CORRELATIVE PROGRESSION FEATURES AND STOCHASTIC DEEP NEURAL NETWORK FOR HUMAN ACTIVITY RECOGNITION

BagavathiLakshmi[1], S.Parthasarathy[2,*]

[1]Research and Development Centre, Bharathiar University, Coimbatore, India

[2]Department of Computer Applications, Thiagarajar College of Engineering, Madurai India

Abstract: ***Human activities are encoded in a sequence for continuous real time activity classification. However, typical convolutional neural network perform recognition tasks without considering discriminative features between input data samples. Recurrent Neural Networks (RNN) addressed this issue by deep learning model throughcapturing temporal patterns from input data. However, with the existence of spatiopattern, the RNN restricts the captured range of patterns between data samples. To address these issues related to the discriminative feature extraction and recognizing human activity related to spatiotemporal patterns, in this paper, the application of Stochastic Deep Neural Network is illustrated for building human activity recognition models. Such models are expected to capture the discriminative features with spatiotemporal patterns. End-to-end mapping from regions of interest obtained from orthogonal intersection algorithm is being used in this research work for human activity recognitionsdepending on stochastic approximation and its effectiveness on public dataset was also evalued. It shows the implementation required for low complexity and accuracy of human activity recognition time.***

Keywords: Human Activity Recognition, Convolutional Neural Network, Recurrent Neural Network, Correlative Progression.

Introduction

Human Activity recognition has been an influential research domain in the last decagon due to its appropriateness in several domains, to name a few, overwhelming requirement for home mechanization, benefit services for the elderly and so on. Among them, human activity recognition with the application of pervasive sensors, has received a lot of awareness in the field of accelerated living technologies, for improving quality of life of elderly inside the home environment.

Convolutional Neural Network architecture (CNN architecture) A Ignatov (2018) investigated a user-independent deep learning-based approach for classifying human activity. To achieve this objective, local feature extraction was performed using Convolutional Neural Networks with the aid of simple and precise statistical features. The advantage of using the precise statistical features was preservation of information pertaining to global form of time series. Besides, impact of time series was also analyzedwith respect to recognition accuracy that in turn had a positive influence on real-time activity classification. Despite achieving classification accuracy with minimum running time, compromising discriminative feature resulted in sacrificing the accuracy of human activity recognition. Also with the application of CNN, the architecture was found to be computationally expensive (i.e. involves higher memory consumption). To address this issue, in this work extraction of discriminative features is performed by applying correlative information based on information theory.

Yet another, Long Short Term Memory (LSTM) Recurrent Neural Network S Deepik et al., (2017) classified human activities without the aid of prior knowledge. The classification of certain human activities like, cooking, bathing and sleeping was performed using LSTM. With the main component of LSTM being memory block, information sharing between each block was carried out in an efficient manner. This was performed with the aid of gate and cell state assisted by multiplicative equations. This in turn avoided the issue of denoting uncertainty in deep learning without sacrificing either computational complexity or the accuracy of data being classified. Despite reducing computational complexity and classification accuracy, only temporal patterns were analyzed. However, with several spatial patterns being in existence, with increase in computational complexity for measuring spatial patterns, human activity recognition accuracy was also said to be compromised. To address this aspect, in this work, the investigation of Stochastic Deep Neural Network with training of regions of interest acquired through spatiotemporal patterns is performed.

Based on this work, the combination of discriminative feature extraction and regions of interest through spatiotemporal patterns to recognize human activity is explored. Based on above said works, three issues are explored for activity recognition. First, a Correlative Progression Feature Extraction is designed to extract discriminative features. Second, the mechanism of intersection angles from multiple body intersections to reduce the human activity recognition time is presented. Third, the discriminative features with spatiotemporal patterns are combined with Stochastic Deep Neural Network to improve the accuracy and robustness of activity recognition. The contributions of our work are summarized as follows:

A CPF-SDNN method is proposed that recognize human activity by using KARD and MSR Action3D dataset to evaluate the performance of existing activity recognition systems. Orthogonal Intersection algorithm and Stochastic Deep Neural Network HAR algorithms are employed to verify the effectiveness of the CPF-SDNN method.

Discriminative features are extracted based on the information theory using the correlative information between the preceding and succeeding sequences. Moreover, Correlative Feature Extraction algorithm is leveraged to extract discriminative features according to progression and reduce the computational complexity of activity recognition.

Intersection angles are then evolved based on spatiotemporal feature matrix, to reduce the time taken for activity recognition.

Stochastic Deep Neural Network is implemented to sense the activity and solve the limitations of discriminative feature extraction and spatiotemporal patterns.

The rest of this paper is organized as follows. Related previous research works are presented in Section 2. Section 3 describes the proposed Correlative Progression Features and Stochastic Deep Neural Network for Human Activity Recognition (CPF-SDNN). Section 4 presents our experimental setup and our evaluation of the performance of the proposed approach. Section 5 presents our findings and discussions on the obtained results, followed by the conclusion in Section 6.

Related previous research

One of the imperative research problems in social life, pervasive computing and monitoring fields is human activity recognition. Subsequence Time Series (STS) clustering was analyzed in SuraRodpongpunet al., (2012) to identifying the clusters of interesting sub-sequences involving time series data. A mechanism of subcarrier selection to reach robustness of human activity recognition was analyzed in LinlinGuo et al., (2018) However, by only combining simple features, recognition of human activities remained complicated task. To provide solutions related to this issue, a Sequence of the Most Informative Joints (SMIJ) FerdaOfli et al., (2013) provided an insight into discriminative human activity recognition.

Though fourth industrial revolution is hitherto in advancement and improvements have been made in automating factories, absolutely automated facilities are still not addressed. In G Rene et al.,(2017) Convolutional Neural Network that employed temporal patterns was designed resulting in the improvement in human activity recognition. But, owing to the complexity of human actions, efficient action recognition still remains unaddressed. In L C Diogo et al.,(2017) skeleton sequences were used as the basis for human action recognition. For segmented images and analyzing videos, a highly accurate semi automatic method for surveillance videos was presented in Yi Wang et al., (2017) Yet another multimodal detection system was presented in A Alireza et al., (2017) using Bounding Box (BB) detection fused by an Artificial Neural Network (ANN).

