
Quasi-periodic time series clustering for human activity recognition

A. V. Grabovoy^{1,*} and V. V. Strijov^{1,**}

¹Moscow Institute of Physics and Technology (National Research University), Moscow, Russia

Received June 13, 2018

Abstract—This paper analyses the periodic signals in the time series to recognize human activity by using a mobile accelerometer. Each point in the timeline corresponds to a segment of historical time series. These segments form a phase trajectory in phase space of human activity. The principal components of segments of the phase trajectory are treated as feature descriptions at the point in the timeline. The paper introduces a new distance function between the points in new feature space. To reveal changes of types of the human activity the paper proposes an algorithm. This algorithm clusters points of the timeline by using a pairwise distances matrix. The algorithm was tested on synthetic and real data. This real data were obtained from a mobile accelerometer.

2010 Mathematical Subject Classification: 12345, 54321

Keywords and phrases: Time series, Clustering, Segmentation, Recognition of physical activity, Principal component method

1. INTRODUCTION

Analysis of the human physical activity is carried out by using mobile phones, smart watches and another wearable devices [1, 2]. These devices use an accelerometer, gyroscope and magnetometer. The main purpose of this work is to construct an informative low-dimensional representation to generate features of a human behaviour [3, 4]. This representation also defines the phase of the movements [5, 6]. Examples of the physical action segment are a step, a step of running, a single squat, a single jump, etc. This work considers a sequence that consists of at least two consecutive segments that correspond to the same type of human activity.

Time series are objects of complex structure. The method of constructing feature vectors for points is very important for their clustering. In this work, the object of analysis is a point on the time axis. Any sample set is collected as samples, taken before this point. This work investigates the clustering problem: all points on the time line are labeled with a label from some finite set of labels. Each label corresponds to physical action. A segment of time series corresponds to physical action. For example: a step with two legs when walking, or a step with two legs when running. A sequence of segments that forms a quasi-periodic sequence is called the chain. A sequence of points $\{b_t\}_{t=1}^N$ is called quasi-periodic with period of T , if for all t there is a Δ , such that:

$$b_t \approx b_{t+T+\Delta}, \quad |\Delta| \ll T. \quad (1)$$

An example of clustering and splitting a series into segments is shown in fig. 1.a. Fig. 1.a shows an example of time series clustering and segmentation. The time series is splitted into two characteristic type of segments: Type 1 and Type 2. This time series contains two quasi-periodic chains of actions.

The proposed solution of the clustering problem consists of two stages. First, an algorithm for local approximation of time series by using the principal component method [11] obtains a feature description of points in the time series. Local information of the time series means that only a certain neighborhood of a point is used to describe the features of their points. A few main components of the phase trajectory segment are considered as a feature description of a time series

