

An Efficient Activity Detection System based on Skeleton Joints Identification

Abdul Lateef Haroon P.S, U. Eranna

Department of Electronics & Communication Engineering, Ballari Institute of Technology and Management, India

Article Info

Article history:

Received Mar 31, 2018

Revised Jun 29, 2018

Accepted Aug 12, 2018

Keyword:

Computer Vision

Depth Image

Human Activity Recognition

Joints

Skeleton

ABSTRACT

The increasing criminal activities in the current world has drawn lot of interest activity recognition techniques which helps to perform the sophisticated analytical operations on human activity and also helps to interface the human and computer interactions. From the existing review analysis it is found that most of the existing systems are not emphasize on computational performance but are more application specific by identifying specific problems. Hence, it is found that all the features are not required for accurate and cost effective human activity detection. Thus, the human skeleton action can be considered and presented a simple and accurate process to identify the significant joints only. From the outcomes it is found that the proposed system is cost effective and computationally efficient activity recognition technique for human actions.

Copyright © 2018 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

Abdul Lateef Haroon P.S

Department of Electronics & Communication Engineering,

Ballari Institute of Technology and Management,

Karnataka, India.

E-Mail: abdulbitm@gmail.com

1. INTRODUCTION

The advancement in sensing mechanisms it is observed that automated sensing techniques are evolved in research state of art [1]. In order to capture the various informations based on different kinds of commercial products that work works on advanced sensing technologies [2]. The human detection system is one of the sensing techniques that utilize different kinds of sensors [3]. The human activity detection systems are applicable for human behavior from a given scene. Also, these systems are used to interface the human activities with a system and perform a specific task. But, most of the applications demands better analysis of complex motions of human and have user friendly interface with the system. Most of the data captured from visual sensors are in the form of depth maps, RGB data, nodal points as joints of skeletal, etc [4]. Kinect sensor is an action detection approach that offers all the forms of information from the input image in different combination [5]. More number of researches have considered Kinect sensor to identify different set of problems [6]-[10]. All these informations are acts as input for applications like video surveillance system, computer and human interaction; contextual attribute based multimodal retrieval, etc. It is also observed that various kinds of abstraction are existing to perform human activity analysis and is also known as activity, gesture, action, etc. These approaches can be varied with potential features. Basically, gestures can be termed as primary motion and is highly specific to the part of human body. However, action relates to activities undertaken by one perform with combination of different gestures. The action is also involved with temporal factor associated with generation of multiple gestures and represents movement of complete human body. However, from computer vision prospective, the recognition system for both action and gesture are found to be interchangeably utilized over different set of recognition problem. The fundamental methodology applied for recognition of human action is to obtain the features from the motion aspect of video or image sequences

in order to facilitating forecasting of one or specific set of actions. However, one of the bigger set of dependencies associated with recognition of human activity is that of feature extraction process. More the number of features will always ensure good accuracy of recognition system. However, such mechanism of capturing higher number of feature will call for two potential problems e.g. i) large processing time and ii) high resource utilization. Both of these two points are detrimental to ensure computational performance of system. Although, an significant growth in research-based techniques related to human activity recognition is exist, but decision to selections effective features using cost effective computational model is highly overlooked. Hence, there is a need of a system which can ensure the significant identification for given input. This paper introduces a novel mechanism which is cost-effective identification of features in the form of significant joints considering skeletal modalities for effective design of recognition system of human actions.

