A two-stage approach for commonality-based temporal localization of periodic motions

Costas Panagiotakis[1,2] and Antonis Argyros[1,3]

[1] Institute of Computer Science, FORTH, Greece
[2] Department of Management Science and Technology, Hellenic Mediterranean University, Greece
[3] Computer Science Department, University of Crete, Greece
{cpanag,argyros}@ics.forth.gr

Abstract. We present an unsupervised method for the detection of all temporal segments of videos or motion capture data, that correspond to periodic motions. The proposed method is based on the detection of similar segments (commonalities) in different parts of the input sequence and employs a two-stage approach that operates on the matrix of pairwise distances of all input frames. The quantitative evaluation of the proposed method on three standard ground-truth-annotated datasets (two video datasets, one 3D human motion capture dataset) demonstrate its improved performance in comparison to existing approaches.

Keywords: Periodicity Detection · Repetitive Motions Detection · Commonalities Discovery · Video Segmentation.

1 Introduction

Periodic or repetitive motions are very common in natural and man-made environments [3]. Therefore, their detection constitutes an important step towards the segmentation and the high level interpretation of video and motion capture (mocap) data. This is an interesting problem in computer vision and pattern recognition whose solution has several applications in action and activity recognition, medical diagnosis [1], detection of machine failures [13], repetition counting [14], etc.

In this work, we address the problem of temporal segmentation of periodic segments in videos and mocap data and we propose a solution which neither requires prior knowledge nor imposes constraints on the speed, the number or the length of the periods of the periodic segments. Moreover, the method tolerates variations of the period of the periodic motions. The input to the proposed method is the $N \times N$ symmetric matrix D that contains the pairwise distances of all N frames of the input sequence. Figure 1 provides an example of such a matrix D, where warm (red) and cold (blue) colors correspond to high and low values in D, respectively. The main diagonal of this matrix contains zeros, since these points hold the distance of each frame of the sequence to itself.

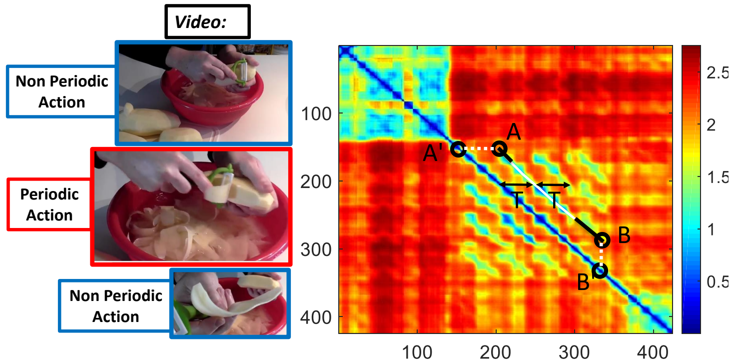


Fig. 1. An example of the distance matrix D of the frames of a video, in which a periodic action appears between two non-periodic actions. The periodic nature of the middle action gives rise to low values in D (i.e., similar frames) along straight line segments that are parallel to the diagonal of D. The goal of the proposed method is to detect such lines and to estimate the period of the corresponding periodic action (i.e., the offset from the diagonal). The periodic segment detected by the first and second stage of the proposed method are plotted with a white and a black straight line, respectively. The projection of the endpoints (A, B) of the black line on the main diagonal, give the start (A') and the end (B') of this periodic segment.

Let us consider that a periodic segment $[A', B']$ with period T starts at frame A' and ends at frame B' of the video. Due to the definition of periodicity, we know that there is a strong similarity between the segment $[A', B']$ and another part of the sequence that is temporally displaced by T (see Figure 1). This means that D contains a straight line segment $\overline{AB}$ where $A = (A', A' + T)$ and $B = (B' - T, B')$ that is parallel to the main diagonal, along which D has very low values. In practice, due to deviations from perfect periodicity, the path connecting A and B might deviate from straightness and might not be perfectly parallel to the diagonal. Detecting periodic segments and estimating their period, amounts to detecting and localizing straight lines of minimum cost that are located off the main diagonal, as for example the line $\overline{AB}$ in Figure 1. The sum of the entries of D under such a line/path is inversely proportional to the similarity of the two parts of the sequence. A low (high) cost path corresponds to a high (low) confidence on the existence of a periodic segment. Short paths correspond to periodic segments of a few frames, which are not that interesting. As paths increase in length, their cost also increases. Thus, the trade-off between the length of the path (the duration of the segments) and its cost should be balanced.