* E-mail: grabovoy.av@phystech.edu

** E-mail: strijov@ccas.ru

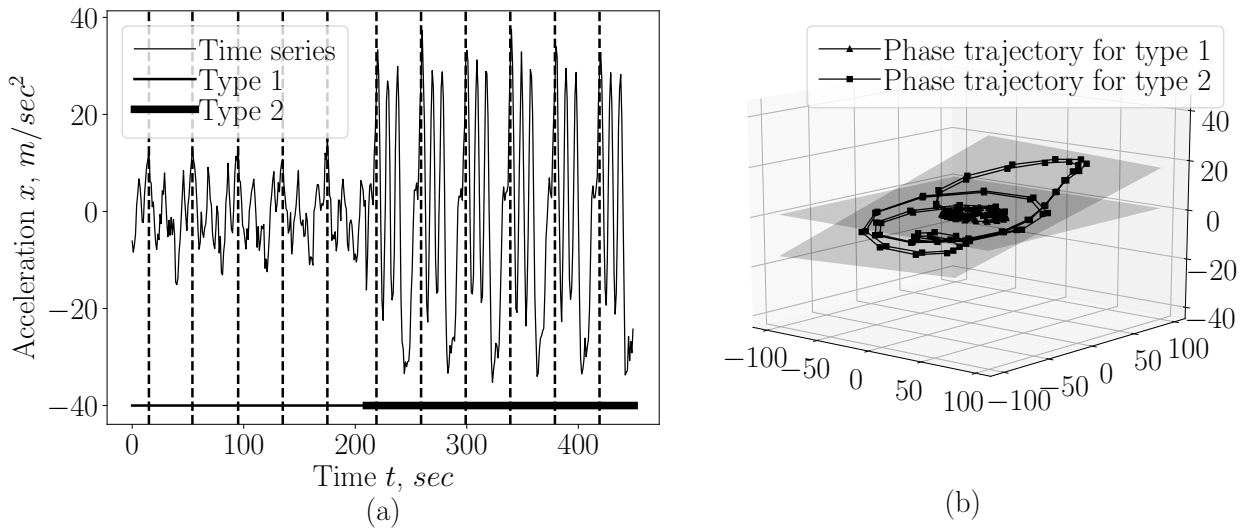


Figure 1. A time series with clustering: a) time series with assessor markings on clusters and markings at beginning of each quasi-periodic segment; b) projection of phase trajectories on the two principal components

point. Fig. 1.b shows the two first principal components of a phase trajectory, and the projection of the phase trajectory to these components. Two trajectories relate to different physical actions: Type 1 and Type 2, in the time series. The planes that are generated by these principal components are different. It separates Type 1 and Type 2 as two different actions. Second the distance function between points in the new feature space. This function is the distance between two bases of some subspaces within the phase space of a time series. It can be explained by using fig 1.b. The function is considered between two planes, which are defined by two different bases for the Type1 and Type2 segments. Points clustered by using a pairwise distance matrix. The segmentation problem is solved by using the main components [6] of the phase trajectory in each cluster separately.

The method assumes that periods of different segments vary slightly. Minimum and maximum periods of segments are given. The number of different segments in the time series is known too. It is also assumed that the type of segments in time does not change often. For example, a human runs for 10 seconds and then walks for 15 seconds and so on.

The analysis of the proposed clustering method is carried out on synthetic and real data. The experiment on the segmentation of the time series was carried out on simple sinusoidal signals with random amplitude and frequency. Synthetic dataset constructed by using the sum of the first few terms of the Fourier series with random coefficients. Real data was received by using a mobile accelerometer, which took readings during exercise of a person. All time series consist of two or three type of action such as walking, running, squats.

2. RELATED WORK

The paper [1] describes a method for constructing a feature description based on expertly defined generating functions. The paper [7] proposes a method for constructing features based on the data generation hypothesis. A combined feature description is introduced in the paper [8] based on these methods. The paper [9] consider the problem of construction a feature space. The article proposes a criterion for the redundancy of the selected features.

The paper [3] is the closest one to our research. It proposes a method for human physical activity recognition. To extract the fundamental period they construct the phase trajectory matrix, applying a technique of the principal component analysis. This method allows you to find and classify segments in time series with great accuracy. But this method can work only with time series, in which all segments belong to the same characteristic action.

The paper [6] is one of the closest to our research. It searches for the beginning of a segment within a quasiperiodic signal, which consists of only one chain of actions. Their method is based on the study of the phase space, namely, the search for a stable hyperplane that divides the phase space

into two equal parts. The points that are close to this hyperplane are selected as the beginning and ending of the segment. The article [6] proposes to project the phase space into two principal components. The beginning of each segment should be highlighted. This method finds the beginning of a segment in the case when the time series consists of one type of signal.

The paper [5] is also close to our research. The article proposes a method for searching for a periodic structure in a series by using the LSTM model with attention mechanism. This method has the same disadvantage as the previous method [6]. This method finds the beginning of a segment in the case when the time series consists of one type of signal.

3. TIME SERIES CLUSTERING PROBLEM

There given a time series

$$\mathbf{x} \in \mathbb{R}^N, \tag{2}$$

where N is a number of samples in the time series. The time series consists of a sequence of segments,

$$\mathbf{x} = [v_1, v_2, \dots, v_M], \tag{3}$$

where v_i is a segment from the set of segments \mathbf{V} . It is supposed that the time series \mathbf{x} contains some v_i from \mathbf{V} . For all i either $[v_{i-1}, v_i]$ or $[v_i, v_{i+1}]$ is a chain of actions. Fig. 2.a shows an example of segments v_i . Each segments v_i has no more than T elements, $|v_i| \leq T$. The set V has K different types of segments, $|\mathbf{V}| = K$. Fig. 2.a shows an example of segments, which are constructed the chain of action $[v_{i-1}, v_i, v_{i+1}]$.