The recent past has witnessed lot of researches towards addressing the human activity recognition system related problems. The current section gives some of the recent researches performed in this field. In the work of Chen et al. [11] and [12] given a wavelets based algorithm which helps to identify and classify the detected human activity by using supervised learning and training approach for training. In De et al. [13], a dictionary learning-based technique by utilizing sparse signal representation for activity recognition. A qualitative and experimental-based human activity analysis by Fullerton et al. [14] in which k-nearest neighboring classifier is used to enhance accuracy score. Author Gavrilova et al. [8] performed an investigation on similar activity from the v sensor. In, Hbali et al. [15] skeleton based approach is utilized for joints for similar activity recognition system. Further, Bayesian based approach is introduced in Hernandez et al. [16] to construct a segmentation technique for gait recognition system. The work of Jain and Kanhangad [17] has used gyroscope and accelerometers to address the classification problems in human activity recognition system. Correlation-based approach was found in Khan et al. [18] to minimize the dimensional attributes involved in it along with usage of feature vector. Manzi et al. [19] have offered a Skeleton-based human activity identification system. A unique context towards approach is given in Noor and Uddin [20] to improve the accuracy level of the activity recognition by using neural network based training. Further, Savvaki et al. [21] utilized a hankel structure to represent image streams for facilitating better classification performance. Sikder and Sarkar [22] utilized a distance-based approach on motion data along with linear regression approach to identify the dynamic human activities and achieved higher accuracy in its classification performance. The work of Ulhaq et al. [23] used a space-time correlation based mechanism for three dimensional tensor structures to perform identification of the human actions. With the experimentation on video dataset it has achieved increased accuracy. Vishwakara and Singh [24] used an energy-factors associated with silhouette image which uses temporal contents of the dataset using transform-based approach. The work associated with Wang et al. [25] gives context-based methodology based on predictive approach for activity monitoring system. The network channel based human activity identificant system was discussed in Wang et al. [26]. The work of Xu et al. [27] given hierarchical-based approach using both distance and time for human activity monitoring system. In, Yang et al. [28] super normal vector is used to perform aggregation of the different discriminative variables for identification of human activity. The human motion tracking is given in Zhao et al. [29]. Hence, there are various cadres of recent techniques towards human activity recognition system. The next section discusses about problems encountered in the above discussed existing literatures.

The research problems associated with activity identification system are as follows:

- a. The existing studies doesn't emphasize on computational complexities associated with the selection of large number of features from a given set of data of human actions.
- b. The possibility of minimizing the computational effort of feature extraction using joint-based attribute in skeleton system has achieved less attention among research communities.
- c. At present, there are very less number of literatures dedicated for exploring an effective number of joints responsible for performing recognition of human activity.
- d. Optimization of usage of depth map has less number of researches and more number of research are towards applying machine learning to offer higher precision.

Therefore, the problem statement of this work is "To develop a cost effective system to minimize the computational effort of feature extraction to leverage the performance of recognition system of human action is computationally challenging".

The proposed system aims to ensure a computationally cost effective mechanism to identify the significant joints in human activity recognition system. The significance of this mechanism is that the proposed system targets to achieve a significant balance among higher accuracy in recognizing human activities with extremely low computational efforts. The computational effort minimization is only feasible if the solution aims for better form of feature extraction mechanism. The pictorial representation of Figure 1 indicates the proposed system.

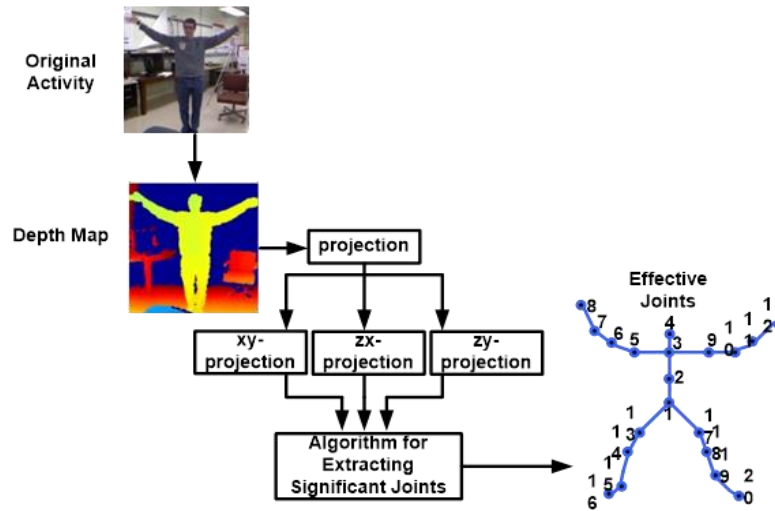


Figure 1. Proposed system

With the implementation of the analytical research approach in the proposed system, the input image sequences are considered with defined standard human actions. Then the depth map of the image will be extracted from input image dataset which is further resumed for projected segments extraction on three different planes of x-y-z axis. An algorithm is developed to assess all the joints of skeleton of depth image by using identity-based attributes. The process outcomes with effective number of joints which helps in significant way of human activity recognition with higher accuracy. The proposed system offers a faster joint processing irrespective of any selection of motion patterns as well as it doesn't offer any form of dependencies towards computing unwanted number of joints for carrying of recognition of an effective human action. The following section idealizes the algorithm implementation of proposed system.

