ISSN 2347 - 3983
Volume 8. No. 7, July 2020
International Journal of Emerging Trends in Engineering Research
Available Online at http://www.warse.org/IJETER/static/pdf/file/ijeter100872020.pdf
https://doi.org/10.30534/ijeter/2020/100872020

CNN-LSTM Hybrid model based human action recognition with skeletal representation using joint movements based energy maps

M. Teja Kiran Kumar[1], P.V.V. Kishore[2], M.V.D.Prasad[3]
[1]Koneru Lakshmaiah Educational Foundation, Guntur, A.P, India, mtejakiran@kluniversity.in
[2]KoneruLakshmaiah Educational Foundation, Guntur, A.P, India, pvvkishore@kluniversity.in
[3]KoneruLakshmaiah Educational Foundation, Guntur, A.P, India, mvd_ece@kluniversity.in

ABSTRACT

In recent years 3D skeleton based human action recognition has become very popular in action classification tasks. Even though action recognition is still less successful due to complexity in actions performed when compared to image recognition. Due to huge success in deep learning concepts for recognizing images, we present a new representation of 3D human sequences into joint movements based energy maps (JME). JME's projects the spatio information embedded in to single image and are inputted to Convolutional Neural Networks (CNN) for classification. Temporal information that is x,y and z position vectors are executed by the Time series LSTM for recognition both spatio CNN and temporal LSTM scores are fused for final classification. The experimentation have done by our own dataset KLHA3D-102. In order to evaluate our proposed algorithm Convolutional Long Short Term (C-LSTM) we considered publicly available action datasets namely HDM05, CMU and NTU RGB-D. The results shown a better performance due to spatio temporal modelling of our proposed hybrid model at both the representation and the recognition stages when compared to other state of the arts.
.

Key words :3D skeleton data, Human action recognition, CNN, LSTM.

1. INTRODUCTION

Human activities in a video sequence frames have appeared to have extensive applications in security observation and surveillance monitoring. Therefore, these applications include esteem, if a specific degree of start to finish mechanization is accomplished. Be that as it may, convolutional neural networks (CNN) has demonstrated an expanding nearness around there of exploration. Already, CNNs performed incredibly well on pictures and fixed length video successions. In any case, for variable length video arrangements they indicated a reduction in execution because of the conventional one versus all expectation models. Here, we propose a start to finish trainable intermittent CNN with Long Short Term Memory(LSTM) system to improve spatio worldly picture acknowledgment task on skeletal human activity acknowledgment..

Human activity acknowledgment has been explored widely in the previous decade utilizing pictures, recordings, depth and skeletal information as sources of info. The preparing algorithms generally utilized sailency identification, human extraction, following kalman and particle filters; displaying the extractions utilizing highlights lastly acknowledgment with models in AI, for example, trees, svm, kernels and HMMs. In any case, the preparing is very serious and requires a great deal of calculations to work simultaneously bringing about exact recognition. Then again this computationally concentrated methodology has offered away to productive deep learning structures that acknowledged the crude video information to settle on choices that are increasingly precise individually.

Convolutional neural systems (CNNs) were generally considered for the filed of image processing, for example, recognition and classification. They quickly broaden their essence into video information Search and recognizable proof applications. Though, video RGB information appears to contain sensor inconsistencies, which turned into an obstruction in the recognition of human activities. Consequently, the scientists turned their concentration towards multi modular information acquired from Kinect sensor which contained RGB video, depth and skeletal information. The utilization of each of the three information has improved the identification with CNNs, RNNs and deep models.
Inevitably, the enormous multi modular information has expanded computational multifaceted nature of the calculations which upscaled equipment necessities. Rather than the past models on human activity acknowledgment with multi modular information, we propose to utilize single modular information to accomplish higher exactnesses. This work presents a CNN put together repetitive system with respect to pictures created from skeletal joint separation changes. The proposed CNN based Long Short Term Memory (LSTM) design can possibly stamp spatial and temporal changes in the spatio temporal pictures.

In the proposed methodology, each action is being recognized based on the proposed JME maps and the joints data. Here JME is given to the CNN's and remaining joints data is given to the LSTMs these two networks train separately and scores will be fused at last for recognizing actions.
Our proposed methodology is divided in to three significant steps.