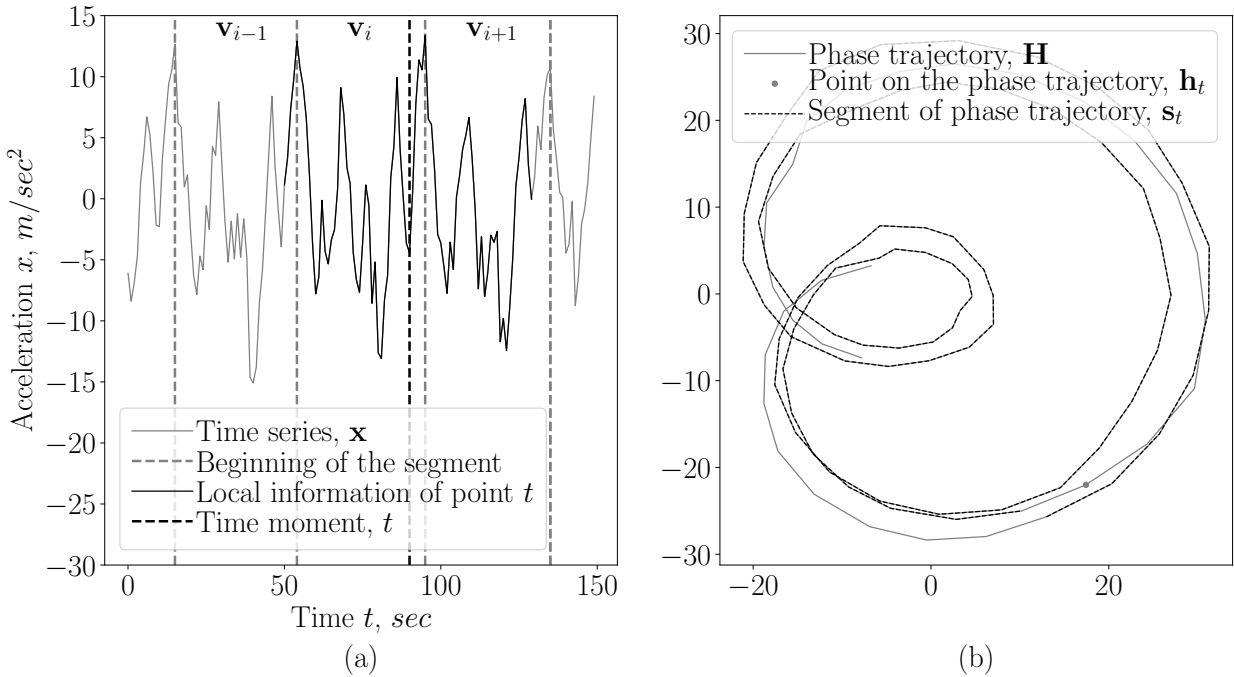


Figure 2. A time series and example of local information of points: a) time series example of local information of time point $t = 90$; b) projection of phase trajectories on the two principal components and example of segment s_t in time point $t = 90$

Construct the mapping of points into classes for their clustering

$$a : t \rightarrow \mathbb{Y} = \{1, \dots, K\}, \tag{4}$$

where $t \in \{1, \dots, N\}$ is a point in the timeline which defines the time series x . Let the map a satisfy the following properties:

$$\begin{cases} a(t_1) = a(t_2), & \text{if time moments } t_1, t_2 \text{ relate to similar type of segments,} \\ a(t_1) \neq a(t_2), & \text{if time moments } t_1, t_2 \text{ relate to different type of segments.} \end{cases} \quad (5)$$

Each point t corresponds to a label y_t from the set $\mathbb{Y} = \{1, \dots, K\}$:

$$y \in \mathbb{Y}^N. \quad (6)$$

The error for the time series clustering is

$$S(x) = \frac{1}{N} \sum_{t=1}^N [y_t = a(t)], \quad (7)$$

where t is a point on the time line, y_t is its label corresponds to the point t for time series x .

4. TIME SERIES SAMPLES CLUSTERING

To cluster the points on the timeline, we construct a matrix of pairwise distances. The matrix is constructed in a few steps.

Construct the phase trajectory \mathbf{H} with the time series x :

$$\mathbf{H} = \{h_t | h_t = [x_{t-T}, x_{t-T+1}, \dots, x_t], T \leq t \leq N\}, \quad (8)$$

where h_t is a point in the phase trajectory.

