Improved Behavior Monitoring and Classification Using Cues Parameters Extraction from Camera Array Images

Ahmad Jalal, Shaharyar Kamal*

Department of Computer Science and Engineering, Air University, Islamabad (Pakistan)

Received 26 February 2018 | Accepted 6 July 2018 | Published 20 July 2018

unir
LA UNIVERSIDAD EN INTERNET

Abstract

Behavior monitoring and classification is a mechanism used to automatically identify or verify individual based on their human detection, tracking and behavior recognition from video sequences captured by a depth camera. In this paper, we designed a system that precisely classifies the nature of 3D body postures obtained by Kinect using an advanced recognizer. We proposed novel features that are suitable for depth data. These features are robust to noise, invariant to translation and scaling, and capable of monitoring fast human body-parts movements. Lastly, advanced hidden Markov model is used to recognize different activities. In the extensive experiments, we have seen that our system consistently outperforms over three depth-based behavior datasets, i.e., IM-DailyDepthActivity, MSRDailyActivity3D and MSRAction3D in both posture classification and behavior recognition. Moreover, our system handles subject's body parts rotation, self-occlusion and body parts missing which significantly track complex activities and improve recognition rate. Due to easy accessible, low-cost and friendly deployment process of depth camera, the proposed system can be applied over various consumer-applications including patient-monitoring system, automatic video surveillance, smart homes/offices and 3D games.

Keywords

Activity Recognition, Body Posture Recognition System, Pattern Clustering, SmartCities.

DOI: 10.9781/ijimai.2018.07.003

I. Introduction

IDENTIFICATION, monitoring, classification and recognition of human from behavior images is very necessary as it is very effective to convey subject's situation, identity, emotion, gait and gestures [1-4]. Still human identification and monitoring is not absolutely perfect in various conditions such as position changes, illumination, orientation, noise variations and dark-area places [5-8]. In spite of the research efforts and significant results in the past decade, recognition accuracy of human behavior still remains a challenge because of self-occlusion of human body parts, variation of body size and appearance, un-clear or hidden body parts behind objects and fast human movements during indoor scenes decade [9, 10]. In addition, several researchers mainly focused on recognizing activities from videos captured by conventional cameras which are less effective due to complex backgrounds, light sensitivity and motion ambiguities (i.e. color and texture variability) [11-13]. Thus, to access the high quality imaging and 3D motions, the development of low-cost and easy-processing depth cameras such as Microsoft Kinect or bumblebee, have initiated new era for a variety of image recognition tasks including human behavior recognition (BR) [14-16]. Depth images provide several opportunities to enhance BR such as additional body joints information, spatial continuity, insensitivity to lighting conditions and controlling overlapping issues of different human body parts.

A large number of methods have been designed for efficient BR method and also a lot of comparative studies were evaluated by series of researchers over depth videos [16-18] to examine the best algorithms for recognition. These methods mainly interact with depth data using two different approaches: skeleton joints features and depth silhouette features. For example, Oreifej and Liu [19] proposed a new descriptor for behavior recognition using a histogram capturing the distribution of the surface normal orientation in the 4D space of time, depth, and spatial coordinates. To build the histogram, they created 4D projectors, which quantize the 4D space and represent the possible directions for the 4D normal. In [17], Yang et al described an effective method that project depth maps onto three orthogonal planes and accumulate global activities through entire video sequences to generate the Depth Motion Maps (DMM). Histograms of Oriented Gradients (HOG) are then computed to enhance the activity recognition results. In [20], authors proposed a behavior recognition system that deals with motion features as magnitude and directional angular features from body joints information between consecutive frames to recognize daily routine human activities. In [21], authors designed mid-level features from Kinect skeletons by considering the orientations of human body limbs connected by two skeleton joints and each orientation is encoded into different states. They employed frequent pattern mining to pick the most frequent feature values, relevant states of parts in continuous several frames and recognize different activity/actions.

However, such methods show better performance and contributions, but different factors having negative impact surrounded each method. Those methods just relied on the skeleton data which became unreliable for postures with self-occlusion. Also, some methods were depended on depth silhouettes information which causes low recognition accuracy especially in case of hidden or missing body parts, fast moving human silhouettes and large distance of subject from the source (i.e. depth camera). Therefore, we elaborate some novel features along with

* Corresponding author.
E-mail address: shaharyar.kamal@mail.au.edu.pk

advanced HMM to overcome the above mentioned problems and improve recognition accuracy.

In this paper, we propose a novel behavior recognition framework based on cues-parameters, which has an improved accuracy over existing algorithms. At the start of the BR framework, we handle the noisy input posture and unclear background data by designing a set of reliability measurement to extract true silhouettes and tracked joint values. These true data is examined to extract human silhouette by considering spatial/temporal continuity, constraints of human motion information and frame differentiation. These data are further processed to get feature representation by considering cues-parameters including angular direction, spatiotemporal velocity and invariant features which provide compact and sufficient feature values for better BR performance. While, all feature values are mapped into codewords and recognized each behavior via advanced Hidden Markov model (HMM). We evaluate our method according to the standard experimental protocols definition on three challenging depth behavior datasets: IM-DailyDepthActivity, MSRDailyActivity3D and MSRAction3D. Our experimental results show that the proposed method is able to achieve better recognition accuracy than the state-of-the-art methods. Since our system is well-organized, affordable and easily installable, therefore, it is the preferable solution for reliable body tracking, orientation, smart environments and behavior recognition systems.

The organization of this paper is as follows. Section II elaborates the system architecture of the proposed method starting from depth image preprocessing, evaluation of feature extraction by cues-parameters and behavior training/recognition via advanced HMMs. Experimental results and comparisons between proposed and state of the art methods are described in Section III. Finally, we conclude this work in Section IV.

II. System Architecture and Description

The proposed behavior recognition system consists of various steps as: 1) raw depth data captured by RGB-D video sensor, 2) noisy background removal, 3) human detection, tracking and identification from the time-sequential behavior video images. 4) feature extraction based on cues-parameters techniques, clustering using Linde, Buzo, and Gray (LBG)'s algorithm and training/recognition using advanced HMM. Fig. 1 shows the overall flow of our proposed BR system.

