International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-4, April 2019

Adaptive Kernels for Skeleton Based Action Recognition using Geometric Feature Score Fusion

K. Rajendra Prasad, P. Srinivasa Rao

Abstract*: This paper presents a novel adaptive kernel based method for classifying the human actions from the skeletal data. Three types of geometric features: joint positions, joint relative distances and joint relative angles were used for action recognition using the information sensed by Microsoft Kinect sensors. Recently, kernel-based methods mostly focused on the action recognition on spatio-temporal data. In this work, the obtained scores from the individual adaptive kernels which are inputted three different geometric feature namely joint positions, joint relative distances and joint relative angles features were fused to detect the action by using simple kernel matching technique which evaluates the similarity between the features. This method is used to rank the actions from the database according to the highest ranked query action. The experimental results conducted over three publicly available datasets, i.e., NTU RGB-D, G3D and UTD-MHAD. The proposed technique has been tested and compared with other state-of-the-art methods on above datasets. The proposed method performed well on all datasets and good recognition rates were recorded.

***Index Terms*: Geometric features, human action recognition, Microsoft Kinect sensor, adaptive kernels.**

I. INTRODUCTION

Human action recognition (HAR) and detection has remained as a challenging and interesting topic in computer vision and pattern recognition. In last few decades, research has been done widely in the area of human action recognition using 2D color video streams. Many researchers published their work based on motion and shape features. However, segmentation of RGB images is complex and missing the depth information. Recently, low cost Microsoft Kinect sensor is playing virtual role in HAR from RGB, RGB-D and 3D skeleton. The sample 2D RGB image associated with RGB-D (Depth maps) and 3D skeleton joint positions captured by Microsoft Kinect V.2 sensor are shown in fig.1.

However, the human action recognition techniques still have problems to representing the structure of actions. To overcome this problem, many researchers have proposed different approaches to extract spatial and temporal features form a set of skeleton joints/points in 3D coordinates [1]. In the present work, we focus on recognizing human actions from skeleton inputs other than RGB and Depth maps from above problems. The skeletons are high level data describing human's motions only.

In recent, joint coordinate, edges, distance, angle, velocity, surface, joint-joint distance, joint-line distance, joint-plane distance, line-plane angle, tree traversal, joint trajectories, spherical angle [1-4] and fusions of these features were used to designate the spatio-temporal information of an actions in 3D video streams. These features are used directly or color coded into images for human action recognition task. These features were inputted to convocational neural networks (CNN) [5], recurrent neural networks (RNN) [2] and several machine learning techniques [4,5] for performing the recognition activity. Even through their outstanding performance was consistent with large training sets, the training and testing times are the considerable disadvantages.

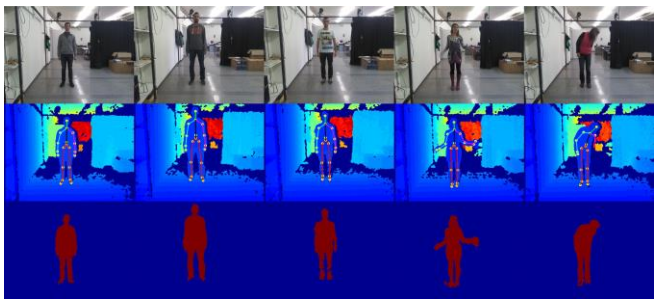

Fig.1. RGB images (top), 3D skeleton joints (center) and depth map (bottom) of sample actions from a publicly available dataset.

In this work, we are proposed to use adaptive kernel for human action recognition using a sequence of skeleton data. As a first step, we extract the three geometric features on skeleton joints 3D coordinates. During features extraction methods, a set of geometric features are considered as joint position, joint distance features and joint angle features are visualized in fig.2. These geometric features are inputted to the adaptive kernels for finding similarity between the query action and the database actions. The proposed algorithm tested against the state-of-the-art methods over three publicly available datasets such as NTU RGB-D [6], G3D [7], UTD-MHAD [8] datasets. The remaining of the paper is structed in the following way: in session 2, literature review on skeleton based action recognition is presented. Feature extraction methods, and adaptive kernel techniques are explained in session 3.

Manuscript published on 30 April 2019.
***** Correspondence Author (s)
K. Rajendra Prasad, Research Scholar, Department of Computer Science and Systems Engineering, College of Engineering, Andhra University, Visakhapatnam, India.
P. Srinivasa Rao, Professor, Department of Computer Science and Systems Engineering, College of Engineering, Andhra University, Visakhapatnam, India.

In session 4, description of action datasets, experimental results obtained in proposed method and state-of-the-art techniques were discussed. Finally, session 5 presents the conclusion and feature work.