Split the phase trajectory into segments by using the maximum length of segments:

$$\mathbf{S} = \{s_t | s_t = [h_{t-T}, h_{t-T+1}, \dots, h_{t+T-1}], T \leq t \leq N - T\}, \quad (9)$$

where s_t is a segment of phase trajectory. Fig. 2.b shows an example of a segment of the phase trajectory. The segments have all local information about the time series, as it contains the information on the period up to some point t and the information about the period after the point t .

The principal components W_t for T -dimensional segments s_t is the feature description at point t . The segment s_t is projected onto subspace with few dimension by using the principal component method $z_t = \hat{W}_t s_t$:

$$\begin{aligned} \mathbf{W} &= \{W_t | W_t = [w_t^1, w_t^2, \dots]\}, \\ \hat{\mathbf{W}} &= \{\hat{W}_t | \hat{W}_t = [\lambda_t^1 w_t^1, \lambda_t^2 w_t^2, \dots]\}, \\ \mathbf{\Lambda} &= \{\lambda_t | \lambda_t = [\lambda_t^1, \lambda_t^2, \dots]\}, \end{aligned} \quad (10)$$

where $[w_t^1, w_t^2, \dots]$ and $[\lambda_t^1, \lambda_t^2, \dots]$ are the basis vectors and eigenvalues in the principal component method for the phase trajectory segment s_t .

Construct the distance function between points $\mathbf{W}_{t_1}, \mathbf{W}_{t_2}$ in the time series x for their clustering:

$$\rho(W_1, W_2) = \max \left(\max_{e_2 \in W_2} d_1(e_2), \max_{e_1 \in W_1} d_2(e_1) \right), \quad (11)$$

where e_i is the basic vector of space W_i , and $d_i(e)$ is the distance from vector e to the subspace W_i .

If all subspaces W_t have dimension two, the distance function $\rho(W_1, W_2)$ has the following interpretation:

$$\rho(W_1, W_2) = \max_{\{a,b,c\} \subset W_1 \cup W_2} V(a, b, c), \quad (12)$$

where $W_1 \cup W_2$ is a concatenation of bases, $V(a, b, c)$ is the volume of parallelepiped built on vectors a, b, c , which are columns of matrix $W_1 \cup W_2$. Construct the distance function between eigenvalues:

$$\rho(\lambda_1, \lambda_2) = \sqrt{(\lambda_1 - \lambda_2)^T (\lambda_1 - \lambda_2)}. \quad (13)$$

Construct the distance between two points in time t_1, t_2 by using equations (12-13), and consider the matrix of pairwise distances between pairs of points in the time series:

$$\rho(t_1, t_2) = \rho(W_1, W_2) + \rho(\lambda_1, \lambda_2), \quad M = \mathbb{R}^{N \times N}, \quad (14)$$

where M is a matrix of pairwise distances between all pairs of points t in the time series x . The pairwise distance matrix M is used for clustering points t of the time series (4).