1. 3d skeleton data is subjected for creating JME's .
2. These JME's and 3D joints data is subjected to proposed Hybrid model and study the performance.
3. Both the CNN and LSTM network scores are fused for recognizing actions.

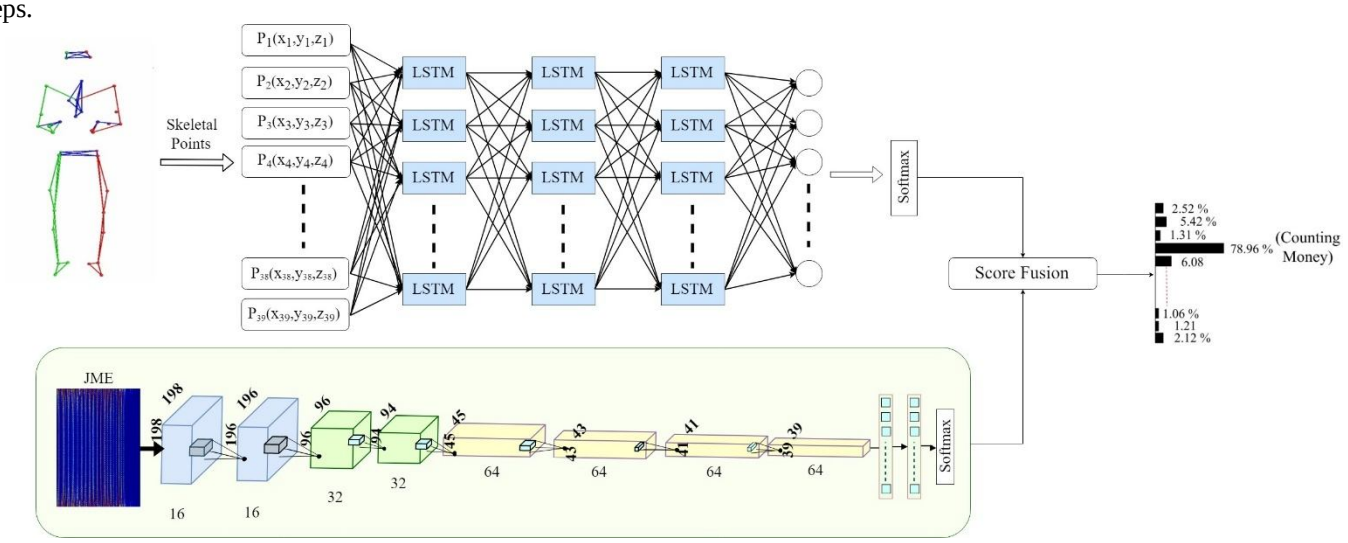

Figure 1: Architecture of proposed CNN-LSTM hybrid architecture

To test the proposed CNN -LSTM hybrid architecture, we use HDM05 [1], CMU [2] and NTU RGB-D [3] human action datasets. The rest of the paper is organized as follows. Section 2 describes the literature review related to the proposed framework. Section 3 gives the information regarding proposed method and the details of datasets used for experimentation. Section 4 discusses about the results of experimentation and finally section 5 with conclusion.

2. LITERATURE REVIEW

Form the past studies human action recognition have shown better results using deep learning techniques using 2D and 3D data respectively. In 2D domain most of the researchers concentrated on bag of visual words[4] these bag of visual words mainly extracted from spatio-Temporal encoded descriptors of a video data and given as input for classifying actions. Laptev et al[5] proposed spatial time based interest points by enhancing Harris corner in to spatial time dimensions. Later Kviatkovsky et al [6] proposed a method for online action recognition namely Covariance Descriptors. There are other methods such as HOF,HOG, TBC and SIFT are very popular local descriptor methods for representation of videos [20][21][22]. These feature are inputted to the efficient classifiers HMM and GMM . Knowledge based human action recognition(KBAR) and context aware techniques have improved recognition compared to previous techniques by prior knowledge of activity for recognition task. Apart from there success there is a difficulty to provide enough knowledge for action for classification. On the other side depth and skeletal captured from Kinect and other vision based stereo cameras has gained more popularity due to three dimensional measurements in real-time.

