



Applied Human Action Recognition Network Based on SNSP Features

M Shujah Islam¹ · Khush Bakhat² · Rashid Khan¹ · Nuzhat Naqvi¹ · M Mattah Islam³ · Zhongfu Ye¹

Accepted: 8 July 2021 / Published online: 9 April 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Recognition of human action is a daunting challenge considering action sequences' embodied and dynamic existence. Recently designed material depth sensors and the skeleton estimation algorithm have developed a renewed interest in human skeletal action recognition. This paper performed human action recognition by using a novel SNSP descriptor that acquired complex spatial information among all skeletal joints. In particular, the proposed SNSP determines combined and unite details using the prominent joint. Our features are calculated using the standard normal, slope, and parameter space features. The neck is proposed as a super-joint, SNSP is utilizing features and a prominent joint. We evaluate the proposed approach on three challenging action recognition datasets i.e., UTD Multimodal Human Action Dataset, KARD- Kinect Activity Recognition Dataset, and SBU Kinect Interaction Dataset. The experimental results demonstrate that the proposed method outperforms state-of-the-art human action recognition methods.

Keywords Human action recognition · SNSP · Super-joint · K-nearest neighbor classifier

✉ M Shujah Islam
mshujahislam@mail.ustc.edu.cn

Khush Bakhat
khush.bakhat.awar@gmail.com

Rashid Khan
rashidkhan@mail.ustc.edu.cn

Nuzhat Naqvi
nuzhatnaqvi@mail.ustc.edu.cn

M Mattah Islam
m.mattahislamsameem@gmail.com

Zhongfu Ye
yezf@ustc.edu

¹ University of Science and Technology of China, Hefei, China

² Air University, Islamabad, Pakistan

³ National University of Computer and Emerging Sciences, Islamabad, Pakistan

1 Introduction

Human action recognition is a widespread study area [1]; it can be implemented in many applications in e-health [2], video surveillance systems [3], and strategic war situations [4], etc. Amid significant work initiatives over the past few decades, recognition of action remains a daunting issue because of human movements' dynamic embodied nature.

It is usually seen that the skeletal joints are useful for understanding actions [5]. Newly implemented depth sensors combined with the real-time skeleton estimation algorithm to render the process of action detection far more comfortable. Skeletal joints' location in the frame and their mobility in changing frames can be used to perform an action recognition task. Such new developments have contributed to a revived attention in recognition of human skeletal action.

Many existing approaches for interpreting skeletal action recognize human skeletal joints or joints' usage to identify specifics of activity [9–22]. These methods do not adapt exceptionally well to identify skeletal joint action because they cannot distinguish complex details across various joints of the body that can reflect the dynamic spatial variations in human behavior.

Two-dimensional skeleton body joints are used as input of the target method; we accomplished an action recognition task using the least amount of data. Initially, the super-joint is observed in the body, which is an essential factor in detecting behavior. The usage of super-joint functions is determined by using standard normal, slope, and space parameters features. In particular, standard normal and slope determine the unite, while parameters space features obtained combined details using joint. All these considerations help us to formulate an action descriptor, known as SNSP, which comprehends Super-joint, standard Normal, Slope & Parameters space features. Our action recognition system collects all complicated spatial details acquired from all skeleton joints points. Anticipated SNSP is based on three commonly established and freely accessible action databases, i.e., UTD Multimodal Human Action Dataset [6], KARD- Kinect Activity Recognition Dataset [7], and SBU Kinect Interaction Dataset [8]. Classification is performed by using the K-nearest neighbor classifier [9]. SNSP outperforms the aforementioned tactics as it collects detailed and necessary spatial information between all skeleton joints; in this research paper, our contributions are as follows:

- Capture all the necessary spatial details of the joints by using 2D joints.
- Super-joint is developed to track the behavior of certain joints.
- Combined and unite spatial information is acquired using SNSP.

The presented article is split up into segments: Sect. 2 is related work, sketches preceding human action recognition approaches. Section 3 outlines our proposed SNSP; Sect. 4 explains experimental analysis, and Sect. 5 addresses evaluation in the projected descriptor action. Results & discussion are clarified in Sect. 6, while our proposed study is eventually concluded in Sect. 7.

2 Related Work

Several valuable surveys compare the skeletal joints designed aimed at human action recognition; many researchers perform recognitions of action using different methods. This section discusses some of the state-of-the-art techniques based on the skeletal joint points [5, 9–22], and other relative approaches [23–31].

Skeletal joints are used frequently in the literature to perform action recognition tasks. Several of the traditional HAR methods at the skeleton joint are: Graph-based approach for 3D human skeletal action recognition [5], recognition of actions using 3-D Posture Data and Hidden Markov Models is presented in [9], Skeleton-based human action recognition with global context-aware attention LSTM networks [10]. Capturing the micro-temporal relations between the joints using CNN and then exploiting their macro-temporal connections by computing the Fourier temporal pyramid [11]. A human activity recognition system using skeleton data from RGB-D sensors is proposed in [12]. The enhanced trajectory is proposed using codebook and local bags of words [13], Co-occurrence features obtained using a deep LSTM network [14], and an end-to-end Spatial-temporal attention model is presented in [15]. Global context-aware attention using LSTM [16] and pose conditioned spatiotemporal attention is explained in research [17]. In [18], the author performed action recognition using skeleton net and deep learning [18]. Obtaining temporal information from frames using deep CNN and multi-task learning networks [19]. Action recognition is performed by using a hand gesture recognition system [20], CNN grounded joints distance maps [21], and joint trajectory maps [22].

