

## Research Article

# Human Depth Sensors-Based Activity Recognition Using Spatiotemporal Features and Hidden Markov Model for Smart Environments

Ahmad Jalal,<sup>1</sup> Shaharyar Kamal,<sup>2</sup> and Daijin Kim<sup>1</sup>

<sup>1</sup>Pohang University of Science and Technology (POSTECH), Pohang, Republic of Korea

<sup>2</sup>KyungHee University, Suwon, Republic of Korea

Correspondence should be addressed to Ahmad Jalal; [ahmadjalal@postech.ac.kr](mailto:ahmadjalal@postech.ac.kr)

Received 30 June 2016; Accepted 15 September 2016

Academic Editor: Liangtian Wan

Copyright © 2016 Ahmad Jalal et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Nowadays, advancements in depth imaging technologies have made human activity recognition (HAR) reliable without attaching optical markers or any other motion sensors to human body parts. This study presents a depth imaging-based HAR system to monitor and recognize human activities. In this work, we proposed spatiotemporal features approach to detect, track, and recognize human silhouettes using a sequence of RGB-D images. Under our proposed HAR framework, the required procedure includes detection of human depth silhouettes from the raw depth image sequence, removing background noise, and tracking of human silhouettes using frame differentiation constraints of human motion information. These depth silhouettes extract the spatiotemporal features based on depth sequential history, motion identification, optical flow, and joints information. Then, these features are processed by principal component analysis for dimension reduction and better feature representation. Finally, these optimal features are trained and they recognized activity using hidden Markov model. During experimental results, we demonstrate our proposed approach on three challenging depth videos datasets including IM-DailyDepthActivity, MSRAction3D, and MSRDailyActivity3D. All experimental results show the superiority of the proposed approach over the state-of-the-art methods.

## 1. Introduction

Human tracking and activity recognition are defined as recognizing different activities by considering activity feature extraction and pattern recognition techniques based on specific input data from innovative sensors (i.e., motion sensors and video cameras) [1–5]. In recent years, advancement of these sensors has boosted the production of novel techniques for pervasive human tracking, observing human motion, detecting uncertain events [6–8], silhouette tracking, and emotion recognition in the real-world environments [9–11]. In these domains, the term which is most commonly used to cover all these topics is technically termed as human tracking and activity recognition [12–14]. In the motion sensors-based activity recognition, activity recognition is based on classifying sensory data using one or more sensor devices. In [15], Casale et al. proposed a complete review about the state-of-the-art activity classification methods using data

from one or more accelerometers. In this work, classification approaches are based on RFs features which classify five daily routine activities from bluetooth accelerometer placed at breast of the human body, using a 319-dimensional feature vector. In [16], fast FFT and decision tree classifier algorithm are proposed to detect physical activity using biaxial accelerometers attached on different parts of the human body. However, these motion sensors-based approaches are not feasible methods for recognition due to uncomfot of the users to wear electronic sensors in their daily life. Also, combining multiple sensors for improvement in recognition performance causes high computation load. Thus, video-based human tracking and activity recognition is proposed where the depth features are extracted from a RGB-D video camera.

Depth silhouettes have made proactive contributions and are the most famous representation for human tracking and activity recognition from which useful human shape features

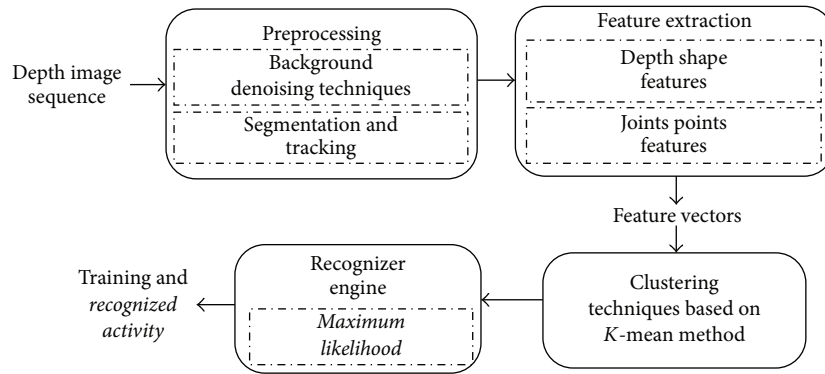


FIGURE 1: System architecture of the proposed human activity recognition system.

are extracted. These depth silhouettes explore research issues and are used as practical applications including life-care systems, surveillance system, security system, face verification, patient monitoring systems, and human gait recognition systems. In [17], several algorithms are developed for feature extraction from the silhouette data of the tracked human subject using depth images as the pixel source. These parameters include ratio of height to weight of the tracked human subject. Also, motion characteristics and distance parameters are used as features for the activity recognition. In [14], a novel life logging translation and scaling invariant features approach is designed where 2D maps are computed through Radon transform which are further processed as 1D feature profiles through R transform. These features are further reduced by PCA and symbolized by Linde, Buzo, and Gray (LBG) clustering technique to train and recognize different activities. In [18], a discriminative representation method is proposed as structure-motion kinematics features including the structure similarity and head-floor distance based on skeleton joint points information. However, these effective trajectory projection based kinematic schemes are learnt by a SVM classifier to recognize activities using the depth maps. In [19], an activity recognition system is designed to provide continuous monitoring and recording of daily life activities. The system includes depth silhouettes as an input to produce skeleton model and its body points information. This information is used as features and is computed using a set of magnitude and direction angle features which are further used for training and testing via hidden Markov models (HMMs). These state-of-the-art methods [14, 17–19] proved more efficiency for recognition accuracy using depth silhouette. However, it is still difficult to find best features from limited information such as joint points information especially during occlusions. It shows bad impact over recognition accuracy. Therefore, we needed to develop methodology which provides combined effects of full-body silhouettes and joints information to improve activity recognition performance.