In human action recognition skeletal and depth information have shown better results when compared to video-based action recognition. Skeletal data is converted in to image representation for human action recognition[7], by this process action recognition problem is converted in to computer vision problem. The 3D skeletal data comprises of joint positional vectors with respect to time series in a 3D space. Based on the skeletal joint data researchers have developed representation methods such as distances [8], angles[9], angular distances[10], acceleration[11] e.t.c., to decrease the effects of camera coordinates and human bio-structure. These handcrafted features are given to classifiers such as HMM, ANN [12], KNN, FIS, dictionary learning, and SVM. In spite of their success these methods shown limitation while dealing with large datasets. DTW[13] and Fourier temporal pyramid representation has shown better result for shorter sequence videos, as it solves the problems of varying temporal durations. Due to recent advances in deep learning, researchers have shown interest to apply deep learning models such as RNN, LSTM[14] and CNN[10] for action recognition, succeeded in effectively applying to depth, RGB and skeletal data. Multi stream CNN is another method which effectively transforms spatial and temporal data to the corresponding labels by fusing the scores of each streams at last [15]. This multi-stream CNN's have shown better results when compared to single CNN, but this method has increased the expense in computational cost.

Deep learning algorithms have shown incredible performance on images for recognition. So, researchers have modelled the 1D or 3D data in to 2D tensor data to form an Image . More-over at image level data lengths and durations are scalable effectively. The state of the art 2D tensor mapped models such as Joint trajectory maps (JTM's), joint distance maps (JDM's)[8], joint angle maps(JAM's)[9], and joint angular displacement maps(JADM's)[10]. These color coded maps highlights the local joint variations based on the action these models are applicable even for complex over lapping actions also.

Mostly skeletal data is captured by the motion capture setup other than Kinect because Kinect is noisy and for recognizing complex human actions is very difficult. Very less mocap datasets are available due to its taut structure. The datasets which are available and mostly used are HDM05 and CMU, among these CMU is noisy compared to HDM05 and NTU RGB-D dataset is Kinect based action dataset. Inspired with these datasets we created our own dataset namely KLHA3D-102 which is less noisy and large dataset with 39 skeletal points used for human action recognition problems.

The above mentioned maps are having limitations of modelling geometrical relations between joints. In this work we are implementing quad joint volumetric energy maps to explore the joint to joint geometric relations and their energies these are unique maps compared to the mentioned maps. To classify the actions we are implementing a new architecture named hybrid CNN LSTM architecture, were Volumetric energy maps are given to CNN and trajectory points are given to LSTM and both the scores are finally fused for recognition.

3. PROPOSED METHODOLOGY

In past there are several color coded maps [8],[9],[10] and [16]. These were unable to provide the joint to joint relations to accurately predict the Human action. Actually color coded maps are basically spatio temporal features that defines a human skeleton joint's inter and intra frame relationships across 3D action video frames. The human skeleton is represented digitally with J joints which convey their spatial location with respect to the camera coordinates. These spatial locations are positional vectors defined as

$$P^i \in (X, Y, Z) \text{ where } P \text{ represent positional vector} \qquad (0.1)$$

3.1. Joint movement based Energy maps

Rather importing spatio temporal information embedded in a single RGB image, our idea is to calculate the energy of the moving joint from frame to frame which only contains the spatial information. These joints are now moved to calculate the fast fourier transform (FFT)[17]. after calculating FFT the matrix is now passed to check for absolute values of matrix to remove complex numbers in matrix and to make it as real numbers. These absolute valued matricesis squared to give energy matrix and the energy matrix is color coded to give joint movement based energy maps (JME). This process is shown in figure.1 and the formulas are stated below