Action recognition is performed by using a two-fold transformation model [1], hexagonal volume local binary pattern [2], and deep learning and wearable sensors [23]. S2DDI model developed using sequences of depth and CNN collecting motion from regional to fine-grained levels [24]. Activity is recognized by using the image relating history technique suggested in [25]. Spatial rank pooling is developed to gather heatmap development as an image of body form [26]. In particular, spatial-temporal activations created by a stack of pose estimation layers are reprojected in space and time using a stack of three-dimensional convolutions learned directly from the data [27]. Collaborating body part contrast mining for human interaction recognition [28]. Each depth frame is projected onto three Cartesian orthogonal planes in a vague boundary sequence using depth maps. The absolute value of the difference between two consecutive projected maps is accumulated [29]. Human action recognition is performed through silhouette and silhouette skeletons [30], and geometrical patterns and optical flow in [31]. Authors, Yu et al. performed image recognition by using: Click prediction for web image reranking using multimodal sparse coding [32], Learning to rank using user clicks and visual features [33], and Hierarchical deep click feature prediction [34]. Recently developed methods for performing human action recognition are hierarchical learning approach [35], uncovering deep learning approach [36], positioning sensor and deep learning technique [37], a robust framework for abnormal act detection [38], Kinect sensor-based interaction monitoring system using the BLSTM neural network [39], skeletal data-based system [40], compressive sensing-based recognition of human upper limb motions with Kinect skeletal data [41], a unified deep framework for joint 3D [42], global co-occurrence feature learning and active coordinate system [43], multi-stream Spatio-temporal fusion network [44], and image representation of pose-transition feature for 3D skeleton-based action recognition [45].

3 Proposed Action Descriptor

The goal is to use the least amount of data to conduct human behavior identification, so two-dimensional skeletal joints are used. Our feature vector is constructed in such a way that joints that are showing identical movement characteristics are collected together, allowing a more discriminating video representation for action recognition. This is

accomplished using the information provided by the human pose. The aim is to link skeleton joints together and determine a descriptor for SNSP operation. Following these factors, we proposed an action descriptor:

3.1 Super-Joint

A point from which the weight of a body or system may be considered to act is known as the center of gravity [46]. Whereas skeleton body joints action recognition, that unique human organ is the neck. If any action sequences have to be identified, it is imperative to observe the neck joint because the neck is a frame of reference for other joints. Its normal range of motion is 40 towards 80 degrees [47], it's observed that the body organ neck is the least changing organ. This unique and vital joint is called a super-joint, while other joints change their super-joint position. It is observed that all others joints are doing action against the super-joint, the same as the sun and planets movement. The super-joint is notated by $j_n(x_n, y_n)$, and it is shown as a red joint in Fig. 1. Features explained in Sects. 3.2, 3.3, and 3.4 are all based on utilizing super-joint.

3.2 Standard Normal Feature

The standard normal distribution is a wave form of a mean of zero and a standard deviation of one. The standard normal distribution is positioned at nil (0), and the grade of assumed dimension deviates from the mean is provided by the standard deviation, as shown in Eq. 1.

$$P(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad (1)$$

We have altered the standard normal distribution formula to permit it to join in features among two joints. This transformed comparison is pulling out coordinate's differences amongst super and any joint. Let us suppose that we have any particular *i*th joint $j_i(x_i, y_i)$,

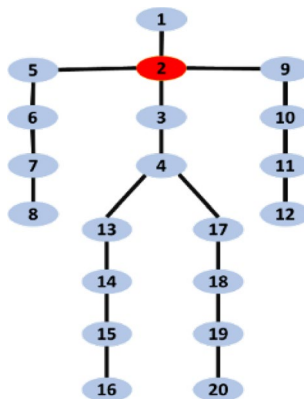


Fig. 1 Super-joint detected on the human body

if we calculate standard normal features sandwiched between j_n and j_i . Following Eq. is describing an association amid j_n and j_i , characterized by $P(x_n, x_i, y_n, y_i)$.

$$P(x_n, x_i, y_n, y_i) = \frac{1}{\sqrt{2\pi}} e^{-((x_i - x_n)^2 + (y_i - y_n)^2)/2} \tag{2}$$

3.3 Slope Feature

Slope or cartesian coordinates angle (θ_s) is calculated between the super-joint $j_n(x_n, y_n)$ and any particular i th joint $j_i(x_i, y_i)$, it validates which angles are connected to two joints. The value of measuring the slope of the slope is that it is used separately as a function and allows us to measure specific features. Equation 3 is demonstrating the slope between j_n and j_i .

$$\theta_s = \tan^{-1}\left(\frac{(y_i - y_n)}{(x_i - x_n)}\right) \tag{3}$$

3.4 Parameter Space Features

We are utilizing parameter space features to calculate individual components of j_n and j_i . The slope angle is used to help us incorporate mutual information at the body joints. Alteration of joint $j_n(x_n, y_n)$ and $j_i(x_i, y_i)$ into parameter space is performed using Eqs. 4 and 5. The ρ_n and ρ_i is obtained by using joints j_n and j_i . Figure 2a represents the transformation of the (x, y) plane and Fig. 2b (ρ, θ) plane. Our feature vector is also incorporating the obtained spatial statistics.

