Synchronization among Groups of Spectators for Highlight Detection in Movies

Muszynski, Michal; Kostoulas, Theodoros; Lombardo, Patrizia; Pun, Thierry; Chanel, Guillaume

Synchronization among Groups of Spectators for Highlight Detection in Movies

Michal Muszynski
Computer Vision and Multimedia Laboratory
University of Geneva
Switzerland
michal.muszynski@unige.ch

Theodoros Kostoulas
Computer Vision and Multimedia Laboratory & Swiss Center for Affective Sciences
University of Geneva
Switzerland
theodoros.kostoulas@unige.ch

Patrizia Lombardo
Department of Modern French & Swiss Center for Affective Sciences
University of Geneva
Switzerland
patrizia.lombardo@unige.ch

Thierry Pun
Computer Vision and Multimedia Laboratory & Swiss Center for Affective Sciences
University of Geneva
Switzerland
thierry.pun@unige.ch

Guillaume Chanel
Swiss Center for Affective Sciences & Computer Vision and Multimedia Laboratory
University of Geneva
Switzerland
guillaume.chanel@unige.ch

ABSTRACT

Detection of emotional and aesthetic highlights is a challenge for the affective understanding of movies. Our assumption is that synchronized spectators' physiological and behavioral reactions occur during these highlights. We propose to employ the periodicity score to capture synchronization among groups of spectators' signals. To uncover the periodicity score's capabilities, we compare it with baseline synchronization measures, such as the nonlinear interdependence and the windowed mutual information. The results show that the periodicity score and the pairwise synchronization measures are able to capture different properties of spectators' synchronization, and they indicate the presence of some types of emotional and aesthetic highlights in a movie based on spectators' electro-dermal and acceleration signals.

Categories and Subject Descriptors

I.5.4 [**Pattern Recognition**]: Applications—*signal processing*; I.5.2 [**Pattern Recognition**]: Design Methodology—*pattern analysis*

Keywords

Synchronization; Grassmann manifolds; Persistent homology; Physiological and behavioral signals; Affective computing; Highlight detection

1. INTRODUCTION

Many studies have focused on matching spectators' physiological signals with affective states and the appearance of highlights in movies [5]. Since physiological reactions are considered to be an important component of emotions [15], [4], their measurements provide insight into spectators' aesthetic experience elicited by particular scenes [24]. In the field of affective computing, researchers have attempted to investigate emotion recognition in responses to multimedia content using electroencephalography (EEG) signals, peripheral physiological signals and facial expressions [14], [22]. The combination of spectators' physiological signals has been proposed in [5]. Because the spectators were watching separately a movie without any social context, they could not interact among themselves as it is the case in our studies.

Spectators can display similar behaviors or have similar physiological reactions when they are watching a movie together because: (i) the aesthetic choices of the filmmaker are made to elicit specific emotional reactions (e.g. special effects, empathy and compassion toward a character, etc.) and (ii) watching a movie together causes spectators' affective reactions to be synchronized through processes of emotional contagion [10]. For these reasons we gain insight on the impact of synchronization among group of spectators.

The first concept of synchronization came from the rhythm adjustment of oscillating objects [19]. In social sciences interpersonal synchrony consists of three components: rhythm, simultaneous movement and smooth meshing of interactions [2], [7]. One important step was made by [21] to introduce generalized synchronization of coupled chaotic systems. In [17], [18] the authors proposed that a level of the periodicity score can measure the amount of pattern repetitions in signals.

In this paper we propose to use periodicity score and baseline pairwise synchronization measures, such as the nonlinear interdependence [20] and the windowed mutual information [13] to uncover relations between occurrence of emotional and aesthetic highlights in films and spectators' physiological and behavioral reactions. Our goal is to verify if a level of synchronization that is computed over time windows of spectators' electro-dermal and acceleration signals indicates the occurrence of different types of emotional and aesthetic highlights in movies. We then evaluate our results with respect to the annotations made by a movie critic.

In section 2 we detail the adaptation of synchronization measures to process spectators' physiological and behavioral signals. In section 3 we describe a movie watching in a ecological situation. In section 4 we present all the results. In section 5 we discuss and interpret the obtained results. In section 6 we provide the conclusions of our studies.

2. SYNCHRONIZATION MEASURES

In this section we propose to apply three synchronization measures: the periodicity score, the nonlinear interdependence and the windowed mutual information to spectators' physiological and behavioral signals in order to detect emotional and aesthetic highlights in movies. A high level of all synchronization measures reveals the synchronized reactions of spectators while watching a movie.

