

Dense RGB-D Map-Based Human Tracking and Activity Recognition using Skin Joints Features and Self-Organizing Map

Adnan Farooq¹, Ahmad Jalal¹ and Shaharyar Kamal²

¹Department of Biomedical Engineering, Kyung Hee University
[e-mail: aadnan@khu.ac.kr , ahmjal@yahoo.com]

²School of Electronics and Information, Kyung Hee University
1-Seocheon-dong Giheung-gu Yongin-si, Gyeonggi-do 446-701, Republic of Korea
[e-mail: shka@khu.ac.kr]

*Corresponding author: Shaharyar Kamal

*Received October 28, 2014; revised December 29, 2015; accepted April 19, 2015;
published May 31, 2015*

Abstract

This paper addresses the issues of 3D human activity detection, tracking and recognition from RGB-D video sequences using a feature structured framework. During human tracking and activity recognition, initially, dense depth images are captured using depth camera. In order to track human silhouettes, we considered spatial/temporal continuity, constraints of human motion information and compute centroids of each activity based on chain coding mechanism and centroids point extraction. In body skin joints features, we estimate human body skin color to identify human body parts (i.e., head, hands, and feet) likely to extract joint points information. These joints points are further processed as feature extraction process including distance position features and centroid distance features. Lastly, self-organized maps are used to recognize different activities. Experimental results demonstrate that the proposed method is reliable and efficient in recognizing human poses at different realistic scenes. The proposed system should be applicable to different consumer application systems such as healthcare system, video surveillance system and indoor monitoring systems which track and recognize different activities of multiple users.

Keywords: Human silhouette tracking, body skin joints features, activity recognition, self-organized maps, human activity recognition system

A preliminary version of this paper appeared in IEEE ICC 2009, June 14-18, Dresden, Germany. This version includes a concrete analysis and supporting implementation results on MICAz sensor nodes. This research was supported by a research grant from the IT R&D program of MKE/IITA, the Korean government [2005-Y-001-04, Development of Next Generation Security Technology]. We express our thanks to Dr. Richard Berke who checked our manuscript.

1. Introduction

Human tracking and activity recognition have become one of the active research area in the field of computer vision due to promising applications such as surveillance, health care, multimedia contents, security systems and smart home systems [1-6]. During human tracking, human motion analysis based on his/her silhouette is extracted from the background and noisy regions. The concept of silhouettes tracking concept is merged with the understanding of human behavior to explore larger term which is generally called human activity recognition (HAR). The task for the HAR can be generally defined as, a sequence of data which identifies the action performed by the subjects [7-9]. Although it is easy for the human being to identify each class of the human activity, currently there are very limited intelligent HAR systems which can robustly and efficiently recognize each class of human activity. However, most of the HAR systems consist of difficulties in human tracking and recognition because of various aspects as: Firstly, span of human motion is expanded at very high dimensional space. Secondly, image data captures from the traditional cameras are sensitive to lighting conditions. However, the restriction of sensing devices limits the previous methods to track and recognize the human activity. Furthermore, the human bodies have complex physical structure, so the information loss due to some sensing devices causes major problems in recognition. Thus, with the availability of faster computer hardware and better digital cameras, video based applications have become more and more popular among the researchers.

For instance, in case of video sensors based HAR system, feature sets are generated from video sequence using both RGB and depth video sensors. In RGB cameras, binary and color (digital) images [10-12] are used to recognize human activities [13]. In [13], Iosifidis et al. presented view-invariant activity representation scheme which exploits the global human information in the sense of binary body marks. They used three optimal discriminant subspaces in order to use the activity video for human identity, activity class, viewing angle classification and extracted seven different activities. In [14], Wang et al. used binary silhouettes to recognize the human activities. Using the data set of five human activities, it is processed by applying R transform to extract dominant directional features from binary silhouettes. Features extracted from R transform are used by Principal Component Analysis (PCA) to compact and reduce dimension which ultimately train/test using Hidden Markov model to recognize the activity. In [15], Chin et al. proposed HAR technique using visual stimuli along with manifold learning which allow the characterization of the binary silhouette activity manifold, and performing activity recognition requires distinguishing between manifolds based on ten vastly different activities. However, binary silhouette itself has very limited information due to flat pixel intensity value's (i.e. 0 and 1) which causes low recognition rate especially in case of complex poses and self-occlusions. Thus, an improvement is needed in the field of HAR.