The significant research problem of Human Motion Analysis (HMA) is action recognition. In recent years, action recognition based on 3D model has been receiving increasing interest with the modern emergence of cost-effective sensors. In Sheng Li et al., (2018) early recognition of 3D human actions was analyzed. In H Amir et al.,(2013) multiple space specific representations were made using decoupled fusion of multiple representations resulting in the improvement of human classification accuracy. Yet another novel approach for gesture recognition using space-time descriptors was presented in H Amir et al., (2013) to ensure better recognition accuracy.

A Human activity recognition method is used to identify of human activities in a video, such as a person is walking, running, jumping, jogging etc are important activities in video surveillance was initiated in G P Jay, S Nishant , D Pushkar , S B Vijay and D R Shiv et al (2013)

A human action recognition method was designed in A A Chaaraouli et al. (2013), where pose representation was presented on the basis of contour points of the human silhouette and actions were further learned using sequences of multi-view key poses. Human activity recognition in real video data was analyzed in J Manuel et al., (2013) Nowadays, You Tube are accompanied by both textual and video information. To exploit both video and textual information, for recognizing human activity, correlations between words and activities were analyzed in C Sunyoung et al., (2012)

In contemporary years, human activity recognition from body sensor data become a substantial research from both academia and health industry. In H M Mohammad et al., (2018) a Deep Belief Network (DBN) model was investigated for human activity recognition. A method for recognizing based on relevant joints of human body was designed in G Salvatore et al.,(2015)that in turn recognized the activities in real time. In Fabio A. Storm et al., (2015) seven different physical activities were monitored using angular velocity signals. To reliably monitor day to day activities, real time location system was analyzed in J Gaby et al., (2017)

The analysis of vision-based human activities in videos is an area with increasingly important consequences from security and surveillance to public place and personal archiving. Several challenges at various levels of processing-robustness against errors in low-level processing, view and rate-invariant representations at mid-level processing, and semantic representation of human activities at higher-level processing make this problem hard to solve was analyzed in S K Alok et al.,(2018)

From the above observations on the related previous research, two important facts are clearly visible. First, both spatial and temporal patterns are highly significant for human activity recognition. Second, discriminative features are required so that irrelevant or insignificant features can be discarded. In our work, we demonstrate that discriminative features when extracted and combined with spatiotemporal patterns yields higher recognition accuracy with less complexity.

Proposed work

In this paper, we propose a new method named as "Correlative Progression Features and Stochastic Deep Neural Network (CPF-SDNN) for Human Activity Recognition". This is aimed at reducing the memory consumption and further improving the accuracy during human activity recognition. This is performed by modifying the existing statistical feature extraction approach A Ignatov(2018) with correlative progression feature extraction. The proposed Correlative Progression Feature Extraction algorithm provides various advantages such as extracting high discriminative features, reduces feature extraction time and memory consumption for human activity recognition. Figure 1 shows the block diagram of Correlative Progression Features and Stochastic Deep Neural Network (CPF-SDNN) for Human Activity Recognition.

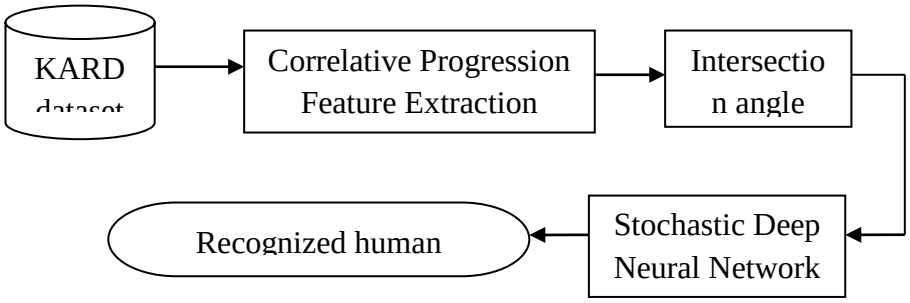


Figure 1 Block diagram of CPF-SDNN

Followed by the correlative and discriminative features being extracted, orthogonal intersection angles are utilized from various body intersections transposing in time that are denoted as a spatiotemporal feature matrix. Here, hip center intersection is employed for calculating intersection angles and to recognize the human activity at a faster rate.

Finally, the actual human activity recognition is performed with the spatiotemporal feature matrix by applying the Stochastic Deep Neural Network model. This proposed CPF-SDNN method implemented using MATLAB simulation and proved that the proposed algorithm provides better performance when it is compared with the other existing methods.

Correlative Progression Feature Extraction

Convolutional Neural Network architecture – CNN architecture A Ignatov(2018) with simple statistical features used convolutional neural networks for local feature extraction that in turn preserved the information about the global form of time series. However, discriminative or selective features are deliberately required that are extracted from videos or a small sequence of poses. In this work, to start with, Correlative Progression Feature Extraction model is performed based on the information theory. A key measure in information theory is entropy. In the proposed work, correlative progression is used as a key measure that quantifies the uncertainty involved between progressions. Figure 2 shows the block diagram of Correlative Progression Feature Extraction.

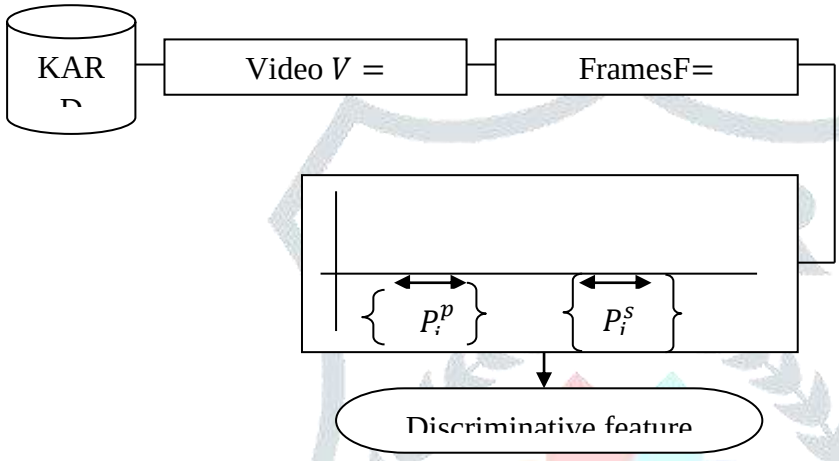


Figure 2 Block diagram of Correlative Progression Feature Extraction

As illustrated in the figure, formally, given an input dataset 'D', Kinect Activity Recognition Dataset (KARD) S Gaglio et al., (2012), comprising several videos '$V = v_1, v_2, \dots, v_n$', split into frames '$F = f_1, f_2, \dots, f_n$', the task of Correlative Progression Feature Extraction (CPFE) lies in extracting the discriminative features based on the information theory. With this base objective, a time-series factor is obtained to evaluate the significance of each progression. Given a progression of a time-series '$P = p_{i+1}, p_{i+2}, \dots, p_{i+m}$', the complexity is formed as the correlative information between the preceding and succeeding sequences, and is mathematically formulated as given below.