5. COMPUTATION EXPERIMENT

5.1. Points clustering

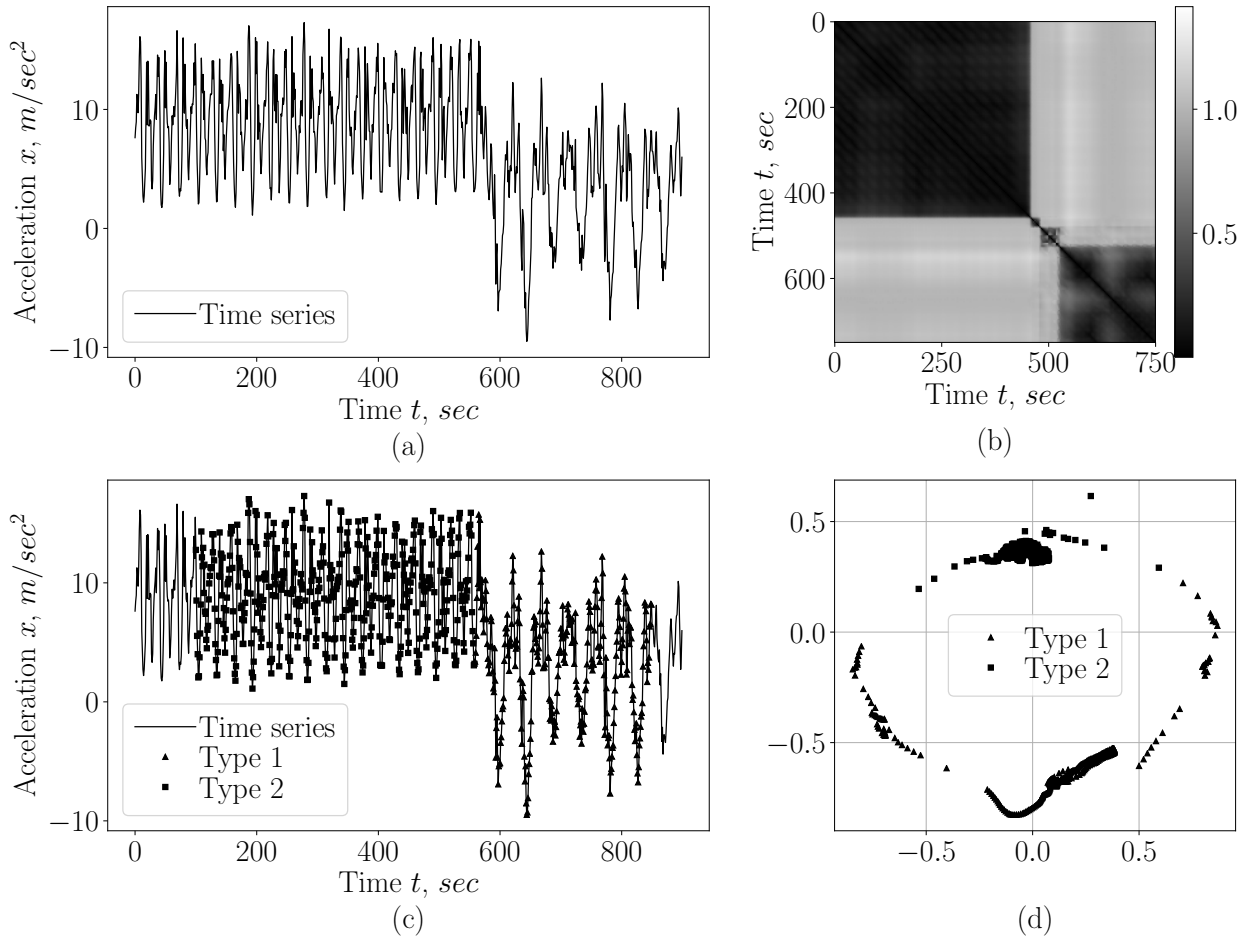


Figure 3. The result of the clustering algorithm for the time series "Physical Motion 2": a) a time series; b) the pairwise distance matrix for the time series; c) an example of clustering; d) Multidimensional Scaling of the points into two dimensional subspace by using pairwise distance matrix

In this experiment, the clustering of points in the time series was carried out using matrices of pairwise distances (14). The experiment was carried out on real and synthetic data, which are described in Table 1. "Physical Motion" is a time series obtained by using a mobile accelerometer. The synthetic time series were constructed by using the first few terms of the Fourier series with random coefficients from the standard normal distribution. The generation of synthetic time series consisted of two stages. At the first stage, the short segments v were generated to build a set of all

segments \mathbf{V} . The second stage is the following random process:

$$\mathbf{x} = [v_1, v_2, \dots, v_M] + \varepsilon, \quad \begin{cases} v_1 \sim \mathcal{U}(\mathbf{V}), \\ v_i = v_{i-1}, & \text{with probability } \frac{3}{4}, \\ v_i \sim \mathcal{U}(\mathbf{V}), & \text{with probability } \frac{1}{4} \end{cases} \quad (15)$$

where $\mathcal{U}(\mathbf{V})$ is a uniform distribution on objects from the set \mathbf{V} , and ε is gaussian noise.