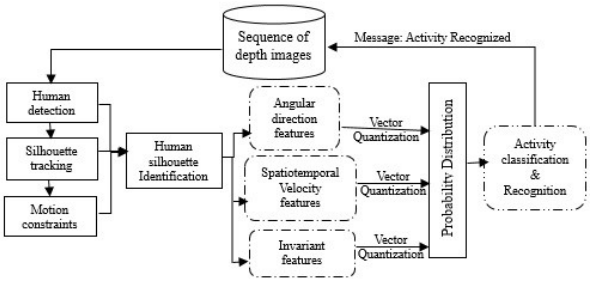

Fig. 1. Overall flow of the proposed human behavior recognition system.

A. Depth Image Processing

For the human identification in depth sequential maps, background subtraction routine is applied which consists of least squares method for estimating the angle and center point of the floor in a real world coordinate system [22]. The depth value y in a spaced grid having least value is used to ignore floor from noisy background. Then, we have localized different objects in the scene and segmented them by computing the modified connected component labeling (CCL) method and examining degree of freedom [23, 24]. It is used to label all candidates pixels separately. Finally, we extract the human moving silhouettes, temporal depth intensity differentiation [22] is applied as expressed in (1) to obtain the depth human silhouettes from consecutive frames.

$$D=\sqrt{(I_i^x-I_{i-1}^x)^2+(I_i^y-I_{i-1}^y)^2+(I_i^z-I_{i-1}^z)^2} \quad (1)$$

To properly track the entire human body, we performed disparity segmentation. Ignoring the 0's, we consider the average of the disparity values in the detected parts and compare neighboring pixels surrounded by the detected moving parts to add the pixels with closure disparity values that make a separate region (i.e. human silhouettes). Thus, the disparity segmentation is employed to find the target human silhouette candidates and subjects which are free to move more naturally.

Overall system's framework allows a smart environment to analysis what the user is doing from the noisy data obtained from the depth camera. We implement the depth motion database targeting at different behaviors. The database includes correctly and incorrectly performed skeleton models for different activity/action purposes, annotated posture information, as well as depth and color images obtained from the RGB-D camera. Some of the examples of the database are shown in Fig. 2.

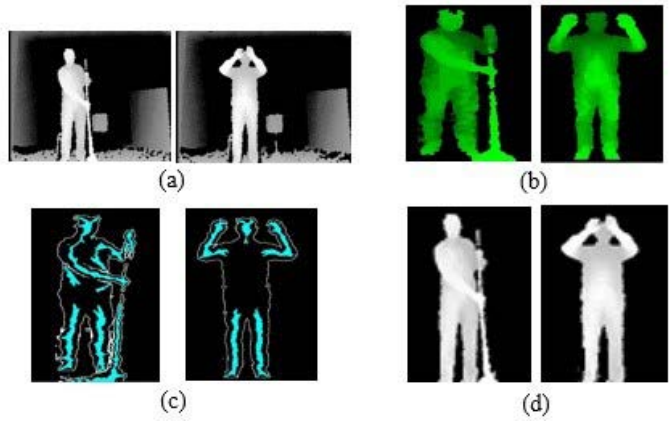

Fig. 2. Human silhouette identification. (a) Noisy background, (b) labeled silhouette, (c) ridge information and (d) depth human silhouette.

B. Evaluation of Cues-parameters Techniques

In this section, we revive human motion and joint tracking information by considering unique cue evaluation techniques using OpenNI environment, and point out why these methods are applied efficiently and produce vital results.

For initialization, we utilized several body silhouettes and joints values cues for local body parts motions, temporal frames information, spatial/temporal depth silhouettes characteristics and speed measurements of human body shape using depth datasets. Following are the sub-sections that describe the features under our evaluation.

1) Angular Direction Feature

Each posture P in the database is represented by a vector of 3D joint points as

$$P=[C_1,C_2,\ldots,C_n] \quad (2)$$

The angular direction feature $\varphi_{\cos}$ is defined as the difference of angular movements between similar joints of two different frames (i.e., consecutive frames) at time t_1 and t_2. It captures the temporal movements of different body parts of human silhouette. It is defined as

$$\varphi_{\cos} = \cos^{-1}\left(\frac{C_k^{t_1} \times C_k^{t_2}}{\left|C_k^{t_1}\right|\left|C_k^{t_2}\right|}\right) \tag{3}$$

where C^{t_1} and C^{t_2} are the joints information with respect to consecutive frames and k indicates the all three coordinates axis (i.e., x, y, z) of respective joints [22]. Fig. 3 shows an example of the directional angular features with different activities. However, the joints angular values are quite effective in order to improve the performance accuracy during feature discrimination and recognition.

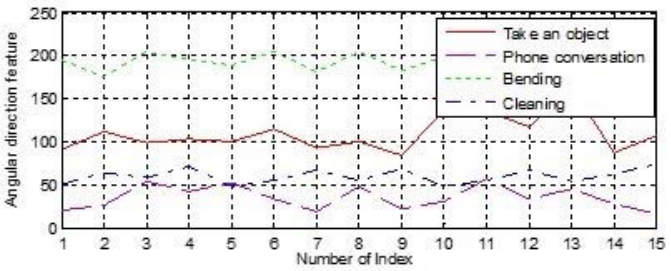

Fig. 3. Human silhouette identification. (a) Noisy background, (b) labeled silhouette, (c) ridge information.

2) Velocity Feature

The spatiotemporal velocity feature f_v captures the velocity of each joint in the direction of the normal vector of the plane from starting frame till ending frame. It is defined as

$$f_v(t,t+1) = \frac{1}{\left|t_e - t_s + 1\right|}\sum_{t=t_s}^{t_e}\left(v(j_{t+1}^k) - v(j_t^k)\right) \tag{4}$$

where t_s and t_e are the starting and ending frames of overall data sequence and j^k deals with the coordinate axis of human body joint information. These features deal with the intensity differentiation and spatiotemporal motion values of body parts. Fig. 4 explains the detail description of spatiotemporal velocity features using depth dataset.