A related work [8], developed *P-MUCOS*, a method that reduces the problem of periodicity detection to the problem of finding common sub-sequences in a sequence. This is achieved by employing *MUCOS* [9] a graph-based search algorithm for finding common sub-sequences (commonalities) between two different sequences. *MUCOS* has a complexity of $O(N^2)$ and it is an efficient alternative to employing Dynamic Time Warping (DTW) [10]. As shown in [9,8], the computational complexity of a DTW-based exhaustive algorithm that evaluates all paths and keeps the best one is $O(N^6)$.

P-MUCOS is an one-stage approach that applies *MUCOS* to the distance matrix D of the input sequence. This results in the detection of the main diagonal of D as the major commonality. To avoid this trivial solution, *P-MUCOS* employs a filtering technique that enhances the off-diagonal commonalities that correspond to periodic segments. However, this enhancement is not strong close to the commonality endpoints, a fact that influences negatively the accuracy of periodicity detection.

Thus, in this work, we present *P-MUCOS-S2*, an improvement of *P-MUCOS*, which addresses the drawback of *P-MUCOS* by introducing a two-stage approach for commonality detection. In the first stage, the strongest part of the off-diagonal commonality is detected in an improved version of the enhancement performed by *P-MUCOS*. In the second stage, a hysteresis-thresholding-like operation extends the initial detections by optimizing an appropriately defined objective function. *P-MUCOS-S2* is evaluated quantitatively in comparison to the *P-MUCOS* and another, Fourier-based baseline approach [8] and is shown to improve substantially the accuracy of periodicity detection. The experimental evaluation is performed on the publicly available video datasets employed in [8], but also on a relevant mocap dataset.

In summary, the contributions of this work are: the improvement of the *P-MUCOS* algorithm for temporal localization of periodic segments by (a) improving the commonalities enhancement filtering approach of the first state and (b) by introducing a hysteresis-thresholding-based second stage, as well as the quantitative evaluation of the new algorithm *P-MUCOS-S2* on standard datasets on motion captured and video datasets in comparison to existing approaches.

2 Related work

The problem of detecting periodic segments in time-series has been well studied. In [4], the problem of periodicity detection in time series is addressed using the time warping algorithm WARP. A given time series is transformed to time-stamped events drawn from a finite set of nominal event types. 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. More recently, Karvounas et al. [5] 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. Another related challenging problem concerns the problem of periodicity detection from incomplete observations [7]. In [7], the authors propose a probabilistic model for periodic behaviors that was successfully applied on real human movement data.

The periodicity detection in videos is a more challenging problem, due to the high variability of the video content. In [11], Polana and Nelson devise an extension of the Fourier formula to detect periodicity in videos based on normal flow variation between successive image frames. The authors show that periodicity is an inherent low-level motion cue that can be exploited for robust detection of periodic phenomena without prior structural knowledge. Cutler and Davis [3] address the problem of periodicity detection for both 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 [11], the objects are tracked and aligned before the periodicity analysis. The baseline method (see Section 4), that is presented in [8], is a natural extension of the power spectrum method [3]. According to this method, 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. 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 [6].

Wang et al. [17] 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. The approach proposed in [15] 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. There exist several supervised techniques that attempt to identify sequences in similarity/distance matrices [2]. In [2], for loop-closure were detected based on a classifier trained on similarity matrices. The proposed methodology can be also applied to such problems.

Levy and Wolf [6] use a deep learning approach to count the number of repetitions of approximately the same action in an input video sequence. In [14], the problem of visual repetition from realistic video has been formulated and solved, improving the results derived by Levy and Wolf [6]. The authors derive three periodic motion types by decomposition of the 3D motion field into its fundamental components and employ the continuous wavelet transform and combine the power spectra of all representations to support viewpoint invariance.

