Dynamical Regularity for Action Analysis

Vinay Venkataraman[1]
http://www.public.asu.edu/~vvenka18/
Ioannis Vlachos[2]
ivlachos@latech.edu
Pavan Turaga[1]
http://www.public.asu.edu/~pturaga/

[1] School of Arts, Media + Engineering, School of Electrical, Computer and Energy Engineering, Arizona State University, Arizona, USA

[2] Center for Biomedical Engineering and Rehabilitation Science, Louisiana Tech University, Louisiana, USA

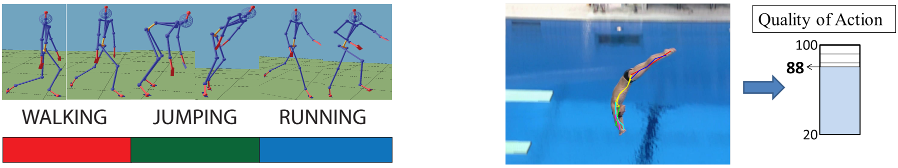
Abstract

In this paper, we propose a new approach for quantification of 'dynamical regularity' as applied to modeling human actions. We use approximate entropy-based feature representation to model the dynamics in human movement to achieve temporal segmentation in untrimmed motion capture data and fine-grained quality assessment of diving actions in videos. The principle herein is to quantify regularity (frequency of typical patterns) in the dynamical space computed from trajectories of action data. We extend conventional ideas for modeling dynamics in human movement by introducing multivariate and cross approximate entropy features. Our experimental evaluation on theoretical models and two publicly available databases show that the proposed features can achieve state-of-the-art results on applications such as temporal segmentation and quality assessment of actions.

1 Introduction

The computer vision community has been interested in modeling human activities for many applications including video surveillance, automatic video annotation and health monitoring [1]. Modeling the underlying dynamics in an activity forms the core idea in many systems. An activity can be seen as a resultant of coordinated movement of body joints and their respective interdependencies to achieve a goal-directed task. This idea is further supported by Johansson's demonstrations that visual perception of the entire human body motion can be represented by a few bright spots which holistically describe the motion of important joints [10]. Traditional dynamical modeling approaches usually operate on the level of individual joints of the human body, lacking any information about the interdependencies between joints [4]. Only recently, researchers have started exploring relationships between body joints, using rotations and translations in 3D space [25], which lacks dynamical information. In this paper, we propose a novel approach for dynamical modeling by extending conventional ideas to quantify the interdependencies between body joints. Towards this end, we propose a new approach – approximate entropy-based feature representation to model the dynamics in human movement by quantifying dynamical *regularity*.

Pages 67.1-67.12
DOI: https://dx.doi.org/10.5244/C.29.67



(a) Temporal segmentation of actions using motion capture data.

(b) Quality assessment of diving actions using videos.

Figure 1: A visual representation of our applications of interest in this work. In (a), our aim is to achieve temporal segmentation of actions from continuous untrimmed motion capture data in an unsupervised manner. In (b), we use a supervised learning framework to assess the quality of diving actions from videos.

Our use of the term *regularity* represents the frequency of repetition of typical patterns in the data. The main principle in our work is that different actions correspond to different levels of regularity, and quantification of regularity can be used for human activity analysis. For instance, *walking* is inherently periodic and hence corresponds to a higher level of regularity when compared to *dancing*, which is more towards randomness due to multiple movement strategies. From the system complexity perspective, *walking* can be represented by simple dynamical systems, while more complex systems with a large number of variables may be required to represent *dancing*. Quantifying regularity and system complexity is a well-studied problem in the field of signal processing. Correlation dimension [2] and largest Lyapunov exponent [28] are examples of invariant measures proposed in the literature to quantify complexity of dynamical systems. It was found that robust estimation of these invariant measures requires large number of data samples (of the order of 10^d), where d is related to the dimension of the dynamical system's state space used in the estimation procedure, with typical values of 3 and above. Later, a probabilistic measure called approximate entropy was proposed to overcome the drawbacks of the above traditional measures for quantification of system complexity [17]. Approximate entropy assigns lower values for ordered time series and higher values for time series towards randomness. In this paper, we utilize the algorithmic framework of [17] for estimating approximate entropy from time series data and extend it to model the dynamics in human activities for applications such as temporal segmentation and fine-grained quality assessment of actions.