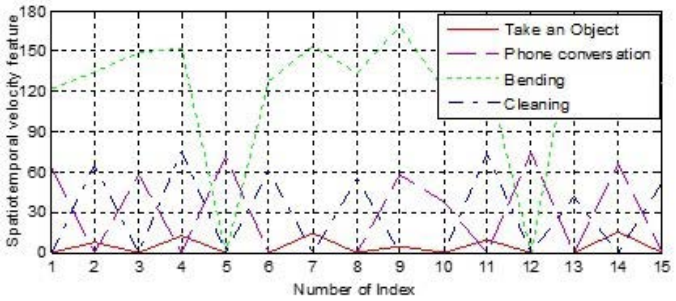

Fig. 4. 2D plot of spatiotemporal velocity features from RGB-D video dataset.

3) Invariant Feature

To observe the invariant characteristics of human body silhouette, we measure the integral of a function over a line as the Radon transform. Basically, it extracts some interested regions in the Euclidean plane as

$$E(\theta) = \int_{\infty}^{-\infty} R_z^2(\rho,\theta)d_\rho \tag{5}$$

where $R_z(\rho,\theta)$ is 2D Radon map that is the line integral of depth data. During extraction of human behavior silhouettes, radon transform is used to represent the distance and local directional movements in human body parts motion. During the computation of radon transform, the line integrals of human silhouettes amplify low frequency components which are useful in behavior recognition. It is quite suitable for BR because there is a variation in the angles of different human body parts such as feet, hands, head, hips and shoulders. Thus, it represents the maximum energy of the human behavior that appears in specific coefficients which vary considerably through time. Fig. 5 shows some human behaviors and their corresponding Radon transforms.

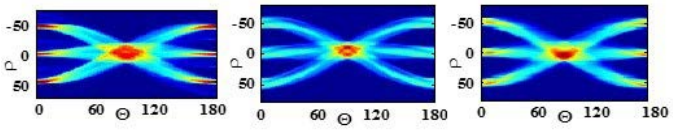

Fig. 5. Distance and local directional movements of different human behaviors using Radon transform.

C. Features Symbolization

Now, these cues-parameters vectors are further symbolized based on vector quantization technique known as Linde, Buzo, and Gray (LBG)'s clustering algorithm [15]. Initially, LBG initializes with a codebook size of 1 and recursively splits the centroids of feature vectors (i.e., datasets) to get an optimally sized codebook. We used the optimal codebook size of 64 after experimenting over different depth datasets. In addition, the codebook size and codevector are directly related with the precision value of a locally global feature values and parallelly intact with the size of the source image. This optimization of the centroids is done to reduce the distortion. While, these code-values are generated per each behavior sequence and stored by considering buffer strategy. Fig. 6 shows the procedure of code-values generation and symbolization of proposed features.

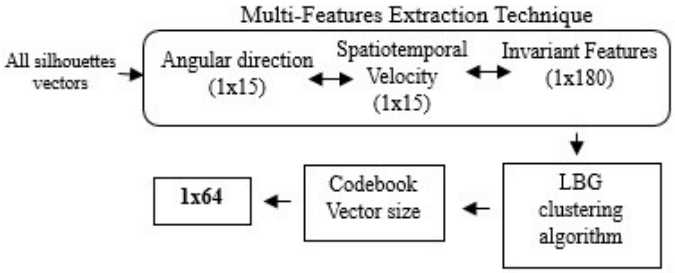

Fig. 6. Overall procedure of code-values generation and symbolization of cues-parameters.

D. Advanced HMMs

To train and recognize different depth datasets, we modified conventional HMM into advanced HMM technique. In conventional HMM approach, each behavior makes its specific HMM based on finite states having transition and symbol observation probability [21]. Such method consists of redundant information in the form of whole body silhouette and less active moving body parts [25]. Also, HMMs are mainly dependent on the feedback of possible transitions and need the priori knowledge to manage the parameters which cause over-fitting in each class [26, 27]. Such kind of unnecessary information causes reduction at overall performance of accuracy results. Thus, advanced HMM is developed which focused on active areas of human body parts such as hands, feet, head, hips and shoulders. For training phase, we need to train (*N+1*) distinct HMMs for N different human activities. During testing phase, maximum likelihood [28-30] value of specific sequential data is chosen to recognize distinct behavior as expressed in (6).

$$H_a = arg\ max_k \{ P(O \mid \lambda_h) \quad (6)$$

where $P(O \mid \lambda_h)$ denotes the probability of likelihood of the h behavior HMM among a number of activities. Fig. 7 shows active features regions of overall human silhouettes to calculate specific likelihood of each behavior.

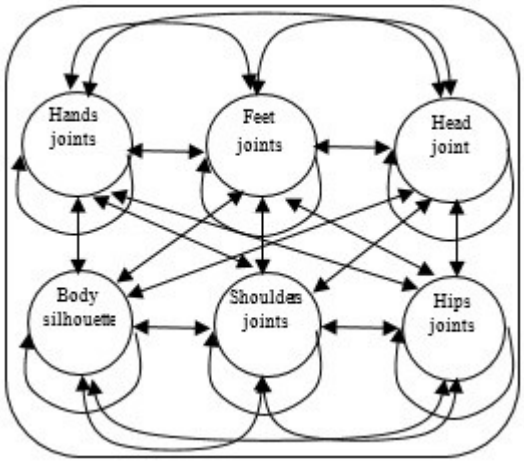

Fig. 7. Structural view of advanced hidden Markov model.

III. Experimental Results

In this section, we describe how our skeleton models and human postures are represented in the database, and details about what kind of behaviors are included to create the database. We evaluate the proposed cues-parameters approach on a newly collected depth-based behavior recognition datasets (i.e. IM-DailyDepthActivity) and two public datasets (i.e., MSRDailyActivity3D and MSRAction3D).