2. ALGORITHM IMPLEMENTATION

The algorithm is responsible to read the input and apply a simple processing to identify the significant joints. This algorithm addresses the problems associated with the different literature and it is also observed that there is a greater amount of dependencies towards using feature extractor mainly when image area is quite large. This results in increased computational complexity. Moreover, it was also explored that level of precision is very specific to different kinds of dataset and is found quite low. Hence, the developed algorithm utilizes the extraction concept for significant regions of the image area. The targeted advantage is to obtain inclusion of lesser image area involved in extraction of feature. It is because not all the joint aspect information is useful to perform the human activity identification. Following are the steps involved in proposed algorithm:

Algorithm for Extracting Significant Joints

Input: a, j, J

Output: D_{joint}

Start

1. $\text{init } a, j, J$
2. $\theta \rightarrow \text{read}(a_{\text{id}}, s_{\text{id}}, e_{\text{id}})$;
3. $d_{\text{ma}} \rightarrow f(\theta)$
4. $C \rightarrow g(\theta)$
5. Extract (x, y, z) from C
6. **For** $s=1:(X)$
7. $S \rightarrow [X, Y, Z]$
8. **End**
9. **For** $s=2:(X)$
10. $[S_p, S] \rightarrow [X, Y, Z]$
11. $D_{\text{joint}} \rightarrow [\text{Euc}(S)]$
12. **End**

End

The algorithm of the proposed system considers the input of a (set of activity), j (joint tag), J (connection of skeleton joints) that after processing results in output of D_{joint} (Display of joints) (Line-1). The implementation is related to defining the action set as well as name of joints as following. Figure 2 show the defined set of activity considered for human activity recognition and Figure 3 show the defined name of joints considered for human activity recognition.