In this paper, we proposed a novel method to recognize activities using sequence of depth images. During preprocessing steps, we extracted human depth silhouettes using background/floor removal techniques and tracked human

silhouettes by considering rectangular box having body shape measurements (i.e., height and width) to adjust the box's size. During spatiotemporal features extraction technique, a set of multifused features are considered as depth sequential history, motion identification, optical flow, joints angle, and joints location features. These features are further computed by principal component analysis (PCA) for global information and reduce dimensions. Then, these features are applied over  $K$ -mean for clustering and fed into a four-state left-to-right HMM for training/testing human activities. The proposed system is compared against the state-of-the-art approaches thus achieving best recognition rate over three challenging depth videos datasets as IM-DailyDepthActivity, MSRAction3D, and MSRDailyActivity3D datasets.

The rest of the sections of this paper are structured as follows. Section 2 describes the system architecture overview of the proposed system where depth maps preprocessing, feature extraction techniques, and training/testing human activities using HMM are explained. In Section 3, we explain experimental results by considering proposed and state-of-the-art methods. Finally, Section 4 presents the conclusion.

## 2. Proposed System Methodology

**2.1. System Overview.** The proposed activity recognition system consists of sequence of depth images captured by RGB-D video sensor, background removal, and human tracking from the time-sequential activity video images. Then, feature representation based on spatiotemporal features, clustering via  $K$ -mean, and training/testing using recognizer engine are processed. Figure 1 explains the overall steps of proposed human activity recognition system.

**2.2. Depth Images Preprocessing.** During vision-based image preprocessing, we captured video data (i.e., digital and RGB-D) that retrieve both binary and depth human silhouettes from each activity. In case of binary silhouette, we received color images from digital camera which are further converted into binary images. In case of using depth silhouettes, we obtained the depth images from the depth cameras (i.e., PrimeSense, Bumblebee, and Zcam) to extract  $320 \times 240$



FIGURE 2: Images comparison as (a) binary silhouettes and (b) depth silhouettes of exercise and sit-down activities.

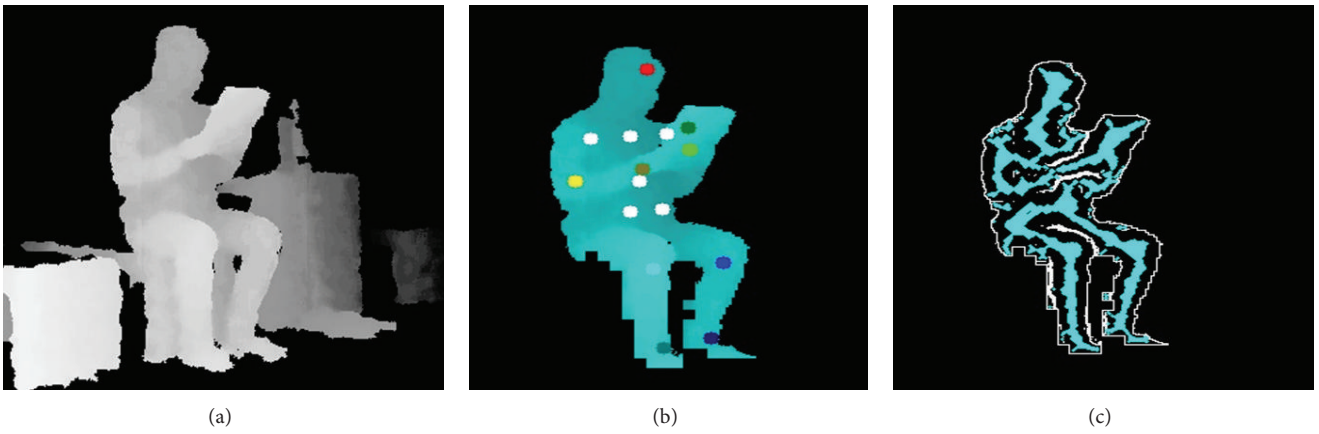


FIGURE 3: Depth images preprocessing as (a) raw depth images having noisy background, (b) labelled human silhouette, and (c) ridge body information.

depth levels per pixel [20, 34, 35]. These cameras deal with both RGB images and depth raw data.

For comparative study, it is examined that, in case of binary images, we can only obtain minimum information (i.e., black or white), while significant pixels values are dropped especially hands movement in front of chest or both legs crossing each other. However, in case of depth silhouettes, we received maximum information in the form of intensity values and additional body parts information (i.e., joint points), controlled during self-occlusion (see Figure 2).

Therefore, to deal with depth images, we remove noisy effects from background by simply ignoring ground line (i.e.,  $y$  parameters) which acts as lowest value (i.e., equal to zero) corresponding to a given pair of  $x$ - and  $z$ -axis for floor removal. Next, we partitioned all objects in the frame using variation of intensity values in between consecutive frames. Then, we differentiate the depth values of corresponding neighboring pixels within a specific threshold and extract depth human silhouettes using depth center values of each object from the scenes. Finally, we apply human tracking by considering temporal continuity constraints (see Figure 3) between consecutive frames [21, 27], while human silhouettes

are enclosed within the rectangular bounding box having specific values (i.e., height and width) based on face recognition and motion detection [36–38].