During experiment, we used the leave one subject out (LOSO) cross validation method [26].

A. Datasets Description

Following sub-sections are used to describe each dataset, its experimental setting and size of each dataset, respectively. We use the Microsoft Kinect to capture the skeleton data, its joints location and posture data for the database, as it is one of the known running and developed device of depth camera based motion sensors. Apart from that, we manually handled descriptions such as the natural movements of the human behavior, risky injury during captured sequence and slotting of each video sequence.

1) IM-DailyDepthActivity

We capture the RGB images, depth images, labeled data and skeleton joints information of 15 different activities as: *sit down*, *both hands waving*, *phone conversation*, *kicking*, *reading an article*, *throwing*, *bending*, *clapping*, *right hand waving*, *take an object*, *exercise*, *eating*, *boxing*, *cleaning and stand up*, respectively. The dataset is captured in indoor environments (i.e., labs, halls and classrooms) having multiple background scenes. However, the dataset includes 45 segmented videos of each activity for training and 30 unsegmented continuous videos for testing performed by 15 different subjects.

Both of the RGB videos and depth maps have the resolution of 640 x 480 pixels and the skeleton contains 15 joints per person. Fig. 8 gives some examples of IM-DailyDepthActivity dataset.

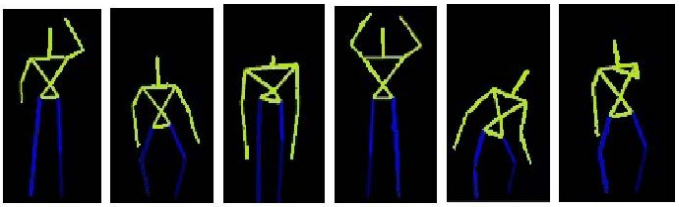

Fig. 8. Some skeleton images of annotated IM-DailyDepthActivity dataset.

2) MSRDailyActivity3D

The dataset has sixteen activities including: drink, eat, read book, call cellphone, write on a paper, use laptop, use vacuum cleaner, cheer up, sit still, toss paper, play game, lay down on sofa, walk, play guitar, stand up and sit down. The total dataset includes 320 video sequences which are performed by 10 subjects. Each subject performs each activity twice, one in standing position and the other in sitting position. Fig. 9 shows some depth images of MSRDailyActivity3D dataset. In addition, most activities involve human-object interactions which make this dataset more challenging.

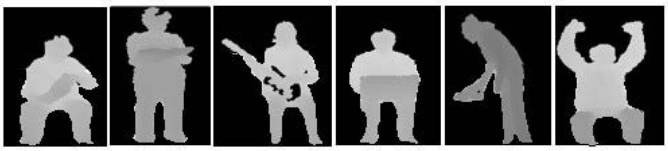

Fig. 9. Sample depth silhouettes images of MSRDailyActivity3D dataset.

3) MSRAction3D

The MSRAction3D dataset [27] is an actions dataset of depth map sequences captured by a depth camera. It contains 20 different actions performed by 10 subjects. These actions include: high arm wave, horizontal arm wave, hammer, hand catch, forward punch, high throw, draw x, draw tick, draw circle, hand clap, two hand wave, side boxing, bend, forward kick, side kick, jogging, tennis swing, tennis serve, golf swing and pick up & throw. There are 567 depth videos sequences. Furthermore, the background of the dataset is clean and mostly actions involve specific body parts movements (i.e., head, arm and leg), which makes the dataset quite challenging. Fig. 10 shows different sequences of depth maps actions used in MSRAction3D dataset.

Fig. 10. Depth sequential images of hand catch and bend actions used in MSRAction3D dataset.

B. Recognition Results

1) IM-DailyDepthActivity

Table I, II and III contain the confusion matrices obtained. Specifically, Table I shows the confusion matrix of 15 typical human activities on the annotated IM-DailyDepthActivity dataset.

In the newly collected dataset, we evaluated the proposed cues-parameters method and compared the results with the state-of-the-art methods. Table IV shows the comparison of recognition results. Here, the existing methods such as motion templates [31] proposed new methods for automatic classification and retrieval of motion capture data facilitating the identification of logically related motions in specific dataset. In [17], Yang et al. describes the depth motion maps

TABLE I. Confusion Matrix 15 Typical Human Activities on the Annotated IM-DailyDepthActivity dataset

SD= *sit down*, **BH**= *both hands waving*, **PC**= *phone conversation*, **KG**= *kicking*, **RA**= *reading an article*, **TG**= *throwing*, **BG**= *bending*, **CG**= *clapping*, **RH**= *right hand waving*, **TA**= *take an object*, **EX**= *exercise*, **EA**= *eating*, **BO**= *boxing*, **CL**= *cleaning*, **SU**= *stand up*

Activities	SD	BH	PC	KG	RA	TG	BG	CG	RH	TA	EX	EA	BO	CL	SU
SD	**72.5**	4.0	0	6.50	4.0	0	0	5.0	0	0	1.0	2.5	1.5	0	3.0
BH	2.5	**77.0**	3.0	0	0	1.5	4.0	0	1.5	3.5	0	0	1.5	4.5	1.0
PC	11.0	3.5	**64.5**	6.0	2.5	0	0	2.5	3.5	1.0	2.0	0	1.5	0	2.0
KG	2.0	1.5	0	**81.0**	2.5	3.0	0	2.5	0	3.5	0	1.5	0	1.0	1.5
RA	0	13.0	0	4.5	**69.5**	1.5	1.0	0	2.5	5.0	1.0	0	0	2.0	0
TG	1.5	3.5	3.0	0	0	**73.0**	3.5	4.0	3.5	0	1.5	3.0	2.0	0	1.5
BG	7.0	2.5	0	1.0	0	8.5	**5.85**	3.5	1.0	2.0	4.0	1.0	0	9.5	1.5
CG	1.5	0	1.0	2.5	0	2.0	2.5	**81.0**	2.50	1.0	0	0	3.5	0	2.5
RH	5.0	7.5	2.0	0	1.0	0	1.5	0	**6.75**	3.5	2.0	1.5	2.0	1.5	5.0
TA	3.5	1.0	4.5	5.0	0	3.0	1.0	0	3.5	**72.5**	0	1.5	1.0	0	3.5
EX	8.5	3.5	0	5.5	6.5	0	2.0	1.5	0	2.5	**55.0**	7.5	0	5.5	2.0
EA	3.0	7.5	2.0	0	11.5	2.5	0	2.0	4.0	0	0	**52.5**	8.5	0	6.5
BO	6.5	0	5.0	0	1.5	0	0	3.5	0	1.5	0	0	**70.5**	9.0	2.5
CL	3.5	12.5	4.5	3.0	0	2.0	0	0	3.5	5.5	1.0	0	0	**57.0**	7.5
SU	1.5	4.0	0	0	3.5	6.5	1.0	0	4.5	0	2.0	1.0	2.5	0	**73.5**
Mean Recognition Rate = 68.4%															