$$\rho_n = x_n \cos \theta_s + y_n \sin \theta_s \tag{4}$$

$$\rho_i = x_i \cos \theta_s + y_i \sin \theta_s \tag{5}$$

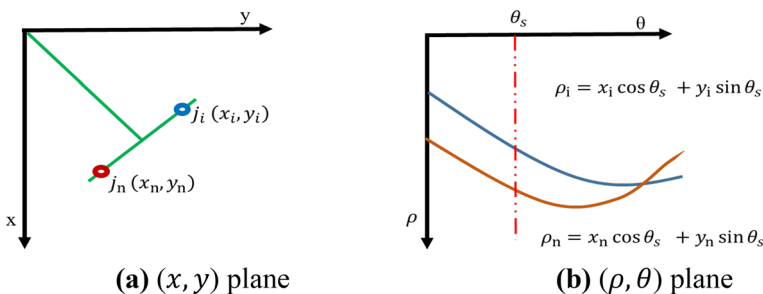


Fig. 2 Body joints illustration on the parameter plane

3.5 Action Descriptor: SNSP

The relationship of individual body joint and super-joint is drawn and called a Correlation of skeletal joints and super-joint. Proposed correlation is very important in collection action details, which are represented in Fig. 3. Twenty joints dataset representations are: eight (8), five (5), and six (6) correlations are shown in Fig. 3a–c. The number of correlations relies upon the number, the number of joints i.e., 20, 15, and 30.

Proposed SNSP is utilizing minimal joints information to perform a human action recognition task, we have used 2D instead of using 3D skeletal joints points. At each correlation, we have calculated standard normal, slope, and parameter space features. In particular, standard normal and slope gained the unite, while parameters space features acquired combined details using joints. By linking all calculated features, we developed an action descriptor, termed SNSP. Comprehensive spatial action information is gained from 2D- skeletal joints, the action recognition algorithm is represented in Table number one (1).

SNSP algorithm will be used to calculate respective features at each frame. Coupling frame by frame features will form a feature matrix. Classification of function matrixes is achieved using the K-nearest neighbor classifier [37]. Figure 4 is representing the system diagram of our proposed SNSP descriptor.

4 Experimental Analysis

This segment labels the used system requirements, and information of human action recognition databases. Details of joints of three databases in this unit.

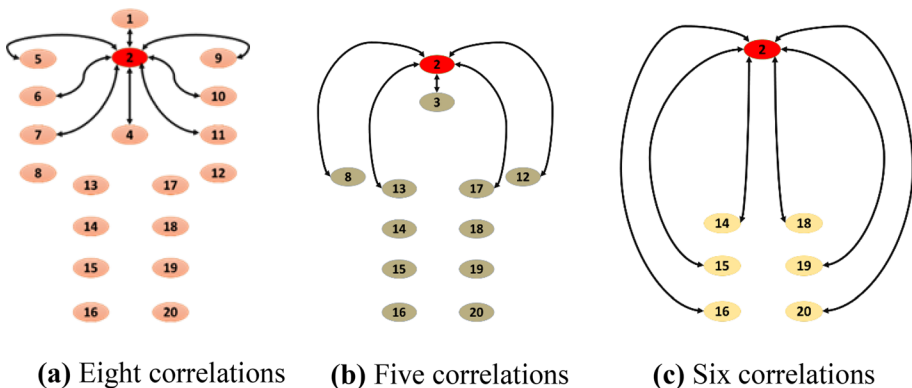


Fig. 3 Correlation of skeletal joints and super-joint

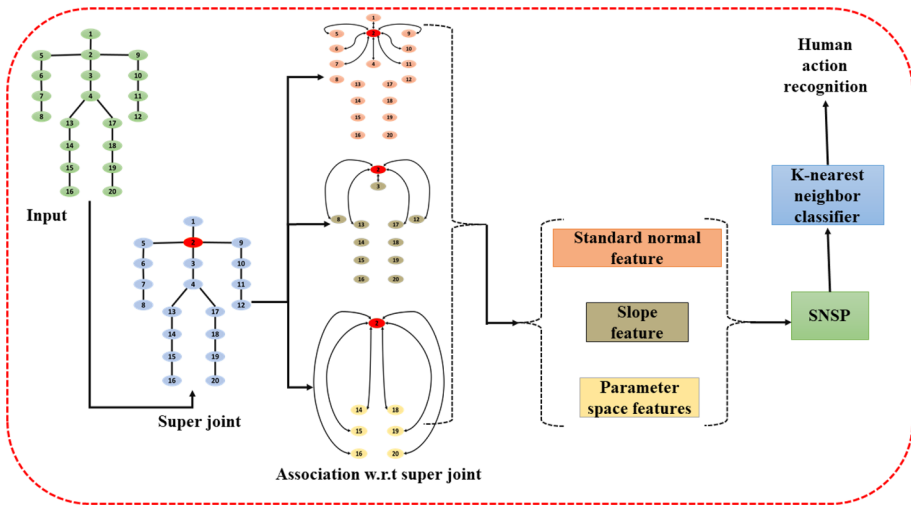


Fig. 4 SNSP system diagram

4.1 System Stipulations

We used a computing computer to facilitate numerical processing, exposed in subsequent:

Fig. 5 UTD Multimodal Human Action Dataset skeletal joints

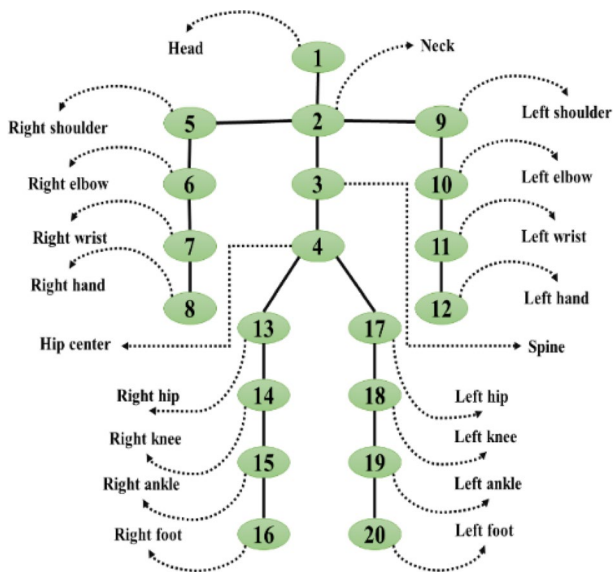


Fig. 6 KARD- Kinect Activity Recognition Dataset skeletal joints

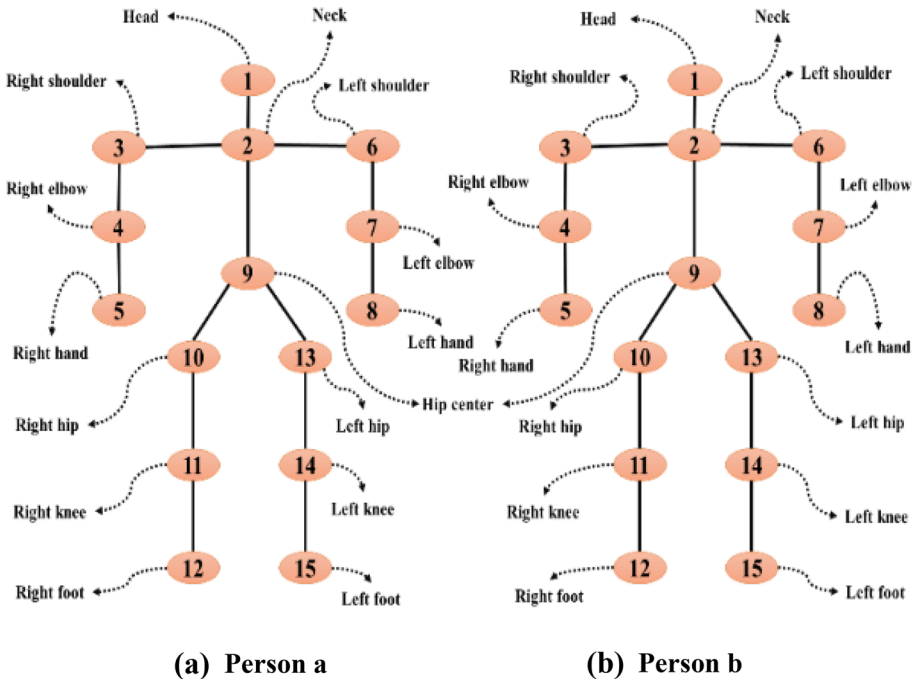
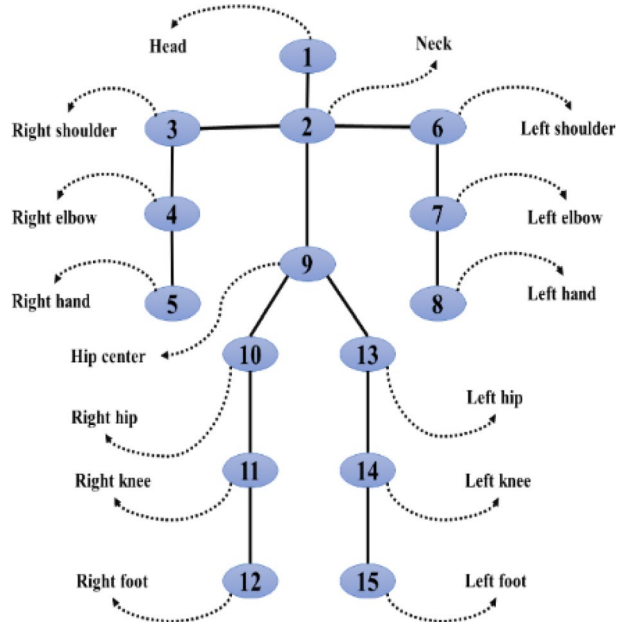


Fig. 7 SBU Kinect Interaction Dataset skeletal joints

4.2 Dataset Joint's Orientations

UTD Multimodal Human Action Dataset comprises twenty joints to represent the human body [6]. Figure 5 is demonstrating what is the order of twenty joints, are their name according to their representation.

The KARD- Kinect Activity Recognition Dataset consists of 15 joints representing the human body [7]. Figure 6 describes 15 variations of integrated databases for the joints.

SBU Kinect Interaction Dataset contains thirty (30) skeletal joints per frame [8]. A person's joints are represented in Fig. 7a; fifteen joints are used to describe person a. Fifteen joints of person b are presented in Fig. 7b. SBU Kinect Interaction skeletal joints can be seen in Fig. 7.