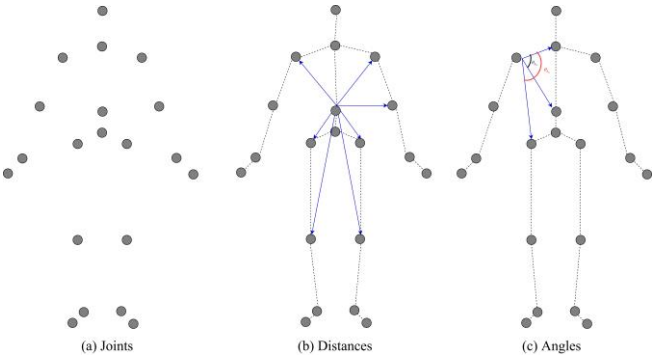

Fig.2. Feature representation of a Kinect 3D human skeleton representing, (a) joint positions, (b) joint distances, (c) joint angles.

II. LITERATURE REVIEW

In the following session, we briefly review different action recognition techniques. Over the last few decades, RGB images has been extensively used in human action recognition techniques. These techniques depends on various features extracted from the techniques such as histogram of oriented gradients (HOG), scale-invariant feature transform (SIFT), local binary patterns (LBP), optical flow, histogram of flow (HOF), locally binary flow patterns (LBFP) features as input to the classifiers such as neural network, support vector machine (SVM), hidden Markov model (HMM) and other deep learning methods [9-11] for human action recognition. There are still unsolved problems in RGB based human action recognition. In particular depth and skeleton models were shown greater recognition rates compared to RGB based models. Recently, with the advent of low-cost sensors such as Microsoft Kinect sensors, research on human action recognition has shifted to sequence of depth maps and skeleton joints. The advantages of RGB-D sensors over 2D color image cameras are their invariance to color and lighting environments. Using both depth maps and skeleton joints, Althloothi et al. [12] proposed a method based on two sets of features, shape features, kinematic structure. Shape features were extracted from depth maps and the kinematic features extracted from the 3D joint positions of a human body and fused together using multiple kernel learning (MKL) methods for human action recognition. The skeleton based HAR is now become a major focus in computer vision research due to its demand in real time applications. Basically the skeleton based HAR has been divided into three categories : 1) joints, that are captures the interdependence of the human body joints by extracting geometric descriptor [1,2], 2) spatial descriptor, a group of joints aims to trying to learn subset of joints to discriminate among the actions, 3) joint dynamics focusing on modelling the dynamics of all joints or subsets of a skeleton in time series.The earlier algorithms were based on the joint parameters of a human body for recognize actions, e.g., positions, distance, angle, velocity, acceleration, surface, area, perimeter. Wang et al. [2] extracted three geometric features form 3D skeleton positions, i.e., joints, edges and surfaces and using recurrent neural networks (RNN) algorithm for detecting and recognizing the actions. Further to increase the recognition, zhang songyang et al. [1] extracted eight geometric features: joint coordinate, joint-joint orientation, joint-joint distance, joint-line distance, joint-plane distance, line-line angle, line-plane angle and plane-plane angle and are enumerated in [1]. The authors combined these features to perfectly identify the motion and pose representations. Lu Xia et al. [13], proposed a novel approach for human action recognition used histograms of 3D joint locations (HOJ3D) via 3D skeletal joint locations to achieve view invariant posture representation. For classification of skeleton based human action recognition, popular classifiers like artificial neural network (ANN), adaptive graph matching (AGM), dynamic time warping (DTW), weighted graph matching (WGM), support vector machine (SVM), histograms and dictionary learning (DL) were used in the past. Despite of their success, these algorithms showed boundaries towards large dataset. Current deep learning algorithms like recurrent neural network (RNN) [2], convolutional neural network (CNN) [5], Long short-term memory (LSTM) [1] were effectively applied to action recognition. Even though the performance was constant, the noticeable disadvantages are requirement of large training sets and training times. To overcome this problem, kernel-based techniques have been proposed by many researchers. These kernel-based techniques [14] perform well and can also be trained using the SVM based kernel learning approaches. These adaptive kernels performs the similarity search between each action sequence in dataset and query action. In this work, we propose an effective adaptive kernel based method for recognizing the human actions using sequence of skeleton frames. We extract three geometric features on 3D skeleton coordinates and are applied over an adaptive kernels for human action recognition by performing similarity check. The proposed technique is compared against the state-of-the-art methods to know the novelty. It has been observed that the proposed method is giving good recognition rates on various skeleton based datasets when compared to others.