TABLE II. Confusion Matrix 16 Different Human Actions on the MSRDailyActivity3D Dataset

DK= drink, **ET**= eat, **RB**= read book, **CC**= call cellphone, **WP**= write on a paper, **UL**= use laptop, **UV**= use vacuum cleaner, **CU**= cheer up, **SS**= sit still, **TP**= toss paper, **PG**= play game, **LD**= lay down on sofa, **WK**= walk, **PR**= play guitar, **ST**= stand up, **SI**= sit down.

Activities	DK	ET	RB	CC	WP	UL	UV	CU	SS	TP	PG	LD	WK	PR	ST	SI
DK	**87.0**	2.5	1.0	0	1.5	0	3.0	0	1.0	0	0	1.5	0	0	1.0	1.5
ET	0	**94.5**	0	0	1.0	2.5	0	0	1.0	0	0	0	0	1.0	0	0
RB	1.5	0	**93.0**	1.0	0	0	1.5	0	0	2.0	0	0	1.0	0	0	0
CC	2.5	4.0	0	**84.5**	0	4.5	0	0	1.0	0	1.0	0	0	2.5	0	0
WP	1.5	2.0	1.5	0	**89.0**	0	0	1.0	0	0	0	2.5	0	0	1.0	1.5
UL	0	0	0	1.0	0	**98.0**	0	0	0	1.0	0	0	0	0	0	0
UV	1.5	0	1.0	0	0	0	**94.5**	0	0	0	1.5	0	0	1.5	0	0
CU	3.5	0	0	1.5	0	0	2.0	**86.5**	1.5	0	1.5	0	0	2.5	0	1.0
SS	2.0	1.0	0	2.5	0	0	1.5	0	**88.0**	0	0	2.0	1.5	0	1.5	0
TP	0	0	1.5	0	0	1.0	0	0	0	**97.5**	0	0	0	0	0	0
PG	0	2.0	0	0	1.0	0	1.0	0	1.0	0	**94.0**	0	0	1.0	0	0
LD	1.5	0	2.0	0	0	2.5	0	1.5	0	1.0	0	**89.5**	0	1.0	0	1.0
WK	0	0	1.0	0	0	0	1.5	0	0	0	0	0	**97.5**	0	0	0
PR	0	2.0	0	1.5	0	0	3.5	0	0	1.0	0	0	0	**91.0**	0	1.0
ST	1.5	0	2.0	1.0	1.5	0	0	1.0	0	1.0	0	1.5	1.0	0	**89.5**	0
SI	3.0	0	0	1.5	0	2.5	1.5	0	1.0	0	1.5	0	0	2.0	1.0	**86.0**
Mean Recognition Rate = 91.2%																

as feature representation and HOG to characterize the local appearance for recognizing data. While, [32] deals with the positions of joints to locally define reference system and multi-part bag-of-poses approach is then defined, which permits the separate alignment of body parts through a nearest-neighbor for classification and recognition. In [33], multi-modality fusion scheme is developed based on spatio-temporal interest points and motion history images features to recognize different activities. From results in Table IV, it is clearly seen that the proposed method improves the recognition results as compared to state-of-the-art methods.

TABLE III. CONFUSION MATRIX 20 DIFFERENT HUMAN ACTIONS ON THE MSRACTION3D DATASET

HW= high arm wave, **HA**= horizontal arm wave, **HM**= hammer, **HC**= hand catch, **FP**= forward punch, **HT**= high throw**, DX**= draw x, **DT**= draw tick, **DC**= draw circle, **HC**= hand clap, **TH**= two hand wave, **SB**= side boxing, **BD**= bend, **FK**= forward kick, **SK**= side kick, **JO**= jogging, **TS**= tennis swing, **TE**= tennis serve, **GS**= golf swing, **PU**= pick up & throw.