5 Evaluation

Our proposed SNSP is compared interim of accuracy (%) by state-of-the-art approaches implemented by using: UTD Multimodal Human Action Dataset, KARD- Kinect Activity Recognition Dataset SBU Kinect Interaction Dataset.

Figure 8 demonstrated the assessment between our proposed system and the current eleven solutions to the UTD Multimodal Human Action Dataset precision index. Rendering to the certain results, authors [6] gained seventy-nine percentage as least correctness, while authors [27] attained all-out of 90% accurateness amongst nominated methods. Our presented SNSP accomplished 91.8% accuracy.

Assessment of our proposed SNSP method with the existing methodologies to assess KARD- Kinect Activity Recognition Dataset accuracy as explained in Fig. 9. According to the given values, the author in [11] obtained a maximum of 96.3 & 97.41%, whereas scholar [7] attained 84.8 & 84.5% as minutest accurateness among whole related methods. Conversely, anticipated SNSP conquered 97.8% intended for KARD- Kinect Activity Recognition Dataset, which significantly amended thru maximum previous efficiency.

The comparison of our proposed method by the current 16 methods in contradiction of the SBU Kinect Interaction Dataset accurateness was shown in Fig. 10. Giving to specific standards, the Author in [8] achieved bottom-line efficiency, while our previous proposed

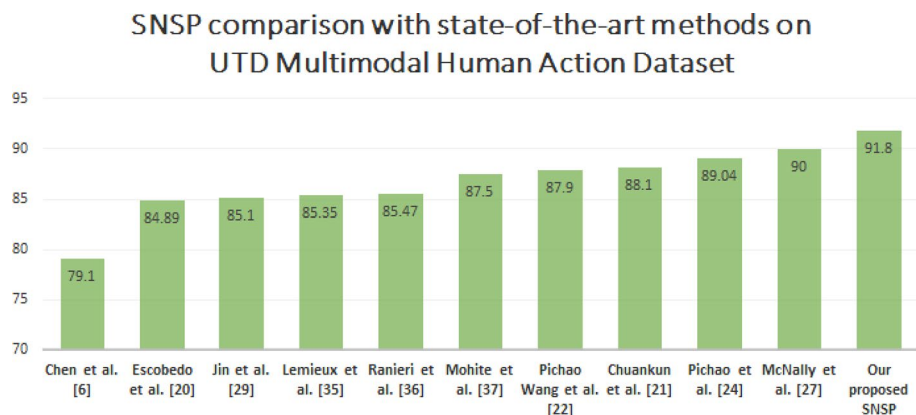


Fig. 8 SNSP evaluation by state-of-the-art approaches arranged UTD Multimodal Human Action Dataset

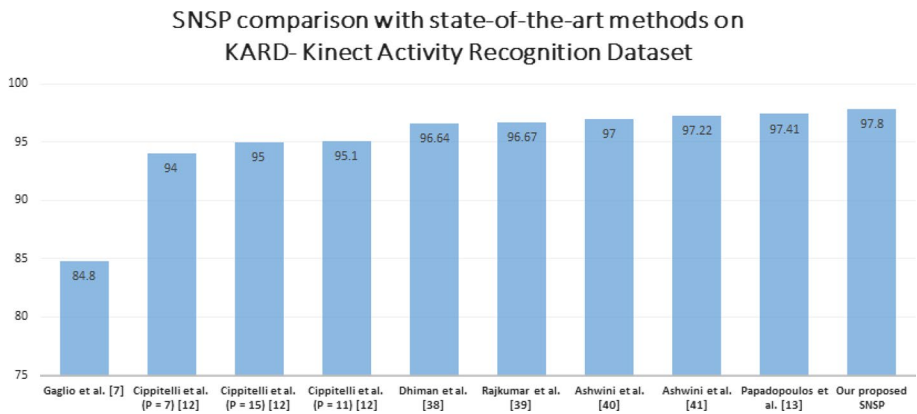


Fig. 9 SNSP evaluation by state-of-the-art approaches arranged KARD- Kinect Activity Recognition Dataset

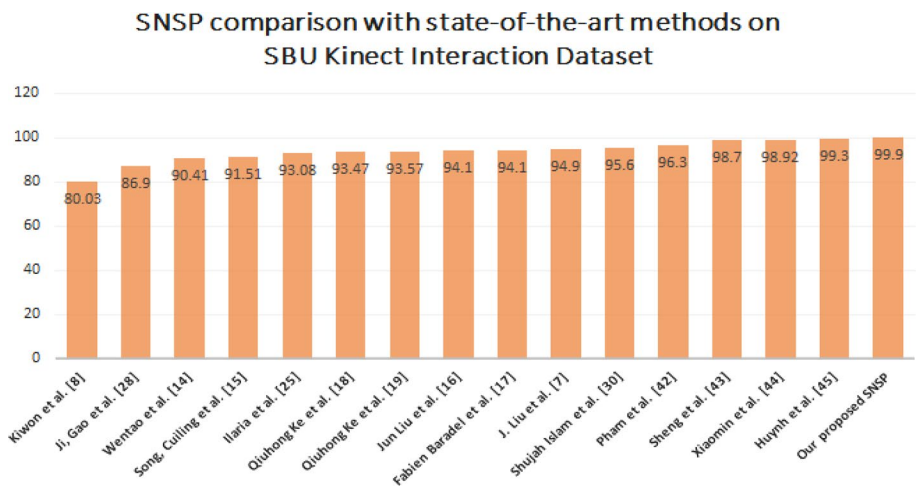


Fig. 10 SNSP evaluation by state-of-the-art approaches arranged SBU Kinect Interaction Dataset

approach, Huynh et al. [45], gained an extreme of 99.3% precision amongst particular approaches. Anticipated SNSP conquered 99.9% for SBU Kinect Interaction Dataset, which was enhanced by earlier supreme accurateness.