Most of the aforementioned approaches cannot handle the problem of periodicity detection under any video content and without some type of supervision. The method presented in this paper improves the results of [8] that was the first that makes no such assumption and is fully unsupervised. Additionally, the proposed method has been also applied to motion captured data and it can be also used for period tracking and in repetition estimation [14].

3 *P-MUCOS-S2*: Commonality-based periodicity detection

Let A be a sequence of N frames and D the $N \times N$ symmetric matrix of the pair-wise distances of these frames. The proposed method can assume a variety of frame representations and corresponding frame distance metrics.

An example of such a distance matrix D of the frames of a sequence, in which a periodic action appears between two non-periodic actions, is shown in Figure 1. The periodic nature of the middle action gives rise to low values in D (i.e., similar frames) along straight line segments that are parallel to the diagonal of D and at horizontal offset T from the diagonal, where T is the period of the periodic action. This is because the distance between frames f_i and f_{i+T} of a periodic action with period T is expected to be very low. The goal of the proposed method is to detect such lines and to estimate the period T of the corresponding periodic action. Such a straight line is shown in Figure 1 in black color. The projection of the endpoints (A, B) of the black line on the main diagonal, give the start (A') and the end (B') of the corresponding periodic segment. By detecting such segments, we can segment the periodic parts of a sequence and estimate the period of each of them. This can be achieved by employing a method that detects commonalities between two sequences. The *MUCOS* [9] method suits this purpose, as it discovers all commonalities (similar segments) of two sequences v_1 and v_2, given their distance matrix D. The application of *MUCOS* on the square matrix D of distances of a single sequence will result in the detection of the diagonal as the major commonality. This trivial solution can be excluded from consideration, as performed in [8] where the *P-MUCOS* algorithm was proposed. *P-MUCOS* is improved by introducing, *P-MUCOS-S2*, an approach that operates in two stages.

P-MUCOS-S2, stage 1: *P-MUCOS* applies the following symmetric filter H_p at point $p = (i, j)$ of the distance matrix D to emphasize the commonalities to be detected:

$$H_p(q) = -a \cdot cos\left(\frac{2\pi d}{\tau}\right) \cdot (\tau - d), \tag{1}$$

where $q = (u, v)$, $d = |v - u|$, $\tau = j - i$, and a is determined by the constraint $\sum_q |H_p(q)| = 1$. The response of filter H_p on point p is given by $D_H(p) = \sum_{q \in R(\tau)} H_p(q) \cdot D(p - q)$ where the square region R is defined as $R(\tau) = \{(u, v)| - \tau/2 \leq u \leq \tau/2, -\tau/2 \leq v \leq \tau/2\}$. Intuitively, H operates as follows. At a point $p = (i, j)$ the distance matrix is convolved with a filter whose width is τ and whose coefficients are a sinusoidal pattern, evolving perpendicularly to the main diagonal of D, with a minimum at p. Thus, in case that a commonality path c passes through p, the locally minimum value at p will be further pronounced. Figure 2 illustrates the positioning of filter H at two different points p and p' of a distance matrix. The dotted rectangle denotes the

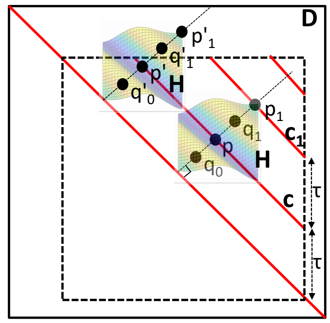


Fig. 2. A schematic illustration of the filtering operations in [8] and in the current work (see text for details).

periodic part of the video corresponding to this distance matrix. The rectangles located at p and p' denote the color-coded coefficients of filter H. Finally, the red diagonal lines besides the main diagonal denote commonality paths.

