2012, pp. 250–259.

[8] A. Balasubramanian and B. Prabhakaran, "Flexible exploration and visualization of motifs in biomedical sensor data," in *Proc. Worksh. on Data Mining for Healthcare, in conj. with ACM KDD*, 2013.

[9] T. Fu, "A review on time series data mining," *Eng. App. of Artificial Intell.*, vol. 24, no. 1, pp. 164–181, 2011.

[10] J. Lin, E. Keogh, S. Lonardi, and B. Chiu, "A symbolic representation of time series, with implications for streaming algorithms," in *Proc. 8th ACM SIGMOD Worksh. on Res. Issues in Data Mining and Knowledge Discovery*, 2003, pp. 2–11.

[11] C. G. Nevill-Manning and I. H. Witten, "Identifying hierarchical structure in sequences: A linear-time algorithm," *J. Artificial Intell. Res.*, vol. 7, pp. 67–82, 1997.

[12] Y. Li, J. Lin, and T. Oates, "Visualizing variable-length time series motifs," in *Proc. SIAM Intl. Conf. on Data Mining*, 2012, pp. 895–906.

[13] J. Eckmann, S. O. Kamphorst, and D. Ruelle, "Recurrence plots of dynamical systems," *Europh. Lett.*, vol. 4, no. 9, p. 973, 1987.

[14] H. Josiński et al., "Recurrence plots and recurrence quantification analysis of human motion data," in *Proc. AIP Conf.*, vol. 1738, no. 1.

[15] M. Bisi et al., "Nonlinear analysis of human movement dynamics offers new insights in the development of motor control during childhood," *J. Biomech. Eng.*, vol. 140, no. 11, p. 111002, 2018.

[16] J. Dingwell and J. Cusumano, "Nonlinear time series analysis of normal and pathological human walking," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 10, no. 4, pp. 848–863, 2000.

[17] K. Nakano and B. Chakraborty, "Effect of dynamic feature for human activity recognition using smartphone sensors," in *Proc. IEEE 8th Intl. Conf. on Awareness Sci. and Tech. (iCAST)*, 2017, pp. 539–543.

[18] L. Ye et al., "Physical violence detection with movement sensors," in *Proc. Intern. Conf. on Machine Learning and Intell. Comm.* Springer, 2018, pp. 190–197.

[19] R. F. Albuquerque, P. de Oliveira, and A. Braga, "Adaptive fuzzy learning vector quantization (AFLVQ) for time series classification," in *Proc. North American Fuzzy Inf. Process. Society Ann. Conf.* Springer, 2018, pp. 385–397.

[20] J. Zbilut and C. Webber, "Embeddings and delays as derived from quantification of recurrence plots," *Phys. Lett. A*, vol. 171, pp. 199–203, 1992.

[21] N. Marwan, M. Romano, M. Thiel, and J. Kurths, "Recurrence plots for the analysis of complex systems," *Physics Reports*, vol. 438, no. 5-6, pp. 237–329, 2007.

[22] A. Fraser and H. Swinney, "Independent coordinates for strange attractors from mutual information," *Phys. Rev. A*, vol. 33, pp. 1134–1140, Feb. 1986.

[23] M. B. Kennel, R. Brown, and H. D. Abarbanel, "Determining embedding dimension for phase-space reconstruction using a geometrical construction," *Phys. Rev. A*, vol. 45, p. 3403, 1992.