III. PROPOSED METHODOLOGY

The proposed methodology consists of three major components, as shown in fig.3. We extract three geometric features in each frame. First we consider the skeleton joint features alone as input features for the adaptive kernel and the obtained scores were recorded. Second, the Euclidian distances between the joint pairs were calculated and are inputted to the second kernel for matching action. Thirdly, we consider the calculated joint angles among all joints as an input to the third kernel. A similarity score between query action and target action in the dataset is calculated by using adaptive kernels. The score fusion from the three features is performed for final action classification. The operation of these three components is detailed as follows.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

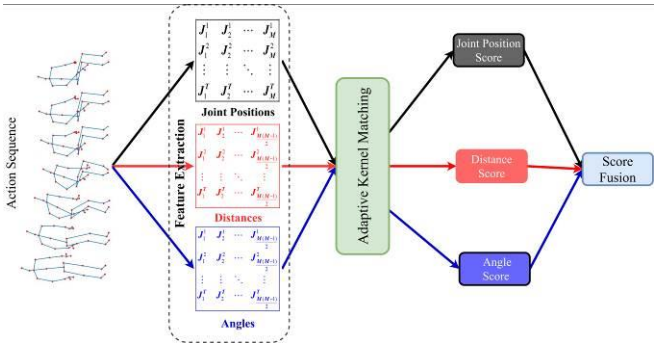

Fig.3. Framework of the proposed method.

A. Extraction of Geometric Features

This section discusses about the three types of geometric features used in this work.

Joint Positions (JP):

The motion sequence of a human action consists of 3D locations of the skeleton joints as x, y, z coordinates in a 3D space. The action sequence $S = \{F_1, F_2, ..., F_T\}$ is consists of T frames and every skeleton frame consists of M joints, for example, in this work 20 joints is used, which are captured by Kinect sensor as shown in fig.2(a). Here $F_t = \{J_1^t, J_2^t, ..., J_M^t\}$ indicates the tth frame of the skeleton; let $J_i^t = [x_i^t, y_i^t, z_i^t] \in \mathbb{R}$ be the ith $(1 < i < M)$ joint at the tth $(1 < t < T)$ skeleton sequence. The dimensions of the joint position matrix $T \times M \times 3$. The frame sequence T is different for various actions. The features for joint positions $J_P(t)$ are quantified as

$$J_P(t) = J_i^t \;\; \forall \; i = 1:M, \; t = 1:T \tag{1}$$

Joint Distance Features (JDF):

Each skeleton joint J is represented with a 3D coordinate joint positions $J_i^t = [x_i^t, y_i^t, z_i^t] \in R^{3\times M}$ $\forall \; i = 1:M$ on every frame t. When searching the literature, most of the work is normalized the joint positions by using mean and variance. However, in this paper we use adaptive kernels for recognizing the action, which was build based on difference in joint positions rather than joint positions itself.

The joint distance features (JDF) measures the relational joint features between two joint positions J_i^t and J_{i+1}^t by calculating Euclidian distance. Where i is the skeleton joint index. The joint distance features $J_D(t)$ is mathematically described as,

$$J_D(t) = \left\| J_i^t - J_{i+1}^t \right\|_2^2 \;\; \forall \; i = 1 \; to \; M, \; t = 1 \; to \; T$$

(2)

The joint distance features of action sequence can be represented as J_D with matrix dimensions $\frac{M(M-1)}{2} \times T$. The matrix values represent the spatial change between joints in an action sequence. The visualization of JDF is given in fig.2(b).

Joint Angle Features (JAF):

Joint angle features (JAF) are calculated by choosing three joint J_1, J_2 and J_3, we can form two projection vectors, i.e. $\overrightarrow{P_{12}} = d(J_1, J_2) \in R^3$ and $\overrightarrow{P_{23}} = d(J_2, J_3) \in R^3$, where $d(J_i, J_j)$ is the distance between joints J_i and J_j. We can measure the angle θ_{J_2} at joint J_2 as

$$\theta_{J_2} = \cos^{-1} \frac{\overrightarrow{P_{12}} \overrightarrow{P_{23}}}{\sqrt{\overrightarrow{P_{12}}} \sqrt{\overrightarrow{P_{23}}}}$$

(3)

The joint angle features J_A for all the joints over all frames in an action can be expressed as

$$J_A(t) = \begin{bmatrix} \theta_{J_1}^1 & \theta_{J_2}^1 & \theta_{J_3}^1 & \dots & \theta_{J_M}^1 \\ \theta_{J_1}^2 & \theta_{J_2}^2 & \theta_{J_3}^2 & \dots & \theta_{J_M}^2 \\ \theta_{J_1}^3 & \theta_{J_2}^3 & \theta_{J_3}^3 & \dots & \theta_{J_M}^3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \theta_{J_1}^T & \theta_{J_2}^T & \theta_{J_3}^T & \dots & \theta_{J_M}^T \end{bmatrix}$$

(4)

The size of the J_A matrix is $\frac{M(M-1)}{2} \times T$. Fig.2(c) gives the pictorial representation of measuring an angle.

B. Adaptive Kernel Matching

In the present work, skeleton based action recognition using adaptive kernels was implemented. We used three geometric features to recognize an action: joint positions, joint distance and joint angle features. The main idea of kernels is to find the similarity between query action with the database actions to convert input skeleton sequence into text labels. We fuse three kernels for recognizing an action in the action database. Fig.4 shows the block diagram of adaptive kernel matching proposed in this paper.

