UNSUPERVISED DETECTION OF PERIODIC SEGMENTS IN VIDEOS

Costas Panagiotakis[1,2], *Giorgos Karvounas*[1,3] *and Antonis Argyros*[1,3]

[1]Institute of Computer Science, FORTH, Greece
[2]Business Administration Department (Agios Nikolaos), TEI of Crete, Greece
[3]Computer Science Department, University of Crete, Greece
Email: {cpanag,gkarv,argyros}@ics.forth.gr

ABSTRACT

We present a solution to the problem of discovering all periodic segments of a video and of estimating their period in a completely unsupervised manner. These segments may be located anywhere in the video, may differ in duration, speed, period and may represent unseen motion patterns of any type of objects (e.g., humans, animals, machines, etc). The proposed method capitalizes on earlier research on the problem of detecting common actions in videos, also known as commonality detection or video co-segmentation. The proposed method has been evaluated quantitatively and in comparison to a baseline, power-spectrum-based approach, on two ground-truth-annotated datasets (**MHAD202-v**, **PERTUBE**). From those, **PERTUBE** has been compiled specifically for the purposes of this study and includes a collection of youtube videos that have been shot in the wild, with several periodic segments. The results of this evaluation demonstrate that the propose method outperforms the baseline considerably, especially in the more challenging **PERTUBE** dataset.

Index Terms— periodicity detection, commonalities discovery, video co-segmentation, temporal video segmentation.

1. INTRODUCTION

Periodic motions are very common in natural and man-made environments [1, 2]. Perhaps the most prevalent periodic motions are the ambulatory motions of humans and animals in their gaits [3]. Thus, the automatic detection of periodic motions in videos is considered as an important problem in computer vision and pattern recognition with several applications. Periodic motion can aid in low level tasks such as tracking or in higher level tasks such as recognition (i.e., recognizing individuals based on the analysis of their periodic gait [3]). Periodic motions may involve multiple interleaving periods, partial time span as well as spatiotemporal noise and outliers [1]. In this work, we are interested in the detection of video segments with periodic motions, without any prior knowledge on the number of periods, their duration, or the semantics of the observed scene. We also opt for a solution that handles motions with a period that is not constant over time.

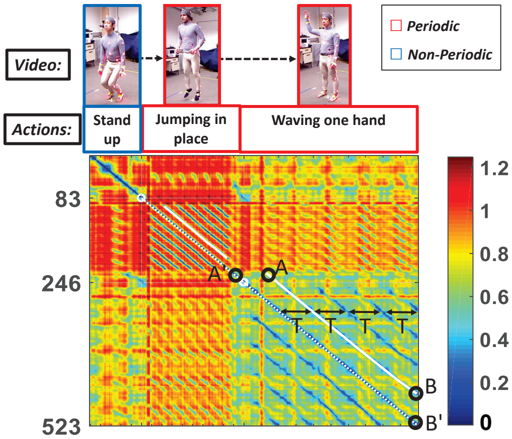


Fig. 1. Visualization of the matrix D of pairwise distances of the video frames, in which two periodic motions (jumping, waiving) appear after a non-periodic one (stand up).Periodic segments are manifested as straight line segments in D (in white) that are parallel to the main diagonal and are associated with low sum of D values. The projections (A', B') of the endpoints (A, B) of such segments on the main diagonal, identify the start and the end of these periodic segments. Our method detects them and estimates their period automatically, without any prior information on the contents of the video.

The proposed method employs the 2D, square, symmetric matrix D that contains the pairwise Euclidean distances of all frames of the input video, as computed in in [4]. Figure 1 visualizes such a matrix D in the form of a heat map. Warm (cold) colors represent large (small) pairwise distances, respectively. The main diagonal of this matrix contains zeros, because this is the distance of each frame to itself. Consider now a straight segment $\overline{AB}$ that is parallel to the main diagonal, along which matrix D has very small entries. Let also T be the horizontal displacement between the main diagonal and $\overline{AB}$. The existence of $\overline{AB}$ signifies a very strong similarity between a part of a video, and another part of the video that is temporally displaced by T. The two parts of the video can be identified by horizontally and vertically projecting the endpoints A and B of $\overline{AB}$ onto the main diagonal of D (points A' and B', respectively). The displacement T corresponds to the temporal offset between the two similar video parts. Thus, a periodic segment with n periods each of length T, will show up in D as n straight diagonal line segments, equally displaced by T. An example of such periodic segment with $n = 4$ is depicted in Fig. 1. Thus, the discovery of periodic segments in a video, amounts to finding diagonal straight line segments of minimum total cost that are located off the main diagonal.