**2.3. Spatiotemporal Features Extraction.** For spatiotemporal features extraction, we composed features as depth history silhouettes, standard deviation, motion variation among images, and optical flow for depth shape features, while joints angle and joints location features are derived from joints points features. Combination of these features explores more spatial and temporal depth-based properties which are useful for activity classification and recognition. All features are explained below.

**Depth Sequential History.** Depth sequential history feature method is used to observe pixel intensity information in overall sequence of each activity (see Figure 4). It contains temporal values, position, and movement velocities. Therefore depth sequential history is defined as

$$H = \frac{1}{T} \sum_{t=t_i}^{t_f} I_t, \quad (1)$$



FIGURE 4: Depth sequential history features applied over exercise, kicking, and cleaning activities.

where  $t_f$  and  $t_i$  are the initial and final images  $I$  of an activity and  $T$  is the duration of activity period.

*Different Intensity Values Features.* Standard deviation is computed as the sum of all the differences of the image pairs  $P$  with respect to the time series (see Figure 5). It provides quite disperse output and hidden values (i.e., especially  $z$  coordinates) having large range of intensities values

$$SD(x, y, z) = \sqrt{\frac{\sum_{t=t_i}^{t_f} S_t^2}{P} - \left(\frac{1}{P} \sum_{t=t_i}^{t_f} S_t\right)^2}. \quad (2)$$

*Depth Motion Identification.* Motion identification feature mechanism is used to handle intra-/intermotion variation and temporal displacement (see Figure 6) among consecutive frames of each activity.

$$MV(x, y, z) = \frac{SD(x, y, z)}{H}. \quad (3)$$

*Motion-Based Optical Flow Features.* In order to make use of the additional motion information from depth sequence, we applied optical flow technique based on the Lucas Kanade method. Basically, it calculates the motion intensity and directional angular values between two images. Figure 7 shows some samples of optical flows calculated from two depth silhouettes images.

*Joints Angle Features.* Due to similar or complex postures of different activities, it is not sufficient to just deal with silhouettes features; therefore, we developed skeleton model having 15 joints' points information (see Figure 8).

However, joints angle features measure the directional movements of the  $i$ th joints points between consecutive frames [39, 40]  $t$  and  $t - 1$  as

$$\theta_{\cos} = \cos^{-1} \left( \frac{J_i^t * J_i^{t-1}}{|J_i^t| |J_i^{t-1}|} \right), \quad (4)$$

where  $J_i$  indicates all three coordinate axes ( $x, y, z$ ) of the body joints with respect to consecutive frames [41–43].

*Joints Location Features.* Joints location features measure the distance between the torso joint point  $J_C$  and all other fourteen joints' points  $J_o$  in each frame  $t$  of sequential activity as

$$F_{Loca}(t) = \sqrt{(J_C^x - J_o^x)^2 + (J_C^y - J_o^y)^2 + (J_C^z - J_o^z)^2}. \quad (5)$$

Finally, we obtained the feature vector size of joints angle and joints location features as  $1 \times 15$  and  $1 \times 14$  dimensions. Figures 9(a) and 9(b) shows the 1D plots of both joints angle and joints location features for exercise, kicking, and cleaning activities.

*2.4. Feature Reduction.* Since spatiotemporal feature extraction using depth shape features consists of larger number of features dimension, thus PCA is introduced to extract global information [44, 45] from all activities data and approximate the higher features dimension data [46] into lower dimensional features. In this work, 750 principal components of the spatiotemporal features are chosen from the whole PC feature space and the size of feature vector becomes  $1 \times 750$ .

*2.5. Symbolization, Training, and Recognition.* Each feature vector of individual activity is symbolized based on  $K$ -mean clustering algorithm. However, a HMM consists of finite states where each state is involved in transition probability and symbol observation probability [47, 48]. During HMM, the underlying hidden process is observable by another set of stochastic processes that provides observation symbols. In case of training each activity, initially, HMM is trained having a size of codebook of 512. During HAR, trained HMMs of each activity are used to choose maximum likelihood of desired activity [49–52]. However, sequence of trained data is generated and maintained by buffer strategy [31, 53]. Figure 10 describes the transition and emission probabilities of cleaning HMM after training.





FIGURE 5: Different intensity values features applied over exercise, kicking, and cleaning activities.