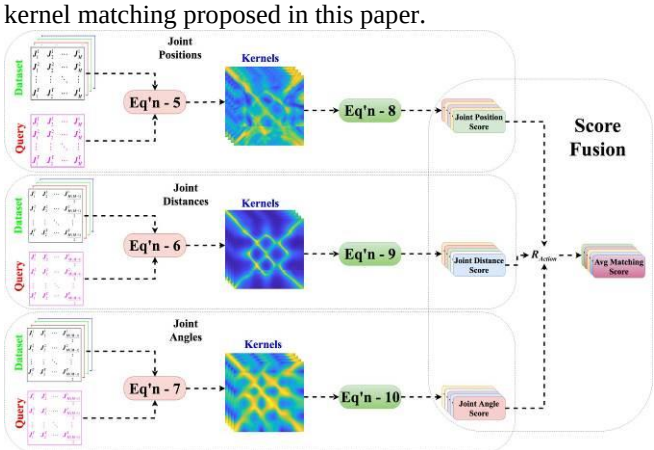

Fig.4. Adaptive kernel construction module.

Let $Q(J_P, J_D, J_A)$ and $D^i(J_P^{'}, J_D^{'}, J_A^{'})$ are the query and the dataset actions respectively. Where i is the action index. Here J_P, J_D and J_A are the metrics of JP, JDF and JAF for each action. Where the JP, JDF and JAF adaptive kernels in an action data for two actions from query and database is represented as $K_P(J_P, J_P^{'})$, $K_D(J_D, J_D^{'})$ and $K_A(J_A, J_A^{'})$ respectively. Here J_P and $J_P^{'}$ are the joint position values of the query action and the database action.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

Similarly, $(J_D, J_D^{'})$ and $(J_A, J_A^{'})$ are joint distance and joint angles of the query and the action from database. The 3D trajectory position kernel is defined as

$$K_P(J_P, J_P^{'}) = \exp\left(\frac{\left\|J_P - J_P^{'}\right\|_2^2}{2\sigma_1^2}\right)$$

(5) Where J_P and $J_P^{'}$ are the action 3D trajectory positions for query and database action sequence Q and D^i. The gaussian kernel parameter σ_1 is very small $(\sigma_1 > 0)$. The distance kernel is given by

$$K_D(J_D, J_D^{'}) = \exp\left(\frac{\left\|J_D - J_D^{'}\right\|_2^2}{2\sigma_2^2}\right)$$

(6)

Where J_D and $J_D^{'}$ are the joint distance features in query and database action. Here $\sigma_2 > 0$ is a gaussian scale parameter. The angle kernel is defined as

$$K_A(J_A, J_A^{'}) = \exp\left(\frac{\left\|J_A - J_A^{'}\right\|_2^2}{2\sigma_3^2}\right) \quad (7)$$

Where J_A and $J_A^{'}$ are the orientations of the query and database actions. $\sigma_3 > 0$ is the variance. The constructed three kernels $K_P(J_P, J_P^{'})$, $K_D(J_D, J_D^{'})$ and $K_A(J_A, J_A^{'})$ for an action with the kernel matrix of size $T_Q \times T_D$. Where T_Q and T_D are the number of frames in query action and dataset action respectively. Cross value analysis shows the matching between query and database. The perfect match gives the maximum score. The proposed algorithm has two advantages: (i) independent of action location in the video frames and (ii) independent of action frames. The results from the one to many frames kernel matching between query and dataset all frames.

The classification of action label gives the maximum kernel value. The decision boundary for kernel matching scores can be given by

$$R_r^{J\_Trj} = \frac{1}{T_Q}\sum_{b \in T_Q} \arg \max_{r \in T_D}\left(K_P^r\left(J_P, J_P^{'}\right)\right) \quad (8)$$

$$R_r^{J\_Dis} = \frac{1}{T_Q}\sum_{b \in T_Q} \arg \max_{r \in T_D}\left(K_D^r\left(J_D, J_D^{'}\right)\right) \quad (9)$$

$$R_r^{J\_Ang} = \frac{1}{T_Q}\sum_{b \in T_Q} \arg \max_{r \in T_D}\left(K_A^r\left(J_A, J_A^{'}\right)\right)$$

(10)

Where $R_r^{J\_Trj}$, $R_r^{J\_Dis}$ and $R_r^{J\_Ang}$ are recognition matching scores for joint positions (trajectories), distance and angles of the action dataset respectively. $R_r^{J\_Trj}$, $R_r^{J\_Dis}$ and $R_r^{J\_Ang}$ are in the range of $[0,1]$. The value '0' and nearest to '0' denotes the nonmatching and the value '1' indicates perfect matching. Then to extract perfect action class of the unknown query action, we apply average score fusion among these three scores obtained from the three proposed kernels (Joint positions, Joint distances, joint angles) as

$$R_{Action} = \frac{R_r^{J\_Trj} + R_r^{J\_Dis} + R_r^{J\_Ang}}{3}. \quad (11)$$

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

In this session, the experimental discussion is carried out in three ways. First, the experiments were conducted on three benchmark datasets NTU RGB-D [6], G3D [7] and UTD-MHAD [8] which were captured from the Microsoft Kinect sensor. For every dataset, the state-of-the-art methods were applied and the obtained recognition rates were compared with the proposed kernel based method. Secondly, we compare our proposed features against various features using our proposed adaptive kernel matching algorithm. Next, we compare the performance of various state-of-the-art algorithms with our proposed method.