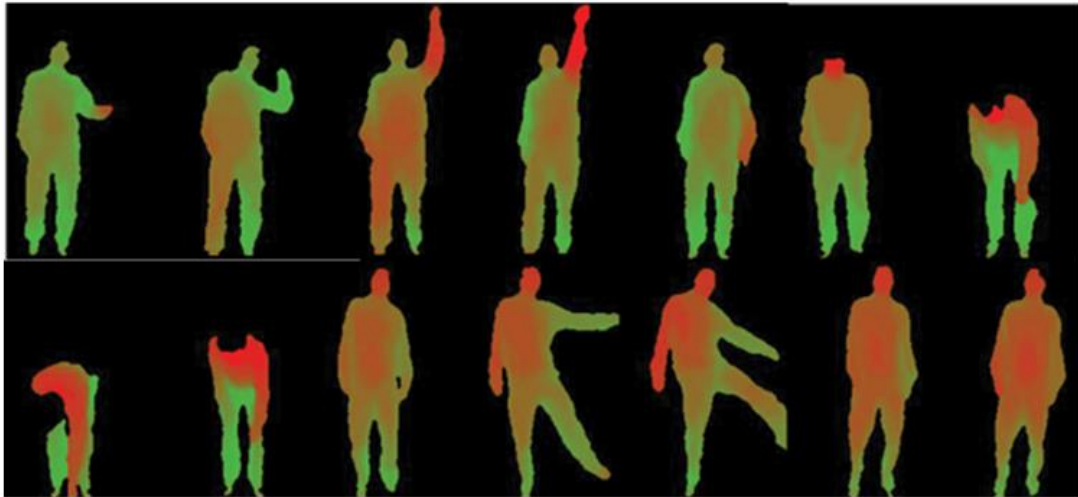


Figure 2. Defined set of activity considered for human activity recognition

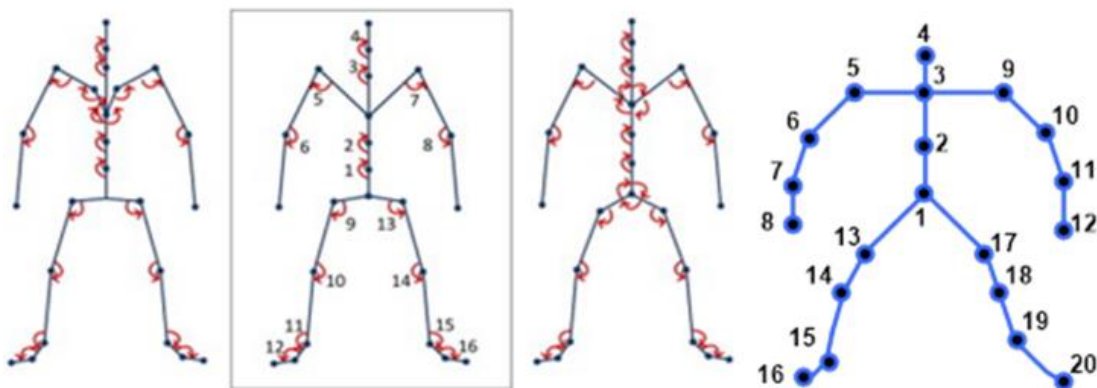


Figure 3. Defined name of joints considered for human activity recognition

The next implementation is associated with defining the identities associated with name of action (a_{id}), identity of the subject (s_{id}), identity of the example (e_{id}). All these identity-related information is read and stored in a matrix θ (Line-2). Thus, it can be said that a data file is constructed on the basis of this identities and are used to extract depth information from the given image file. Further, a new function $f(x)$ is implemented which takes the input of location of all the image file and result in depth image as an outcome. The function takes the input of the newly obtained matrix θ (Line-3) that finally results in an array of depth matrix d_{ma} (Line-3). The input J that is initialized in Line-1 is basically used for specifying the connections among all the joints of skeleton followed by construction an input file on the basis of an action. A simple normalization process of all the coordinates were carried out followed by reshaping all the coordinates in order to obtain four row elements along with transposition operation. It is further followed by reshaping to total number of joints.

The complete process of extracting the coordinates from the input image is highlighted in Figure 4. According to this process, the algorithm reads the identity-based information and constructs a new file, which is checked for its non-zero elements. In case of non-zero elements, the algorithm constructs a data B from the

text file using a function g (Line-4) followed by series of matrix-processing functions $func$ to the obtained matrix B that finally results in a temporary matrix C (Line-4). The matrix C is used for obtaining all the coordinate information (Line-5). Then, depth is visualized following by evaluating the skeleton of each associated frames. The algorithm reads all the frames considering the X coordinates (Line-6) in order to obtain s^{th} frame in all the three directions (Line-7). The next part of the algorithm computes all the frames in order to compute the displacement of the joint coordinates (Line-9). A new matrix S_p is constructed in order to extract all the coordinates of the prior frame as well as coordinates of the present frame i.e. S (Line-10). This operation is followed by computation of a Euclidean distance between the prior and current frame i.e. S_p and S that gives the final displacement. This computation is carried out for all the frames so that the algorithm could dynamically compute the joints and display it for all the frames (Line-11). Therefore, the algorithm exhibits a most simplest and cost effective mechanism for significant joints identification which are the primary backbone information required for every application of human activity recognition system. The analysis of the obtained results from this algorithm is given below.