$$energy = \{abs[FFT(P^i)]\}^2 \qquad (0.2)$$

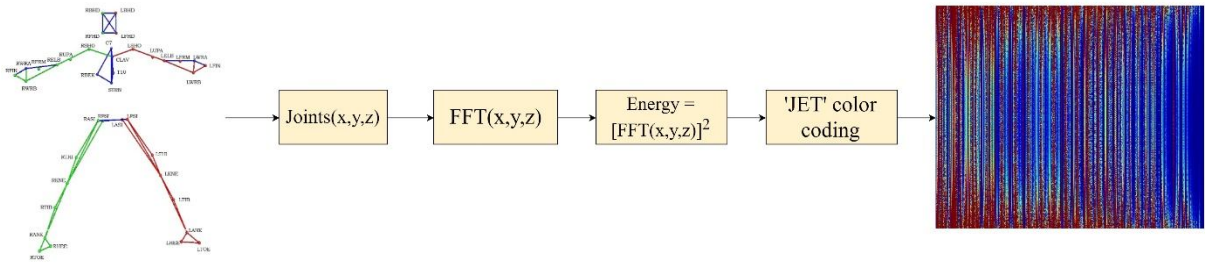

Figure 2: Calculation of Volumetric energy maps from volumes

3.2. Comparison of feature maps

To overcome these network implications for action recognition, a rich spatio temporal feature representation in the form of RGB color images. These RGB color maps characterize a particular skeletal action across a set of 3D video frames. Consequently, the proposed spatio temporal images are found to be independent of length of the video sequences as well as number of joints. These spatio temporal features represent spatial relationships among joints within a 3D action frame and temporal changes between frames as we move horizontally representing temporal patterns. The proposed spatio temporal features are joint positional maps (JPM)[18], Joint Distance Maps (JDM)[8], joint Angular maps (JAM)[9], Joint Angular displacement maps(JADM)[10], Joint Velocity maps (JVM) [16]. These maps are spatio temporal embedded features and to analyzing the networks performance separately in terms of spatial temporal these maps fails and hence limited scope for improving network performance. So, we created JME's this is a spatial feature which is empowered with a energy calculated by the joint movements.

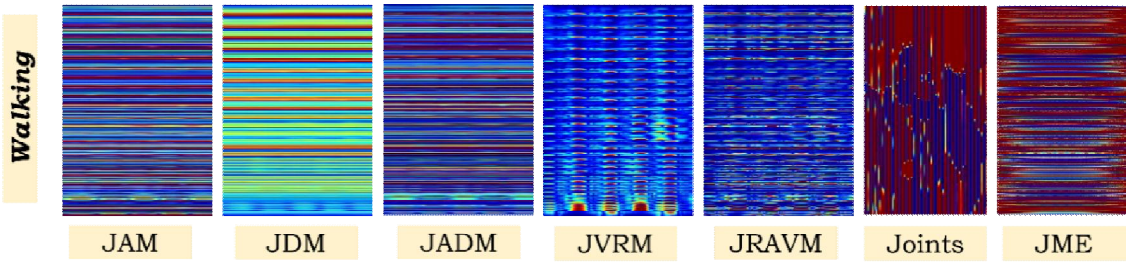

Figure 3: Comparison of different feature maps

3.3. Long short term memory

Long short term memory (LSTM) is the another deep learning technique which gained popularity in dealing with time series data. LSTM's are updated form of recurrent neural network (RNN) where RNN's have a serious vanishing gradient problem. LSTM has solved the vanishing gradient problem as its main idea is memory cell as it can maintain the cell state over time . LSTM nonlinear gating system manages the in/out flow of cell as LSTM contains three of gate namely input gate, output gate and forget gate. Researchers have developed different variations of LSTM in order toanalyze the complex variations in normal LSTM's.

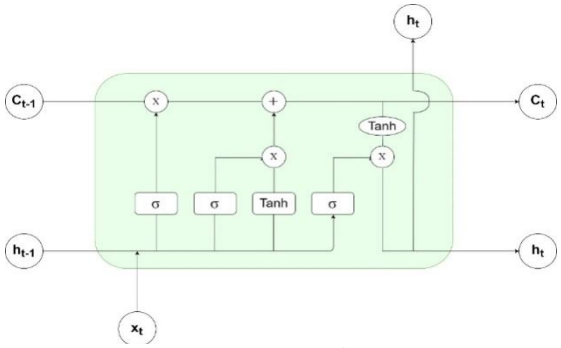

Figure 4: Architecture of LSTM cell

$$I_t = \sigma((x_t + h_{t-1})W_I + b_I)$$
$$F_t = \sigma((x_t + h_{t-1})W_F + b_F)$$
$$O_t = \sigma((x_t + h_{t-1})W_O + b_O)$$
$$G = \tanh((x_t + h_{t-1})W_G + b_G)$$