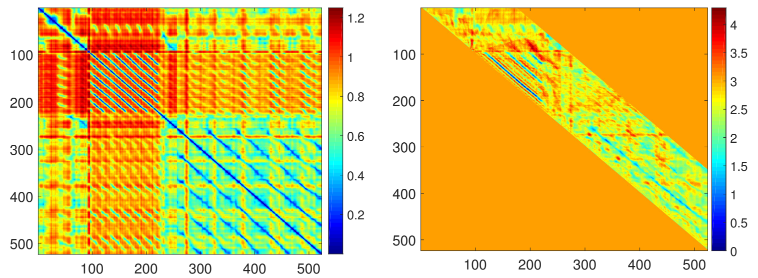
In this work, we simplify the filter H significantly. Essentially, the realized improvement is that at a certain point p, the response is obtained by combining collinear distance matrix values in a direction perpendicular to the main diagonal. More specifically, for the point p in Figure 2, the new filter response is given by

$$D_f(p) = 2 \cdot D(p) - D(q_0) - D(q_1) + D(p_1), \tag{2}$$

where $v = [\frac{\tau}{4}, -\frac{\tau}{4}]^T$, $q_0 = p + v$, $q_1 = p - v$, $p_1 = p - 2 \cdot v$. If p belongs to a commonality path c, then the point p_1 will belong to the next commonality path c_1 that is parallel to c. The point q_0 is located halfway between the main diagonal and commonality c, and q_1 is located halfway between commonalities c and c_1. So, the new filter response can be explained by the fact that D should get low values on p and p_1 and high values on q_0 and q_1. Finally, we subtract from D_f its minimum value, so that D_f becomes a non-negative matrix.

Let S be the part of the upper right triangle of D that is restricted by the minimum (T_m) and the maximum (T_M) duration of a period. In our implementation, we set $T_m = 3$ 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. The new filter H is applied to all points in S, to emphasize the commonalities that are close to the diagonal of the distance matrix D (see Figure 3(b)).

The computational cost of the application of the new filter is constant and equal to 5 operations per point, while the recursive computation of the filter H proposed in [8] has computational cost $O(\tau \cdot N^2) = O(N^3)$ ($4 \cdot \tau$ operations per point). Thus, the new filter results in a significant reduction of the computational cost. Moreover, as demonstrated by experimental results, the application of this filter improves the quantitative metrics of periodicity detection compared to the filter employed in [8].



(a) (b)

Fig. 3. (a) An example of a distance matrix D in which two periodic motions (jumping, waiving) appear after a non-periodic one (stand up). (b) The enhanced distance matrix D_f which is fed into *P-MUCOS-S2*. The two orange triangular parts (top-right, bottom left) are excluded from the commonality search space.

P-MUCOS-S2, stage 2: The filtering operation of stage 1 enhances the distance matrix so that a commonality can be detected more easily by *MUCOS*. However, this enhancement fails close to the endpoints of a commonality path. This is because the filtering operation close to the end-points of a commonality path involves, inevitably, values that are outside the commonality rectangle (see for example points q_1' and p_1' when the filter is applied to point p in Figure 2). To deal with this issue, the second stage of *P-MUCOS-S2* operates as follows. For each detected commonality c of the first stage, its endpoints are extended in the direction that is locally parallel to c, by measuring the following objective function $\omega(c)$

$$\omega(c) = \frac{A(c)}{\sum_{p \in c} D(p) + \epsilon}, \tag{3}$$

where $A(c)$ is equal to the area of the bounding rectangle of the commonality path c, $P(c) = \sum_{p \in c} D(p)$, and ϵ is a small constant preventing division by zero. This objective function captures the trade-off between the terms $A(c)$ and $P(c)$. Specifically, commonalities c with large $A(c)$ are preferable. At the same time, as $A(c)$ increases, $P(c)$ also increases. The selected commonality endpoints are the ones that optimize the objective function $\omega(c)$. Example detections of the second stage are plotted as black lines in Figure 1.

Finally, the period T of a periodic segment that corresponds to a detected commonality c, can be estimated as the average of the quantities $j - i$, for all points $(i, j) \in c$.

Fig. 4. Snapshots from videos of the PERTUBE dataset.

4 Experimental Results

Datasets: The proposed method was evaluated on one dataset that contains motion capture (mocap) data and on two datasets that contain conventional RGB videos.