$$C(P_i) = I\left(P_i^p, P_i^s\right) \quad (1)$$

$$P_i^p = p_i, p_{i+1}, \dots\dots, p_{i+\frac{n}{2}-1} \quad (2)$$

$$P_i^s = p_{i+\frac{n}{2}}, \dots\dots, p_{i+n-1} \quad (3)$$

From the above equation (2) and (3), the correlative information between the preceding progression 'P_i^p' and succeeding progression 'P_i^s' are evolved with progression involving different time series. Based on the complexity measure as provided in equation (1), significant progressions are further used for representing intersection angles. Algorithm 1 shows the proposed Correlative Feature Extraction. Here, the threshold, 'τ' is dependent on the dataset.

Input: Dataset 'D', videos '$V = v_1, v_2, \dots, v_n$', frames '$F = f_1, f_2, \dots, f_n$', Threshold 'τ'
Output: Features extracted 'R'
1: **Begin** 2: **For** each Dataset 'D' with videos '$V = v_1, v_2, \dots, v_n$' 3: **For** each frame 'F' 4: Measure correlative information between the preceding and succeeding using equation (1) 5: **If**'$C(P_i) > \tau$' **then** 6: Measure '$R \rightarrow R \cup C(P_i)$' 7: **Else** 8: Discard $C(P_i)$ 9: **End if** 10: **End for** 11: **End for** 12: **End**

Algorithm 1 Correlative Feature Extraction algorithm

By assessing the benefits of progressions and by discriminatorily using these progressions, the Correlative Feature Extraction algorithm yield more relevant results. A meaningful progression requires a significant amount of information. On the other hand, a meaningless progression requires less information. To address this issue and obtain meaningful and discriminative information, complexity measure is evolved. The complexity measure evaluates the meaningfulness of a progression, where the complexity in the proposed work is viewed as a measure of the amount of discriminative features or information in a set of data. With this discriminative features or information, complexity involved in human activity recognition is said to be improved.

Orthogonal Intersection Angle

With the extracted features, the proposed work utilizes intersection angles from multiple body intersections with respect to time represented in the form of spatiotemporal feature matrix.Figure 3 given below shows the sample representation of intersection angle with the extracted features.

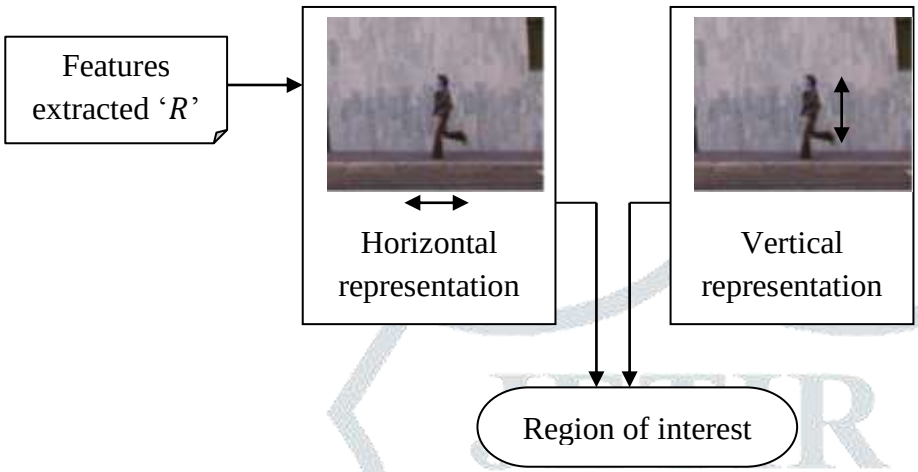


Figure 3 Representation of Intersection angle

Let us consider the horizontal vector 'α' to denote the vector from left to right of the hip center and the vertical vector 'θ' as the vector that is perpendicular to the horizontal vector. The horizontal vector 'α′' is then mathematically formulated as given below.

$$\alpha' = \begin{bmatrix} HC_a \cup & LH_a \cup & RH_a \\ HC_b \cup & LH_b \cup & RH_b \\ HC_c \cup & LH_c \cup & RH_c \end{bmatrix} \quad (4)$$

From the above equation (4), the horizontal vector represents the intersection position of the hip center 'HC', left hip 'LH' and right hip 'RH' with 'a', 'b' and 'c' forming three dimensions respectively. Followed by the horizontal vector representation, the vertical vector 'θ′' representation that forms the vector perpendicular to horizontal vector is mathematically formulated as given below.

$$\theta' = a \cup b \cup c \quad (5)$$

$$a = a_\alpha \alpha' + b_\beta \beta' + c_\gamma \gamma' \quad (6)$$

$$b = a_\alpha \alpha' + b_\beta \beta' + c_\gamma \gamma' \quad (7)$$

$$c = a_\alpha \alpha' + b_\beta \beta' + c_\gamma \gamma' \quad (8)$$

From the above equation (6), (7) and (8), 'α', 'β' and 'γ' denotes the three dimension form and 'α′', 'β′' and 'γ′' representing the corresponding vertical vectors. Unlike classical approaches that use Long Short-Term Memory classifier Deepika Singh et al(2017) that focus on the temporal sequences, the proposed method uses orthogonal intersection angles that form spatiotemporal matrix. This spatiotemporal matrix is represented via hip center of the horizontal and vertical vector. The pseudo code representation of Orthogonal Intersection is given below.