4.1. 3D Action Dataset – KLHA3D-102

The KLHA3D-102 is captured with an 8 camera Vicon motion capture technology [19]. The human skeleton in our dataset has 39 joints from head to toe. The joints in 3D mocap are placed manually by pre-determining the highly articulated joints on the human body. The 3.2.MP vireo series cameras from Vicon have a frame rate of 320fps and can capture subjects at a distance of up to 15 meters. The setup used for creating our action dataset (KLHA3D-102) is shown in figure 8. The camera heights and subject's positions were generalized in accordance with the ten human subjects who participated in the data collection. The origin is set as shown in figure 5(a). Figures 5(b) and (c) show actions being captured by two human subjects. The KLHA3D-102 dataset comprises of 102 action classes

$$C_t = C_{t-1} * F_t + G * I_t$$
$$h_t = O_t * \tanh(C_t)$$
$$Y = soft\max(h_t)$$

3.4. CNN-LSTM Hybrid Architecture

The proposed architecture is shown in figure 4.The proposed architecture is a 2 stream CNN LSTM model with 12 layers in CNN stream and 3 layers in LSTM stream. There are 4 layers of convolution, max pooling in CNN stream. In addition to that there is batch normalization before the 1st convolutional layer. The output scores from the two streams are fused after SoftMax layers and There are no dropouts in both the streams and in the dense layers. The proposed architecture eliminated overfitting and vanishing gradients problems that are very much prevalent in regular CNNs. The activation function in all layers is ReLu. The dropouts in features at the end of the network is preferred in traditional CNNs to improve non linearity which there by increases accuracy.

4. EXPERIMENTATION AND RESULTS

This section describes and evaluates the proposed method for 3D action recognition. First, we present our 3D action dataset recorded using 8 cameramotion capture technology, along with other publicly available motion capture and Kinect action datasets. Secondly, we validate the proposed method through experimentation and comparing the obtained results with similar state-of-the-art methods. Finally, illustration of the analysis gives an insight into the applicability of the proposed CNN-LSTM hybrid architecture for 3D human action recognition.

Figure 5: KLHA3D-102 dataset capture setup.

These 102 classes are divided across multitude of human actions performed by ten different subjects. Due to the intensity in actions and subject's mental stimuli, the within class actions were indifferent in time and space. Accordingly, each action class is comprised of five 3D action videos. Using 3D scaling and rotation properties, each action video was converted into 109 using nexus software package associated with Vicon mocap system. Nine scales with a range of 0.95 to 0.5 in 12 different rotations at incremental 30 degrees creates

the additional 108 3D action data within an action class. In total KLHA3D-102 action dataset consists of action samples. In order to perform cross data validation, KLHA3D-102 has 70% actions from publicly available mocap and Kinect datasets.

Table 1: Performance analysis of proposed model

Architecture	JDM		JAM		JADM		JAVM		Joints		JME + Joints	
	Cross Subject	Cross View	Cross Subject	Cross View	Cross Subject	Cross View	Cross Subject	Cross View	Cross Subject	Cross View	Cross Subject	Cross View
VGG [10]	73.09	71.51	76.25	73.38	80.61	78.25	84.58	80.21	85.59	82.53	89.99	86.57
CNN+LSTM [9]	72.21	68.24	72.54	69.52	73.39	71.54	78.21	75.94	82.63	80.51	88.55	84.31
CNN+RNN [8]	73.88	69.11	74.38	71.02	76.76	73.78	82.19	78.04	83.17	80.67	89.23	84.73
Multi-Stream CNN [6]	73.65	71.51	78.25	73.38	83.88	78.25	83.82	80.21	85.64	84.24	90.16	88.24
GoogLeNet	80.57	77.24	81.49	78.43	85.29	82.31	83.69	79.19	84.59	82.19	90.35	89.31
Connived ResNet	83.37	80.82	84.18	82.27	87.36	84.35	85.32	83.65	85.73	83.57	93.07	92.17
Proposed	90.02	86.14	89.15	85.21	92.27	88.31	94.27	92.14	92.63	92.46	98.12	96.15