Temporal Segmentation: Human motion recognition from untrimmed videos is a challenging problem due to large variations in the temporal scale of actions and extremely large number of possible movement combinations [9, 22]. Traditionally, one assumes temporal segmentation of videos is a step which has been done beforehand, resulting in pre-segmented videos containing individual action sequences [3]. However, in a real world scenario, applications such as surveillance require automatic recognition of action sequences from continuous untrimmed videos. In this work, as shown in Figure 1a, we develop a framework using approximate entropy-based features for temporal segmentation of actions from untrimmed motion capture data in an unsupervised manner.

Quality Assessment: With adequate success in recognizing actions from videos, researchers in the computer vision community have shown growing interest in fine-grained analysis of human activities by developing frameworks for quantification of movement quality [19, 26]. Quality assessment of human activities has recently been used in the field of sports [19], healthcare and rehabilitation [26]. In this work, as shown in Figure 1b, we use an approximate entropy-based feature representation and show its utility to assess action quality to match human expert ratings on diving actions.

2 Related Work

Our current work is focused on dynamical modeling of human actions for temporal segmentation and fine-grained quality assessment of actions, and therefore we restrict our discussion to related methods focused on our applications of interest.

Temporal Segmentation: Some of the early approaches for temporal segmentation of actions include learning representations for motion primitives using the theory of linear dynamical systems [12, 23, 24], thereby segmenting the human motion into its constituent action sequences. Oh *et al.* [13] utilized switching linear dynamical system to learn and infer motion patterns. Other approaches have been proposed in literature for temporal segmentation of human actions based on hidden Markov models (HMMs). Bregler *et al.* [6] utilized HMMs to model complex human gestures as successive phases of simple movements. Brand *et al.* [5] applied coupled HMMs demonstrating superiority to conventional HMMs towards classifying two-handed human motion. Spriggs *et al.* [22] used HMMs for temporal segmentation of activities in a kitchen using a wearable camera and inertial measurement units.

Recent work by Zhou *et al.* [30] proposed hierarchical aligned cluster analysis (HACA) for temporal segmentation by extending standard kernel k-means clustering combined with dynamic time warping for unsupervised temporal segmentation of human motion. HACA was proposed as an extension to their previous work of aligned cluster analysis [29] by reducing the computational complexity. We note here that the input to both these algorithms is a frame kernel matrix (recurrence matrix), and it is apparent that the performance of these clustering approaches depends on the *quality* of the recurrence matrix. In this paper, we show the utility of the approximate entropy-based feature representation to estimate a recurrence matrix which is better suited for clustering temporal actions as validated by our experiments.

Quality Assessment: Even though researchers have been working towards automatic recognition of human actions for decades, the task of automatically quantifying the quality of a given action has remained unexplored until recently. Such automated frameworks for quality assessment of actions will find real-world applications in sports and healthcare. Hamed *et al.* [19] used a regression model to predict the scores given by human expert judges on diving actions using spatio-temporal pose features. A similar approach using a regression model learned from shape-based dynamical features to quantify the quality of movement has been proposed for stroke rehabilitation [26]. In [14], authors quantified team performance in a multi-player basketball activity context using Bayesian networks. In this paper, we utilize the approximate entropy-based feature to quantify the quality of diving actions and show that using a dynamical measure performs better than the previously used frequency domain representation using discrete cosine transform (DCT).

Contributions: Our work has the following contributions: (1) We propose a feature representation to model human motion by quantification of regularity using approximate entropy. The novelty in the proposed feature representation is that it encodes both the dynamics of individual joints and cross-coupling information (interaction) between joints. (2) We show the utility of the approximate entropy features to produce improved recurrence matrices for temporal segmentation of actions. (3) We also show its usage in fine-grained quality assessment. Our experimental evaluation on two publicly available databases show that the proposed framework achieves state-of-the-art performance.