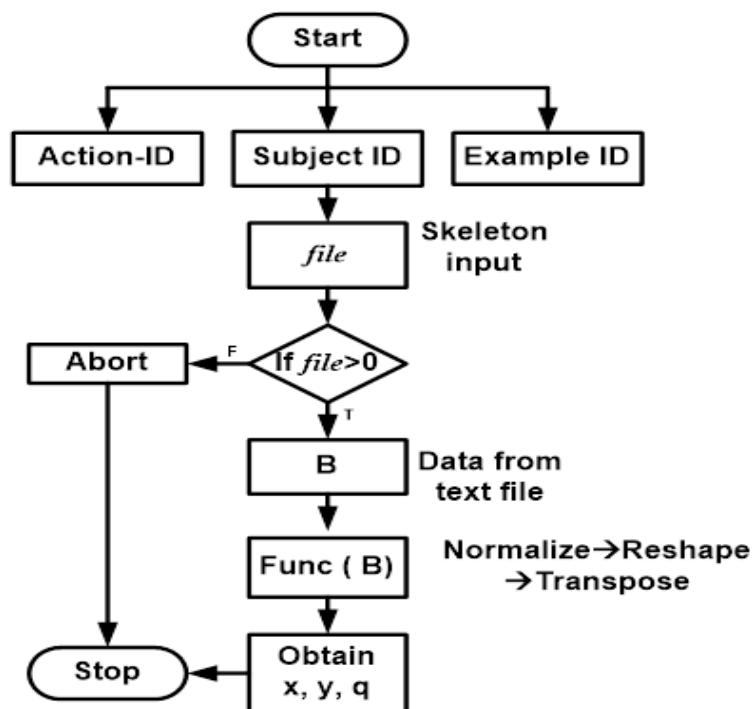
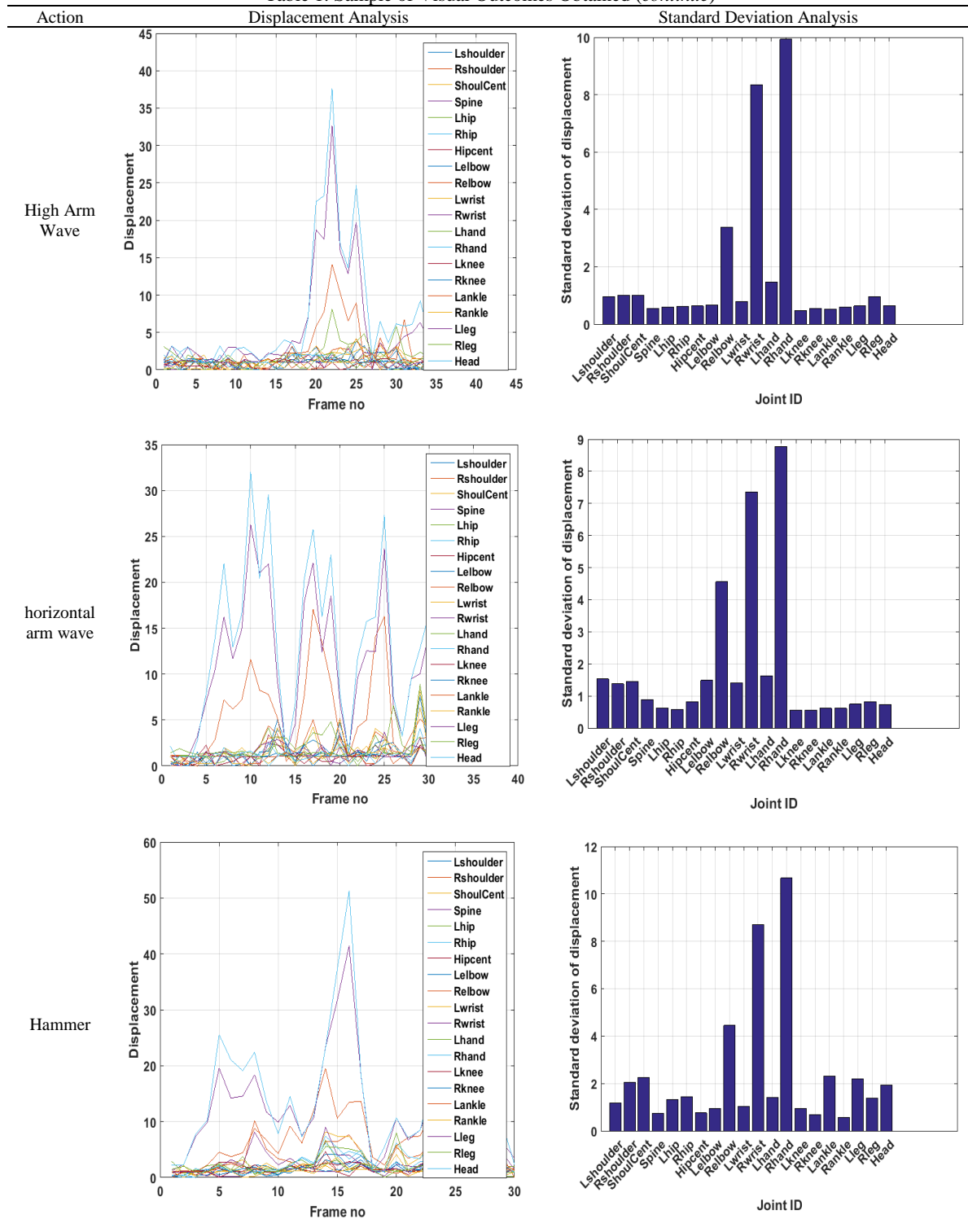