Actions	HW	HA	HM	HC	FP	HT	DX	DT	DC	HC	TH	SB	BD	FK	SK	JO	TS	TE	GS	PU
HW	**91.5**	1.0	0	0	2.5	0	0	2.0	0	0	1.0	0	0	0	1.0	0	0	1.0	0	0
HA	2.0	**86.0**	1.0	0	0	2.5	0	1.0	1.5	0	3.0	0	0	1.0	0	1.0	0	0	0	1.0
HM	0	0	**91.5**	1.5	0	2.0	0	0	1.0	2.0	0	1.0	0	0	0	0	1.0	0	0	0
HC	0	0	1.0	**97.0**	0	0	0	0	0	1.0	0	0	1.0	0	0	0	0	0	0	0
FP	1.0	0	0	0	**96.0**	1.0	0	0	0	0	0	0	0	0	0	1.0	0	0	1.0	0
HT	2.0	1.0	0	0	1.0	**88.5**	3.0	0	1.0	0	2.5	0	0	0	0	0	0	1.0	0	0
DX	0	0	2.0	0	1.5	0	**90.0**	0	1.5	0	2.0	1.0	1.0	0	1.0	0	0	0	0	0
DT	0	0	0	0	0	0	0	**97.0**	0	1.0	0	0	0	1.0	0	0	1.0	0	0	0
DC	0	0	1.0	0	1.5	0	0	0	**94.5**	0	0	0	0	1.0	0	0	0	1.0	0	1.0
HC	0	1.0	0	2.5	0	2.0	0	0	0	**91.0**	0	1.5	0	0	1.0	0	0	0	1.0	0
TH	2.0	0	0	0	3.5	0	1.5	0	0	1.0	**90.5**	0	1.5	0	0	0	0	0	0	0
SB	0	1.0	0	1.0	0	0	0	2.0	0	0	0	**96.0**	0	0	0	0	0	0	0	0
BD	0	0	0	0	0	1.5	0	0	2.5	0	0	0	**95.0**	0	0	0	1.0	0	0	0
FK	0	0	0	0	1.5	0	0	0	0	2.0	0	0	0	**93.5**	0	1.0	0	1.0	0	1.0
SK	1.0	0	1.0	0	0	0	1.5	0	0	0	0	2.5	0	0	**92.0**	0	1.0	0	1.0	0
JO	0	1.5	0	0	2.5	0	0	1.0	0	1.0	0	0	1.5	0	0	**92.5**	0	0	0	**0**
TS	1.0	0	0	1.0	0	0	1.0	0	0	0	1.5	0	0	0	0	0	**95.5**	0	0	**0**
TE	0	0	1.0	0	0	0	0	0	0	1.0	0	0	0	0	0	0	0	**97.0**	0	**1.0**
GS	0	1.0	0	0	2.0	1.0	0	0	1.5	0	0	0	0	0	0	0	0	0	**94.5**	**0**
PU	0	2.5	0	3.0	0	0	0	1.5	0	0	1.0	0	2.0	0	0	2.0	0	0	0	**88.0**
Mean Recognition Rate = 92.9%																				

TABLE IV. RECOGNITION RESULTS USING IM-DAILYDEPTHACTIVITY

Methods	Recognition Accuracy
Motion templates [31]	38.7
Depth motion maps [17]	42.3
Naive-Bayes-Nearest-Neighbor [32]	47.6
Color-depth fusion features [33]	51.6
Proposed Cues-parameters	**68.4**

TABLE V.
RECOGNITION ACCURACY COMPARISON USING MSRDAILYACTIVITY3D

Methods	Recognition Accuracy
Motion templates [31]	54.0
Eigenjoints [36]	58.1
Spatiotemporal features [37]	63.7
HON4D [19]	80.0
Actionlet ensemble [26]	85.7
Multilayer perceptron [45]	87.6
Cuboid Similarity Feature [38]	88.2
Deep learning approach [46]	90.5
Proposed Cues-parameters	**91.2**

2) MSRDailyActivity3D

We evaluate our proposed cues-parameters method using MSRDailyActivity3D dataset [34, 35]. We performed the experiment using the depth silhouettes and the joint information together in cues-parameters. We used existing methods [19, 31, 36] where [31] and [36] mainly deal with joint points information to monitor the human movements and recognize human activities/actions. While, in [10], a novel approach is proposed for human action recognition with histograms of 4D joint locations (HON4D) as a compact representation to recognize different activities. In [26], Wang et al developed the actionlet ensemble model that used 3D point cloud to model body shape and reported a recognition accuracy of 85.7%. In [38], Xia and Aggarwal proposed a novel depth cuboid similarity feature (DCSF) to describe the local 3D depth cuboid around the depth images to recognize different actions/activities.

Table II demonstrates the confusion matrix of 16 different human activities that is obtained from the proposed cues parameters method on the MSRDailyActivity3D dataset.

Besides, we compare the experimental results of our proposed method with the algorithms defined as state-of-the-arts methods and the results are shown in Table V. As the table shows, our proposed method performs much better than state-of-the-art methods. It is clearly shown that the proposed method achieved significantly better recognition accuracy as 91.2% than the state-of-the-art methods as 54.0%, 58.1%, 72.1%, 80.0%, 85.7% and 88.2%, respectively.

3) MSRAction3D

The confusion matrix of 20 different human actions that is obtained from the proposed cues parameters method on the MSRAction3D dataset is shown in Table III.

We compare the performance of our proposed method with the state of the art methods using MSRAction3D dataset. In the first approach [39], developed by Lv and Nevatia, learning-based algorithm for automatic recognition and segmentation of 3D human actions is defined. In [27] and [36], bag of 3D model and position differences of joints as eigenjoints are analyzed for action recognition. While, actionlet ensemble [26], pose set [42] and moving pose [43] methods deal with body joints information for activity classification and recognition.

To make a fair comparison, we performed cross subject test between proposed method and the state-of-the-art methods because the cross subject is more challenging due to dynamic intra-class differences in actions among different subjects. The proposed cues-parameters method achieves the recognition accuracy of 92.9% which significantly outperforms the existing methods, as listed in Table VI.

TABLE VI.
RECOGNITION ACCURACY COMPARISON OF OUR METHOD AND PREVIOUS METHODS USING MSRACTION3D.

Methods	Recognition Accuracy
Hidden Markov model [39]	63.0
Bag of 3D points [27]	74.7
Shape and motion features [40]	82.1
Eigenjoints [36]	82.3
Hybrid features [41]	83.6
Actionlet ensemble [26] Multilayer perceptron [45]	88.2 88.7
Pose Set [32] Deep learning approach [46]	90.0 91.3
Moving Pose [43]	91.7
Proposed Cues-parameters	**92.9**

IV. CONCLUSION

In this paper, we proposed a novel methodology having multi-joints features along with advanced HMM for BR system using depth sequences. Our proposed BR system utilizes a combination of human silhouettes identification and tracking process using modified connected component labeling method. In-addition, it includes angular direction, spatiotemporal velocity and invariant features which are used to extract local body part information, intensity differentiation and temporal variation properties to reinforce the feature classification and accuracy. Finally, these features are modeled, trained and recognized using recognizer engine. The proposed method is implemented with Matlab (R2009b) using an Intel Core i3 processor with 4 GB RAM in Windows XP platform. We evaluated the proposed method using three different datasets as IM-DailyDepthActivity, MSRDailyActivity3D and MSRAction3D. Experimental results showed some promising performance of the proposed BR method over the state of the art methods.