A.Performance evaluation on benchmark datasets

Evaluating the proposed method on NTU RGB-D dataset:

The NTU RGB-D dataset [6] is a relatively largest 3D skeleton dataset. The dataset is captured with 3 Microsoft Kinect v.2 cameras and recorded 3D skeletons contains the 3D coordinate locations of 25 body joints, at every skeleton frame. It contains 56,880 action samples containing RGB, RGB-D, 3D Skeleton and Infrared videos for each action. The dataset consists 60 different action classes is categorized into three parts: 40 daily actions, 11 mutual actions and 9 medical conditions, which are performed by 40 different subjects age between 10 to 35. The dataset containing two types of standard evaluation protocol, cross subject and cross view. Table I shows the performance of the proposed method with earlier proposed methods along with our method.

Table I. Performance of the proposed method along with the state-of-the-art methods on the NTU RGB-D dataset.

Method	NTU RGB-D	
	Cross Subject	Cross View
Bi-Vector [9]	66.23	66.19
JDM [5]	76.21	82.31
Fusing Geometric Features [1]	76.43	87.69
SPMF [10]	78.89	86.15
Beyond joints [2]	79.51	87.6
Lie Net [11]	61.37	66.97
Li Group [10]	50.08	52.76
Proposed	**82.71**	**91.40**

Table I compares various skeletal based human action recognition techniques in terms of classification rates. Both the CNN and non-CNN based algorithms were applied on a NTU RGB-D skeletal data. Even though the CNN based methods were most popularly used methods, data preparation for them is a typical task and consumes more time. A high speed hardware configuration is required to train the network. Our proposed kernel based algorithms can run on a simple hardware setup with better recognition rates.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

The confusion matrix generated with our proposed method, fusion of trajectory, distance, angle kernel on NTU RGB-D dataset is shown in fig.5. From the confusion matrix, the proposed method recognizes many actions very well.

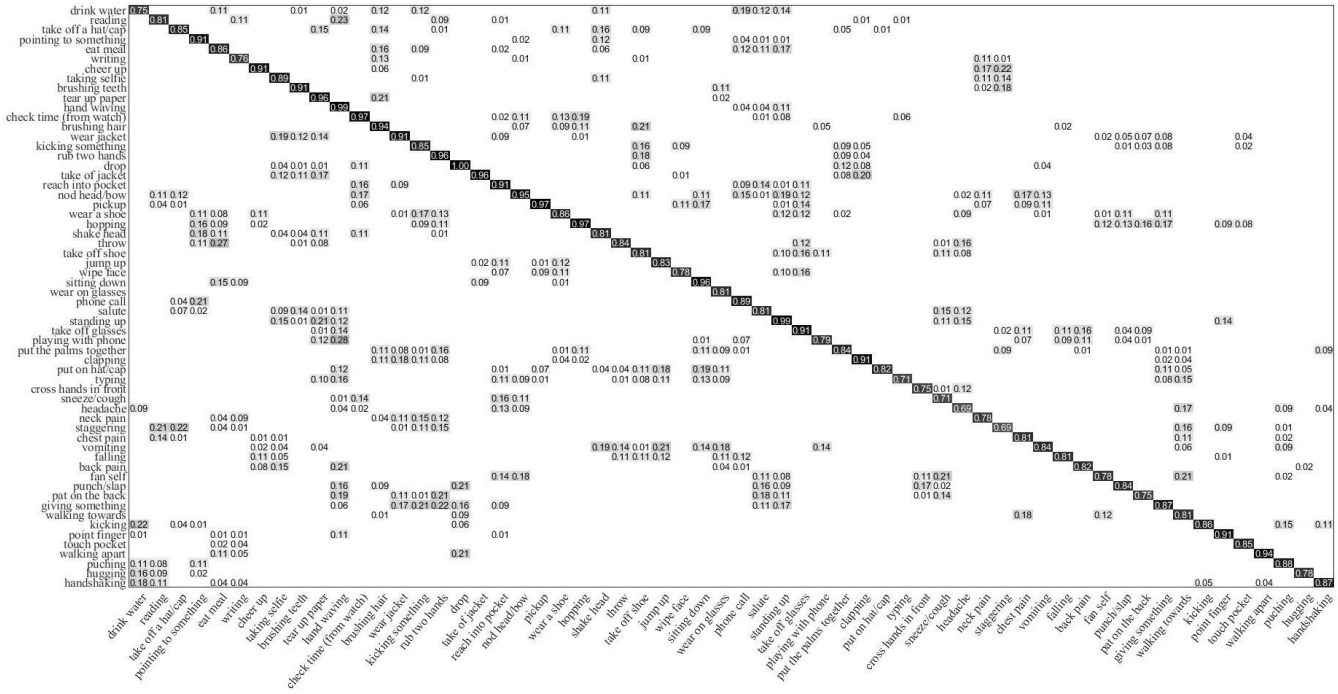

Fig.5. Confusion matrix showing the recognition rates achieved for different actions in NTU RGB-D dataset using our proposed adaptive kernel matching.