3 Approximate Entropy (ApEn)

Approximate entropy is a statistical tool proposed by Pincus [17, 18] for quantification of regularity of time series data and system complexity. It is a probabilistic measure based on the log-likelihood of repetitions of patterns of length m being close within a defined tolerance window that will exhibit similar characteristics as patterns of length $(m+1)$ [16, 17]. It assigns a non-negative number to time series data, with lower values for predictable (ordered) signals and higher values for signals with increased irregularity (or randomness). Ideally, a pure sine wave should have a zero value of approximate entropy. It has an advantage over Shannon's entropy [20] in that it takes into account the temporal order, which makes it more suitable to represent the dynamical evolution of time series data. The development of approximate entropy was motivated to address the drawbacks of traditional measures to quantify system complexity, thereby having a measure to successfully handle noise and address the limitations of data length requirements and other model constraints [18].

It is defined using three parameters: embedding dimension (m), radius (r), and time delay (τ). Here, m represents the length of pattern (also called as embedding vector) in the data which is checked for repeatability, τ is selected so that the components of the embedding vector are sufficiently independent, and r is used for the estimation of local probabilities. Given N data samples $\{x_1, x_2, x_3, \ldots, x_N\}$, we can define embedding vector $\mathbf{x}(i)$ as,

$$\mathbf{x}(i) = [x_i, x_{i+\tau}, x_{i+2\tau}, \ldots, x_{i+(m-1)\tau}]^T; \quad for \quad 1 \leq i \leq N-(m-1)\tau. \tag{1a}$$

The frequency of repeatable patterns of the embedding vector within a tolerance r is given by $\mathbf{C}_i^m(r)$ as

$$\mathbf{C}_i^m(r) = \frac{1}{N-(m-1)\tau} \sum_{<j>} \Theta(r - d(\mathbf{x}(i), \mathbf{x}(j))). \tag{1b}$$

where:

$$\Theta(a) = \begin{cases} 1, & \text{if } a \geq 0 \\ 0, & \text{otherwise.} \end{cases}$$

$$d(\mathbf{x}(i), \mathbf{x}(j)) = \max_{k=1,2,..,m} (|x(i+(k-1)\tau) - x(j+(k-1)\tau)|).$$

Approximate Entropy is given by

$$ApEn(m, r, \tau) = \Phi^m(r) - \Phi^{m+1}(r). \tag{1c}$$

where:

$$\Phi^m(r) = \frac{1}{N-(m-1)\tau} \sum_{i=1}^{N-(m-1)\tau} \ln \mathbf{C}_i^m(r). \tag{1d}$$

In the above equations, $\mathbf{C}_i^m(r)$ represents the frequency of repeatable patterns (local probabilities) in the embedding vector $\mathbf{x}(i)$, $\Theta(a)$ is the Heaviside step function, and $\Phi^m(r)$ represents the conditional frequency estimates. Evident from the above algorithm, the estimation procedure requires parameters m, τ, and r to be specified. In an ideal case, where one has access to an infinite amount of data of infinite accuracy, any set of parameters which can result in smooth embedding would give similar results ([2], chap. 3). With real world data, the choice of these parameters should ensure smooth embedding with components of the embedding vectors being sufficiently independent.

Multivariate Approximate Entropy: Motion capture sensing allows us to observe 3-dimensional time series data per body joint. A trivial solution to model the dynamics would be to consider each dimension of a body joint independently to create the embedding vector (eq. 1a) as in [4, 26]. Recent theoretical and empirical findings have demonstrated that multivariate embedding of time series data by simple concatenation of individual univariate embedding vectors achieves good state space reconstruction as evaluated by the shape and dynamics distortion measures [27]. In this work, we propose to use the multivariate embedding procedure as described by Cao *et al.* [8] per body joint and estimate the approximate entropy feature representation.

Natural human movement involves multiple body joints interacting with each other to together accomplish a particular action task. Hence, it would be beneficial to utilize the cross-coupling information between these joint trajectories. Research carried out by Kavanagh *et al.* [11] using cross approximate entropy to model trunk motion during walking supports our hypothesis that adding information about cross-coupling offers better feature representation to model human motion and will be validated by our experiments.