Input:Dataset 'D', videos '$V = v_1, v_2, \ldots, v_n$', frames '$F = f_1, f_2, \ldots, f_n$', Features extracted 'R', hip center positions 'HC_a, HC_b, HC_c', left hip positions 'LH_a, LH_b, LH_c', right hip positions 'RH_a, RH_b, RH_c', postures 'k'

Output: Region of interest '$ROI = roi_1, roi_2, \ldots, roi_n$'

```
1: Begin
2:   For each Dataset 'D' with videos 'V = v1, v2, ..., vn'
3:     For each frame 'F'
4:       For each features extracted 'R'
5:         Measure horizontal vector 'α′' using equation (4)
6:         Measure vertical vector 'θ′' using equation (5)
7:       End for
8:     End for
9:   End for
10: End
```

Algorithm 2 Orthogonal Intersection algorithm

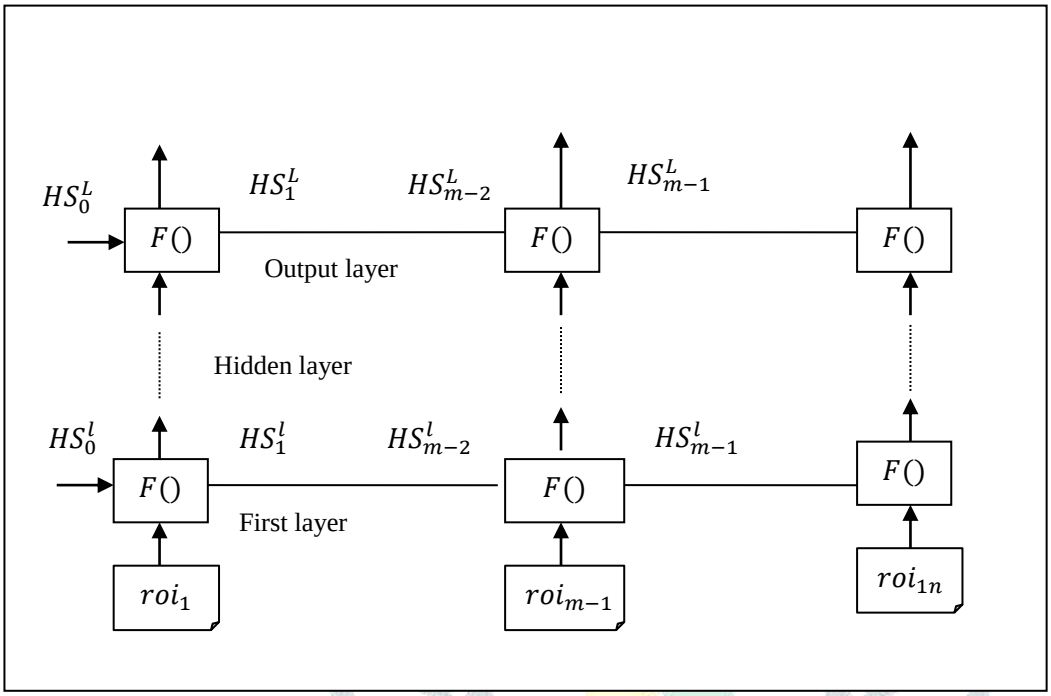
As given in the above Orthogonal Intersection algorithm, the selection of intersection angle features is approximated through spatiotemporal feature matrix into ‘k’ postures. Then encoding spatiotemporal intersection angle features is performed to generate Orthogonal Intersection for each video ‘V’ constituting of ‘n’ frames. Each individual human action is then recognized using deep neural network, discussed elaborately in the following section.

Stochastic Deep Neural Network

Finally,Stochastic Deep Neural Network(SDNN) proposed to adopt the spatiotemporal dynamics of human activity recognition is proposed in order to capture the spatiotemporal dynamics of activity sequences. In contrast to S Deepika et al., (2017) where Recurrent Neural Networks (RNN) was employed for activity recognition, the CPF-SDNN method is based on Stochastic Deep Neural Network (SDNN) that focuses on minimizing the variance on human activity predictions. The schematic diagram of the human activity recognition using Stochastic Deep Neural Network model is illustrated in figure 4.

The Stochastic Deep Neural Network HAR performs direct end-to-end mapping from regions of interest obtained from orthogonal intersection algorithm to human activity recognitions. It recognizes the activity performed during a specific time interval.

As illustrated in the figure, the input is a discrete sequence of regions of interest ‘$\mathrm{ROI} = \{\mathrm{roi}_1, \mathrm{roi}_2, \ldots, \mathrm{roi}_n\}$’, where each region of interest denotes a vector of individual samples observed at different time interval ‘t’ respectively.



These samples are then fed to a Stochastic Deep Neural Network. The method outputs a sequence of activities. Here, stochastic approximation refers to the minimization of an objective function. The minimization function is as given below.

$$O(w) = \frac{1}{n}\sum_{i=1}^{n} F\,[O^i(w)] \quad (9)$$

From the above equation (9), where the parameter ‘w’ that minimizes ‘$O(w)$’ is estimated. Initially, the hidden state ‘HS_0^l’ and the internal state ‘IS_0^l’ of every layer are set to zeros. Followed by this, the first layer uses the input sample ‘roi_t’ at time ‘t’, preceding hidden state ‘HS_{t-1}^1’, and preceding internal hidden state ‘IS_{t-1}^1’ to obtain the output of the first layer ‘O_t^1’ and is mathematically formulated as given below.

$$O_t^1 HS_t^1 IS_t^1 = F[IS_{t-1}^1, HS_{t-1}^1, roi_t] \quad (10)$$

In a similar manner, the output of the ‘nth’ layer is mathematically formulated as given below.

$$O_t^n HS_t^n IS_t^n = F[IS_{t-1}^n, HS_{t-1}^n, roi_t] \quad (11)$$

With the obtained output from the above equation (11), the output of the ‘nth’ layer identifies or recognized the activity. The pseudo code representation of human activity recognition using Stochastic Deep Neural Network model is given below.

Input: Region of interest ‘$\mathrm{ROI} = roi_1, roi_2, \ldots, roi_n$’, Layer ‘$L = l_1, l_2, \ldots, l_n$’, Hidden State ‘HS’, Internal State ‘IS’
Output: Human activity recognition
1: **Begin** 2: **For** each Region of interest ‘ROI 3: **For** each layer ‘L’ with Hidden State ‘HS’ and Internal State ‘IS’ 4: Set ‘HS_0^l’ and ‘IS_0^l’ to zero

5:	Measure the output of the first layer 'O_t^1' using equation (10)
6:	**End for**
7:	**End for**
8:	**End**

Algorithm 3 Stochastic Deep Neural Network HAR algorithm

As given above, the Stochastic Deep Neural Network HAR algorithm, with each hidden state and internal state as input, the objective lies in recognizing the human activity that minimizes the stochastic approximation. Once, by setting the internal state and hidden state to zero, the output of the first layer is obtained at time 't'. Followed by this the output forms the input to the second internal state and hidden state. Once each neural network has been trained on the sequences of several postures, a new sequence is tested with respect to the set of human activity on the basis of the minimization of an objective function. As a result, the HAR recognition accuracy is said to be improved.