- **MHAD202-s dataset**: Contains 202 motion data sequences of the 101 sequence pairs of the MHAD101-s dataset [10]. Each video consists of 3-7 periodic actions (e.g. jumping in place, jumping jacks, bending hands) and non periodic actions (e.g. throwing a ball, sit down). Each frame of a sequence represents the 3D human pose as a 64D vector whose dimensions encode angles of selected body parts. Distance matrices are obtained by estimating the Euclidean distances in pairs of such vectors.
- **MHAD202-v dataset**: Contains the 202 videos that correspond to the 202 motion capture sequences of the MHAD202-s dataset. As suggested in [10] and employed in [8], each frame is represented as a 100D vector that concatenates trajectory shape, HOG, HOF, and MBH descriptors computed on top of Improved Dense Trajectories (IDT) features [16].
- **PERTUBE dataset**: This dataset was introduced in [8] for assessing solutions to the periodicity detection problem. PERTUBE contains 50 videos showing human, animal and machine motions in lab settings or in the wild (see Fig. 4). The representation of video frames is as in MHAD202-v.

Performance metrics: In order to assess the performance of the proposed methods we employed the standard metrics of precision $\mathcal{P}$, recall $\mathcal{R}$, F-measure score F_1 score and overlap $\mathcal{O}$ (intersection-over-union). The reported metrics were computed in each individual sequence and then averaged across all sequences of a dataset.

Obtained results: Table 1 summarizes the results of *P-MUCOS-S2*, *P-MUCOS* and *BASELINE* methods [8] on the MHAD202-s, MHAD202-v and PERTUBE datasets. All algorithms run with the same parameters in different datasets. It can be verified that in all datasets, *P-MUCOS-S2* outperforms *P-MUCOS* and the baseline method. The difference in performance is more striking in the more challenging, real-world PERTUBE dataset where *P-MUCOS-S2* achieves a 10% and 7% improvement in overlap and F_1 score, respectively, compared to *P-MUCOS*. Our interpretation is that the MHAD202-s and MHAD202-v datasets contain simpler data, i.e., each sequence contains clearly periodic and non-periodic parts of the motion of a human recorded in laboratory conditions.

Table 1. Evaluation results on the MHAD202-s, MHAD202-v and PERTUBE datasets.

	MOCAP				VIDEO							
Dataset	MHAD202-s				MHAD202-v				PERTUBE			
Metric	$\mathcal{R}(\%)$	$\mathcal{P}(\%)$	$F_1(\%)$	$\mathcal{O}(\%)$	$\mathcal{R}(\%)$	$\mathcal{P}(\%)$	$F_1(\%)$	$\mathcal{O}(\%)$	$\mathcal{R}(\%)$	$\mathcal{P}(\%)$	$F_1(\%)$	$\mathcal{O}(\%)$
P-MUCOS-S2	**90.4**	93.2	**91.2**	**85.4**	87.8	**98.6**	**92.7**	**87.1**	92.5	**80.2**	**83.9**	**75.9**
P-MUCOS	85.8	**95.7**	89.5	82.1	**94.4**	89.5	90.9	84.7	**97.5**	68.0	76.8	65.8
BASELINE	86.3	92.9	88.8	82.4	93.2	86.2	88.9	81.6	79.3	61.1	66.8	57.3

This is contrasted to the more complex situations encountered in the youtube videos of PERTUBE. As a result, the distance matrices of the MHAD datasets are already of good quality. Therefore, the improved filtering and the two stage approach followed in this paper has more significant impact when applied to the lower-quality distance matrices of the PERTUBE data set.

We have also evaluated the following variant of the proposed method. We kept the original filtering of *P-MUCOS* and on that result, we applied the second stage proposed in this work. This hybrid scheme yields an overlap rate $\mathcal{O} = 70.5\%$ on the PERTUBE dataset. This means that the filtering improvement of this work and the introduction of the second stage contribute approximately equally to the improvement of the *P-MUCOS-S2* over the *P-MUCOS*. Finally, from a computational point of view, *P-MUCOS-S2* is about three times faster than *P-MUCOS*.