Evaluating the proposed method on G3D dataset:

The G3D [7] is a real time action recognition in gaming scenario. The dataset contains 80000 sequences of RGB, RGB-D and skeleton data, captured by single stationary depth sensor. It consisting of 10 subjects performing 20 actions: “walk”, “punch right”, “punch left”, “aim and fire gun”, “kick right”, “kick left”, “tennis serve”, “defend”, “throw bowling ball”, “golf swing”, “jump”, “tennis swing forehand”, “run”, “tennis swing backhand”, “climb”, “crouch”, “steer a car”, “wave”, “flap” and “clap”. Every subject doing each action three times and every frame contains 20 joints. The performance of the proposed method in this paper is compred with other existing methods and the results are shown in Table II.

Table II. Comparison of the proposed method and the existing methods on G3D dataset.

Method	G3D Dataset Accuracy
AdaBoost [7]	71.04
discrete spherical harmonics [4]	92.31
Spherical Harmonics [4]	92.89
JTM [3]	94.24
Proposed	**96.80**

Fig.6 shows the confusion matrix obtained in our method on G3D dataset. From the confusion matrix, we can clearly observe that the performance of the proposed algorithm recognizes very well. Compared to previous approaches, our adaptive kernel method with fusion of three geometric features recognizing the actions with very good classification rates.

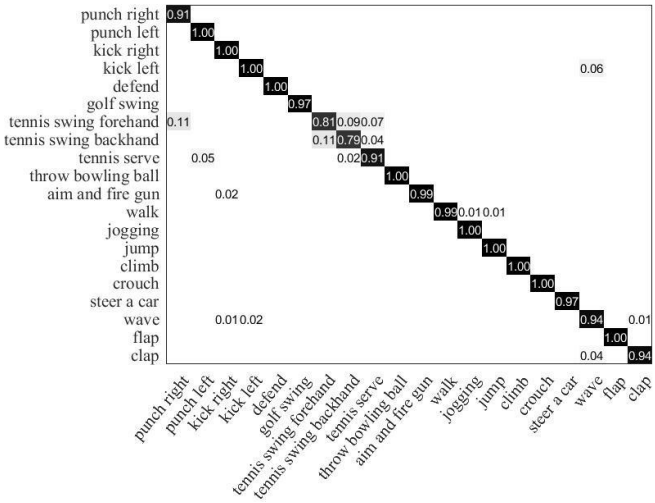

Fig.6. Confusion matrix showing the recognition rates achieved for different actions in Gaming 3D (G3D) dataset using our proposed adaptive kernel matching.

Evaluating the proposed method on UTD-MHAD dataset:

The UTD MHAD [8] dataset is one multimodal action dataset, captured by single Kinect camera and one wearable inertial sensor. It contains 861 frames of 27 actions. Each action performed by 8 subjects (four males and four females) and every subject repeated each action 4 times.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

The actions are: “arm curl (two arms)”, “baseball swing from right”, “basketball shoot”, “bowling (right hand)”, “cross arms in the chest”, “draw triangle”, “forward lunge (left foot forward)”, “front boxing”, “jogging in place”, “right hand know on door”, “right arm swipe to the left”, “right hand pick up and throw”, “right arm swipe to the right”, “right hand draw circle (clockwise)”, “right hand wave”, “right arm throw”, “right hand draw circle (counter clockwise)”, “right hand draw x”, “right hand catch an object”, “sit to stand”, “squat (two arms stretch out)”, “stand to sit”, “tennis serve”, “tennis right hand forehand swing”, “two hand front clap”, “two hand push” and “walking in place”. It covers hand gestures, sport actions, daily activities. The dataset, cross subject protocol is adopted as in [8]. Table III shows the comparison of proposed method with previous methods.

Table III. Comparison of the proposed methods with previously proposed state-of-the-art methods on UTD-MHAD dataset.

Method	UTD MHAD Dataset
	Accuracy
Kinect & Inertial [8]	79.1
JTM [3]	85.81
Cov3DJ [14]	85.5
SPMF [10]	95.3
Proposed	**96.1**

Fig.7 shows the confusion matrix for UTD MHAD dataset. When compare to previous datasets this dataset is more challenging due to its complex actions. From fig.7, we can see that the proposed algorithm unable to discriminate some similar shape actions very well, for example, “knock”, “catch” and “jog”, “walk”. Even though it fails to recognize some actions, we found a good amount of classification rates than the other methods.