Cross Approximate Entropy (XApEn): Cross approximate entropy is defined as the amount of asynchrony between two time series data [15, 16]. Let $\mathbf{u} = [u_1, u_2, \dots, u_N]^T$ and $\mathbf{v} = [v_1, v_2, \dots, v_N]^T$ denote two time series data of length N. The embedding vectors for given parameters m, τ, and r are defined as

$$\mathbf{x}_1(i) = [u_i, u_{i+\tau}, \dots, u_{i+(m-1)\tau}]^T; \quad \mathbf{x}_2(i) = [v_i, v_{i+\tau}, \dots, v_{i+(m-1)\tau}]^T. \tag{2a}$$

The frequency of repeatable patterns within the embedding vectors $\mathbf{x}_1(i)$ and $\mathbf{x}_2(i)$ for a tolerance r is given by $\mathbf{C}_i^m(r)(v||u)$ as

$$\mathbf{C}_i^m(r)(v||u) = \frac{1}{N-(m-1)\tau} \sum_{<j>} \Theta(r - d(\mathbf{x}_1(i), \mathbf{x}_2(j))). \tag{2b}$$

The cross approximate entropy is then given by

$$XApEn(m, r, \tau) = \Phi^m(r)(v||u) - \Phi^{m+1}(r)(v||u). \tag{2c}$$

where:

$$\Phi^m(r) = \frac{1}{N-(m-1)\tau} \sum_{i=1}^{N-(m-1)\tau} \ln \mathbf{C}_i^m(r)(v||u). \tag{2d}$$

We estimate the XApEn feature across all pairs of body joints (after performing multivariate embedding using data available from each body joint). It is evident from the above equations that XApEn is an asymmetric measure. We note here that our initial analysis on exemplar human action data did not show a significant difference in the values of XApEn for forward and backward directions. Hence, we use only one of these values in our feature representation. We then concatenate ApEn and XApEn values to form our final approximate entropy-based feature vector to model actions denoted by ***ApEnFT***.

3.1 Choice of Parameters

Data Length (N): The suggested value for N was typically between 50 and 5000. This constraint was imposed by Pincus in [18] to ensure a homogeneous segment of data under certain experimental conditions, and this range for N was not an algorithmic limitation. Our choice of N depends on the dataset used, and typically ranges between 30 and 50.

Embedding Dimension (m): Through theoretical analysis and extensive experimental validation, it has been shown that both $m = 1$ and $m = 2$ can distinguish data on the basis of regularity [18].

Delay Time (τ): The purpose of delay time τ is to ensure that the components in the embedding vectors are sufficiently independent. A low value of delay time will make adjacent components in the embedding vector to be correlated and hence cannot be considered as independent. On the other hand, a high value of delay time will make adjacent components to become uncorrelated (almost independent). Suggested methods in the literature to estimate an optimum delay time has been first minimum of the lagged auto-mutual information, and the time lag when the autocorrelation drops to 1/e of its initial value or the first zero of the autocorrelation function [2].

Radius (r): The value of r could range anywhere between 0.1 to 0.25 times the standard deviation of the data . A good choice of r should ensure that the conditional frequencies defined in Eq. 1c are reasonably estimated. Smaller values of r may result in poor conditional frequency estimates (not enough data samples), while large values of r cannot capture enough local information of the system (the patterns are not similar).

Baselines: The main contribution of our work is to propose a better way to encode dynamics compared to traditional dynamical modeling approaches. To evaluate the effectiveness of our framework, we provide comparative results in each experiment with univariate approximate entropy estimated on individual dimensions of action data denoted by ***UniAp***. We also compare our performance with a feature vector of traditional chaotic invariants obtained by concatenating largest Lyapunov exponent, correlation dimension and correlation integral (for 8 values of radius) resulting in a 10-dimensional feature vector denoted as ***Dynamics***, which has been recently used in action recognition [4] and natural scene recognition [21].