For spectators' electro-dermal and acceleration signals $\{x_i\}$ we consider time windows $\{x_i(l)\}$, $i = 1, ..., M$, $l = 1, ..., N$, where M is a number of spectators' signals and N is a number of time windows.

2.1 Contribution - Periodicity Score

In this subsection we detail the usage of the periodicity score to measure synchronization of signals [17], [18]. The overview is described in Figure 1. First, we map spectators' physiological or behavioral signals to the geometric framework of real Grassmann manifolds by applying the reduced singular value decomposition (RSVD) to their short time Fourier transform (STFT). We analyze time windows of spectators' signals as a sequence of points encoded on the Grassmann manifold preserving their intrinsic dependencies. Next, we associate a level of the periodicity score with the synchronized spectators' physiological and behavioral signals during emotional and aesthetic highlights in movies.

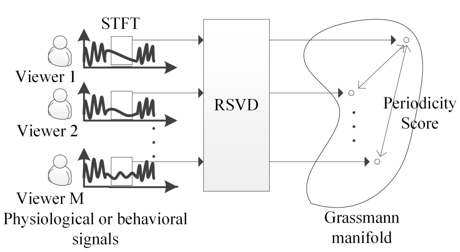

Figure 1: Physiological or behavioral signals are mapped to points on a Grassmann manifold applying RSVD to their STFT over time windows.

STFT. As shown in Figure 1 that we apply STFT to given time windows $x_i(l)$, $i = 1, ..., M$, and we yield $x^{i,l}_{t,f}$ in the time and frequency domain, where t is the time frame index and f is the frequency band index. Each time window $x_i(l)$ is split into segments with an overlap of 50% to apply STFT in this paper. Let $S^{i,l}_x(t, f)$ be the squared magnitude of the STFT, as follows

$$S^{i,l}_x(t, f) = ||x^{i,l}_{t,f}||^2. \quad (1)$$

RSVD. Then, we map time windows $\{x_i(l)\}$ of all signals on Grassmann manifolds to recover the intrinsic dependencies among them [6]. The real Grassmann manifold $G(k, n)$ parametrizes all $k-$dimensional subspaces of the vector space $\mathbb{R}^n$. A sequence of corresponding matrices $S^{i,l}_x(t, f)$, $i = 1, ..., M$ can be mapped to the points on the manifold $G(k, n)$ using RSVD as shown in the figure 1. If we compute RSVD of matrix $S^{i,l}_x(t, f)$, as follows:

$$S^{i,l}_x(t, f) = U^i \Sigma^i {V^i}^T, \quad (2)$$

then the columns of the $n \times k$ orthogonal matrix U^i are a non-unique basis for the column space of $S^{i,l}_x(t, f)$. Thus, U^i can be used to represent the matrix $S^{i,l}_x(t, f)$, and can be identified with a point on the Grassmann manifold $G(k, n)$. Once the time windows are mapped to a sequence of points on $G(k, n)$, the pairwise distances between these points can be found using a function of the angles between subspaces. Let U^i and U^j be two $k-$ dimensional subspaces, we measure the similarity $d_{min}(U^i, U^j)$ of two points on the Grassmann manifold $G(k, n)$ using the minimum correlation distance [9]

$$d_{min}(U^i, U^j) = \sin \theta_k, \quad (3)$$

where $0 \leq \theta_1 \leq \theta_2 \leq ... \leq \theta_k \leq \frac{\pi}{2}$ are principle angles between two subspaces.

Periodicity Score. Finally, we introduce the basics of persistent homology: filtrations and persistence diagrams [17], [18], [8] plotted in Figure 2. Once the sequence of $S^{i,l}_x(t, f)$, $i = 1, ..., M$ matrices is mapped to $G(k, n)$ and defines a metric space $(U = \{U^1, ..., U^M\}, d_{min}(\cdot, \cdot))$, we recall the definition of the Vietoris-Rips complex $Rips_\alpha(U)$ as the set of the simplices $[U^1, ..., U^q]$ such that $d_{min}(U^i, U^j) \leq \alpha$ for $i, j = 1, ..., q$. There is an inclusion of $Rips_\alpha(U)$ in $Rips_\beta(U)$ for any $\alpha \leq \beta$. The sequences of inclusions are called filtrations $Filt_\alpha(U)$. An example is given in Figure 2(a). Persistence diagrams allow us to study the evolution of the topology of a filtration, and to capture properties of the metric which is used to generate the filtration. Existing connected components are merged for $0-$th homology, when α increases shown in Figure 2(b). Persistent homology tracks the birth (appearance) b and death (disappearance) d of all connected components shown in Figure 2(c).