Figure 4. Process of obtaining coordinates

3. RESULT ANALYSIS

The proposed system is implemented in MATLAB by considering the input dataset from MSR Action 3D dataset [30]. The algorithm is scripted and evaluated using 20 action sets which targets for precise identification of effective joints. Hence, it is individually assessed using performance parameters of displacement and standard deviation with respect to each discrete human action defined in the dataset. By monitoring displacement, it could be possible to assess how uniquely the proposed system is capable of identifying the body in motion and by monitoring standard deviation; it is feasible to determine the significant joints. Table 1 gives only 3 sample visual outcomes obtained from proposed system.

Table 1. Sample of Visual Outcomes Obtained (continue)



Apart from the individual outcome, the proposed system is also benchmarked using standard training algorithm access its accuracy performance. The benchmarking process is as follows: The proposed technique is allowed for training using two frequently used classifiers e.g. K-nearest neighbor algorithm (KNN) and Support Vector Machine (SVM). The analysis of accuracy is performed by considering Sensitivity and Specificity as performance parameter with respect to increasing training ratio.

The Figure 5 and Figure 6 indicates the proposed system which gives higher accuracy (with respect to specificity and sensitivity) when these are trained using KNN algorithm for all the significant joints. From this outcome, it can be said that it is absolutely not necessary to select all the joints for performing human activity as identification process is performed by using only the selected joints offer better accuracy performance. The performance of accuracy using KNN is much significantly better as compared to conventional SVM algorithm.

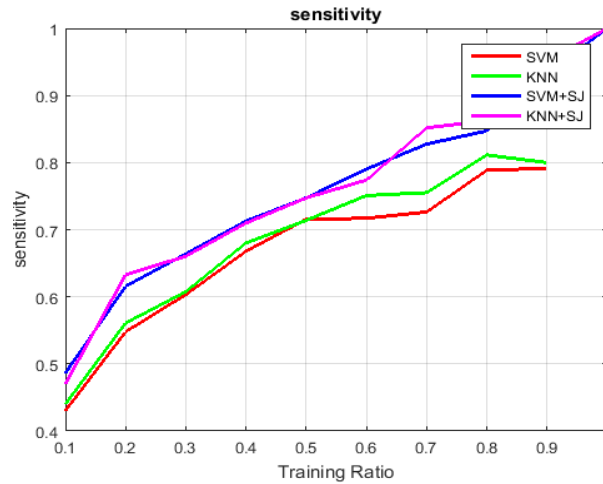


Figure 5. Analysis of sensitivity

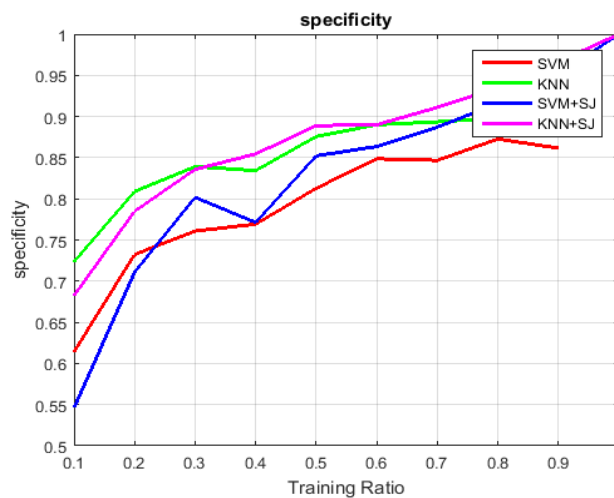


Figure 6. Analysis of Specificity

4. CONCLUSION

This paper gives a very simple and yet novel approach of enhancing the human activity recognition system performance. The prime basis of the work carried out is that existing mechanism performs similar objective by selection all the features in either form of depth map image or skeleton image, whereas, the proposed theory imposes that it could strike a computational complexity by considering so many feature points. This problems can be sorted if a mechanism is designed that could identify only the significant points in the skeleton image. The proposed study is reflected with significant outcomes that indicates higher accuracy performance when trained using KNN.