Recently, with the development of the information technology systems and video sensors devices, many researchers in the field of HAR used depth video sensor such as Microsoft Kinect [16-18]. With the expansion of depth sensors and algorithms for the HAR, new chances have emerged in this field. Also, depth sensor technologies have made it feasible and low-cost for the researcher to work on color images as well depth maps. In [18], Shotton et al. described body parts representation, discriminative feature approach, review decision forests and body part recognition using single depth images captured by utilizing the Kinect sensor. The main contribution of Shotton's work is to estimate and recognize different human poses based on

body joints localization. In [19], Oreifej and Liu presented a novel descriptor which captures motion and geometry cues jointly using a histogram of normal orientation in the 4D space of depth, time, and spatial coordinates for activity recognition from depth sequences. In [20], Jalal et al. described random forest (RFs) approach to train a set of randomly selected features based on some quality measurements (i.e., information gain). They created a DB of synthetic depth silhouettes and their corresponding pre-labelled silhouettes using the Kinect camera to recognize the motion features treating with Hidden Markov Model (HMM). In [21], Jalal et al. proposed a life logging HAR method which extracts human body joints information to perform magnitude and directional angular features generation using the depth images. These features are further modelled, trained and recognized human activity in real time using indoor environment. In [22], Karg and Kirsch developed two different approaches such as spatio-temporal plan representation (STPRs) and hierarchical hidden markov model (HHMMs) using depth cameras to perform activity recognition utilizing context dependent spatial regions. However, STPR represents the human activity as a sequence of spatial regions visited throughout the task and HHMMs used lower/higher level to estimate the posterior marginal over all activities.

Depth silhouettes-based HAR systems are mainly dealing with the marker-based HAR and markerless-based HAR system. In marker-based approaches [23,24], subjects need to wear specific suit, in which markers are attached to the designated body parts and special (i.e., depth) cameras have been used to detect these markers. In [23], Ganapathi et al. designed motion capture system which includes model-based hill climbing search, inverse kinematics, and GPU-accelerated approach in order to track and recognize different activities using multiple depth cameras and 3D markers attached to subject's body. However, the system is quite inconvenient in the real time applications because marker motion during the movement of the object is not smooth. Also, the self-occlusion of the object parts causes low accuracy rate. In [24], Zhao et al. used semi-supervised learning which makes it possible to use both labeled and unlabeled human joints position data. Semi-supervised discriminant analysis with global constraint (SDG) optimizes by treating labeled training data with linear discriminant analysis (LDA) and unsupervised algorithms like locality preserving projection (LPP) and PCA using all trained data to estimate better data distribution. Then, to classify data, k-Nearest Neighbors (k-NN) method is employed. They used sixteen markers on human joints using five different human actions (box, gesture, jog, throw-catch, and walk). However, these systems consist of expensive equipment which are not feasible during natural movements of subjects.

Depth markerless-based HAR system along with body parts segmenting and labeling is also an important factor in the field of depth video-based HAR system. Recently, different approaches have been introduced for segmentation and labeling the human body parts such as RFs, in which multiple classifiers are used to label each pixel to appropriate position. For instance, Simari et al. [25] proposed a method for segmenting human silhouettes using the centroid of body based on k-means clustering. In [26], Jalal et al. showed an example for body part labeling using depth silhouettes with Gaussian contour classifier to segment the human body and labelling. In [27], Buys et al. used RGB-D data for human body detection and pose estimation without background subtraction. Using the single depth image, a pixel-wise approach is used to label the body parts. These pixel-wise body parts used random decision forest (RDF) classifier to assign the labels. Then, kinematic search tree method is used for the final skeleton configuration.

In this paper, we present an effective methodology to track and recognize the human activity based on depth silhouettes and features of body skin color joints. Initially, raw depth maps are captured using Kinect depth camera where human silhouettes are extracted from

noisy background. These silhouettes are tracked based on bounding box and estimate the boundary of each silhouettes using chain coding concept. Then, each depth silhouette is converted into skeleton representation using skin color detection algorithm and thus producing joint points. These joint points are computed as feature extraction using distance position features and centroid distance features. Applying the feature vector, we used k-means clustering to cluster n objects based n attributes into k partitions where $k < n$. Finally, we train and test our depth datasets using self-organizing maps to reduce the dimension of the data to map and recognize different activities.

The rest of the paper is organized as follows. In section II, the methodology has been explained which includes; depth silhouettes preprocessing in which we track human silhouettes, feature generation, k-mean clustering for symbol selection followed by activity training and recognition using self-organized map (SOM). In section III, the detail of our experimental procedure and results have been described. Section IV explains the conclusion of our presented work and also discuss the possible future directions.