Such a line (more generally, a path in D) and its associated cost can be estimated by employing Dynamic Time Warping (DTW) [5, 6, 7]. However, considering all possible paths, evaluating them and keeping the best one has a prohibitively high computational complexity of $O(N^6)$ [4], where N denotes the number of frames of the video. The problem can also be seen as an instance of the problem of finding commonalities/common subsequences between two different videos. In our previous work [4] we proposed *MUCOS*, an effective and efficient solution to this problem. In this work, we capitalize on *MUCOS* to solve the periodicity detection problem. Towards that direction, we show that if *MUCOS* is applied to an appropriately preprocessed version of matrix D, the detected commonalities correspond to the periodic segments of the input video. Moreover, the period of each periodic segment is efficiently computed by tracking the paths of the detected commonalities.

To evaluate the proposed method we employed two datasets, one with synthetically generated periodic segments and one containing 50 videos with periodic segments acquired in the wild and downloaded from *youtube*. The quantitative analysis of the obtained results shows the effectiveness of the proposed approach, also in comparison to a baseline, power-spectrum-based method.

In summary, the main contributions of this paper are:

- An effective and efficient solution for detecting the periodic segments of a video, which is formulated as an instance of the video co-segmentation problem[1].

- The introduction of **PERTUBE**[2], the first, ground-truth annotated dataset containing 50 *youtube* videos with periodic segments acquired in the wild.

2. RELATED WORK

Discovering periodic segments in time-series: In [6], the problem of periodicity detection in time series is addressed using a time warping algorithm named WARP. The main idea of WARP is that if the time series is shifted by a number of elements equal to the period of the time series, then the original time series and the shifted one will be very similar. Karvounas et al. [8] formulate the detection of a periodic segment as an optimization problem that is solved based on an evolutionary optimization technique. Given a time series representing a periodic signal with a non-periodic prefix and tail, the method estimates the start, the end and the period of the periodic part. The most important limitation of that method is that it assumes a video containing a single periodic segment.
Periodicity in videos: Polana and Nelson [9] devise an extension of the Fourier formula to detect periodicity. Cutler and Davis [3] address the problem of periodicity detection for both the case of stationary and non-stationary periodic signals. For the case of stationary signals, this can be achieved by a Fourier Transform followed by a Hanning filter. For the non-stationary case, Short-Time Fourier Transform is employed to better handle the shifting spectrum. As in [9], the objects are tracked and aligned before the periodicity analysis. Such spectral domain methods have the limitation that the action frequency should be almost constant and it would emerge as a discernible peak at a time frequency graph. However, the amount of variation in appearance between repetitions and the variation in action length means that in certain cases, no such clear peak may be identifiable [10].

In [11], the analysis of multiple periodic motions over a static background is motivated by the observation that repetitive patterns have distinct signatures in frequency space. Wang et al. [12] proposed a method for retrieval of social games that are characterized by repetitions, from unstructured videos. Each frame is mapped to the nearest keyframe, yielding a sequence of keyframe indices that are used to mine recurring patterns. In a more recent work, Levy and Wolf [10] use a deep learning approach to count the number of repetitions of approximately the same action in an input video sequence. The approach proposed in [13] combines ideas from nonlinear time series analysis and computational topology, by translating the problem of finding recurrent dynamics in video data, into the problem of determining the circularity of an associated geometric space.

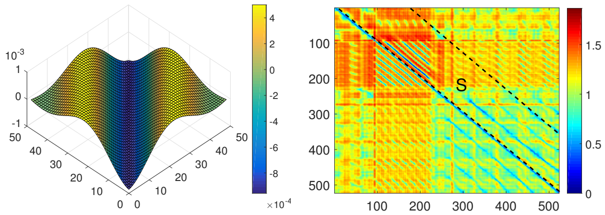
All aforementioned approaches cannot handle the problem of periodicity detection under any video content (e.g., regardless of the type of moving object and pattern) and without some type of supervision (e.g., require some knowledge regarding the number of periods, their duration or the location of periodic segments in a given video). To the best of our knowledge, the method presented in this paper is the first that makes no such assumption and is fully unsupervised.