4 Experimental Evaluation

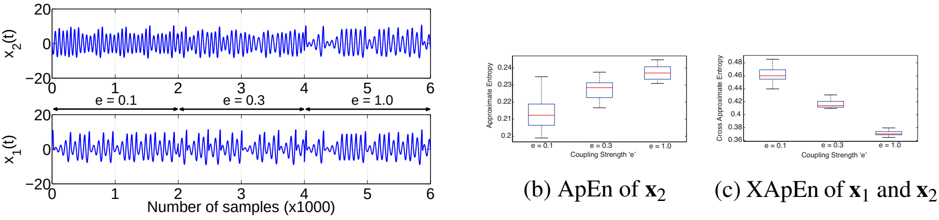
In this section, we evaluate the performance of our feature representation on (1) synthetic data generated from coupled Rossler oscillators, (2) temporal segmentation on motion capture dataset, and (3) quality assessment of diving actions.

4.1 Coupled Rossler Model

In order to demonstrate the utility of the proposed feature representation for quantifying regularity and cross-coupling in time series data, we use two coupled Rossler oscillators given by the equations shown below. The main motive behind this experiment is to provide an analogy to human actions as coupled systems with changing coupling strengths to accomplish different actions.

$$\begin{aligned} \dot{x}_1 &= -w_1 y_1 - z_1 \\ \dot{y}_1 &= w_1 x_1 + \alpha y_1 \\ \dot{z}_1 &= \beta + z_1(x_1 - \gamma) \end{aligned} \tag{3a}$$

$$\begin{aligned} \dot{x}_2 &= -w_2 y_2 - z_2 + e(x_1 - x_2) \\ \dot{y}_2 &= w_2 x_2 + \alpha y_2 \\ \dot{z}_2 &= \beta + z_2(x_2 - \gamma) \end{aligned} \tag{3b}$$



(a) Time series data

(b) ApEn of $\mathbf{x}_2$

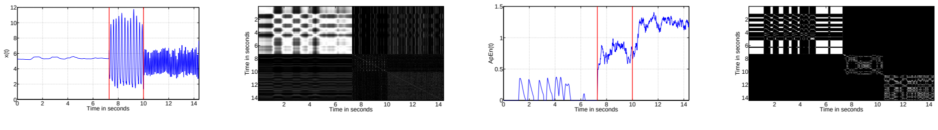
(c) XApEn of $\mathbf{x}_1$ and $\mathbf{x}_2$

Figure 2: Illustration of utility of approximate entropy feature representation for quantifying regularity and cross-coupling on coupled Rossler model. (a) shows exemplar time series data synthesized from the coupled Rossler model for three different coupling strength $e = 0.1, 0.3, 1.0$. (b) and (c) respectively show the distribution of ApEn values of $x_2(t)$ and the distribution of XApEn values of $x_1(t)$ and $x_2(t)$ for 20 trials each for different values of e.

Here, the Rossler system in Eq. 3a *drives* the Rossler system in Eq. 3b. 'e' is the coupling strength between the two Rossler oscillators. As the coupling strength is increased, the two oscillators become synchronized. For this configuration of Rossler oscillators, the parameters were chosen as $\alpha = 0.2$, $\beta = 0.2$, $\gamma = 5.7$, $w_1 = 1$, and $w_2 = 0.2$. We choose three values of coupling strength, $e = 0.1, 0.3$, and 1.0 to demonstrate the sensitivity of cross approximate entropy measure to coupling strength. For each value of e, we generate 20 data segments from the coupled Rossler system, with each segment having 2000 samples. Figure 2 shows exemplar time series of $\mathbf{x}_1$ and $\mathbf{x}_2$ for different coupling strengths. From Figure 2a, we see that as e approaches 1.0, $\mathbf{x}_2$ becomes more synchronized with $\mathbf{x}_1$. In a coupled Rossler system where one oscillator drives the other, the dynamics of the receiver oscillator depend on the coupling strength and the receiver becomes more synchronized with the driver as coupling strength increases. From Figure 2b, we see the changes in distribution of ApEn values for different e, showing that univariate ApEn can capture the change in dynamics (or regularity). Similarly, Figure 2c shows the changes in distribution of XApEn values for different e, indicating that as the two oscillators become more synchronized, the cross approximate entropy value decreases, thereby capturing the amount of asynchrony between two time series data. The dynamics in human actions can be considered as analogous to the dynamics of such coupled systems in that different coupling strength between body joints corresponds to different actions, and we believe that the proposed feature can be used to model dynamics.