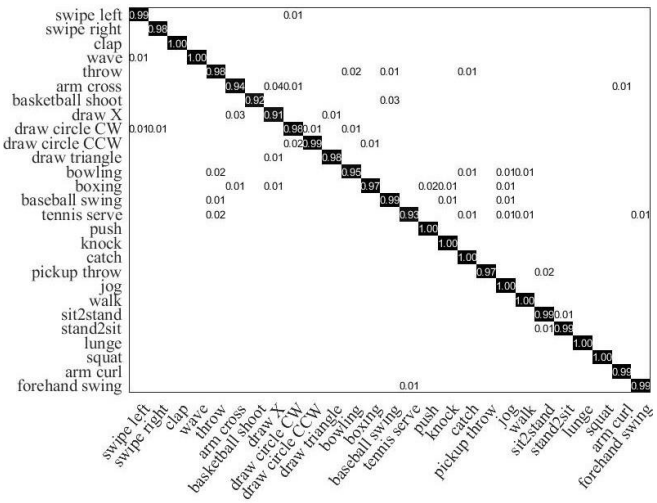

Fig.7. Confusion matrix showing the recognition rates achieved for different actions in UTD-MHAD dataset using our proposed adaptive kernel matching.

B. Performance comparison of the proposed method with various geometric features as input

In this section, various features like joints, edges, distance, angle, surface, area, perimeter, velocity and our proposed average fusion of joints, distance and angle (JP+JDF+JAF) is evaluated on above datasets using adaptive kernel matching algorithm. Table IV shows the average recognition accuracies of various features with our proposed method on above datasets. In our observation, the fusion of geometric feature gives satisfactory results compared to other features.

Table IV. Recognition rates of various geometric features on NTU RGB-D, G3D and UTD-MHAD datasets.

Features	% of Recognition			
	NTU RGB-D		G3D	UTD MHAD
	Cross subject	Cross View		
Joint Positions (JP)	69.4	73.8	71.8	71.2
Joint distance (JDF)	77.8	89.2	80.2	79.7
Joint angles (JAF)	79.5	89.7	84.6	81.9
Joint surface (JSF)	68.1	73.4	79.9	75.1
Joint velocity (JVF)	69.7	74.5	71.1	65.9
JP+JDF+JSF	80.1	85.7	90.1	87.3
JP+JDF+JVF	81.7	90.7	91.8	89.7
JP+JSF+JVF	80.9	91.1	90.7	88.6
JP+JSF+JAF	81.1	90.1	91.9	90.8
JP+JVF+JAF	80.2	89.2	93.8	94.7
JDF+JSF+JVF	81.4	90.8	93.7	91.5
JDF+JSF+JAF	81.9	91.1	95.4	93.6
Proposed (JP+JDF+JAF)	**82.7**	**91.4**	**96.8**	**96.1**

Fig.8 visualizes the recognition rates of each action from the G3D dataset, we compare the accuracies of different individual geometric features like joint, distance, angles and our fused geometric features JP+JDF+JAF. From the fig.8, the recognition rates for fusion of joints, distance and angle gives better recognition rates compared to individual joints, distance and angle features.

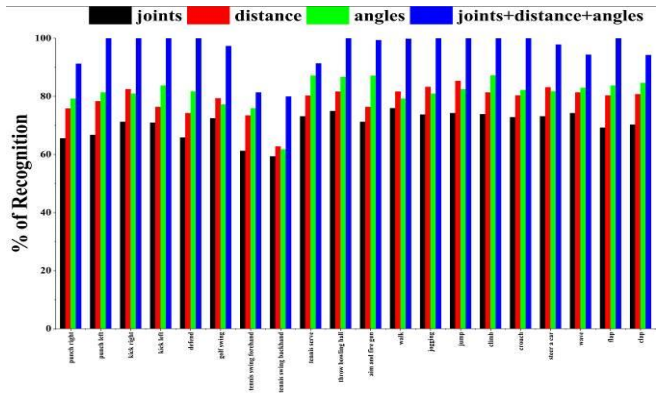

Fig.8. Recognition rates of each action on G3D dataset with individual kernel matching and fused kernels matching.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

C. Performance evaluation of the proposed against state-of-the-art methods

As shown in fig.9, three datasets were used to compare the performance of the proposed features with proposed method and existing state-of-the-art methods namely weighted graph matching (WGM), dynamic time wrapping (DTW), adaptive graph kernels (AGK), locally preserving positions (LPP)+ bag of words (BoW), histogram, support vector machine (SVM) and extreme learning machine (ELM). Recognition rates were computed for the above algorithms on three skeleton datasets against the proposed method.