3. *P-MUCOS*: DETECTING PERIODICITY BY DETECTING COMMONALITIES

We assume a video sequence of N frames, and the $N \times N$ symmetric matrix D of the pair-wise distances of these frames (see Fig. 1). Depending on the nature of the sequence, differ-

[1] Implementation available online at: https://sites.google.com/site/costaspanagiotakis/research/pd

[2] Available onine at: http://www.ics.forth.gr/cvrl/pd



(a) F (b) $D - D'_f$

Fig. 2. (a) The filter F used to enhance D of Fig. 1, and (b) the resulting the $D - D'_f$ which is fed into *P-MUCOS*. The set S of candidate commonality points is defined between the two dashed lines.

ent frame representations and distance functions can be employed [4, 7]. A candidate periodic segment can be represented as a connected path of points (x_i, y_i) on D for which it holds that (a) $\forall x_i, y_i,\ x_i \leq x_{i+1}$ and $y_i \leq y_{i+1}$ and (b) the sum of the values at the points of the path is very small. The 1^{st} condition guarantees that matched frames are ordered in time, and the 2^{nd} that the similarity between corresponding frames is high. Besides these constraints, paths can start and end anywhere above[3] the diagonal of D meaning that $y_i > x_i$. Then, the period T_i of frame x_i is given by:

$$T_i = y_i - x_i. \tag{1}$$

Thus, the start and the end times of the periodic video segment is given by the minimum x-coordinate and the maximum y-coordinate of the path, respectively. Figure 1 gives an example of a distance matrix D for a video showing one non-periodic action (stand up) and two periodic ones (jumping in place, waving one hand). The two identified paths shown in white correspond to the two periodic segments. By detecting such segments, we can segment the periodic parts of a video and estimate the period of each of them. In order to do this, we employ *MUCOS* [4], a method that has been designed to co-segment commonalities in two videos.

The *MUCOS* algorithm: To discover all commonalities of two videos V_1, V_2 *MUCOS* operates as follows: (1) Compare pairwise all frames of V_1, V_2 and estimate their distance matrix D, (2) compute the sets of potential commonality end points and midpoints, (3) define a graph G whose nodes are the end points and midpoints, (4) compute all shortest paths in G, (5) associate shortest paths with commonalities and discard those that don't meet certain criteria, and (6) employ an objective function to evaluate and accept/reject the remaining commonalities. More details on *MUCOS* can be found in [4].

From *MUCOS* to *P-MUCOS*: The application of *MUCOS* on the matrix D of pairwise distances of all frames of a certain video will result in the detection of the diagonal as the major commonality. This is because the main diagonal represents the longest possible path with the minimum cost. However, in the context of our problem, we should exclude this trivial solution from consideration and only identify commonalities that are different than the main diagonal.

The set S of candidate commonality points are the ones in the upper right triangle of D above its main diagonal. S is also restricted by the minimum (T_m) and the maximum (T_M) duration of a period. In our implementation, we set $T_m = 6$ and $T_M = \lfloor N/3 \rfloor$ frames, so as to ensure that there exist at least three periods of the periodic part of the video. Thus,

$$S \subseteq \{(x, y) : 1 \leq x \leq N - T_m \wedge x + T_m \leq y \leq x + T_M\}. \tag{2}$$

Next, we design a filter H, applied to S, to emphasize the commonalities that are close to the diagonal of the distance matrix D. The symmetric filter H_{ij} is defined as:

$$H_{ij}(u, v) = -a \cdot cos\left(\frac{2\pi d}{\tau}\right) \cdot (\tau - d), \tag{3}$$

where $d = |v - u|$, $\tau = j - i$, and a is given by the constraint:

$$\sum_{u,v} |H_{ij}(u, v)| = 1.$$

The filter size ($\tau \times \tau$) is dynamically adapted based on the coordinates of the point it is applied to. Additionally, the filter's response is maximized for commonalities passing through the point (i, j) the filter is applied to. In [14], a similar filtering technique is used for partial curvilinear structure enhancement, based on a family of Gabor-like, location invariant filters. In this work, the defined filter is location variant. In order to reduce the computational cost, the response of this filter to point (i, j) can be estimated recursively based on its response at point $(i - 1, j - 1)$, since both filters are equally sized and are applied with the same coefficients to the same matrix subregion of size $(\tau - 2) \times (\tau - 2)$. Thus recursive computation requires 4τ operations which is much less than the τ^2 operations that are required without recursion.