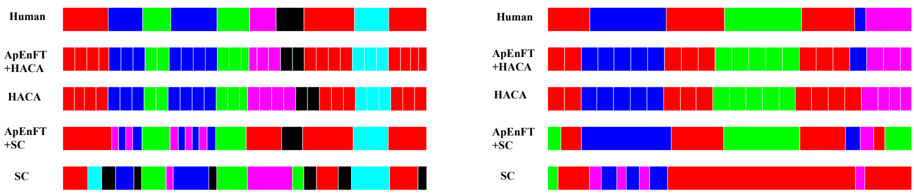
4.2 Temporal Segmentation

In this experiment, we use the publicly available Carnegie Mellon University motion capture database [1]. As in [30], we use the data collected from subject 86 with 14 markers placed on the most informative body joints with the motion capture system recording at 120 Hz. The dataset is a collection of 14 action sequences, each sequence containing multiple natural actions such as walking, punching, drinking, running. The main idea in [29, 30] is that such natural actions are inherently periodic, and this periodicity can be observed in the recurrence matrix showing block structures. Clustering methods such as spectral clustering can be used to cluster (segment) these blocks to achieve temporal segmentation of actions, and hence the clustering accuracy will greatly depend on the *quality* of the recurrence matrix. In this work, we demonstrate that quantifying regularity in actions using approximate entropy-based features can be used to improve the quality of recurrence matrix. We calculate the approximate entropy features as explained in section 3 over a sliding window and the estimated feature values are indexed to the center of the sliding window. The recurrence



(a) (b) (c) (d)

Figure 3: Illustration of utility of approximate entropy feature for quantifying regularity and improving quality of recurrence matrix. (a) shows exemplar time series data collected from hip joint of a subject performing *DANCE, JUMP* and *RUN* actions, (c) shows the corresponding ApEn feature values, (b) and (d) respectively show the recurrence matrix estimated on raw time series data in (a) and ApEn feature values in (c).



(a) (b)

Figure 4: Comparison of temporal clustering methods on the CMU motion capture dataset. Different colors indicate different actions. Ground truth motion segmentation was provided by human observers.

matrix is now calculated on the approximate entropy feature values instead of the time series data collected from the mo-cap system. Figure 3 shows an illustration of our proposed idea using one-dimensional time series data, where we clearly see that the recurrence matrix in (d) calculated from approximate entropy feature values looks more suitable to segment the three actions than recurrence matrix in (b) calculated directly from mo-cap raw time series data. We follow the evaluation protocol as in [30] using the Hungarian algorithm to find the optimum cluster correspondence and to compute clustering accuracy [7]. We compute the confusion matrix between the segmentation provided by the algorithm and the ground truth such that each entry C_{c_1,c_2} in the confusion matrix represents the total number of frames that belong to the cluster segment c_1 that are shared by the cluster segment c_2 in the ground truth. The accuracy is then given by the equation

$$accuracy = \mathbf{max} \frac{tr(C\mathbf{P})}{tr(C1_{k\times k})} \tag{4}$$

where $\mathbf{P} \in \{0,1\}^{k\times k}$ is a permutation matrix.

Figure 4 shows exemplar segmentation results obtained using the approximate entropy-based features along with Spectral Clustering (SC) and HACA on two action sequences. Different colors mark different actions and the ground truth segmentation was obtained from human observers. In both these examples we see that using approximate entropy features provides better segmentation than just using SC or HACA on mo-cap time series data. Due to space constraints, we only show the segmentation results on two sequences. We report the average segmentation accuracy using various features in Table 1, which further supports our claim that using the proposed approximate entropy-based features along with a clustering approach will provide better segmentation accuracy compared to using a clustering approach on mo-cap time series data.

Method	Avg. Accuracy
ApEnFT+HACA	**0.93**
HACA	0.91
ApEnFT+SC	**0.86**
SC	0.75

Baseline	Avg. Accuracy
UniAp+HACA	0.67
UniAp+SC	0.56
Dynamics+HACA	0.65
Dynamics+SC	0.63

Table 1: Comparison of average temporal segmentation accuracy for various methods.