REFERENCES

- [1] B Vinod, Richard Voyles, Prahlad Vadakkepat, M. Sundaram, K S Sujatha, J Joe Brislin, "Advancements in Automation, Robotics and Sensing," *Springer-Computers*, 2016
- [2] Chong-Min Kyung, Hiroto Yasuura, Yongpan Liu, Youn-Long Lin, "Smart Sensors and Systems: Innovations for Medical, Environmental, and IoT Applications," *Springer-technology and Engineering*, 2016
- [3] John Vetelino, Aravind Reghu, "Introduction to Sensors," *CRC press*, 2017
- [4] Ling Shao, Jungong Han, Pushmeet Kohli, Zhengyou Zhang, "Computer Vision and Machine Learning with RGB-D Sensors," *Springer-Computers*, 2014
- [5] Mansib Rahman, "Beginning Microsoft Kinect for Windows SDK 2.0: Motion and Depth Sensing for Natural User Interfaces," *Apress-Computer*, 2017
- [6] B. Ni, Y. Pei, P. Moulin and S. Yan, "Multilevel Depth and Image Fusion for Human Activity Detection," in *IEEE Transactions on Cybernetics*, vol. 43, no. 5, pp. 1383-1394, Oct. 2013.
- [7] Y. Yang, F. Pu, Y. Li, S. Li, Y. Fan and D. Li, "Reliability and Validity of Kinect RGB-D Sensor for Assessing Standing Balance," in *IEEE Sensors Journal*, vol. 14, no. 5, pp. 1633-1638, May 2014.
- [8] M. L. Gavrilova, Y. Wang, F. Ahmed and P. Polash Paul, "Kinect Sensor Gesture and Activity Recognition: New Applications for Consumer Cognitive Systems," in *IEEE Consumer Electronics Magazine*, vol. 7, no. 1, pp. 88-94, Jan. 2018
- [9] B. Dehbandi *et al.*, "Using Data From the Microsoft Kinect 2 to Quantify Upper Limb Behavior: A Feasibility Study," in *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 5, pp. 1386-1392, Sept. 2017.
- [10] J. Han, L. Shao, D. Xu and J. Shotton, "Enhanced Computer Vision With Microsoft Kinect Sensor: A Review," in *IEEE Transactions on Cybernetics*, vol. 43, no. 5, pp. 1318-1334, Oct. 2013.
- [11] Y. Chen and C. Shen, "Performance Analysis of Smartphone-Sensor Behavior for Human Activity Recognition," in *IEEE Access*, vol. 5, pp. 3095-3110, 2017.
- [12] Z. Chen, Q. Zhu, Y. C. Soh and L. Zhang, "Robust Human Activity Recognition Using Smartphone Sensors via CT-PCA and Online SVM," in *IEEE Transactions on Industrial Informatics*, vol. 13, no. 6, pp. 3070-3080, Dec. 2017.
- [13] P. De, A. Chatterjee and A. Rakshit, "Recognition of Human Behavior for Assisted Living Using Dictionary Learning Approach," in *IEEE Sensors Journal*, vol. 18, no. 6, pp. 2434-2441, Mar 2018.
- [14] E. Fullerton, B. Heller and M. Munoz-Organero, "Recognizing Human Activity in Free-Living Using Multiple Body-Worn Accelerometers," in *IEEE Sensors Journal*, vol. 17, no. 16, pp. 5290-5297, Aug. 2017.
- [15] Y. Hbali, S. Hbali, L. Ballihi and M. Sadgal, "Skeleton-based human activity recognition for elderly monitoring systems," in *IET Computer Vision*, vol. 12, no. 1, pp. 16-26, 2018.
- [16] U. Martinez-Hernandez, I. Mahmood and A. A. Dehghani-Sanij, "Simultaneous Bayesian Recognition of Locomotion and Gait Phases with Wearable Sensors," in *IEEE Sensors Journal*, vol. 18, no. 3, pp. 1282-1290, 2018.
- [17] A. Jain and V. Kanhangad, "Human Activity Classification in Smartphones Using Accelerometer and Gyroscope Sensors," in *IEEE Sensors Journal*, vol. 18, no. 3, pp. 1169-1177, Feb 2018.
- [18] M. U. S. Khan *et al.*, "On the Correlation of Sensor Location and Human Activity Recognition in Body Area Networks (BANs)," in *IEEE Systems Journal*, vol. 12, no. 1, pp. 82-91, Mar 2018.