2. Methodology

Our HAR system consists of incoming dense depth maps using a depth video camera (i.e., PrimeSense Kinect camera), which includes: preprocessing step to track depth silhouettes followed by feature generation. During feature generation, centroids are calculated from the contour of each depth silhouettes. Then, body joint points are extracted using body parts skin color detection. These skeletons joint points deal with distance position and centroid distance features of body parts for feature extraction and training/testing through implementing SOMs. **Fig. 1** shows the system architecture flow of the proposed HAR system.

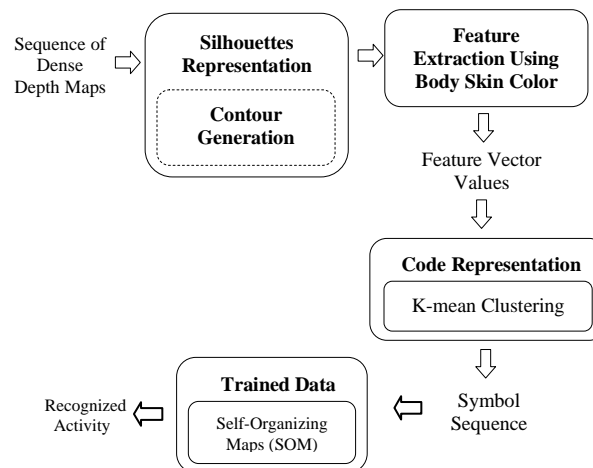


Fig. 1. Overall flow architecture of proposed HAR system

2.1 Depth Silhouette Preprocessing

In order to capture the depth image silhouette, we involve Kinect sensor. It consists of RGB images and raw depth data. **Fig. 2** shows some sample depth silhouettes of different activities such as eating meal, walking, cleaning, and sitting down with the rectangle bounding box (blue in color). However, to track each silhouettes, frame difference can be used to detect overall silhouettes of human body. Also, we performed disparity segmentation to extract the entire body silhouettes using modified flood fill algorithm. Ignoring the background area (i.e., black

as 0's), we compute the disparity pixel values in the moving regions and calculate surrounding neighboring pixel to extract the human silhouettes [28, 29]. However, certain threshold (i.e., height and width) are used to control our bounding box size [30, 31]. Finally, the bounding box is used to extract the desired human silhouettes region for further feature extraction.

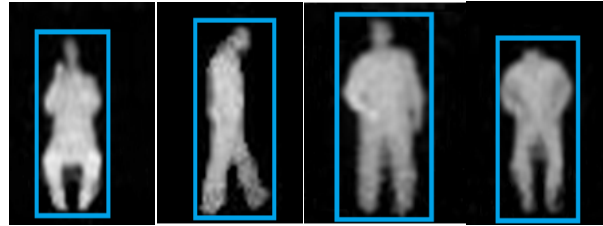


Fig. 2. Some examples of human activities using depth silhouettes tracked by bounding box

During human tracking mechanism, depth maps contain objects in the scenes which are labeled their specific pixels using connected component labeling (CCL) method [17]. During CCL, the difference of pixel intensity in an image is monitored. Similarly, the color variation and intensity values of background are very less, therefore, we remove all the non-subject components (i.e., background). While, to consider the monitoring of different connected components, every depth pixel intensity of the connected components are used to differentiate the depth features data into two ranks (i.e., subject or human and objects like table, chair, etc.). **Fig. 3** shows (a) complex background along with human activities, (b) CCL help to differentiate all components along with background removal and (c) human silhouettes are extracted which perform daily activities.