The maximum persistence $mp(dgm(x_i(l)))$ of a persistence diagram $dgm(x_i(l))$ is defined as follows [18]

$$mp(dgm(x_i(l))) = \max_{(b,d) \in dgm(x_i(l))} pers(b, d), \quad (4)$$

where $pers(b, d) = d - b$ for $(b, d) \in dgm(x_i(l))$, and as ∞ otherwise. Finally, we can provide the periodicity score $S(x_i(l))$ [18]

$$S(x_i(l)) = \frac{mp(dgm(x_i(l)))}{\sqrt{3}}. \quad (5)$$

The normalized maximum persistence $mp(dgm(x_i(l)))$ of a persistence diagram $dgm(x_i(l))$ can help us to quantify synchronization among signals because it is capable of measuring their intrinsic geometric dependencies. The persistent score can measure synchronization among groups of signals based on the connectivity of signal clusters. Our approach to synchronization contains the multivariate measure which ascribes a single value to all signals in comparison with univariate measures, such as the nonlinear interdependence and the windowed mutual information which can only be computed over pairs of signals. When $S(x_i(l))$ equals 0, it means that we can not explore any structure in our data. If a value of $S(x_i(l))$ rises close to 1, we find some strong connectivity structure of data (synchronization).

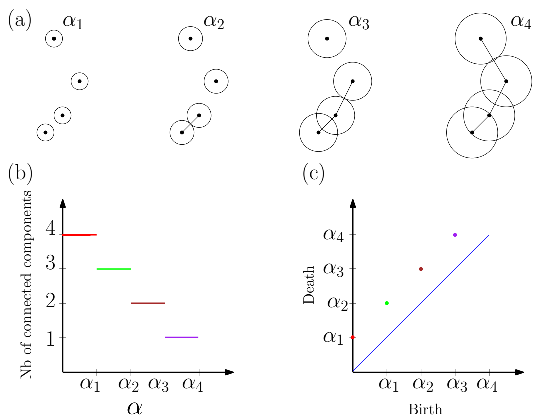

Figure 2: We show an example of : (a) the filtered Vietoris-Rips complex, (b) a number of connected components for different values of filtration parameter α, (c) the persistent diagram.

2.2 Nonlinear Interdependence

The nonlinear interdependence measures the geometrical similarity between the state space trajectories of two dynamical systems that are reconstructed from two time series $\{x_i\}$ and $\{x_j\}$ $i, j = 1, ..., M$ using time-delay embedding [23]. For each time window $x_i(l)$ the mean square Euclidean distance to its K nearest neighbours $x_i(n)$, $n = 1, ..., K$ is

$$R^K(x_i(l)) = \frac{1}{K}\sum_{n=1}^{K}(x_i(l) - x_i(n))^2, \quad (6)$$

and the mean squared Euclidean distance conditioned by the equal time partners of the K nearest neighbours of $x_j(l)$ is

$$R^K(x_i(l)|x_j(l)) = \frac{1}{K}\sum_{n=1}^{K}(x_i(l) - x_j(n))^2. \quad (7)$$

The nonlinear interdependence measure is defined as [20]

$$S^K(x_i(l)|x_j(l)) = \frac{R^K(x_i(l))}{R^K(x_i(l)|x_j(l))}. \quad (8)$$

To make the nonlinear interdependence symmetric, we consider $S^K(x_j(l)|x_i(l))$ and we then average these parameters.

2.3 Windowed Mutual Information

From an information theory viewpoint, any signal can be treated as a collection of random variables which describes the evolution of a system over time. In this context, the windowed mutual information may capture nonlinear dependencies between signals that are not revealed in the covariance of signals [13]. The straightforward approach to estimation of the windowed mutual information consists of partitioning the supports of two time windows $x_i(l)$, $x_j(l)$ into finite size bins, and the approximation by the finite sum

$$I(x_i(l), x_j(l)) = \sum_{w,q} p(w,q)\log(\frac{p(w,q)}{p(w)p(q)}), \quad (9)$$

where $p(w,q)$ is the joint probability density function of $x_i(l)$ and $x_j(l)$, $p(w)$ and $p(q)$ are the marginal probability density functions, respectively.