Let D_f be D after being filtered by H. Another enhancement of paths that are close to the diagonal of D can be achieved by using the expected low values (probably local minima) of D_f that appear in the middle between two (probably) local maxima, the point (i, j) and its projection to main diagonal $(\frac{i+j}{2}, \frac{i+j}{2})$, according to the following equation:

$$D'_f(i, j) = D_f(i, j) - D_f\left(m - \frac{T}{4}, m + \frac{T}{4}\right),\ m = \frac{i + j}{2}. \tag{4}$$

Points (i, j) that belong to paths close to the diagonal are enhanced because $D_f(i, j) \gg D_f(m - \frac{T}{4}, m + \frac{T}{4})$. Finally, we subtract from D'_f its maximum value, so that $-D'_f$ becomes a non-negative matrix. The application of *MUCOS* to $D - D'_f$ (see Fig. 2) yields commonality paths that correspond to periodic actions (white curves in Fig. 1). Additionally, the

[3] The matrix D is symmetric, therefore all computations can be focused on the upper right triangle of D.

Fig. 3. Indicative snapshots from videos of **PERTUBE**.

Table 1. Evaluation results on the **MHAD202-v** dataset.

Methods	$\mathcal{R}(\%)$	$\mathcal{P}(\%)$	$F_1(\%)$	$\mathcal{O}(\%)$
P-MUCOS	88.2 (82.2)	95.2 (**98.5**)	**90.9** (89.2)	**84.2** (81.3)
BASELINE	**93.2**	86.2	88.9	81.6

Table 2. Evaluation results on the **PERTUBE** dataset.

Methods	$\mathcal{R}(\%)$	$\mathcal{P}(\%)$	$F_1(\%)$	$\mathcal{O}(\%)$
P-MUCOS	**84.1** (80.4)	**75.7** (75.4)	**77.0** (76.0)	**67.7** (67.2)
BASELINE	79.3	61.1	66.8	57.3

period of each frame of a periodic segment can be estimated by Eq.(1), supporting variable periodicity over the time.

Baseline method: We implemented a baseline method that is used for comparisons with *P-MUCOS*. This can be seen as a natural extension of the power spectrum method [3], according to which, a given signal is periodic if the peak of its spectrum is greater than $\mu + 3\sigma$, where μ and σ denote the mean and the standard deviation of the signal spectral power. In this work, three sizes of time windows are used, since the period value as well as the start and end frame of the periodic segments are not known. So, for each frame i of the video, we consider three signals $d_j(i), j \in \{1, 2, 3\}$, each centered at i and having 51, 101 and 151 frames, respectively. If at least one of these signals successfully passes the spectrum test, then frame i is added to a set of frames that can be clustered to sets of candidate periodic segments. We ignore segments with insufficient/marginal number of supporting frames, i.e., segments containing less than two periods. The period of each detected periodic segment is given by the period corresponding to the peak of its spectrum.

4. EXPERIMENTAL EVALUATION

The experimental evaluation of the proposed method was conducted using two datasets:

- **MHAD202-v dataset**: Contains the 202 videos of the 101 video pairs of MHAD101-v dataset presented in [7]. Each video consists of 3-7 periodic actions (e.g., jumping in place, jumping jacks, bending hands, waving one hand, waving two hands) or non periodic actions (throwing a ball, sit down, stand up).

- **PERTUBE dataset**: This is a new dataset that we compiled for assessing solutions to the periodicity detection problem. **PERTUBE** contains 50 videos (a total of 40307 frames) showing human, animal and machine motions in lab settings or in the wild. Each video consists of 143 to 2307 frames. Videos form 200 segments, 75 of which are periodic. Each video has from 1 to 4 periodic segments and each segment consists of 30 to 1037 frames. All sequences were collected from *youtube* and were annotated with ground truth, manually. Indicative snapshots can be seen in Fig. 3.