Method	STIP	Hierarchical	Pose+DFT	Pose+DCT	UniAp	Dynamics	Proposed
SVR	0.07	0.19	0.27	0.41	0.05	0.17	**0.45**

Table 2: Mean rank correlation for various methods. Our proposed feature achieves 10% improvement in the correlation coefficient compared to the state-of-the-art. [19] reported correlation coefficient using STIP, hierarchical and pose+DCT features.

4.3 Action Quality Assessment

In the next experiment, we show that the proposed feature can also be used to quantify the quality of diving actions. For this experiment, we use the diving dataset released by Pirsiavash *et al.* [19] which is a collection of videos downloaded from YouTube. The diving dataset consists of 159 videos of diving actions performed by multiple subjects with their respective quality scores given by expert judges. The dataset also provides estimated pose for each frame of the video which is used as input to our framework. The problem of quantifying the quality of diving actions on this dataset is shown to be challenging by the experimental analysis done by Pirsiavash *et al.* in [19], where the best performance achieved was of mean rank correlation of 0.41 between predicted scores and ground truth scores given by judges. We use the same evaluation protocol of generating random training and testing example splits 200 times as introduced in [19] with 100 instances as training examples and the rest as testing examples. Using the estimated pose for each frame, we calculate the approximate entropy features as explained in section 3 for different values of radius ($r = 0.1, 0.12, 0.14, 0.18$) and concatenate to get a high-dimensional feature vector. Using PCA to achieve dimensionality reduction and an SVM regressor to generate real-valued scores indicative of the quality of diving actions, we show that our approximate entropy-based feature performs better than the traditional DCT-based feature. We believe that this is achieved due to the fact that our feature encodes the dynamical information in the time series of poses while DCT does not. In addition, traditional approaches consider each joint independently, while the proposed framework incorporates the interdependency between the joints. The results are tabulated in Table 2 and we achieve a rank correlation of 0.45 in comparison with 0.41 reported in [19].

5 Conclusion

In this paper, we propose the use of an approximate entropy-based feature representation to quantify dynamical regularity in time series of action data for applications in (a) temporal segmentation of actions and (b) quantification of quality of diving actions. The novelty in the proposed feature is in the use of the multivariate embedding approach for approximate entropy to model dynamics in individual body joints and cross approximate entropy to model interaction between body joints. Using nonlinear dynamical models such as the coupled Rossler system, we showed that the proposed feature is sensitive to changes in coupling factor, analogous to interactions between body joints in different actions. Extensive experimental evaluation was presented on two publicly available databases showing better results than the state-of-the-art and the traditional approaches used as baseline measures.

6 Acknowledgments

This work was supported in part by National Science Foundation (NSF) CAREER grant 1452163 and NSF 1320267.

References

[1] Carnegie mellon university motion capture database. http://mocap.cs.cmu.edu. 2012.

[2] Henry Abarbanel. Analysis of observed chaotic data. springer-verlag. 1996.

[3] JK Aggarwal and Michael S Ryoo. Human activity analysis: A review. *ACM Computing Surveys (CSUR)*, 43(3):16, 2011.

[4] Saad Ali, Arslan Basharat, and Mubarak Shah. Chaotic invariants for human action recognition. In *International Conference on Computer Vision*, pages 1–8. IEEE, 2007.

[5] Matthew Brand, Nuria Oliver, and Alex Pentland. Coupled hidden markov models for complex action recognition. In *Conference on Computer Vision and Pattern Recognition*, pages 994–999. IEEE, 1997.

[6] Christoph Bregler. Learning and recognizing human dynamics in video sequences. In *Conference on Computer Vision and Pattern Recognition*, pages 568–574. IEEE, 1997.

[7] Rainer E Burkard, Mauro Dell'Amico, and Silvano Martello. *Assignment Problems, Revised Reprint*. Siam, 2009.

[8] Liangyue Cao, Alistair Mees, and Kevin Judd. Dynamics from multivariate time series. *Physica D: Nonlinear Phenomena*, 121(1):75–88, 1998.

[9] Minh Hoai, Zhen-Zhong Lan, and Fernando De la Torre. Joint segmentation and classification of human actions in video. In *Conference on Computer Vision and Pattern Recognition*, pages 3265–3272. IEEE, 2011.