Apart from our KLHA3D-102, we also used publicly available 3D mocap action datasets HDM05 and CMU. Out of the two, HDM05 has less noise and consists of 70 action classes with 5 subjects performing an action several times. In this work, we used action samples for training and testing. The CMU dataset in this work is carefully crafted to avoid missing and noisy marker information. Consequently, the CMU dataset used for training and testing has samples, with 30 actions classes, 10 subjects and 30 variations per subject. On the contrary to our 39-joint skeleton, HDM05 and CMU are captured with 41-joint skeletons. Finally, to discover the usefulness of the proposed maps and the ML algorithm, we investigated Kinect skeletal action data with 25 joints from NTU RGB D dataset. Our refabricated NTU RGB D dataset has action samples. Besides, these datasets were selected to have a 30 to 40% overlap among action classes. For example, walking, drinking, jumping, eating etc. are to name a few, which are a part of all selected datasets. Finally, training, validation and testing are initiated on the CCNN and other DNNs using the QjRVMs fabricated from the 3D action data.

4.3. Evaluation of proposed Hybrid architecture across state of the art model and publicly available datasets

In this section, we evaluated our proposed model with state of the art models while trained of the publicly available datasets of cross view and cross subject validation. Here proposed model has shown increased performance while trained on KLHA3D dataset because it is a clean dataset which we created in our research center.

Table 2: Recognition rates of proposed model

Architecture	Datasets	Validation Error (%)	Cross Subject	Cross View
VGG	HDM05	5.54	83.18	81.27
	CMU	6.67	78.54	76.22
	NTU RGB-D	5.92	80.17	79.52
	KLHA3D	4.98	89.99	86.57
CNN+LSTM	HDM05	4.84	85.24	83.48
	CMU	5.55	80.19	78.62
	NTU RGB-D	4.96	82.63	80.82
	KLHA3D	4.67	88.55	84.31
CNN+RNN	HDM05	4.59	85.61	82.64
	CMU	5.29	81.35	79.21
	NTU RGB-D	4.77	82.66	79.83
	KLHA3D	3.92	89.23	84.73
Multi-Stream CNN	HDM05	4.21	88.14	85.91
	CMU	5.03	82.37	79.36
	NTU RGB-D	4.42	84.43	83.57
	KLHA3D	3.76	90.16	88.24
GoogLeNet	HDM05	4.18	89.21	87.54
	CMU	4.93	84.19	82.4
	NTU RGB-D	4.37	86.11	84.09
	KLHA3D	3.42	90.35	89.31
Connived	HDM05	3.53	90.54	88.35

4.2. Evaluation of proposed Hybrid architecture across state of the art model and Maps

In this section, we evaluated our proposed maps with the different computed feature maps namely Joint XYZ, JDM, JADM and proposed JME's. Table-1 shows the performance evaluation of state of the art models compared with different feature maps in terms of cross view and cross subject validation. Here proposed model has shown increased performance while trained with JME's and joints data.

ResNet	CMU	4.95	89.34	87.63
	NTU RGB-D	3.26	89.97	86.39
	KLHA3D	2.57	93.07	92.17
Proposed Mode	HDM05	2.06	94.52	93.34
	CMU	3.21	91.24	89.75
	NTU RGB-D	2.17	93.65	92.21
	KLHA3D	1.54	98.12	96.15

4.4. Evaluation of proposed Hybrid architecture across multi model datasets

In this section, we evaluated our proposed model with state of the art models while trained on the multi model dataset. In here multi model dataset means for training the network we considered mocap data and tested with Kinect data next vice versa. Also combined mocap and Kinect training and evaluated on the any of the mocap or Kinect and Vice versa. The performance is surprising well on our proposed model and is tabulated below