5.2. Time series segmentation

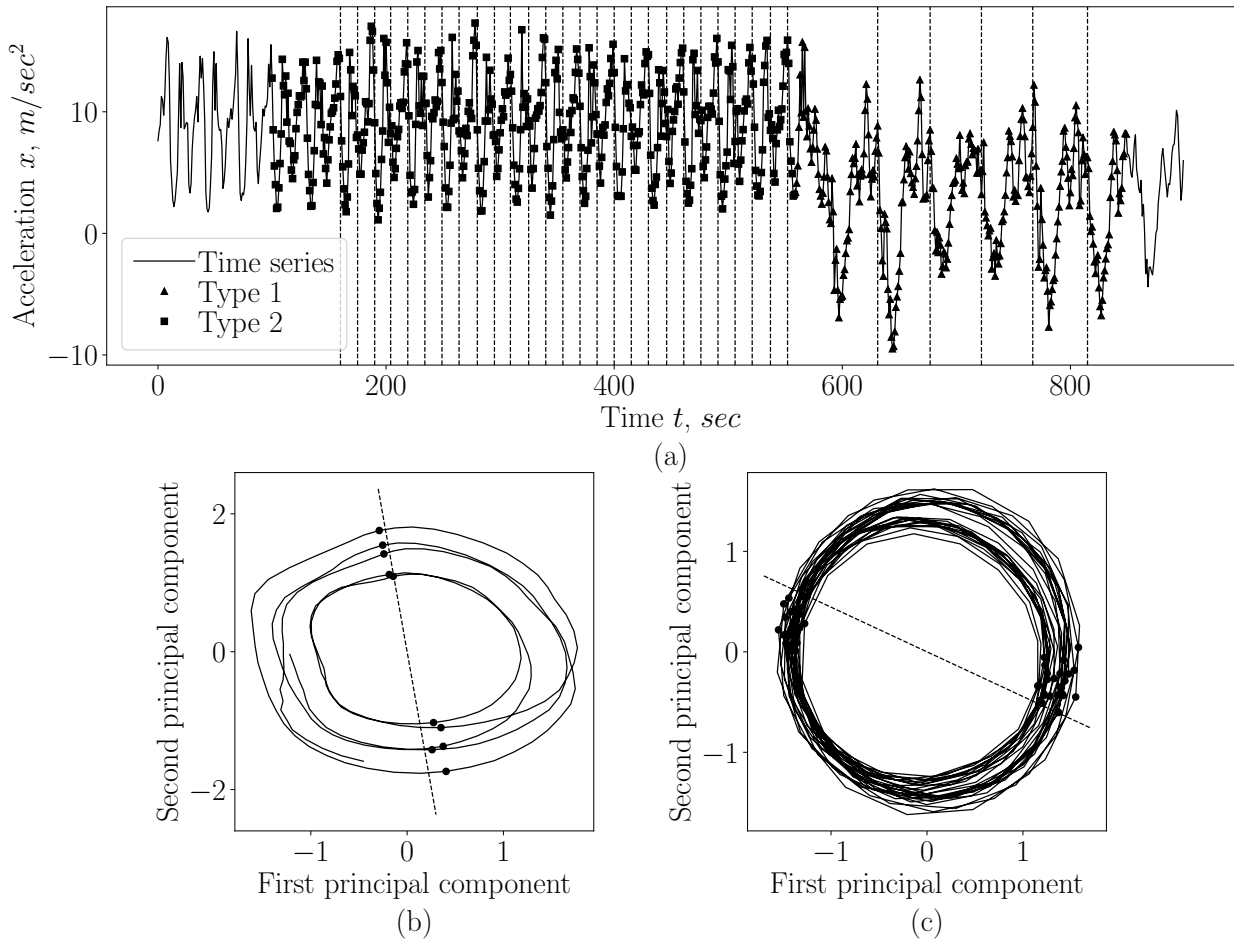


Figure 4. The result of the segmentation algorithm for the time series "Physical Motion 2": a) a time series segmentation; b) a phase trajectory for cluster "Type 1"; c) a phase trajectory for cluster "Type 2"

The result of the clustering algorithm for the real time series is shown in the fig. 3. The fig 3.b shows the matrix of pairwise distances for points in the timeline. Time series points can be easily visualised on a plane by using the pairwise distance matrix and the Multidimensional Scaling method [10]. The visualisation is shown in the fig. 3.d. Fig 3.c shows the result of clustering points. More results of clustering are presented in Table 1.