[10] Gunnar Johansson. Visual perception of biological motion and a model for its analysis. *Perception & psychophysics*, 14(2):201–211, 1973.

[11] Justin J Kavanagh. Lower trunk motion and speed-dependence during walking. *Journal of neuroengineering and rehabilitation*, 6(1):9, 2009.

[12] ChunMei Lu and Nicola J Ferrier. Repetitive motion analysis: segmentation and event classification. *Transactions on Pattern Analysis and Machine Intelligence*, 26(2):258–263, 2004.

[13] Sang Min Oh, James M Rehg, Tucker Balch, and Frank Dellaert. Learning and inferring motion patterns using parametric segmental switching linear dynamic systems. *International Journal of Computer Vision*, 77(1-3):103–124, 2008.

[14] Matej Perše, Matej Kristan, Janez Perš, and Stanislav Kovačič. *Automatic Evaluation of Organized Basketball Activity using Bayesian Networks*. Citeseer, 2007.

[15] Steve Pincus and Rudolf E Kalman. Not all (possibly) "random" sequences are created equal. *Proceedings of the National Academy of Sciences*, 94(8):3513–3518, 1997.

[16] Steve Pincus and Burton H Singer. Randomness and degrees of irregularity. *Proceedings of the National Academy of Sciences*, 93(5):2083–2088, 1996.

[17] Steven M Pincus. Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences*, 88(6):2297–2301, 1991.

[18] Steven M Pincus. Irregularity and asynchrony in biologic network signals. *Methods in enzymology*, 321:149–182, 2000.

[19] Hamed Pirsiavash, Carl Vondrick, and Antonio Torralba. Assessing the quality of actions. In *European Conference on Computer Vision*, pages 556–571. Springer, 2014.

[20] Claude Elwood Shannon. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(1):3–55, 2001.

[21] Nitesh Shroff, Pavan Turaga, and Rama Chellappa. Moving vistas: Exploiting motion for describing scenes. In *Conference on Computer Vision and Pattern Recognition*, pages 1911–1918. IEEE, 2010.

[22] Ekaterina H Spriggs, Fernando De La Torre, and Martial Hebert. Temporal segmentation and activity classification from first-person sensing. In *Conference On Computer Vision and Pattern Recognition Workshops*, pages 17–24. IEEE, 2009.

[23] Pavan Turaga, Ashok Veeraraghavan, and Rama Chellappa. Unsupervised view and rate invariant clustering of video sequences. *Computer Vision and Image Understanding*, 113(3):353–371, 2009.

[24] Pavan K Turaga, Ashok Veeraraghavan, and Rama Chellappa. From videos to verbs: Mining videos for activities using a cascade of dynamical systems. In *Conference On Computer Vision and Pattern Recognition*, 2007.

[25] Raviteja Vemulapalli, Felipe Arrate, and Rama Chellappa. Human action recognition by representing 3d skeletons as points in a lie group. In *Conference on Computer Vision and Pattern Recognition*, pages 588–595. IEEE, 2014.

[26] Vinay Venkataraman, Pavan Turaga, Nicole Lehrer, Michael Baran, Thanassis Rikakis, and Steven L Wolf. Attractor-shape for dynamical analysis of human movement: Applications in stroke rehabilitation and action recognition. In *Conference on Computer Vision and Pattern Recognition Workshops*, pages 514–520. IEEE, 2013.

[27] I Vlachos and D Kugiumtzis. State space reconstruction from multiple time series. In *Topics on Chaotic Systems: Selected Papers from Chaos 2008 International Conference*, page 378. World Scientific, 2009.

[28] A Wolf. Quantifying chaos with lyapunov exponents. *Chaos*, pages 273–290, 1986.

[29] Feng Zhou, F Torre, and Jessica K Hodgins. Aligned cluster analysis for temporal segmentation of human motion. In *Conference on Automatic Face & Gesture Recognition*, pages 1–7. IEEE, 2008.

[30] Feng Zhou, Fernando De la Torre, and Jessica K Hodgins. Hierarchical aligned cluster analysis for temporal clustering of human motion. *Transactions on Pattern Analysis and Machine Intelligence*, 35(3):582–596, 2013.