3. EXPERIMENT

The spectators' physiological and behavioral signals were recorded with a sampling frequency of 10 Hz during a movie projection (Taxi Driver, 1976) in a theater (Grütli cinema, Geneva) [12]. The duration of the movie is 113 min. In this paper we use 12 spectators' electro-dermal activity and acceleration (x,y, z axes, accelerometer attached to the hand of the spectator) signals ($M = 12$). All signals are filtered by third order lowpass Butterworth filter with cutoff frequency 0.3 Hz, and they are segmented into overlapping time windows with a time step and a window length equal 2 s and 5 s, respectively. The selection of these parameters was done to indicate highlights in meaningful time period for the whole duration of the movie.

Annotation of the movie was performed offline by a movie critic, who annotated the movie based on the following five types of emotional and aesthetic highlights [1], [3].
"Form-highlights" (the manner in which the subject is presented in the film):
- *H1*: Spectacular (technical choice, special effects);
- *H2*: Subtle (use of camera, lighting, music).
"Content-highlights" (the presented subject in the film):
- *H3*: Character development (characters' emotions and responses to dramatic events);
- *H4*: Dialogue (motivation of actions and tensions among characters);
- *H5*: Theme development (unusual close up, urban theme).
For each type of highlights, non-highlights are scenes without the particular type of the highlights (possibly, containing the other highlight type).
In these studies the periodicity score (PS) and pairwise synchronization measures: the nonlinear interdependence (NI) and the windowed mutual information (WMI) are applied to the spectators' physiological and behavioral signals to uncover different properties of their synchronization.

4. RESULTS

We employ the two side Welch's t-test at the significance level ($\alpha = 0.1$) to test the hypothesis. We verify if an increase/a drop of the synchronization of the spectators' physiological and behavioral reactions might appear during a particular type of highlights [16], [11].

Figure 3(a) shows the mean values of the PS which are computed over the spectators' physiological signals for all types of highlights. The values of the PS increase marginally significantly for the scenes in the movie containing spectacular highlights *H1* (t=1.77, p<0.1, r=0.12) and theme development highlights *H5* (t=1.90, p<0.1, r=0.14) in comparison to the scenes without those particular types of highlights. Figure 3(b) plots the results for the behavioral signals. The values of the PS rise significantly and marginally significantly for the scenes consist of spectacular highlights *H1* (t= 2.55, p=0.01, r=0.15) and subtle highlights *H2* (t=1.77, p<0.1, r=0.09), respectively. Moreover, the values of the PS drop marginally significantly for character development highlights *H3* (t=-1.71, p<0.1, r=0.11).

In Figure 3(c) the mean values of the NI for the electrodermal signals are plotted. The synchronization increases

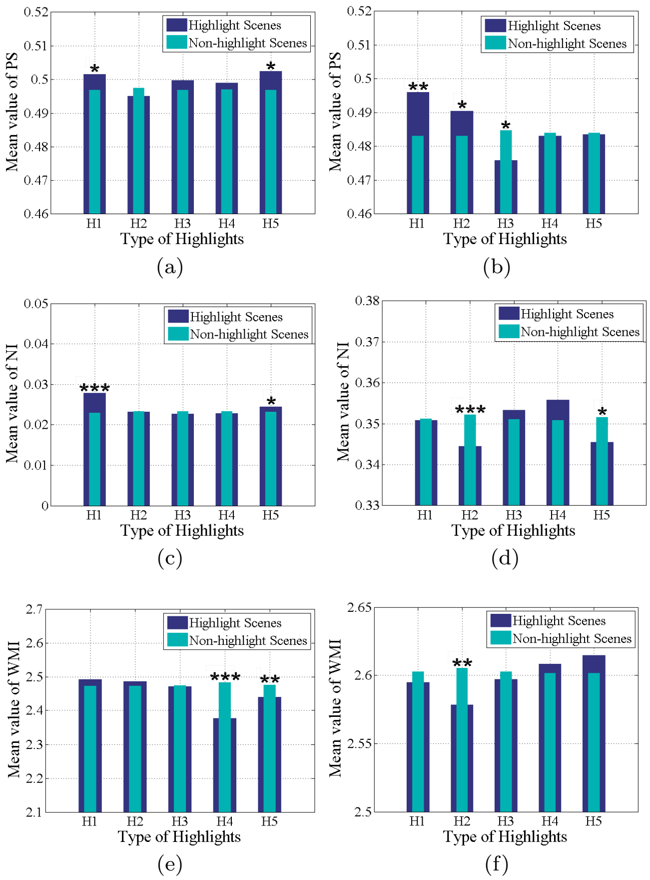

Figure 3: Mean values of synchronization measures: PS ((a), (b)), NI ((c), (d)), WMI ((e), (f)) for one particular type of highlight scenes (*H1*, *H2*, *H3*, *H4*, *H5*) versus scenes without the type of highlights in the movie. Left charts correspond to elctro-dermal activity signals and right charts correspond to acceleration signals. $\star$ stands for a p-value < 0.1, $\star\star$ for a p-value < 0.05 and $\star\star\star$ for a p-value < 0.01.