The time series segmentation is shown on the example of real data obtained by using a mobile accelerometer. Segmentation is carried out by using the method that is presented in the work [6]. The method is used for each action within the time series separately. Fig. 4 shows the result of the segmentation algorithm for the time series "Physical Motion 2". The segmentation result for a cluster "Type 1" is much better than for a cluster "Type 2". This result is easily explained by using

Table 1. Description of time series in the experiment

Series, x	Length, N	Number of segments, K	Period, T	Error, S
Physical Motion 1	900	2	50	0.03
Physical Motion 2	900	2	35	0.08
Physical Motion 3	900	2	30	0.09
Physical Motion 4	800	2	50	0.01
Synthetic 1	2000	3	40	0.008
Synthetic 2	2000	2	40	0.06
Synthetic 3	2000	2	40	0.03
Synthetic 4	2000	2	40	0.03
Synthetic 5	2000	2	40	0.04
Simple	1000	2	135	0.14

fig. 4.b and fig. 4.c. Fig. 4.b shows a phase trajectory for the cluster "Type 1". This phase trajectory has no self-intersections in contrast to the phase trajectory shown in the fig. 4.c. As a result, for time series with a simple phase trajectory, this method performs the time series segmentation well.

6. CONCLUSION

The paper investigated the problem of finding periodic structures in time series. A method based on local reduction of the phase space dimension was proposed. An algorithm for searching for segments was proposed. The algorithm is based on the principal component method for local dimension reduction. Proposed the the function of the distance between local basis. The local bases were interpreted as a features description of a point in the time series.

During the experiment on the real and synthetic data set showed that the proposed method for measuring the distance between the basis well separates points that are related to different type of action, which leads to good clustering of time series points. The results of the experiment are shown in Table 1. The experiment was carried out segmentation of time series by using the method [6] for each cluster separately.

The proposed method has few disadvantages associated with a large number of assumption on a time series. These restrictions will be relaxed in subsequent papers. It is also planned to solve the problem of finding the minimum dimension of the phase space for which the phase trajectory will not have self-intersections.

Acknowledgments. At the end of the articles are written thanks.

For example:

We thank A. A. Surname1 and B. B. Surname2 for their participation in discussions of the results. We are grateful to C. C. Surname3 and the reviewers for careful reading of the manuscript and helpful remarks.

Links to grants are also listed here.

REFERENCES

1. J. R. Kwapisz and G. M. Weiss and S. A. Moore, "Activity Recognition using Cell Phone Accelerometers," Proceedings of the Fourth International Workshop on Knowledge Discovery from Sensor Data (12), 74–82 (2010).
2. W. Wang and H. Liu and L. Yu and F. Sun, "Human activity recognition using smart phone embedded sensors: A linear dynamical systems method. In: Neural Networks (IJCNN)," International Joint Conference on Neural Networks , 1185–1190 (2014).

3. A. D. Ignatov and V. V. Strijov, "Human activity recognition using quasiperiodic time series collected from a single tri-axial accelerometer," *Multimedial Tools and Applications* 75 (12), 7257–7270 (2016).
4. A. Olivares and J. Ramirez and J. M. Gorris and G. Olivares and M. Damas, "Detection of (in)activity periods in human body motion using inertial sensors: A comparative study," *Sensors (Basel)* 12 (5), 7257–7270 (2012).
5. Y. G. Cinar and H. Mirisae, "Period-aware content attention RNNs for time series forecasting with missing values," *Neurocomputing* 312 , 177–186 (2018).
6. A. P. Motrenko and V. V. Strijov, "Extracting fundamental periods to segment biomedical signals," *Journal of Biomedical and Health Informatics* 20 (6), 1466 - 1476 (2015).
7. Y. P. Lukashin, *Adaptive methods for short-term forecasting* (Finansy and Statistik, Moscow, 2003).
8. M. P. Kuznetsov and N. P. Ivkin, "Time series classification algorithm using combined feature description," *Machine Learning and Data Analysis* 11 (1), 1471–1483 (2015).
9. A. M. Katrutsa and V. V. Strijov, "Stress test procedures for features selection algorithms," *Chemometrics and Intelligent Laboratory System* 142 , 172–183 (2015).
10. I. Borg and P. J. F. Groenen, *Modern Multidimensional Scaling* (Springer, New York, 2005).
11. D. L. Danilov and A. A. Zhiglovsky, *Main components of time series: method "Gesensitsa"* (St. Petersburg University, St. Petersburg, 1997).