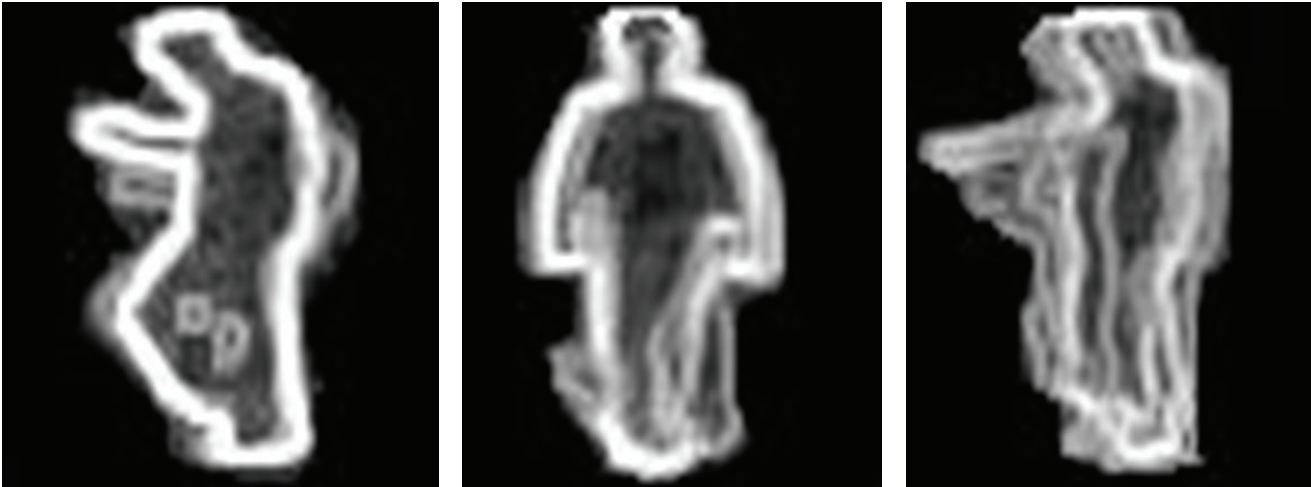


FIGURE 6: Depth motion identification features applied over exercise, kicking, and cleaning activities.

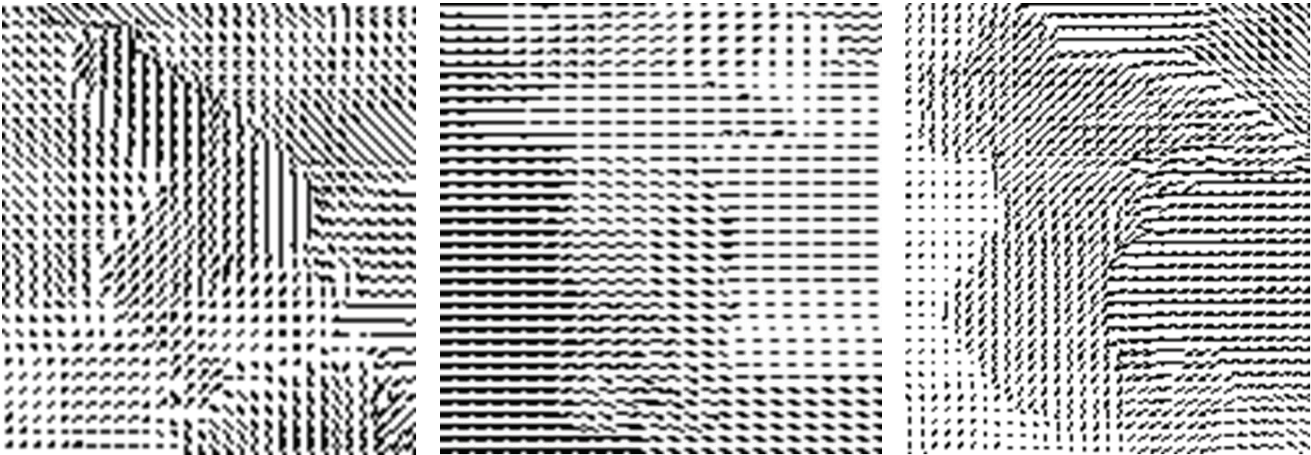


FIGURE 7: Examples of depth silhouettes images using motion-based optical flow.

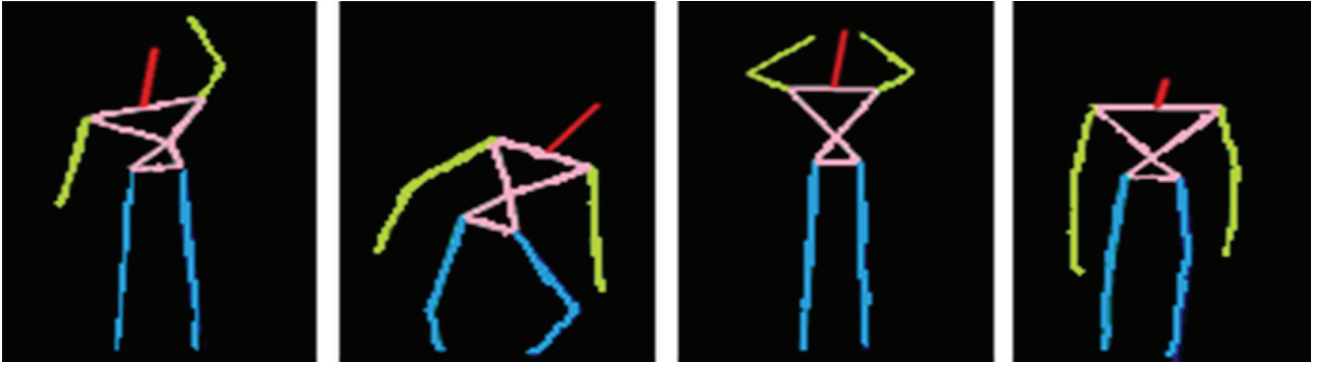


FIGURE 8: Samples of human skeleton model for different activities.