Experimental Setup

In this section we compare our proposed human activity recognition method, Correlative Progression Features and Stochastic Deep Neural Network (CPF-SDNN) (described in Section 3) using JAVA platform against the baseline human activity recognition methods, Convolutional Neural Network architecture (CNN architecture) A Ignatov (2018) and Long Short Term Memory (LSTM) Recurrent Neural Network Singh et al. 2017. section 4.1 test the outlined features of dataset.section 4.2 provides the experimental results.

4.1 Dataset

We evaluate the performance of each feature representation described above on two different human action datasets. One of the dataset, KARD dataset S Gaglio et al., (2014) comprises 18 activities, where each activity performed 3 times by 10 different subjects. Altogether, the dataset have 4 (files) x 18 (activities) x 3 (repetitions) x 10 (subjects), that is 2160 files. Each file contains 15xF lines. Here, F represents the number of frames for that sequence. Each line represents three numbers: real world coordinates (x, y, z) for realworld.txt, or screen coordinates and depth value (u, v, depth) for screen.txt.Besides, the dataset included 540 sequences for about a total of 1 hour of videos captured at a resolution of 640x480 pixels at 30fps.

Second, the human activity recognition was also evaluated on the MSR Action3D dataset W Li Zhang et al (2010). The dataset comprises skeleton data extracted from a depth sensor similar to the Microsoft Kinect. For conducting experiments, a subset of 17 actions was selected that were performed by 8 subjects, with 3 repetitions of each action. In total the subset included 379 action sequences, with the duration of the sequences ranging from 14 to 76 frames. Some of the included set of actions were, side kick, forward punch, horizontal arm wave, draw tick, draw x, high throw, side-boxing, jogging, tennis swing, hammer, hand catch, draw circle, hand clap, two hand wave, forward kick, tennis serve and high arm wave.

Each dataset has entirely unique action set with differing frame rates, different skeleton extraction method, and hence, of various dynamic properties. The objective of including such different input data was to investigate how discriminative the proposed CPF-SDNN method is with respect to the different characteristics of these datasets. The diversity is pertinent in the first set of experiments where we aim to evaluate the performance of the human activity recognition accuracy on a wide range of actions using KARD and MSR Action3D dataset. For the second set of experiments, where we evaluate the human activity recognition time acrossdatasets, we select a small subset of actions that are shared between the first two datasets. Finally, for the third set of experiment, computational complexity was evaluated with the aid of KARD dataset.

4.2 Performance measure of human activity recognition accuracy

The first goal of the experiment remains in measuring the human activity recognition accuracy for two datasets, namely, KARD dataset and MSR Action3D dataset. The human activity recognition accuracy measures the ratio of successful sequence recognition to the total sequences provided as input and is mathematically provided as given below.

$$HARA = \sum_{i=1}^{n} \frac{SSR}{S_i} * 100 \quad (12)$$

From the above equation (12), the human activity recognition accuracy '$HARA$' is the ratio of successful sequence recognition 'SSR' to the total number of sequences 'S_i' considered for experimentation. The training sequences for conducting experimentation were 440 (using KARD dataset) and 320 (using MSR Action 3D dataset), whereas 100 testing sequences were considered to measure the human activity recognition accuracy. It is measured in terms of accuracy.

4.2.1. Sample calculation (using KARD dataset)

- **CNN architecture**: With '10' number of sequences provided as input and successful sequence recognition being '8', the human activity recognition accuracy is measured as given below.

$$HARA = \frac{8}{10} * 100 = 80\%$$

- **LSTM Recurrent Neural Network**: With '10' number of sequences provided as input and successful sequence recognition being '7', the human activity recognition accuracy is measured as given below.

$$HARA = \frac{7}{10} * 100 = 70\%$$

- **Proposed CPF-SDNN**: With '10' number of sequences provided as input and successful sequence recognition being '9', the human activity recognition accuracy is measured as given below.

$$HARA = \frac{9}{10} * 100 = 80\%$$

4.2.2. Sample calculation (using MSR Action3D dataset)

- **CNN architecture**: With '10' number of sequences provided as input and successful sequence recognition being '7', the human activity recognition accuracy is measured as given below.

$$HARA = \frac{7}{10} * 100 = 70\%$$

- **LSTM Recurrent Neural Network**: With '10' number of sequences provided as input and successful sequence recognition being '6', the human activity recognition accuracy is measured as given below.

$$HARA = \frac{6}{10} * 100 = 60\%$$

- **Proposed CPF-SDNN**: With '10' number of sequences provided as input and successful sequence recognition being '7', the human activity recognition accuracy is measured as given below.

$$HARA = \frac{8}{10} * 100 = 80\%$$

4.3 Performance measure of human activity recognition time

The second goal of the experiment is the evaluation of human activity recognition time or simply the time taken to recognize the human activity and is measured in terms of milliseconds (ms). It is mathematically formulated as given below.

$$HAR_t = \sum_{i=1}^{n} S_i * Time\ [O(w)] \quad (13)$$

From the above equation (13), the human activity recognition time 'HAR_t', is measured on the basis of the number of sequences provided as input 'S_i' and the time taken to obtain the stochastic approximation function '$Time\ [O(w)]$'. Here, the stochastic approximation function refers to the minimization of an objective function with the objective function remains in human activity recognition.