significantly and marginally significantly for spectacular highlights *H1* (t=4.99, p<0.01, r=0.52) and theme development highlights *H5* (t=1.91, p<0.1, r=0.14), respectively. Figure 3(d) gives information on the synchronization among the behavioral signals. The values of the NI falls significantly and marginally significantly for subtle highlights *H2* (t=-3.55, p<0.01, r=0.16) and theme development highlights *H5* (t=-1.78, p<0.1, r=0.13), respectively.

Figure 3(e) for the WMI shows a significant decrease of dependencies among the physiological signals of the spectators for dialogue highlights *H4* (t=-10.9, p<0.01, r=0.55) and theme development highlights *H5* (t=-2.32, p<0.05, r=0.18). In Figure 3(f), we also observe a significant drop of dependencies among behavioral signals but only for subtle highlights *H2* (t=-2.58, p=0.01, r=0.14).

5. DISCUSSION

We can detect some types of highlights applying synchronization measures. The obtained results show that the values of the PS and NI increase marginally significantly and significantly, respectively, for spectacular highlights *H1*. These observations are in line with our previous works [16], [11], [12]. This can be justified by the nature of the scenes since *H1* corresponds to spectacular scenes where the director uses special effects, such as an increasing saturation of red color during final shooting scenes, playing with lights and a location of the camera. These may evoke strong emotional reactions and emotional contagion.

The slow movement of a camera and music during subtle highlights *H2* could decrease the spectators' behavioral reactions. Due to discrepancies among the spectators we observe a drop of the NI and the WMI while the PS might still capture some synchronization in their behaviors.

The PS falls marginally significantly for the spectators' behavioral reactions during character development highlights *H3* while the pairwise synchronization measures could not expose any marginally significant drop/increase of the synchronization. Some group of the spectators could only react similarly to the attitude of the main characters which could be caused by the ambiguity of their personality.

Also, we observe a significant drop of the WMI values for the spectators' physiological signals in the case of dialogue highlights *H4*. This can occur because of two reasons: long average duration of highlights *H4* may cause that the spectators' emotions fade in time, and the main character is also an ambiguous movie character who could elicit different reactions across the audience.

The results also report that the values of the PS and NI increase marginally significantly for the theme development highlights *H5*. The rise of spectators' synchronization might be caused partially by the overlapping of spectacular and theme development scenes in this particular movie. But the WMI also measures a significant drop in the statistical dependencies. It can be explained by a lack of any mutual dependence between signals.

6. CONCLUSIONS

In this paper we suppose that the occurrence of synchronized spectators' behaviors and physiological reactions are linked to the presence of emotional and aesthetic highlights in a movie. To detect these highlights, we propose to apply the periodicity score (PS) to measure synchronization among groups of spectators' signals that can not be identified by other measures. To explore the PS's properties, we compare it with the baseline pairwise synchronization measures: the nonlinear interdependence (NI) and the windowed mutual information (WMI). We show that the PS as well as the NI, can indicate the appearance of spectacular highlights *H1* and theme development highlights *H5* based on the spectators' physiological reactions. These results can be explained by special effects related to spectacular scenes in the urban landscape when strong physiological and behavioral reactions of spectators are supposed to be evoked. Also, the PS depicts the appearance of the spectacular highlights based on the spectators' behavioral signals. The PS as the NI and the WMI could identify the appearance of subtle highlights *H2* based on acceleration signals of the spectators. Moreover, the PS only decreases indicating character development highlights *H3* in comparison with the other methods.

In the future work, we will collect more multimodal signals to verify our preliminary studies. Also, we will attempt to reveal more topological features that could uncover different types of synchronization in spectators' physiological and behavioral data while watching different movies.

7. REFERENCES

[1] A. Bazin. *What is cinema?* University of California Press, 2004.

[2] F. J. Bernieri and R. Rosenthal. Interpersonal coordination: Behavior matching and interactional synchrony. 1991.

[3] D. Bordwell, K. Thompson, and J. Ashton. *Film art: an introduction*. McGraw-Hill New York, 1997.