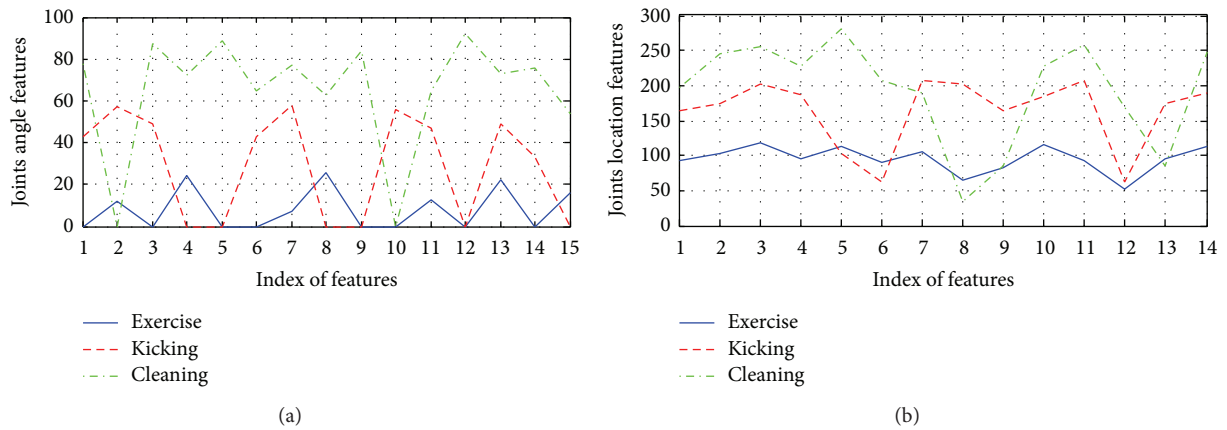


FIGURE 9: 1D plots of (a) joints angle and (b) joint locations features for exercise, kicking, and cleaning activities.

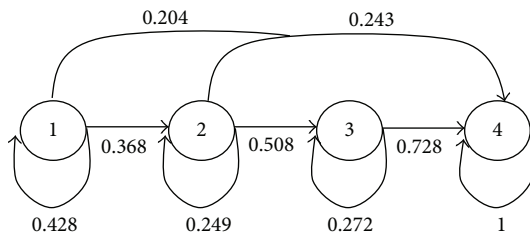


FIGURE 10: Transition and emission probabilities distribution of cleaning activity of trained HMM based on left-to-right HMM approach.

### 3. Experimental Results and Descriptions

**3.1. Experimental Settings.** The proposed method is evaluated on three challenging depth videos datasets. First is our own annotated depth dataset known as IM-DailyDepthActivity [54]. It includes fifteen types of activities as *sitting down*, *both hands waving*, *bending*, *standing up*, *eating*, *phone conversation*, *boxing*, *clapping*, *right hand waving*, *exercise*, *cleaning*, *kicking*, *throwing*, *taking an object*, and *reading an article*. During experimental evaluation, we used 375 videos sequences for training and 30 unsegmented videos for testing. All videos are collected in indoor environments (i.e., labs,

classroom, and halls) performed by 15 different subjects. Figure 11 shows some depth activities images used in IM-DailyDepthActivity dataset.

Second is public depth database as MSRAction3D dataset and third is MSRDailyActivity3D dataset. In the following sections, we explain and compare our method with other state-of-the-art methods using all three depth datasets.

**3.2. Comparison of Recognition Rate of Proposed and State-of-the-Art Methods Using IM-DailyDepthActivity.** We compare our spatiotemporal features method with the state-of-the-art methods including body joints, eigenjoints, depth motion maps, and super normal vector features using depth images. It is cleared in Table 1 that the spatiotemporal features achieved highest recognition rate as 63.7% over the state-of-the-art methods.

**3.3. Recognition Results of Public Dataset (MSRAction3D).** The MSRAction3D dataset is a public dataset captured by a Kinect camera based on game consoles phenomenon. It includes twenty actions as *high arm wave*, *horizontal arm wave*, *hammer*, *hand catch*, *forward punch*, *high throw*, *drawing X*, *drawing tick*, *drawing circle*, *hand clap*, *two-hand wave*, *side boxing*, *bending*, *forward kicking*, *side kicking*, *jogging*, *tennis swing*, *tennis serve*, *golf swing*, and *pickup and*



FIGURE 11: Examples of depth activities images used in IM-DailyDepthActivity dataset.



FIGURE 12: Examples of depth actions images used in MSRAction3D dataset.

TABLE 1: Comparison of recognition accuracy using IM-DailyDepthActivity.

Method	Recognition accuracy
Recognized body parts features [19]	28.4
Eigenjoints [20]	40.3
Super normal vector [1]	51.6
Depth silhouettes context features [21]	57.69
<i>Proposed method</i>	63.7

*throw*. The overall dataset consists of 567 (i.e., 20 actions  $\times$  10 subjects  $\times$  2 or 3 trails) depth map sequences. Also, this dataset is quite complex due to similar postures of different actions. Examples of different actions used in this dataset are shown in Figure 12.

To perform experimentation over MSRAction3D, we evaluated all 20 actions and examined their recognition accuracy performance based on LOSO (leave-one-subject-out) cross-subject training/testing mechanism. Table 2 shows the recognition accuracy of this dataset.

While some other researchers used MSRAction3D [22–26, 28] dataset by dividing it into action set 1, action set 2, and action set 3 as mentioned in [22], we compare the recognition performance of spatiotemporal method with other state-of-the-art methods in Table 3. All methods are implemented by us using similar instructions provided by their respective papers.

TABLE 2: Recognition results of proposed HAR system.