4.3.1 Sample calculation (using KARD dataset)

- **CNN architecture**: With number of sequences being '10', the time taken for human activity recognition for 1 sequence is '$0.020ms$', then the human activity recognition time for '10' sequences is as given below.

$$HAR_t = 10 * 0.020ms = 0.20ms$$

- **LSTM Recurrent Neural Network**: With number of sequences being '10', the time taken for human activity recognition for 1 sequence is '$0.025ms$', then the human activity recognition time for '10' sequences is as given below.

$$HAR_t = 10 * 0.025ms = 0.25ms$$

Proposed CPF-SDNN: With number of sequences being '10', the time taken for human activity recognition for 1 sequence is '$0.018ms$', then the human activity recognition time for '10' sequences is as given below.

$$HAR_t = 10 * 0.018ms = 0.18ms$$

4.3.2. Sample calculation (using MSR Action3D dataset)

- **CNN architecture**: With number of sequences being '10', the time taken for human activity recognition for 1 sequence is '$0.028ms$', then the human activity recognition time for '10' sequences is as given below.

$$HAR_t = 10 * 0.028ms = 0.28ms$$

- **LSTM Recurrent Neural Network**: With number of sequences being '10', the time taken for human activity recognition for 1 sequence is '$0.038ms$', then the human activity recognition time for '10' sequences is as given below.

$$HAR_t = 10 * 0.038ms = 0.38ms$$

- **Proposed CPF-SDNN**: With number of sequences being '10', the time taken for human activity recognition for 1 sequence is '$0.024ms$', then the human activity recognition time for '10' sequences is as given below.

$$HAR_t = 10 * 0.024ms = 0.24ms$$

4.4 Performance measure of computational complexity

The third goal of the experiment is the evaluation of computational complexity or the memory consumed during human activity recognition. It is measured in terms of kilobytes (KB). It is mathematically formulated as given below.

$$CC = \sum_{i=1}^{n} S_i * Mem\ [\mathrm{O(w)}] \quad (14)$$

From the above equation (14), the computational complexity 'CC' is evaluated based on the memory consumed while deriving the objective function '$Mem\ [\mathrm{O(w)}]$' with respect to the number of sequences 'S_i'.

4.4.1. Sample calculation (using KARD dataset)

- **CNN architecture:** With'10' sequences provided as input and the memory consumed while performing objective function being '$242KB$' for obtaining sequences of activities, and when these values are substituted in the above formula to obtain the computational complexity.

$$CC = 10 * 242KB = 2420KB$$

- **LSTM Recurrent Neural Network:** With'10' sequences provided as input and the memory consumed while performing objective function being '$265KB$' for obtaining sequences of activities, and when these values are substituted in the above formula to obtain the computational complexity.

$$CC = 10 * 265KB = 2650KB$$

- **Proposed CPF-SDNN:** With'10' sequences provided as input and the memory consumed while performing objective function being '$225KB$' for obtaining sequences of activities, and when these values are substituted in the above formula to obtain the computational complexity.

$$CC = 10 * 225KB = 2250KB$$

The elaborate discussion of the above three parameters, human activity recognition accuracy, human activity recognition time and computational complexity is provided in the following section.

5. Discussion

In the first set of experiments, human activity recognition accuracy is measured on each of the aforementioned datasets separately. Specifically, we used 100 sequences (with the presence of 440 sequences from KARD dataset and with the presence of 320 sequences of MSR Action3D dataset). Figure 5 shows the human activity recognition accuracy for these two datasets with 10 different sequences provided as input.

Specifically, the plots in different columns correspond to different human activity recognition methods, i.e., CPF-SDNN, CNN architecture AndreyIgnatov(2018) and LSTM Recurrent Neural Network Deepika Singh et al., (2017) from left to right, respectively. The plots in different rows show recognition results from different datasets, i.e., KARD and MSR Action3D, from top to bottom, respectively. In all plots, the vertical axis represents the human activity recognition accuracy performance and the horizontal axis represents the number of sequences, ranging from 10 to 100 with a step size of 10.

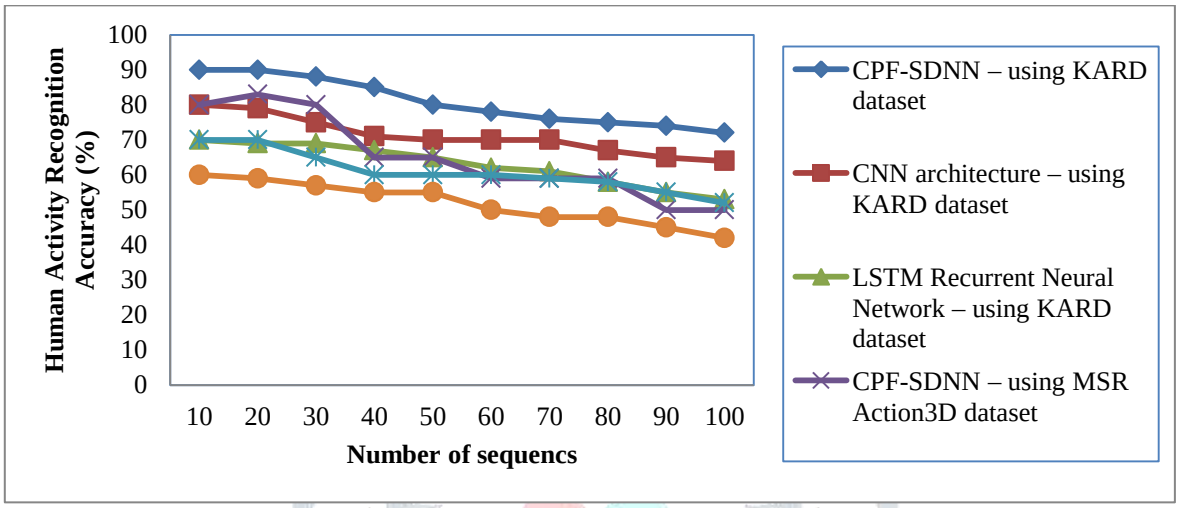


Figure 5 Human Activity Recognition Accuracy results for 10 different sequences

From the figure it is observed that with the increase in the number of sequences, the human activity recognition accuracy first improves and then saturates when the number of sequences is moderately large, for both MSR Action3D dataset and KARD dataset. That is, the human activity recognition accuracy performance in general tends to decrease as we increase the number of sequences but found to be comparatively better than with the MSR Action3D dataset.

These two inspections are homogeneous as increasing the number of sequences corresponds to increasing the time and therefore decreasing the human activity recognition accuracy. However, comparative analysis shows the performance improvement using CPF-SDNN than when compared to CNN architecture A Ignatov (2018) and LSTM Recurrent Neural Network S Deepika et al., (2017).This is because, by applying the Stochastic Deep Neural Network HAR algorithm, human activity recognition is only carried out for the regions of interest based on the orthogonal intersection. With the application of orthogonal intersection, not only multiple features are considered, but also capture both the spatio and temporal patterns amongst the human body parts through hip center. This in turn results in the improvement of human activity recognition accuracy using CPF-SDNN by 14% and 29% compared to CNN architecture A Ignatov (2018) and LSTM Recurrent Neural Network S Deepika et al., (2017), 6% and 24% compared to CNN architecture A Ignatov (2018) and LSTM Recurrent Neural Network S Deepika et al., (2017) (using MSR Action3D dataset) respectively.