- [19] A. Manzi, L. Fiorini, R. Limosani, P. Dario, Fi. Cavallo, "Two-person activity recognition using skeleton data", *IET Computer Vision*, vol. 12, Issue. 1, pp. 27-35, 2017
- [20] S. Noor, V. Uddin, "Using context from inside-out vision for improved activity recognition", *IET Computer Vision*, vol. 12, Issue. 3, pp. 276-287, 2018
- [21] S. Savvaki, G. Tsagakatakis, A. Panousopoulou and P. Tsakalides, "Matrix and Tensor Completion on a Human Activity Recognition Framework," in *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 6, pp. 1554-1561, Nov. 2017.
- [22] F. Sikder and D. Sarkar, "Log-Sum Distance Measures and Its Application to Human-Activity Monitoring and Recognition Using Data from Motion Sensors," in *IEEE Sensors Journal*, vol. 17, no. 14, pp. 4520-4533, Jul 2017.
- [23] A. Ulhaq, X. Yin, J. He and Y. Zhang, "On Space-Time Filtering Framework for Matching Human Actions Across Different Viewpoints," in *IEEE Transactions on Image Processing*, vol. 27, no. 3, pp. 1230-1242, Mar 2018.
- [24] D. K. Vishwakarma and K. Singh, "Human Activity Recognition Based on Spatial Distribution of Gradients at Sublevels of Average Energy Silhouette Images," in *IEEE Transactions on Cognitive and Developmental Systems*, vol. 9, no. 4, pp. 316-327, Dec. 2017.
- [25] L. Wang, X. Zhao, Y. Si, L. Cao and Y. Liu, "Context-Associative Hierarchical Memory Model for Human Activity Recognition and Prediction," in *IEEE Transactions on Multimedia*, vol. 19, no. 3, pp. 646-659, Mar 2017.
- [26] W. Wang, A. X. Liu, M. Shahzad, K. Ling and S. Lu, "Device-Free Human Activity Recognition Using Commercial WiFi Devices," in *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 5, pp. 1118-1131, May 2017.
- [27] W. Xu, Z. Miao, X. P. Zhang and Y. Tian, "A Hierarchical Spatio-Temporal Model for Human Activity Recognition," in *IEEE Transactions on Multimedia*, vol. 19, no. 7, pp. 1494-1509, Jul 2017.
- [28] X. Yang and Y. Tian, "Super Normal Vector for Human Activity Recognition with Depth Cameras," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 5, pp. 1028-1039, May 2017.
- [29] W. Zhao, M. A. Reinthal, D. D. Espy and X. Luo, "Rule-Based Human Motion Tracking for Rehabilitation Exercises: Realtime Assessment, Feedback, and Guidance," in *IEEE Access*, vol. 5, pp. 21382-21394, 2017.
- [30] "Datasets", http://users.eecs.northwestern.edu/~jwa368/my_data.html, Retrieved on 23th July, 2018

BIOGRAPHIES OF AUTHORS

Abdul Lateef Haroon P.S., currently working as an Assistant Professor in the Department of Electronics and Communication Engineering, Ballari Institute of Technology and Management. I have completed B.E from RYMEC College, Ballari during 2008-2012 with aggregate 72%; I have done M. Tech from KSSEM College, Bangalore during 2012-2014 with aggregate 84%. I am Currently Pursuing Ph.D. from VTU with BITM College as a Research centre. I have published 7 IEEE Conference Papers and 20 International Journals. My areas of Interest are Digital Signal Processing, Digital Image Processing, Computer Vision, ARM Microcontrollers, and Neural Networks.



U. Eranna, currently working as Professor and HOD in the Department of Electronics and Communication Engineering, Ballari Institute of Technology and Management. He has published 8 conference papers and 20 journal papers. His area of research was Communication, Control System.