Actions	Accuracy
High arm wave	93.7
Horizontal arm wave	90.4
Hammer	89.2
Hand catch	95.5
Forward punch	91.1
High throw	89.7
Drawing X	93.4
Drawing tick	97.1
Drawing circle	92.3
Hand clap	89.7
Two-hand wave	94.5
Side boxing	92.4
Bending	86.8
Forward kicking	91.3
Side kicking	98.6
Jogging	96.7
Tennis swing	87.1
Tennis serve	86.9
Golf swing	97.8
Pickup and throw	94.3
<i>Mean recognition accuracy = 92.4</i>	

**3.4. Recognition Results of Public Dataset (MSRDailyActivity3D).** The MSRDailyActivity3D dataset is a depth activity dataset collected by a Kinect device based on living room daily routine. It includes sixteen activities as *stand up*, *sit*



FIGURE 13: Some depth images used in MSRDailyActivity3D dataset.

TABLE 3: Comparison of recognition accuracy of proposed method and state-of-the-art methods using MSRAction3D dataset.

Method	Recognition accuracy
Bag of 3D points [22]	74.7
Shape and motion features [23]	82.1
Eigenjoints [20]	82.3
Spatiotemporal motion variation [24]	84.6
STOP features [25]	84.8
Joints plus body features [26]	85.6
Actionlet ensemble [11]	88.2
Cuboid similar features [27]	89.3
Spatial and temporal part-sets [28]	90.2
Pose-based features [29]	91.5
<i>Proposed method</i>	92.4

down, walk, drink, write on a paper, eat, read book, call on cell phone, use laptop, use vacuum cleaner, cheer up, sit still, toss paper, play game, lie down on sofa, and play guitar. This dataset includes 320 (i.e., 16 activities  $\times$  10 subjects  $\times$  2 trails) depth videos activities mostly operated in a room. These activities also involved human-object interactions. Some of the examples of the MSRDailyActivity3D dataset are shown in Figure 13.

Table 4 shows the accuracy performance of 16 different human activities that is obtained from the proposed spatiotemporal features method over the specific dataset.

Finally, we reported the comparison of recognition accuracy over the MSRDailyActivity3D dataset where the proposed method shows superior recognition rate over state-of-the-art methods in Table 5.

#### 4. Conclusions

In this paper, we proposed spatiotemporal features based on depth images derived from Kinect camera for human activity recognition. The features include depth sequential history to represent the spatial-temporal information of human silhouettes in each activity, motion identification to calculate the change among motion in between consecutive frames, and optical flow to represent in the form of partial image to get optimum depth information. During

TABLE 4: Recognition results of proposed HAR system.

Activities	Accuracy
Drink	95.4
Eat	87.8
Read book	96.3
Call phone	89.1
Write	96.7
Use laptop	90.5
Use cleaner	100
Cheer up	88.5
Sit still	94.6
Toss paper	94.2
Play game	86.3
Lay down	98.7
Walk	95.3
Play guitar	88.2
Stand up	96.5
Sit down	93.6
<i>Mean recognition accuracy = 93.2</i>	

TABLE 5: Comparison of recognition accuracy of using MSRDailyActivity3D dataset.

Method	Recognition accuracy
Only joints position features [11]	68.0
Moving pose [6]	73.8
Motion features [30]	79.1
Hybrid features [31]	85.3
Actionlet ensemble [11]	85.7
Super normal vector [1]	86.2
Volumetric spatial features [32]	89.7
Spatiotemporal features learning [33]	90.4
<i>Proposed method</i>	93.2

experimental results, these features are applied over proposed IM-DailyDepthActivity, MSRAction3D, and MSRDailyActivity3D datasets, respectively. Our proposed activity recognition system shows superior recognition accuracy performance as 63.7% over the state-of-the-art methods using our depth annotated dataset. In case of public datasets,



our method achieved accuracy performance as 92.4% and 93.2%, respectively. Our future work needs to explore more enhanced feature techniques for complex activities and multiple person interactions.

## Competing Interests

The authors declare that there are no competing interests regarding the publication of this paper.

## Acknowledgments

The research was supported by the Implementation of Technologies for Identification, Behavior, and Location of Human Based on Sensor Network Fusion Program through the Ministry of Trade, Industry and Energy (Grant no. 10041629). This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (B0101-16-0552, Development of Predictive Visual Intelligence Technology).