5 Conclusions

We proposed a method for discovering periodic segments in motion captured data and videos improving the state-of-the-art method proposed in [8]. The proposed framework is applied to distance matrices of pairwise distances of all frames of a given sequence detecting periodic actions in two stages. The experimental results on challenging datasets showed the effectiveness of the proposed method. As future work, we plan to extend the proposed method in two directions, (a) computation of the number of repetitions of a certain action [14] and (b) monitoring of the variation of the period of periodic motions. Both of these quantitative and qualitative characterizations of periodic motions find important applications in several computer vision applications involving action recognition, anomaly detection (e.g., [12]), performance characterization (e.g. [1]), e.t.c.

Acknowledgments

This work was partially supported by the EU project Co4Robots (H2020-731869).

References

1. Afsar, O., Tirnakli, U., Marwan, N.: Recurrence quantification analysis at work: Quasi-periodicity based interpretation of gait force profiles for patients with parkinson disease. Scientific reports **8**(1), 9102 (2018) 1, 9
2. Bampis, L., Amanatiadis, A., Gasteratos, A.: Fast loop-closure detection using visual-word-vectors from image sequences. The International Journal of Robotics Research **37**(1), 62–82 (2018) 4
3. Cutler, R., Davis, L.S.: Robust real-time periodic motion detection, analysis, and applications. IEEE Transactions on Pattern Analysis and Machine Intelligence **22**(8), 781–796 (2000) 1, 4
4. Elfeky, M.G., Aref, W.G., Elmagarmid, A.K.: Warp: time warping for periodicity detection. In: Data Mining, Fifth IEEE International Conference on. pp. 8–pp. IEEE (2005) 3
5. Karvounas, G., Oikonomidis, I., Argyros, A.A.: Localizing periodicity in time series and videos. In: BMVC (2016) 3
6. Levy, O., Wolf, L.: Live repetition counting. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 3020–3028 (2015) 4
7. Li, Z., Wang, J., Han, J.: eperiodicity: Mining event periodicity from incomplete observations. IEEE Transactions on Knowledge and Data Engineering **27**(5), 1219–1232 (2015) 4
8. Panagiotakis, C., Karvounas, G., Argyros, A.: Unsupervised detection of periodic segments in videos. In: 2018 25th IEEE International Conference on Image Processing (ICIP). pp. 923–927. IEEE (2018) 2, 3, 4, 5, 6, 8, 9
9. Panagiotakis, C., Papoutsakis, K., Argyros, A.: A graph-based approach for detecting common actions in motion capture data and videos. Pattern Recognition **79**, 1–11 (2018) 2, 3, 5
10. Papoutsakis, K., Panagiotakis, C., Argyros, A.: Temporal action co-segmentation in 3d motion capture data and videos. In: IEEE Conference on COmputer Vision and Pattern Recognition (CVPR) (2017) 3, 8
11. Polana, R., Nelson, R.C.: Detection and recognition of periodic, nonrigid motion. International Journal of Computer Vision **23**(3), 261–282 (1997) 4
12. Ramasso, E., Panagiotakis, C., Pellerin, D., Rombaut, M.: Human action recognition in videos based on the transferable belief model. Pattern analysis and Applications **11**(1), 1–19 (2008) 9
13. Ramasso, E., Placet, V., Boubakar, M.L.: Unsupervised consensus clustering of acoustic emission time-series for robust damage sequence estimation in composites. IEEE Transactions on Instrumentation and Measurement **64**(12), 3297–3307 (2015) 1
14. Runia, T.F., Snoek, C.G., Smeulders, A.W.: Repetition estimation. arXiv preprint arXiv:1806.06984 (2018) 1, 4, 5, 9
15. Tralie, C.J., Perea, J.A.: (quasi) periodicity quantification in video data, using topology. arXiv preprint arXiv:1704.08382 (2017) 4
16. Wang, H., Kläser, A., Schmid, C., Liu, C.L.: Dense trajectories and motion boundary descriptors for action recognition. IJCV **103**(1), 60–79 (mar 2013). https://doi.org/10.1007/s11263-012-0594-8, `http://dx.doi.org/10.1007/s11263-012-0594-8` 8
17. Wang, P., Abowd, G.D., Rehg, J.M.: Quasi-periodic event analysis for social game retrieval. In: Computer Vision, 2009 IEEE 12th International Conference on. pp. 112–119. IEEE (2009) 4