Table 3: Recognition rates of proposed model on multi model data

Architecture	Training Model	Testing Model	Average Recognition Rate (%)
VGG	Mocap	Kinect	64.15
	Kinect	Mocap	48.24
	Mocap + Kinect	Kinect	70.42
	Mocap + Kinect	Mocap	79.82
CNN+LSTM	Mocap	Kinect	62.75
	Kinect	Mocap	43.24
	Mocap + Kinect	Kinect	71.64
	Mocap + Kinect	Mocap	77.16
CNN+RNN	Mocap	Kinect	62.91
	Kinect	Mocap	46.12
	Mocap + Kinect	Kinect	72.21
	Mocap + Kinect	Mocap	78.49
Multi-Stream CNN	Mocap	Kinect	62.18
	Kinect	Mocap	47.29
	Mocap + Kinect	Kinect	71.31
	Mocap + Kinect	Mocap	76.34
GoogLeNet	Mocap	Kinect	65.34
	Kinect	Mocap	50.63
	Mocap + Kinect	Kinect	72.34
	Mocap + Kinect	Mocap	77.64
Connived ResNet	Mocap	Kinect	66.18
	Kinect	Mocap	50.31
	Mocap + Kinect	Kinect	72.49
	Mocap + Kinect	Mocap	76.74
Proposed	Mocap	Kinect	70.92
	Kinect	Mocap	56.17
	Mocap + Kinect	Kinect	76.74
	Mocap + Kinect	Mocap	82.91

5. CONCLUSION

In this paper, we proposed a new feature maps namely joint movement based energy maps and also we came up with new hybrid CNN-LSTM for action recognition. In which, CNN layer is inputted with proposed JME maps and LSTM in second layer is inputted with joint data. This framework helped in predicting the actions with much more efficiently with decreased epochs. In this paper we compared our model with other state of the art networks and results shown that our method performed well. We also compared with other publicly available datasets by computing various spatio temporal embedded features and results shown that our method has given improved prediction rates. This proposed method is evaluated by available action database and given better accuracies, based on the action classes recognized by the model at the output. Our model is suitable in predicting the human action much more fast, reliable and efficient and it can be integrating in the smart surveillance systems, self-assessment systems and in human computer interaction models. Further, our aim is to improve the performance and also extend our work to online action recognition.

REFERENCES

1. Müller, Meinard, TidoRöder, Michael Clausen, Bernhard Eberhardt, Björn Krüger, and Andreas Weber. "Documentation mocap database hdm05." (2007).
2. Rogez, Grégory, and Cordelia Schmid. "Mocap-guided data augmentation for 3d pose estimation in the wild." In Advances in neural information processing systems, pp. 3108-3116. 2016.
3. Liu, Jun, Amir Shahroudy, Mauricio Lisboa Perez, Gang Wang, Ling-Yu Duan, and Alex KotChichung. "Nturgb+ d 120: A large-scale benchmark for 3d human activity understanding." IEEE transactions on pattern analysis and machine intelligence (2019). https://doi.org/10.1109/TPAMI.2019.2916873
4. Yang, Jun, Yu-Gang Jiang, Alexander G. Hauptmann, and Chong-Wah Ngo. "Evaluating bag-of-visual-words representations in scene classification." In Proceedings of the international workshop on Workshop on multimedia information retrieval, pp. 197-206. 2007.
5. Laptev, Ivan. "On space-time interest points." International journal of computer vision 64, no. 2-3 (2005): 107-123.