In the future work, we will improve the effectiveness of our system by adding biometric identification techniques [44] such as face reorganization, iris detection, gesture recognition, etc., with the help of involving multi-cameras views or RGB contents in Kinect sensors. Besides, we will also use combined cues of biometric identification cues [44] along with proposed cues-parameters techniques to cover large interventions of worldly-scenarios and act as multi-mode biometric system.

REFERENCES

[1] H. Chaminda, V. Klyuev and K. Naruse, "A smart reminder system for complex human activities," in *International Conference on Advanced Communication Technology*, 2012, pp. 235-240.

[2] V. B. Semwal, N. Gaud and G. C. Nandi, "Human Gait State Prediction Using Cellular Automata and Classification Using ELM," in *International Conference on* Machine Intelligence and signal processing, 2017.

[3] A. Jalal and A. Shahzad, "Multiple facial feature detection using Vertex-Modeling structure," in *International Conference on Interactive Computer Aided Learning (ICL)*, 2007, pp. 1-7.

[4] M. Mehmood A. Jalal, and H. A. Evans, "Facial Expression Recognition in Image Sequences Using 1D Transform and Gabor Wavelet Transform," in *IEEE International Conference on Applied and Engineering Mathematics*, 2018.

[5] M. Raj, V. B. Semwal, and G. C. Nandi. "Bidirectional association of joint angle trajectories for humanoid locomotion: the restricted Boltzmann machine approach," *Neural Computing and Applications*, pp. 1-9, 2016.

[6] A. Jalal and I. Uddin, "Security architecture for third generation (3G) using GMHS cellular network," in *IEEE Conference on Emerging Technologies*, 2007, pp. 74-79.

[7] A. Jalal and Y. Rasheed, "Collaboration achievement along with performance maintenance in video streaming," in *Proceedings of the IEEE conference on Interactive computer aided learning*, pp. 1-8, 2007.

[8] V. B. Semwal and G. C. Nandi, "Toward developing a computational model for bipedal push recovery–a brief," *IEEE Sensors Journal,* vol. 15, no. 4, pp. 2021-2022, 2015.

[9] A. Jalal, and M. A. Zeb, "Security enhancement for e-learning portal," *International Journal of Computer Science and Network Security*, vol. 8, no. 3, pp. 41-45, 2008.

[10] A. Jalal and S. Kim, "Global security using human face understanding under vision ubiquitous architecture system," *World Academy of Science, Engineering, and Technology*, vol. 13, pp. 7-11, 2006.

[11] V. B. Semwal, J. Singha, P. K. Sharma, A. Chauhan and B. Behera, "An optimized feature selection technique based on incremental feature analysis for bio-metric gait data classification," *Multimedia tools and applications,* vol. 76, no. 22, pp. 24457-24475, Nov. 2017.

[12] A. Jalal and S. Kim, S, "Advanced performance achievement using multi-algorithmic approach of video transcoder for low bit rate wireless communication," *ICGST International Journal on graphics, vision and image processing*, vol. 5, no. 9, pp. 27-32, 2005.

[13] A. Jalal and S. Kim, "The Mechanism of Edge Detection using the Block Matching Criteria for the Motion Estimation," *Proc. Human Computer Interaction*, pp.484-489, Jan. 2005.

[14] G. C. Nandi, V. B Semwal, M. Raj and A. Jindal, "Modeling bipedal locomotion trajectories using hybrid automata," *Proc. IEEE Region 10 Conference (TENCON),* pp. 1013-1018, 2016.

[15] A. Jalal, N. Sharif, J. T. Kim and T. S. Kim, "Human activity recognition via recognized body parts of human depth silhouettes for residents monitoring services at smart home," *Indoor and Built Environment*, vol. 22, no. 1, pp. 271-279, January, 2013.

[16] P. Turaga, R. Chellappa, V. S. Subrahmanian and O. Udrea, "Machine recognition of human activities: A survey," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 11, pp. 1473-1488, November, 2008.

[17] X. Yang, C. Yang and Y. Tian, "Recognizing actions using depth motion maps-based histograms of oriented gradients," in *International Conference on Multimedia (ICM)*, 2012, pp. 1057-1060.

[18] A. Jalal, Y. Kim, and D. Kim, "Ridge body parts features for human pose estimation and recognition from RGB-D video data," *in Conference on computing, communication and networking technologies*, 2014, pp. 1-6.

[19] O. Oreifej and Z. Liu, "Hon4d: Histogram of oriented 4d normal for activity recognition from depth sequences," in *Conference on Computer Vision and Pattern Recognition*, 2013, pp. 716-723.

[20] A. Jalal, S. Kamal and D. Kim, "A depth video sensor-based life-logging human activity recognition system for elderly care in smart indoor environments," *Sensors*, vol. 14, no. 7, pp. 11735-11759, July, 2014.

[21] A. Jalal, Y.-H. Kim, Y.-J. Kim, S. Kamal and D. Kim, "Robust human activity recognition from depth video using spatiotemporal multi-fused features, *Pattern recognition*, vol. 61, pp. 295-308, 2017.

[22] A. Jalal and Y. Kim, "Dense Depth Maps-based Human Pose Tracking and Recognition in Dynamic Scenes Using Ridge Data," in *Conference on Advanced Video and Signal-Based Surveillance*, 2014, pp. 119-124.

[23] V. B. Semwal and G. C. Nandi, "Generation of joint trajectories using hybrid automate-based model: a rocking block-based approach," *IEEE Sensors Journal*, vol. 16, no. 14, pp. 5805-5816, May, 2016.