## References

- [1] X. Yang and Y. Tian, "Super normal vector for activity recognition using depth sequences," in *Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '14)*, pp. 804–811, IEEE, Columbus, Ohio, USA, June 2014.
- [2] A. Jalal, J. Kim, and T. 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*, pp. 1–8, 2012.
- [3] G. Okeyo, L. Chen, and H. Wang, "An agent-mediated ontology-based approach for composite activity recognition in smart homes," *Journal of Universal Computer Science*, vol. 19, no. 17, pp. 2577–2597, 2013.
- [4] A. Jalal, S. Lee, J. Kim, and T. Kim, "Human activity recognition via the features of labeled depth body parts," in *Proceedings of the International Conference on Smart Homes and Health Telematics*, pp. 246–249, June 2012.
- [5] A. Jalal, J. Kim, and T. Kim, "Development of a life logging system via depth imaging-based human activity recognition for smart homes," in *Proceedings of the 8th International Symposium on Sustainable Healthy Buildings*, pp. 91–95, Seoul, Republic of Korea, September 2012.
- [6] M. Zafar, M. Leordeanu, and C. Sminchisescu, "The moving pose: an efficient 3D kinematics descriptor for low-latency action recognition and detection," in *Proceedings of the 14th IEEE International Conference on Computer Vision (ICCV '13)*, pp. 2752–2759, Sydney, Australia, December 2013.
- [7] A. Jalal and Y. Kim, "Dense depth maps-based human pose tracking and recognition in dynamic scenes using ridge data," in *Proceedings of the 11th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS '14)*, pp. 119–124, IEEE, Seoul, Republic of Korea, August 2014.
- [8] F. Xu and K. Fujimura, "Human detection using depth and gray images," in *Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS '03)*, pp. 115–121, IEEE, Colorado Springs, Colo, USA, July 2003.
- [9] A. Jalal and IjazUddin, "Security architecture for third generation (3G) using GMHS cellular network," in *Proceedings of the 3rd International Conference on Emerging Technologies (ICET '07)*, pp. 74–79, IEEE, Islamabad, Pakistan, November 2007.
- [10] A. Farooq, A. Jalal, and S. Kamal, "Dense RGB-D map-based human tracking and activity recognition using skin joints features and self-organizing map," *KSII Transactions on Internet and Information Systems*, vol. 9, no. 5, pp. 1856–1869, 2015.
- [11] J. Wang, Z. Liu, Y. Wu, and J. Yuan, "Mining actionlet ensemble for action recognition with depth cameras," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '12)*, pp. 1290–1297, Providence, RI, USA, June 2012.
- [12] A. Jalal, M. Z. Uddin, J. T. Kim, and T.-S. Kim, "Recognition of human home activities via depth silhouettes and R transformation for smart homes," *Indoor and Built Environment*, vol. 21, no. 1, pp. 184–190, 2012.
- [13] A. Jalal, Y. Kim, and D. Kim, "Ridge body parts features for human pose estimation and recognition from RGB-D video data," in *Proceedings of the IEEE International Conference on Computing, Communication and Networking Technologies (ICCCNT '14)*, pp. 1–6, IEEE, Hefei, China, July 2014.
- [14] A. Jalal, Z. Uddin, and T.-S. Kim, "Depth video-based human activity recognition system using translation and scaling invariant features for life logging at smart home," *IEEE Transactions on Consumer Electronics*, vol. 58, no. 3, pp. 863–871, 2012.
- [15] P. Casale, O. Pujol, and P. Radeva, "Human activity recognition from accelerometer data using a wearable device," in *Proceedings of the 5th International Conference on Pattern Recognition and Image Analysis (IbPRIA '11)*, pp. 289–269, Las Palmas de Gran Canaria, Spain, June 2011.
- [16] L. Bao and S. Intille, "Activity recognition from user-annotated acceleration data," in *Proceedings of the International Conference on Pervasive Computing*, pp. 1–17, April 2004.
- [17] S. Kamal and A. Jalal, "A hybrid feature extraction approach for human detection, tracking and activity recognition using depth sensors," *Arabian Journal for Science and Engineering*, vol. 41, no. 3, pp. 1043–1051, 2016.
- [18] C. Zhang and Y. Tian, "RGB-D camera-based daily living activity recognition," *Journal of Computer Vision and Image Processing*, vol. 2, no. 4, pp. 1–7, 2012.
- [19] A. Jalal, N. Sarif, 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, 2013.
- [20] X. Yang and Y. L. Tian, "EigenJoints-based action recognition using Naïve-Bayes-Nearest-Neighbor," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW '12)*, pp. 14–19, Providence, RI, USA, June 2012.
- [21] A. Jalal, S. Kamal, and D. Kim, "Individual detection-tracking-recognition using depth activity images," in *Proceedings of the 12th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI '15)*, pp. 450–455, Goyang, Republic of Korea, October 2015.
- [22] W. Li, Z. Zhang, and Z. Liu, "Action recognition based on a bag of 3D points," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW '10)*, pp. 9–14, IEEE, San Francisco, Calif, USA, June 2010.
- [23] A. Jalal, S. Kamal, and D. Kim, "Shape and motion features approach for activity tracking and recognition from kinect video camera," in *Proceedings of the 29th IEEE International*