[4] G. Chanel and C. Mühl. Connecting brains and bodies: applying physiological computing to support social interaction. *Interacting with Computers*, page iwv013, 2015.

[5] C. Chênes, G. Chanel, M. Soleymani, and T. Pun. Highlight detection in movie scenes through inter-users, physiological linkage. In *Social Media Retrieval*, pages 217–237, 2013.

[6] S. Chepushtanova, M. Kirby, C. Peterson, and L. Ziegelmeier. An application of persistent homology on grassmann manifolds for the detection of signals in hyperspectral imagery. In *Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International*, pages 449–452. IEEE, 2015.

[7] E. Delaherche, M. Chetouani, A. Mahdhaoui, C. Saint-Georges, S. Viaux, and D. Cohen. Interpersonal synchrony: A survey of evaluation methods across disciplines. *IEEE Transactions on Affective Computing*, 3(3):349–365, 2012.

[8] R. Ghrist. Barcodes: the persistent topology of data. *Bulletin of the American Mathematical Society*, 45(1):61–75, 2008.

[9] J. Hamm and D. D. Lee. Grassmann discriminant analysis: a unifying view on subspace-based learning. In *Proceedings of the 25th international conference on Machine learning*, pages 376–383. ACM, 2008.

[10] E. Hatfield, J. T. Cacioppo, and R. L. Rapson. *Emotional contagion*. Cambridge university press, 1994.

[11] T. Kostoulas, G. Chanel, M. Muszynski, P. Lombardo, and T. Pun. Dynamic time warping of multimodal signals for detecting highlights in movies. In *Proceedings of the 1st Workshop on Modeling INTERPERsonal SynchrONy And infLuence*, pages 35–40. ACM, 2015.

[12] T. Kostoulas, G. Chanel, M. Muszynski, P. Lombardo, and T. Pun. Identifying aesthetic highlights in movies from clustering of physiological and behavioral signals. In *Quality of Multimedia Experience (QoMEX), 2015 Seventh International Workshop on*, May 2015.

[13] A. Kraskov, H. Stögbauer, and P. Grassberger. Estimating mutual information. *Physical review E*, 69(6):066138, 2004.

[14] E. Kroupi, J.-M. Vesin, and T. Ebrahimi. Phase-amplitude coupling between eeg and eda while experiencing multimedia content. In *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*, pages 865–870. IEEE, 2013.

[15] T. Li, Y. Baveye, C. Chamaret, E. Dellandréa, and L. Chen. Continuous arousal self-assessments validation using real-time physiological responses. In *International Workshop on Affect and Sentiment in Multimedia (ASM)*, 2015.

[16] M. Muszynski, T. Kostoulas, G. Chanel, P. Lombardo, and T. Pun. Spectators' synchronization detection based on manifold representation of physiological signals: Application to movie highlights detection. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, pages 235–238. ACM, 2015.

[17] J. A. Perea, A. Deckard, S. B. Haase, and J. Harer. Sw1pers: Sliding windows and 1-persistence scoring; discovering periodicity in gene expression time series data. *BMC bioinformatics*, 16(1):257, 2015.

[18] J. A. Perea and J. Harer. Sliding windows and persistence: An application of topological methods to signal analysis. *Foundations of Computational Mathematics*, 15(3):799–838, 2015.

[19] A. Pikovsky, M. Rosenblum, and J. Kurths. *Synchronization: a universal concept in nonlinear sciences*, volume 12. Cambridge university press, 2003.

[20] R. Q. Quiroga, J. Arnhold, and P. Grassberger. Learning driver-response relationships from synchronization patterns. *Physical Review E*, 61(5):5142, 2000.

[21] N. F. Rulkov, M. M. Sushchik, L. S. Tsimring, and H. D. Abarbanel. Generalized synchronization of chaos in directionally coupled chaotic systems. *Physical Review E*, 51(2):980, 1995.

[22] M. Soleymani, S. Asghari-Esfeden, M. Pantic, and Y. Fu. Continuous emotion detection using eeg signals and facial expressions. In *Multimedia and Expo (ICME), 2014 IEEE International Conference on*, pages 1–6. IEEE, 2014.

[23] F. Takens. Detecting strange attractors in turbulence. *Dynamical Systems and Turbulence, Lecture Notes in Mathematics*, 898:366–381, 1981.

[24] H. L. Wang and L.-F. Cheong. Affective understanding in film. *IEEE Transactions on Circuits and Systems for Video Technology*, 16(6):689–704, June 2006.