The second critical factor that remains to be addressed among different parameters is the human activity recognition time which eventually impacts the recognition performance. Hence, in the second set of experiments, human activity recognition time is measured using KARD and MSR Action3D datasets separately. The same 100 sequences amongst 440 sequences from KARD dataset 320 sequences of MSR Action3D dataset were used for conducting experiments.

Figure 6shows the human activity recognition time for these two datasets with 10 different sequences provided as input. As illustrated in both the figures, with the increase in the number of sequences given as input, feature extraction time gets increased and subsequently, the time taken for extraction of regions of interest. Therefore, the human activity recognition time eventually increases with the provided sequences involving different actions from KARD and MSR Action3D datasets respectively.

As illustrated in the figure, with the increase in the number of sequences, the human activity recognition time also increases. The increase in human activity recognition time is not unique for both the datasets. These distinctions indicates that the underlying sequences with the skeleton joint positions in the two datasets have different levels of sensitivity and noise for different joints that in turn has greater influence on the orthogonal intersection and their spatiotemporal orderings.

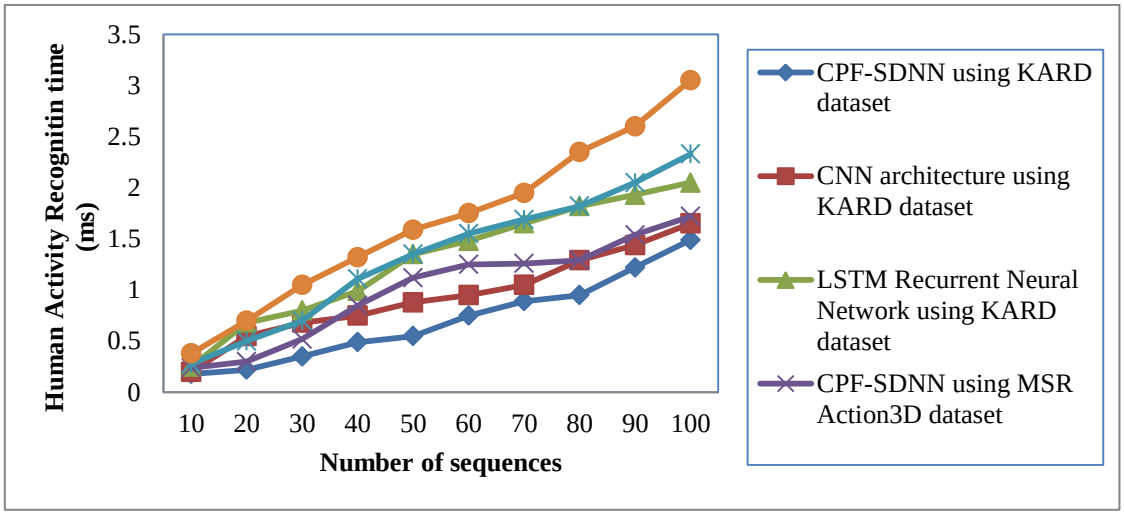


Figure 6 Human Activity Recognition time results for 10 different sequences using KARD and MSR Action3D dataset

Besides the instructions given to the subjects for different action results in activity recognitions variations. This in turn results in differences in the set of the most orthogonal intersection angles or in their spatiotemporal orderings. However, with the both datasets value given as input, the human activity recognition time is found to be reduced by applying the CPF-SDNN method. This is because separate horizontal and vector representation are formed with the hip centre as the orthogonal coordinate for the extracted discriminative features with invariant posture and scale. This in turn minimizes the time taken for human activity recognition using CPF-SDNN method by 28% and 49% compared to CNN architecture A Ignatov (2018) and LSTM Recurrent Neural Network S Deepika et al., (2017), 25% and 40% compared to CNN architecture A Ignatov (2018) and LSTM Recurrent Neural Network S Deepika et al., (2017), (using MSR Action3D dataset) respectively.

Finally, we demonstrate the performance of the aforementioned features in computational complexity setting in which the sequences are trained and tested on KARD dataset. For this purpose, we used the same 440 training and 100 testing subjects from the first set of experiments to model the computational complexity.

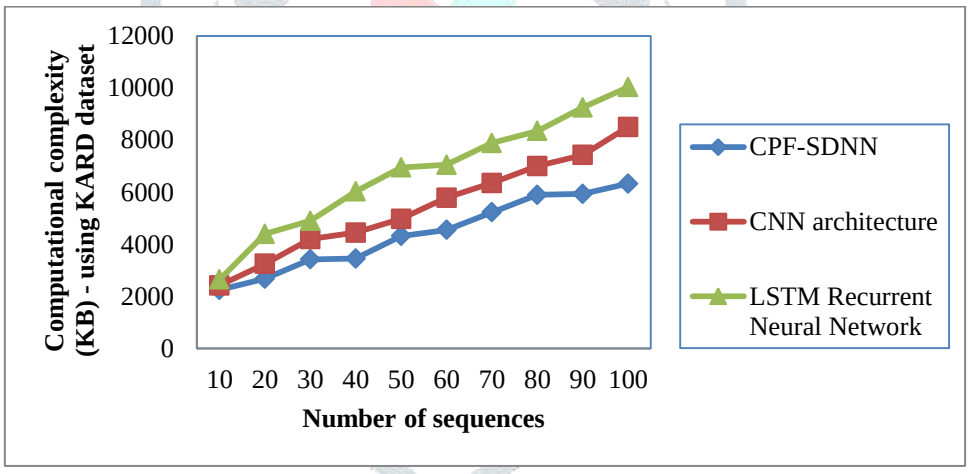


Figure 7 Computational complexities with respect to 100 different sequences

Figure 7 given above illustrates the computational complexities with respect to 100 different sequences by applying KARD dataset. With the increase in the number of sequences, the computational complexity also increases using all the three methods, CPF-SDNN, CNN architecture A Ignatov (2018) and LSTM Recurrent Neural Network S Deepika et al., (2017) respectively. This is because with the increase in the sequence provided as input, correlative information between the preceding and succeeding actions are measured. This in turn increases the complexity involved in feature extraction and therefore the overall computational complexity in human activity recognition. However with the application of Correlative Feature Extraction algorithm, more discriminative features are extracted, where meaningful progression is used and meaningless progression are discarded. This meaningful and meaningless progression is arrived at based on the correlation measure. This when applied to the Orthogonal Intersection algorithm selects the intersection angle features on the basis of spatiotemporal feature matrix. As a result, the computational complexity involved during human activity recognition using CPF-SDNN method is reduced by 18% when compared to CNN architecture A Ignatov (2018) and 34% when compared to LSTM Recurrent Neural Network S Deepika et al., (2017) respectively.