6 Results and Discussion

The preceding segment represents that SNSP outclasses state-of-the-art tactics of three datasets. Each dataset from Figs. 8, 9, and 10 have been clarified inside class results throughout here. We have categorized using the K-nearest neighbor classifier's corresponding classification help to produce confusion matrices for each dataset. The confusion matrix was provided with ten severity levels. The lowest intensity indicates the least

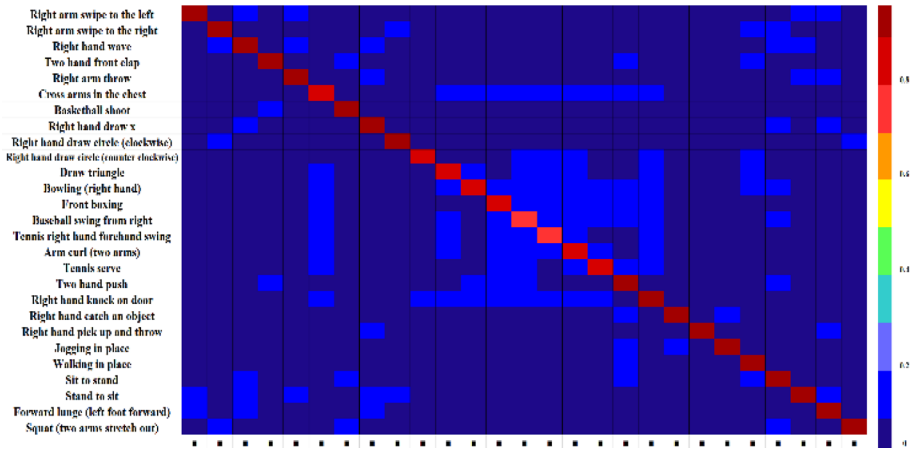


Fig. 11 SNSP results on UTD Multimodal Human Action Dataset

likelihood of error and the highest intensity indicates the greatest probability of error inside the class.

UTD Multimodal Human Action Dataset consisted of twenty-seven sequences of actions [6]. Individual performance between classes, as confusion matrix was seen in Fig. 11. The suggested action descriptor of acts produced outstanding success within the class. Minimum version, according to the record, 76.6% within the category "Tennis right-hand forehand swing" while 97.9% can be seen in the "Right hand pick up and throw" class. Use SNSP, and the average interclass score is 91.8 percent.

Figure 12 displays the uncertainty matrix for KARD- Kinect Activity Recognition Dataset, which includes 18 separate tasks. All maximum amounts of correct decisions can be seen in classes "Take Umbrella" and "Stand up," but then bottom-line results can be seen by 91.1% in activity "Two hand wave," using K-nearest neighbor classifier.

At SBU Kinect Interaction Dataset, SNSP reached 99.9 percent; it includes eight different classes. Figure 13 illustrates the confusion matrix employing the SNSP descriptor.

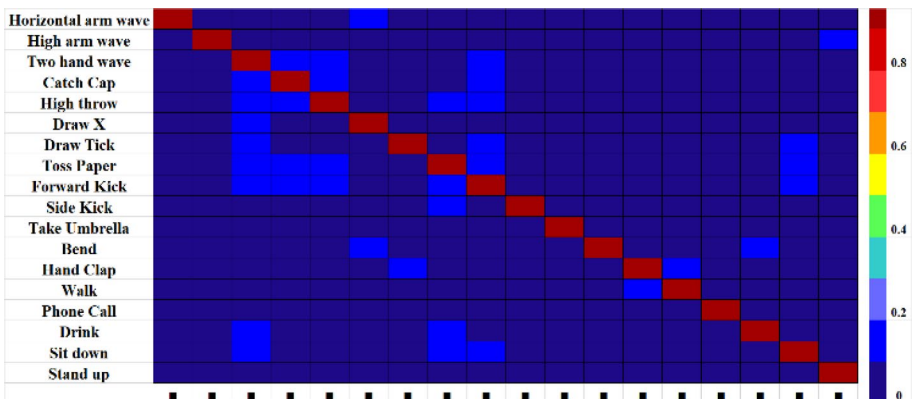


Fig. 12 SNSP results on KARD- Kinect Activity Recognition Dataset

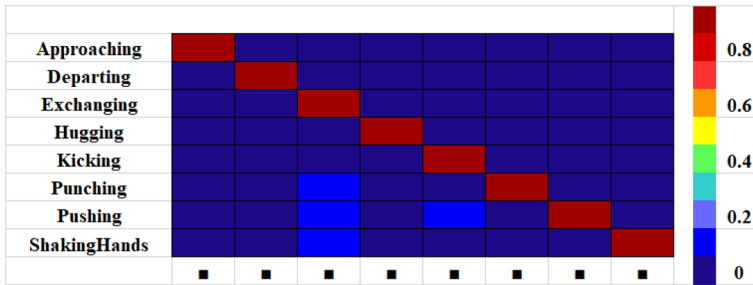


Fig. 13 SNSP results on SBU Kinect Interaction Dataset

Table 1 Proposed action descriptor, SNSP algorithm

Algorithm 1: SNSP Algorithm	
Use 2D skeleton joint's	
for frame = 1 to n do	
Super-joint	
Standard normal feature	
Slope feature	
Parameter space features	
Action descriptor: SNSP	
end for	
K-nearest neighbor classifier	
Return the best solution	