In both datasets, we represented frames based on Improved Dense Trajectories features, as in [7]. To assess the performance of the evaluated methods, we employed the metrics of precision, recall, F_1 score and overlap as in [7, 4].

Tables 1 and 2 summarize the results obtained on the **MHAD202-v** and **PERTUBE** datasets, respectively. The scores are presented as average percentage scores computed over all individual scores per video of a dataset. We report the performance of *P-MUCOS* and *BASELINE* on all aforementioned metrics. In parenthesis, we provide the scores of *P-MUCOS* without the filtering and enhancement steps, to show that these steps increase the *P-MUCOS* performance. Missdetections are mainly due to the failure of the used descriptors (e.g. on fast camera motions) to yield distance matrices with repetitive patterns. *P-MUCOS* clearly outperforms the baseline method in both datasets. The difference in performance is more striking in the more challenging, real-world **PERTUBE** dataset (10% improvement in overlap and F_1 scores compared to the baseline method). Additionally, *P-MUCOS* gives significantly better estimates of the periods.

5. CONCLUSIONS

We proposed a method for discovering periodic segments in videos. To achieve this, we extended an existing solution to the problem of discovering commonalities in two videos [4]. To the best of our knowledge, this is the first method for detecting periodic segments in a video that is totally unsupervised, i.e., it is completely agnostic to the semantic content of the videos, the number of periodic segments, their start and end in the video, as well as the duration and the number of periods. The experimental results on challenging video datasets showed the effectiveness of the proposed method.

Acknowledgments

This work was partially supported by the EU projects ACANTO (H2020-643644) and Co4Robots (H2020-731869).

6. REFERENCES

[1] Zhenhui Li, Bolin Ding, Jiawei Han, Roland Kays, and Peter Nye, "Mining periodic behaviors for moving objects," in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2010, pp. 1099–1108. 1

[2] Ashis Kumar Chanda, Chowdhury Farhan Ahmed, Md Samiullah, and Carson K Leung, "A new framework for mining weighted periodic patterns in time series databases," *Expert Systems with Applications*, vol. 79, pp. 207–224, 2017. 1

[3] Ross Cutler and Larry S. Davis, "Robust real-time periodic motion detection, analysis, and applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 781–796, 2000. 1, 2, 4

[4] Costas Panagiotakis, Konstantinos Papoutsakis, and Antonis Argyros, "A graph-based approach for detecting common actions in motion capture data and videos," *Pattern Recognition*, vol. 79, pp. 1–11, 2018. 1, 2, 3, 4

[5] Lawrence Rabiner and Biing-Hwang Juang, *Fundamentals of Speech Recognition*, vol. Chapter 4, Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993. 2

[6] Mohamed G Elfeky, Walid G Aref, and Ahmed K Elmagarmid, "Warp: time warping for periodicity detection," in *Data Mining, Fifth IEEE International Conference on*. IEEE, 2005, pp. 8–pp. 2

[7] Konstantinos Papoutsakis, Costas Panagiotakis, and Antonis Argyros, "Temporal action co-segmentation in 3d motion capture data and videos," in *IEEE Conference on COmputer Vision and Pattern Recognition (CVPR)*, 2017. 2, 3, 4

[8] Giorgos Karvounas, Iason Oikonomidis, and Antonis A Argyros, "Localizing periodicity in time series and videos.," in *BMVC*, 2016. 2

[9] Ramprasad Polana and Randal C Nelson, "Detection and recognition of periodic, nonrigid motion," *International Journal of Computer Vision*, vol. 23, no. 3, pp. 261–282, 1997. 2

[10] Ofir Levy and Lior Wolf, "Live repetition counting," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 3020–3028. 2

[11] Alexia Briassouli and Narendra Ahuja, "Extraction and analysis of multiple periodic motions in video sequences," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 7, pp. 1244–1261, 2007. 2

[12] Ping Wang, Gregory D Abowd, and James M Rehg, "Quasi-periodic event analysis for social game retrieval," in *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 112–119. 2

[13] Christopher J Tralie and Jose A Perea, "(quasi) periodicity quantification in video data, using topology," *arXiv preprint arXiv:1704.08382*, 2017. 2

[14] Costas Panagiotakis, Eleni Kokinou, and Apostolos Sarris, "Curvilinear structure enhancement and detection in geophysical images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 6, pp. 2040–2048, 2011. 3