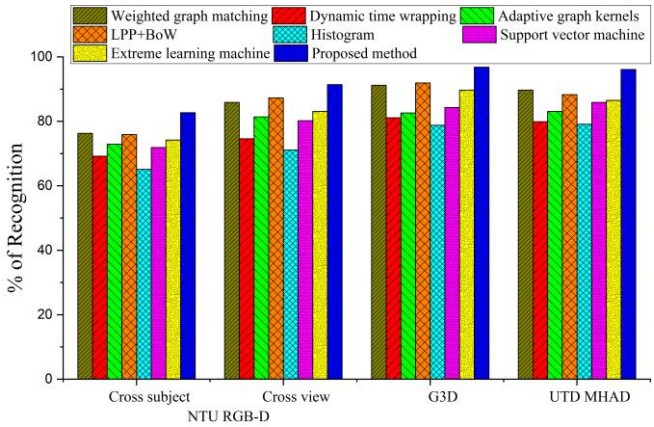

Fig.9. Performance analysis of various methods on proposed fusion geometric features.

The weighted graph matching and LPP+BoW methods came closed to our proposed algorithm. But as it is an action frame by frame matching, the computation time very high in recognition. When limited number of frames present, the recognition rate is good in dynamic time wrapping. Predicting the missing data is not possible via histogram because it is a probability-based method. In some cases, support vector machine and extreme learning machine were failed to identify complex actions.

V.CONCLUSION

In this work, we proposed an effective method for recognizing actions from 3D skeleton data. JP's are joint trajectory positions, JDF are shape features and JAF are orientation features, which are extracted as positions, distance and angles on the 3D joint skeleton. Three feature kernels based on positions, distance and angle are constructed and fused together to find best matching score between query action with the dataset actions. The experimental results shows an improvement in overall recognition rate of the proposed framework due to the fusion of three geometric features. The proposed algorithm is tested on three well known datasets gives good recognition rates when compared to other existing methods.

REFERENCES

1. Zhang, Songyang, Yang Yang, Jun Xiao, Xiaoming Liu, Yi Yang, Di Xie, and Yueting Zhuang. "Fusing Geometric Features for Skeleton-Based Action Recognition using Multilayer LSTM Networks." IEEE Transactions on Multimedia (2018).
2. Wang, Hongsong, and Liang Wang. "Beyond Joints: Learning Representations From Primitive Geometries for Skeleton-Based Action Recognition and Detection." IEEE Transactions on Image Processing 27, no. 9 (2018): 4382-4394.
3. Wang, Pichao, Wanqing Li, Chuankun Li, and Yonghong Hou. "Action Recognition Based on Joint Trajectory Maps with Convolutional Neural Networks." Knowledge-Based Systems 158 (October 2018): 43–53.
4. Adnan Salih, Al Alwani, and Chahir Youssef. "Spatiotemporal Representation of 3D Skeleton Joints-Based Action Recognition Using Modified Spherical Harmonics." Pattern Recognition Letters 83 (November 2016): 32–41.
5. 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 (May 2017): 624–628.
6. Shahroudy, Amir, Jun Liu, Tian-Tsong Ng, and Gang Wang. "NTU RGB+D: A Large Scale Dataset for 3D Human Activity Analysis." 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2016).
7. Bloom, Victoria, Dimitrios Makris, and Vasileios Argyriou. "G3D: A Gaming Action Dataset and Real Time Action Recognition Evaluation Framework." 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (June 2012).
8. Chen, Chen, Roozbeh Jafari, and Nasser Kehtarnavaz. "UTD-MHAD: A Multimodal Dataset for Human Action Recognition Utilizing a Depth Camera and a Wearable Inertial Sensor." 2015 IEEE International Conference on Image Processing (ICIP) (September 2015).
9. Feng, Jiageng, Songyang Zhang, and Jun Xiao. "Explorations of Skeleton Features for LSTM-Based Action Recognition." Multimedia Tools and Applications 78, no. 1 (October 28, 2017): 591–603.
10. Pham, Huy-Hieu, Louahdi Khoudour, Alain Crouzil, Pablo Zegers, and Sergio A. Velastin. "Skeletal Movement to Color Map: A Novel Representation for 3D Action Recognition with Inception Residual Networks." 2018 25th IEEE International Conference on Image Processing (ICIP) (October 2018).
11. Huang, Zhiwu, Chengde Wan, Thomas Probst, and Luc Van Gool. "Deep Learning on Lie Groups for Skeleton-Based Action Recognition." 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (July 2017).
12. Althloothi, Salah, Mohammad H. Mahoor, Xiao Zhang, and Richard M. Voyles. "Human Activity Recognition Using Multi-Features and Multiple Kernel Learning." Pattern Recognition 47, no. 5 (May 2014): 1800–1812. doi:10.1016/j.patcog.2013.11.032.
13. Xia, Lu, Chia-Chih Chen, and J. K. Aggarwal. "View Invariant Human Action Recognition Using Histograms of 3D Joints." 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (June 2012).
14. Brun, Luc, Gennaro Percannella, Alessia Saggese, and Mario Vento. "Action Recognition by Using Kernels on Aclets Sequences." Computer Vision and Image Understanding 144 (March 2016): 3–13.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)