Table 2 Computational time analysis

Elapsed time is 0.039205 seconds per frame
Intel(R) Core (TM) i3-4010U CPU @ 1.70 GHz (x64-based processor)
8-GB RAM
64-bit Windows OS

99.8% results can be seen in "Punching, Pushing, and Shaking Hands" classes. SNSP successfully classified all other classes. SNSP algorithm is shown in Table 1, whereas computational time analysis is represented in Table 2.

7 Conclusion

In research work, we have projected an innovative skeleton-based approach to 2D human skeleton action recognition. At initial, we developed a novel SNSP descriptor with skeleton joints. Features are extracted using super-joints, standard normal, space, and parameter space. Our results of the experiments have shown that the proposed SNSP descriptor has superior performance compared to several state-of-the-art systems on three different action datasets i.e., UTD Multimodal Human Action Dataset, KARD- Kinect Activity

Recognition Dataset, and SBU Kinect Interaction Dataset, besides gained an accurateness of 99.2%, 98.3%, and 99.6% respectively.

Acknowledgements This work is supported by the Fundamental Research Funds for the Central Universities (Grant no. WK235000002).

Data availability Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

References

1. Gao Z, Wang P, Wang H, Mingliang Xu, Li W (2020) A review of dynamic maps for 3D human motion recognition using ConvNets and its improvement. *Neural Process Lett* 52(2):1501–1515
2. Islam MS, Bakhat K, Khan R, Iqbal M, Islam MM, Ye Z (2021) Action recognition using interrelationships of 3D joints and frames based on angle sine relation and distance features using interrelationships. *Appl Intell*, 1–13
3. Liao Z, Haifeng Hu, Liu Y (2020) Action recognition with multiple relative descriptors of trajectories. *Neural Process Lett* 51(1):287–302
4. Mishra SR, Mishra TK, Sanyal G, Sarkar A, Satapathy SC (2020) Real time human action recognition using triggered frame extraction and a typical CNN heuristic. *Pattern Recogn Lett* 135(2020):329–336
5. Li M, Leung H (2017) Graph-based approach for 3D human skeletal action recognition. *Pattern Recogn Lett* 87:195–202
6. Chen C, Jafari R, Kehtarnavaz N (2015) Utd-mhad: a multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor. In: 2015 IEEE international conference on image processing (ICIP), pp 168–172. IEEE
7. Gaglio S, Re GL, Morana M (2014) Human activity recognition process using 3-D posture data. *IEEE Transactions on Human-Machine Systems* 45(5):586–597
8. Yun K, Honorio J, Chattopadhyay D, Berg TL, Samaras D (2012) Two-person interaction detection using body-pose features and multiple instance learning. In: 2012 IEEE computer society conference on computer vision and pattern recognition workshops, pp 28–35. IEEE
9. Keller JM, Gray MR, Givens JA (1985) A fuzzy k-nearest neighbor algorithm. *IEEE Trans Syst Man Cybern* 4:580–585
10. Liu J, Wang G, Duan L-Y, Abdiyeva K, Kot AC (2017) Skeleton-based human action recognition with global context-aware attention LSTM networks. *IEEE Trans Image Process* 27(4):1586–1599
11. Liu J, Akhtar N, Mian A (2017) Skepxels: Spatio-temporal image representation of human skeleton joints for action recognition. *arXiv preprint arXiv:1711.05941*
12. Cippitelli E, Gasparrini S, Gambi E, Spinsante S (2016) A human activity recognition system using skeleton data from rgbd sensors. *Comput Intell Neurosci* 2016:21
13. Papadopoulos K, Antunes M, Aouada D, Ottersten B (2017) Enhanced trajectory-based action recognition using human pose. In: 2017 IEEE international conference on image processing (ICIP), pp 1807–1811. IEEE
14. Zhu W, Lan C, Xing J, Zeng W, Li Y, Shen L, Xie X (2016) Co-occurrence feature learning for skeleton based action recognition using regularized deep LSTM networks. In: Thirtieth AAAI Conference on Artificial Intelligence
15. Song S, Lan C, Xing J, Zeng W, Liu J (2017) An end-to-end spatio-temporal attention model for human action recognition from skeleton data. In: Thirty-first AAAI conference on artificial intelligence
16. Liu J, Wang G, Hu P, Duan L-Y, Kot AC (2017) Global context-aware attention LSTM networks for 3D action recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1647–1656
17. Baradel F, Christian W, Julien M (2017) Pose-conditioned spatio-temporal attention for human action recognition." *arXiv preprint arXiv:1703.10106*
18. Ke Q, An S, Bennamoun M, Sohel F, Boussaid F (2017) Skeletonnet: Mining deep part features for 3-d action recognition. *IEEE Signal Process Lett* 24(6):731–735
19. Ke Q, Bennamoun M, An S, Sohel F, Boussaid F (2017) A new representation of skeleton sequences for 3d action recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3288–3297
20. Escobedo E, Camara G (2016) A new approach for dynamic gesture recognition using skeleton trajectory representation and histograms of cumulative magnitudes. In: 2016 29th SIBGRAPI conference on graphics, patterns and images (SIBGRAPI), pp 209–216. IEEE