[24] M. Raj, V. B. Semwal and G. C. Nandi, "Multiobjective optimized bipedal locomotion," *International Journal of Machine Learning and Cybernetics*, pp. 1-17, 2017.

[25] A. Jalal, S. Y. Lee, J. T. Kim, T. S. Kim, "Human activity recognition via the features of labeled depth body parts," in *International Conference on Smart Homes and Health Telematics*, 2012, pp. 246-249.

[26] J. Wang, Z. Liu, Y. Wu, J. Yuan, "Mining actionlet ensemble for action recognition with depth cameras," in *International Conference on Computer Vision and Pattern Recognition*, 2012, pp. 1290-1297.

[27] W. Li, Z. Zhang and Z. Liu, "Action recognition based on a bag of 3D points," in *International Workshop on Computer Vision and Pattern Recognition*, 2010, pp. 9-14.

[28] A. Jalal, J. T. Kim, and T.-S Kim, "Development of a life logging system via depth imaging-based human activity recognition for smart homes," *in Proceedings of the International Symposium on Sustainable Healthy*

Buildings, 2012, pp. 91-95.

[29] A. Jalal, S. Kamal and D.-S. Kim, "Detecting Complex 3D Human Motions with Body Model Low-Rank Representation for Real-Time Smart Activity Monitoring System," *KSII Transactions on Internet and Information Systems*, vol. 12, no. 3, pp. 1189-1204, 2018.

[30] A. Jalal, J. T. Kim, and T.-S. Kim, "Human activity recognition using the labeled depth body parts information of depth silhouettes," *in Proceedings of the 6th international symposium on Sustainable Healthy Buildings*, 2012, pp. 1-8.

[31] M. Muller and T. Roder, "Motion templates for automatic classification and retrieval of motion capture data," in *SIGGRAPH/Eurographics symposium on computer animation*, 2006, pp. 137-146.

[32] X. Seidenari, C. Varano, Y. Berretti, C. Bimbo and Y. Pala, "Recognizing actions from depth cameras as weakly aligned multi-part bag-of-poses," in *Conference on Computer Vision and Pattern Recognition Workshops*, 2013, pp. 479-485.

[33] B. Ni, G. Wang and P. Moulin, "RGBD-HuDaAct: A color-depth video database for human daily activity recognition," in *Conference on Computer Vision Workshops*, 2011, pp. 1147-1153.

[34] A. Jalal, S. Kamal and D. Kim, "Depth Silhouettes Context: A new robust feature for human tracking and activity recognition based on embedded HMMs," in *International Conference on Ubiquitous Robots and Ambient Intelligence*, 2015, pp. 294-299.

[35] A. Jalal, Y. Kim, S. Kamal, A. Farooq and D. Kim, "Human daily activity recognition with joints plus body features representation using Kinect sensor," in *International Conference on Informatics, electronics and vision*, 2015, pp. 1-6.

[36] X. Yang and Y. Tian, "Eigenjoints-based action recognition using naive-bayes-nearest-neighbor," in *Conference on Computer vision and pattern recognition workshops*, 2012, pp. 14-19.

[37] A. Jalal, S. Kamal and D. Kim, "Human depth sensors-based activity recognition using spatiotemporal features and hidden markov model for smart environments, *J. of computer networks and communications,* pp. 1-11, 2016.

[38] L. Xia and J. Aggarwal, "Spatio-Temporal Depth Cuboid Similarity Feature for Activity Recognition Using Depth Camera," in *International Conference on Computer Vision and Pattern Recognition*, 2013, pp. 2834-2841.

[39] F. Lv and R. Nevatia, "Recognition and segmentation of 3-D human action using HMM and multi-class adaboost," in *European Conference on Computer Vision*, 2006, pp. 359-372.

[40] A. Jalal, S. Kamal and D. Kim, "Shape and motion features approach for activity tracking and recognition from Kinect video camera," *in International Conference on Advanced Information Networking and Applications Workshops*, 2015, pp. 445-450.

[41] Shaharyar Kamal and Ahmad Jalal, "A hybrid feature extraction approach for human detection, tracking and activity recognition using depth sensors," *Arabian J. of Science and Engineering,* vol. 41, no. 3, pp. 1043-1051, 2016.

[42] C. Wang, Y. Wang and A. Yuille, "An Approach to Pose-Based Action Recognition," in *International Conference on Computer Vision and Pattern Recognition*, 2013, pp. 915-922.

[43] M. Zanfir, M. Leordeanu and C. Sminchisescu, "The Moving Pose: An Efficient 3D Kinematics Descriptor for Low-Latency Action Recognition and Detection," in *International Conference on Computer Vision*, 2013, pp. 2752-2759.

[44] V. B. Semwal, J. Singha, P. K. Sharma, A. Chauhan and B. Behera, "An optimized feature selection technique based on incremental feature analysis for bio-metric gait data classification," *Multimedia Tools and Applications*, vol. 76, no. 22, pp. 24457-24475, 2017.

[45] V. B. Semwal, M. Raj, and G. C. Nandi, "Biometric gait identification based on a multilayer perceptron," *Robotics and Autonomous Systems,* vol. 65, pp. 65-75, 2015.

[46] V. B. Semwal, K. Mondal, and G. C. Nandi, "Robust and accurate feature selection for humanoid push recovery and classification: deep learning approach," *Neural Computing and Applications*, vol. 28, no. 3, pp. 565-574, 2017.

Ahmad Jalal

A. Jalal received his M.S. degree in Computer Science from Kyungpook National University, Republic of Korea. He received his Ph.D. degree in the Department of Biomedical Engineering at Kyung Hee University, Republic of Korea. His research interest includes human computer interaction, image processing, and computer vision.

Shaharyar Kamal

S. Kamal received his M.S. degree in Computer Engineering from Mid Sweden University, Sweden. He obtained Ph.D. degree from the Department of Radio and Electronics Engineering, Kyung Hee University, Republic of Korea. His research interest includes 5G, IoT, Cyber Security and Image & Signal Processing.