6. Kviatkovsky, Igor, Ehud Rivlin, and IlanShimshoni. "Online action recognition using covariance of shape and motion." Computer Vision and Image Understanding 129 (2014): 15-26. https://doi.org/10.1016/j.cviu.2014.08.001
7. Laraba, Sohaib, Mohammed Brahimi, JoëlleTilmanne, and Thierry Dutoit. "3D skeleton□based action recognition by representing motion capture sequences as 2D□RGB images." Computer Animation and Virtual Worlds 28, no. 3-4 (2017): e1782.
8. Li, Chuankun, Yonghong Hou, Pichao Wang, and Wanqing Li. "Joint distance maps based action recognition with convolutional neural networks." IEEE Signal Processing Letters 24, no. 5 (2017): 624-628.
9. Uddin, Md Zia, Nguyen Duc Thang, and Tae-Seong Kim. "Human Activity Recognition via 3-D joint angle features and Hidden Markov models." In 2010 IEEE International Conference on Image Processing, pp. 713-716. IEEE, 2010. https://doi.org/10.1109/ICIP.2010.5651953
10. Maddala, Teja Kiran Kumar, P. V. V. Kishore, Kiran Kumar Eepuri, and Anil Kumar Dande. "YogaNet: 3-D Yoga Asana Recognition Using Joint Angular Displacement Maps WithConvNets." IEEE Transactions on Multimedia 21, no. 10 (2019): 2492-2503.
11. Krosshaug, Tron, and Roald Bahr. "A model-based image-matching technique for three-dimensional reconstruction of human motion from uncalibrated video sequences." Journal of biomechanics 38, no. 4 (2005): 919-929.
12. Kishore, P. V. V., MVD Prasad, Ch Raghava Prasad, and R. Rahul. "4-Camera model for sign language recognition using elliptical fourier descriptors and ANN." In 2015 International Conference on Signal Processing and Communication Engineering Systems, pp. 34-38. IEEE, 2015. https://doi.org/10.1109/SPACES.2015.7058288
13. Berndt, Donald J., and James Clifford. "Using dynamic time warping to find patterns in time series." In KDD workshop, vol. 10, no. 16, pp. 359-370. 1994.
14. Greff, Klaus, Rupesh K. Srivastava, Jan Koutník, Bas R. Steunebrink, and Jürgen Schmidhuber. "LSTM: A search space odyssey." IEEE transactions on neural networks and learning systems 28, no. 10 (2016): 2222-2232.
15. Gao, Yuan, Jiayi Ma, Mingbo Zhao, Wei Liu, and Alan L. Yuille. "Nddr-cnn: Layerwise feature fusing in multi-task cnns by neural discriminative dimensionality reduction." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3205-3214. 2019. https://doi.org/10.1109/CVPR.2019.00332
16. Kumar, Eepuri Kiran, P. V. V. Kishore, Maddala Teja Kiran Kumar, Dande Anil Kumar, and A. S. C. S. Sastry. "Three-dimensional sign language recognition with angular velocity maps and connived feature ResNet." IEEE Signal Processing Letters 25, no. 12 (2018): 1860-1864. https://doi.org/10.1109/LSP.2018.2877891
17. Müller, Meinard, TidoRöder, Michael Clausen, Bernhard Eberhardt, Björn Krüger, and Andreas Weber. "Documentation mocap database hdm05." (2007).
18. Brause, Rudiger. "Optimal information distribution and performance in neighbourhood-conserving maps for robot control." In [1990] Proceedings of the 2nd International IEEE Conference on Tools for Artificial Intelligence, pp. 451-456. IEEE, 1990.
19. Kumar, D. Anil, A. S. C. S. Sastry, P. V. V. Kishore, E. Kiran Kumar, and M. Teja Kiran Kumar. "S3DRGF: Spatial 3-D relational geometric features for 3-D sign language representation and recognition." IEEE Signal Processing Letters 26, no. 1 (2018): 169-173. https://doi.org/10.1109/LSP.2018.2883864
20. Petrosov, David Aregovich, Roman Alexandrovich Vashchenko, Alexey Alexandrovich Stepovoi, and N. V. Petrosova. "Application of artificial neural networks in genetic algorithm control problems." International Journal of Emerging Trends in Engineering Research 8, no. 1 (2020): 177-181. https://doi.org/10.30534/ijeter/2020/24812020
21. Khudov, H., I. Khizhnyak, F. Zots, G. Misiyuk, and O. Serdiuk. "The Bayes Rule of Decision Making in Joint Optimization of Search and Detection of Objects in Technical Systems." IJETER 8, no. 1 (2020): 7-12. https://doi.org/10.30534/ijeter/2020/02812020
22. Nawaz, Nishad. "Artificial Intelligence Face Recognition for Applicant Tracking System." International Journal of Emerging Trends in Engineering Research 7 (2019): 12.