21. Li C, Hou Y, Wang P, Li W (2017) Joint distance maps based action recognition with convolutional neural networks. *IEEE Signal Process Lett* 24(5):624–628
22. Wang P, Li Z, Hou Y, Li W (2016) Action recognition based on joint trajectory maps using convolutional neural networks. In: *Proceedings of the 24th ACM international conference on Multimedia*, pp 102–106. ACM
23. Chikhaoui B, and Gouineau F (2017) Towards automatic feature extraction for activity recognition from wearable sensors: a deep learning approach. In: *2017 IEEE international conference on data mining workshops (ICDMW)*, pp 693–702. IEEE
24. Wang P, Wang S, Gao Z, Hou Y, Li W (2017) Structured images for RGB-D action recognition. In: *Proceedings of the IEEE international conference on computer vision*, pp 1005–1014
25. Gori I, Aggarwal JK, Matthies L, Ryoo MS (2016) Multitype activity recognition in robot-centric scenarios. *IEEE Robot Automat Lett* 1(1):593–600
26. Liu M, Junsong Y (2018) Recognizing human actions as the evolution of pose estimation maps. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 1159–1168
27. McNally W, Wong A, McPhee J (2019) STAR-Net: Action recognition using spatio-temporal activation reprojection. *arXiv preprint arXiv:1902.10024*
28. Ji Y, Ye G, Cheng H (2014) Interactive body part contrast mining for human interaction recognition. In: *2014 IEEE international conference on multimedia and expo workshops (ICMEW)*, pp 1–6. IEEE
29. Jin Ke, Jiang M, Kong J, Huo H, Wang X (2017) Action recognition using vague division DMMs. *J Eng* 2017(4):77–84
30. Islam MS, Iqbal M, Naqvi N, Bakhat K, Islam MM, Kanwal S, Ye Z (2019) CAD: Concatenated Action Descriptor for one and two Person (s), using Silhouette and Silhouette's Skeleton. *IET Image Processing*
31. Islam S, Qasim T, Yasir M, Bhatti N, Mahmood H, Zia M (2018) Single-and two-person action recognition based on silhouette shape and optical point descriptors. *SIViP* 12(5):853–860
32. Yu J, Rui Y, Tao D (2014) Click prediction for web image reranking using multimodal sparse coding. *IEEE Trans Image Process* 23(5):2019–2032
33. Yu J, Tao D, Wang M, Rui Y (2014) Learning to rank using user clicks and visual features for image retrieval. *IEEE Trans Cybern* 45(4):767–779
34. Yu J, Tan M, Zhang H, Tao D, Rui Y (2019) Hierarchical deep click feature prediction for fine-grained image recognition. *IEEE transactions on pattern analysis and machine intelligence*
35. Tianjin et al 88 Lemieux N, Noumeir R (2020) A hierarchical learning approach for human action recognition. *Sensors*, 20(17): 4946
36. Ranieri CM, Vargas PA, Romero RAF (2020) Uncovering human multimodal activity recognition with a deep learning approach. In: *2020 International joint conference on neural networks (IJCNN)*, pp 1–8
37. Mohite A, Rege P, Chakravarty D (2021) Human activity recognition using positioning sensor and deep learning technique. In: *Advances in signal and data processing*, Springer, pp 473–489
38. Dhiman C, Vishwakarma DK (2019) A robust framework for abnormal human action recognition using \mathcal{R} -transform and zernike moments in depth videos. *IEEE Sens J* 19(13):5195–5203
39. Saini R, Kumar P, Kaur B, Roy PP, Dogra DP, Santosh KC (2019) Kinect sensor-based interaction monitoring system using the BLSTM neural network in healthcare. *Int J Mach Learn Cybern* 10(9):2529–2540
40. Ashwini K, Amutha R (2020) Skeletal data based activity recognition system. In: *2020 International conference on communication and signal processing (ICCSP)*, pp 444–447
41. Ashwini K, Amutha R (2021) Compressive sensing based recognition of human upper limb motions with kinect skeletal data. *Multimed Tools Appl*, pp 1–19
42. Pham HH, Salmane H, Khoudour L, Crouzil A, Velastin SA, Zegers P (2020) A unified deep framework for joint 3d pose estimation and action recognition from a single rgb camera. *Sensors* 20(7):1825
43. Li S, Jiang T, Huang T, Tian Y (2020) Global co-occurrence feature learning and active coordinate system conversion for skeleton-based action recognition. In: *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pp 586–594
44. Xiaomin P, Huijie F, Yandong T (2020) Two-person interaction recognition based on multi-stream spatio-temporal fusion network. *红外与激光工程*, 49(5), 20190552
45. Huynh-The T, Hua C-H, Ngo T-T, Kim D-S (2020) Image representation of pose-transition feature for 3D skeleton-based action recognition. *Inf Sci (Ny)* 513:112–126
46. Proffitt DR, Gilden DL (1989) Understanding natural dynamics. *J Exp Psychol Hum Percept Perform* 15(2):384
47. Youdas JW, Garrett TR, Suman VJ, Bogard CL, Hallman HO, Carey JR (1992) Normal range of motion of the cervical spine: an initial goniometric study. *Phys Ther* 72(11):770–780