6. Conclusion

In this work, a novel method of human activity recognition using correlative progression and orthogonal intersections, to improve the robustness and accuracy of activity recognition is presented. First, we leverage the moving variance of correlative progression to extract the discriminative features and adopt the discriminative features for further analysis. Moreover, several meaningful progressions are also selected by using Correlative Feature Extraction algorithm to reduce the feature extraction time and computational complexity of human activity recognition. Then, an Orthogonal Intersection Angle is designed on the basis of spatiotemporal feature matrix to recognize activities by leveraging spatiotemporal relationship invariant to posture and scale. Finally, we solve the limitations of spatiotemporal patterns-based

activity recognition using Stochastic Deep Neural Network. Experimental results show that CPF-SDNN method achieves 42% of average recognition accuracy in the KARD dataset.

References

[1] A A Chaaraoui , Perez, C Revuelta, F.F "*Silhouette-based human action recognition using sequences of key poses",* Pattern Recognition Letters, Volume 34, Issue 15, 2013, PP 1799-1807

[2] A Alireza , G Luis , P Cristiano , P Paulo , J N Urbano, "*Multimodal vehicle detection: fusing 3D-LIDAR and color camera data"*, Pattern Recognition Letters, Sep 2017

[3] A Ignatoc, "*Real-time human activity recognition from accelerometer data using Convolutional Neural Networks*", Applied Soft Computing, Volume 62, 2018, PP 915-922.

[4] C Sunyoung, SooyeongKwak, HyeranByun, "*Recognizing human–human interaction activities using visual and textual information",* Pattern Recognition Letters, Nov 2012

[5] Fabio A. Storm, Ben W. Heller, M Claudia, *"Step Detection and Activity Recognition Accuracy of Seven Physical Activity Monitors",* PLOS ONE journal.pone.0118723 March 19, 2015

[6] FerdaOfli, C Rizwan, K Gregorijo, V Rene, RuzenaBajcsy, "*Sequence of the most informative joints (SMIJ): A new representation for human skeletal action recognition",* Journal of Visual Communication and Image Representation, May 2013

[7] G P Jay, S Nishant , D Pushkar , S B Vijay and D R Shiv et al (2013)" Human Activity Recognition Using Gait Pattern" International Journal of Computer Vision and Image Processing ,Volume 3,PP .23

[8] G Rene ,L M Jan, R M Fernando, F Gernot, F Sascha, H Michael *"Deep Neural Network based Human Activity Recognition forthe Order Picking Process"*, ACM, June 2017

[9] G Salvatore, R L Giuseppe,M Marco, "*Human Activity Recognition Process Using 3-D Posture Data",* IEEE Transactions on Human-Machine Systems, VOL 45, NO 5, OCTOBER 2015.

[10] H Amir, Shabani, John S. Zelek, David A. Clausi, (2013) "Multiple scale-specific representations for improved human action Recognition", Pattern Recognition Letters, Elsevier, Volume 34, Issue 15, PP 1771-1779

[11] H M Mohammad, H Shamsu, U MdZa, A Ahmad, A Majed, "*Human Activity Recognition from Body Sensor Data using Deep Learning*", Mobile & Wireless Health Journal of Medical Systems, Volume 42, Issue No 6, Mar 2018

[12] J Gaby, H Jessie de Witts, D Allana, A Robertr, *"The development and validation of a Real Time Location System to reliably monitor everyday activities in natural contexts",* Plos One February 14, 2017

[13] J Manuel, Y Enriques, B P Nicolas, "*Exploring STIP-based models for recognizing human interactions in TV videos*", Pattern Recognition Letters, Volume 34, Issue 15, 1 November 2013, PP. 1819-1828.

[14] L C Diogo, HediTabia, P David, *"Learning features combination for human action recognition from skeleton sequences",* Pattern Recognition Letters, Volume 99, 1 November 2017, Pages 13-20

[15] L Sheng, L Kang, F Yun, "*Early Recognition of 3D Human Actions*", ACM Transactions on Multimedia, Computing, Communications and Applications, Volume 14, Mar 2018

[16] LinlinGuo , W Lei, L Jialin, B Z Wei, "*Human Activity Recognition Using Crowd sourced WiFi Signals and Skeleton Data*", Hindawi Wireless Communications and Mobile Computing, Volume 2018,May 2018,PP. 1-15.

[17] S Deepika, M Erinc, P Isminia, K Johannes, StenHanke, G Matthieu, and H Andreas , *"Human Activity RecognitionUsing Recurrent Neural Networks"*, International Federation for Information Processing, May 2017,Pages 267-274

[18] S. Gaglio, G. Lo Re, M. Morana, "*Human Activity Recognition Process Using 3-D Posture Data",* IEEE Transactions on Human-Machine Systems, May 2014

[19] sk Alok , Kushwaha et al.,(2018) Recognition of Humans and Their Activities for Video Surveillance, Computer Vision: Concepts, Methodologies, Tools, and Applications,PP. 18

[20] SuraRodpongpun, VitNiennattrakul, R Chotirat, *"Selective Subsequence Time Series clustering",* Knowledge-Based Systems, Volume 35, November 2012, PP. 361-368.

[21] UpalMahbub, R T I Hafiz, Md. ShafiurRahman, Md. AtiqurRahmanAhad, "*A template matching approach of one-shot-learning gesture recognition",* Pattern Recognition Letters, Volume 34, Issue 15, 1 November 2013, PP.1780-1788.

[22] W. Li, Z. Zhang, Z. Liu, *"Action recognition based on a bag of 3D points",* Proceedings of Computer Vision and Pattern Recognition Workshops, 2010, PP. 9–14.

[23] Yi Wang, ZhimingLuo, J Pierre-Marc, "*Interactive Deep Learning Method for Segmenting Moving Objects",* Pattern Recognition Letters, Elsevier, Volume 96, 1 September 2017, PP. 66-75.