- Conference on Advanced Information Networking and Applications Workshops (WAINA '15)*, pp. 445–450, IEEE, Gwangju, South Korea, March 2015.
- [24] A. Jalal, S. Kamal, A. Farooq, and D. Kim, “A spatiotemporal motion variation features extraction approach for human tracking and pose-based action recognition,” in *Proceedings of the International Conference on Informatics, Electronics and Vision (ICIEV '15)*, pp. 1–6, Fukuoka, Japan, June 2015.
- [25] A. Vieira, E. Nascimento, G. Oliveira, Z. Liu, and M. Campos, “STOP: space-time occupancy patterns for 3d action recognition from depth map sequences,” in *Proceedings of the 17th Iberoamerican Congress on Pattern Recognition*, pp. 252–259, Buenos Aires, Argentina, September 2012.
- [26] 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 *Proceedings of the International Conference on Informatics, Electronics and Vision (ICIEV '15)*, pp. 1–6, IEEE, Fukuoka, Japan, June 2015.
- [27] L. Xia and J. K. Aggarwal, “Spatio-temporal depth cuboid similarity feature for activity recognition using depth camera,” in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '13)*, pp. 2834–2841, IEEE, Portland, Ore, USA, June 2013.
- [28] C. Wang, Y. Wang, and A. L. Yuille, “An approach to pose-based action recognition,” in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '13)*, pp. 915–922, Portland, Ore, USA, June 2013.
- [29] A. Eweiwi, M. S. Cheema, C. Bauckhage, and J. Gall, “Efficient pose-based action recognition,” in *Computer Vision—ACCV 2014: 12th Asian Conference on Computer Vision, Singapore, Singapore, November 1–5, 2014, Revised Selected Papers, Part V*, vol. 9007 of *Lecture Notes in Computer Science*, pp. 428–443, Springer, Berlin, Germany, 2015.
- [30] 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, 2014.
- [31] A. Jalal and S. Kim, “Algorithmic implementation and efficiency maintenance of real-time environment using low-bitrate wireless communication,” in *Proceedings of the IEEE International Workshop on Software Technologies for Future Embedded and Ubiquitous Systems*, pp. 1–6, April 2006.
- [32] S.-S. Cho, A.-R. Lee, H.-I. Suk, J.-S. Park, and S.-W. Lee, “Volumetric spatial feature representation for view-invariant human action recognition using a depth camera,” *Optical Engineering*, vol. 54, no. 3, Article ID 033102, 8 pages, 2015.
- [33] L. Liu and L. Shao, “Learning discriminative representations from RGB-D video data,” in *Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI '13)*, pp. 1493–1500, Beijing, China, August 2013.
- [34] S. Kamal, A. Jalal, and D. Kim, “Depth images-based human detection, tracking and activity recognition using spatiotemporal features and modified HMM,” *Journal of Electrical Engineering and Technology*, vol. 11, no. 3, pp. 1921–1926, 2016.
- [35] A. Jalal, Z. Uddin, J. Kim, and T. Kim, “Daily human activity recognition using depth silhouettes and R transformation for smart home,” in *Proceedings of the 9th International Conference on Toward Useful Services for Elderly and People with Disabilities: Smart Homes and Health Telematics (ICOST '11)*, pp. 25–32, Montreal, Canada, June 2011.
- [36] A. Jalal and S. Kim, “Advanced performance achievement using multi-algorithmic approach of video transcoder for low bit rate wireless communication,” *ICGST Journal on Graphics, Vision and Image Processing*, vol. 5, no. 9, pp. 27–32, 2005.
- [37] 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.
- [38] M. Turk and A. Pentland, “Face recognition using eigenfaces,” in *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, pp. 586–591, Maui, Hawaii, USA, June 1991.
- [39] A. Jalal and Y. A. Rasheed, “Collaboration achievement along with performance maintenance in video streaming,” in *Proceedings of the IEEE Conference on Interactive Computer Aided Learning*, pp. 1–8, Villach, Austria, 2007.
- [40] A. Jalal and M. A. Zeb, “Security and QoS optimization for distributed real time environment,” in *Proceedings of the 7th IEEE International Conference on Computer and Information Technology (CIT '07)*, pp. 369–374, Aizuwakamatsu, Japan, October 2007.
- [41] L. Chen, P. Zhu, and G. Zhu, “Moving objects detection based on background subtraction combined with consecutive frames subtraction,” in *Proceedings of the International Conference on Future Information Technology and Management Engineering (FITME '10)*, pp. 545–548, IEEE, Changzhou, China, October 2010.
- [42] A. Jalal and S. Kim, “The mechanism of edge detection using the block matching criteria for the motion estimation,” in *Proceedings of the Conference on Human Computer Interaction (HCI '05)*, pp. 484–489, January 2005.
- [43] A. Jalal, S. Kim, and B. J. Yun, “Assembled algorithm in the real-time H.263 codec for advanced performance,” in *Proceedings of the 7th International Workshop on Enterprise Networking and Computing in Healthcare Industry (HEALTHCOM '05)*, pp. 295–298, IEEE, Busan, South Korea, June 2005.
- [44] Y.-H. Taguchi and A. Okamoto, “Principal component analysis for bacterial proteomic analysis,” in *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW '11)*, pp. 961–963, IEEE, Atlanta, Ga, USA, November 2011.
- [45] A. Jalal and S. Kim, “A complexity removal in the floating point and rate control phenomenon,” in *Proceedings of the Conference on Korea Multimedia Society*, vol. 8, pp. 48–51, January 2005.
- [46] A. Jalal and A. Shahzad, “Multiple facial feature detection using vertex-modeling structure,” in *Proceedings of the International Conference on Interactive Computer Aided Learning*, pp. 1–7, September 2007.
- [47] P. M. Baggenstoss, “A modified Baum-Welch algorithm for hidden Markov models with multiple observation spaces,” *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 4, pp. 411–416, 2001.
- [48] A. Jalal, Y. Kim, Y. 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.
- [49] A. Jalal, S. Kamal, and D. Kim, “Depth map-based human activity tracking and recognition using body joints features and Self-Organized Map,” in *Proceedings of the 5th International Conference on Computing, Communication and Networking Technologies (ICCCNT '14)*, pp. 1–6, IEEE, Hefei, China, July 2014.
- [50] A. Jalal and S. Kamal, “Real-time life logging via a depth silhouette-based human activity recognition system for smart

- home services,” in *Proceedings of the 11th IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS '14)*, pp. 74–80, Seoul, South Korea, August 2014.
- [51] S.-S. Jarnng, “HMM voice recognition algorithm coding,” in *Proceedings of the International Conference on Information Science and Applications (ICISA '11)*, pp. 1–7, IEEE, Jeju, South Korea, April 2011.
- [52] A. Jalal and M. Zeb, “Security enhancement for e-learning portal,” *International Journal of Computer Science and Network Security*, vol. 8, no. 3, pp. 41–45, 2008.
- [53] 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 *Proceedings of the 12th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI '15)*, pp. 294–299, Goyang, South Korea, October 2015.
- [54] A. Jalal, “IM-DailyDepthActivity dataset,” 2016, <http://imlab.postech.ac.kr/databases.htm>.





**Hindawi**

Submit your manuscripts at  
<http://www.hindawi.com>