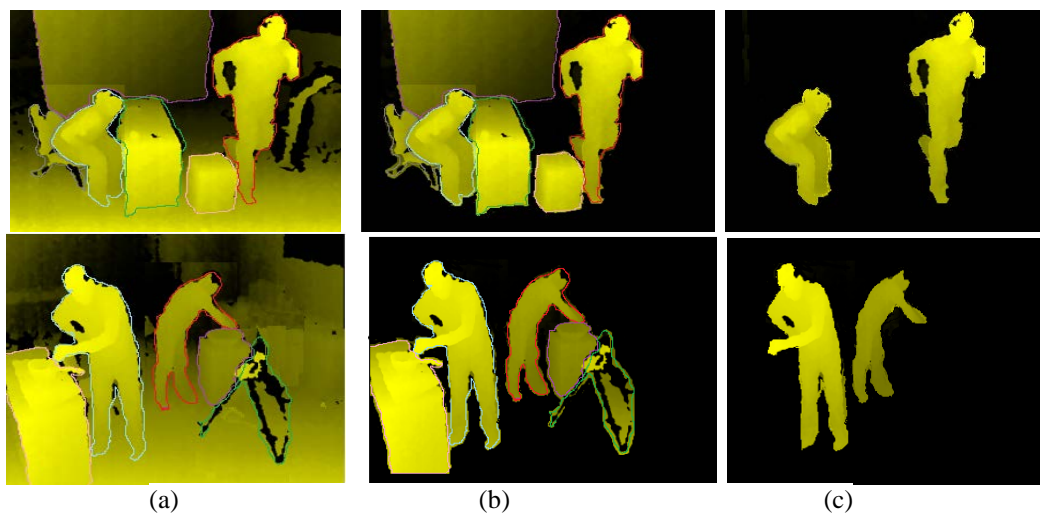


Fig. 3. Human tracking mechanism. (a) Human subjects performed daily activities in the indoor environments (i.e., in a lab environment), (b) background removal and CCL implementation, and (c) extraction of human silhouettes performing different activities (i.e., sit down, walking, prepare food and exercise)

2.2 Feature Generation

During feature generation, we compute centroid points of all human activities and use during body skin joints features for centroid distance features.

2.2.1 Centroid of Depth Silhouettes

After extracting the human depth silhouettes from the noisy background, we performed silhouettes representation method to compute centroid of all human activities for feature generation mechanism. Thus, in order to estimate the boundary length based silhouette contour, we used chain coding mechanism [32]. Assuming that each pixel is connected with its 8 neighbor pixels, therefore, the chain code is composed of sequence of numbers between 0 and 7. Thus, we used 8 chain code vector. In addition, we traverse our silhouette contour in clockwise direction. However, these boundary contour of the moving silhouettes are used to compute its shape centroid as

$$C_J = \frac{1}{F} \sum_{j=1}^{15} C_j(x, y) \quad (1)$$

In Fig. 4, different human activities used chain coding mechanism marked as orange line and centroid is marked as star (red in color).

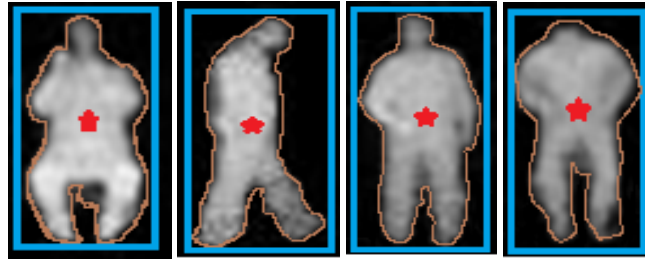


Fig. 4. Representation of human depth silhouettes using chain code mechanism and centroid points extraction

2.2.2 Body Skin Joints Features

From a sequence of depth sequences, the corresponding skeleton body models are produced using body skin color (BSC) detection mechanism. In BSC mechanism, we estimate the probability of a pixel which is acting as skin colored being derived from the probability maps to achieve five joint points. Let's assume, we have image points $I(x, y)$ and color $c(x, y)$. Then, the prior probability $P(b)$ of body skin-color, occurrence of each color c having prior probability $P(c)$ and color c being a skin-color having prior probability $P(b|c)$ can be computed as

$$P(b|c) = P(c|b)P(b)/P(c) \quad (2)$$

Thus, the probability of each image point having skin-color is determined by considering specific range as $t_{\max} > P(b|c) > t_{\min}$ which is further structured as skeleton model having five skin joint points (i.e., head, both hands, and both feet), respectively.



Fig. 5. Body skin joints features representation based on body skeleton model having five joint points using skin color detection technique

Distance Position Features

After obtaining centroid point and five skin joints points, we used feature extraction based on body skin joints features. Here, this approach includes distance position features (DPF) and centroids distance features (CDF).

Initially, using DPF, we measure the distance D_J between respective joints points of the two consecutive frames [33, 34] t and $t-1$ as,

$$D_J = \sqrt{(Jx_{(i-1)} - Jx_{(i)})^2 + (Jy_{(i-1)} - Jy_{(i)})^2 + (Jz_{(i-1)} - Jz_{(i)})^2} \quad (3)$$

However, the size of feature vectors obtained from the distance position features of each joint points become 1×5 .

Centroid Distance Features

While, considering CDF, we measure the distance between joint points coordinates and the centroid of each activity frame. Thus, CDF is expressed as,

$$D_{cen}^i = \sqrt{(P_x^i - P_{cen}^x)^2 + (P_y^i - P_{cen}^y)^2 + (P_z^i - P_{cen}^z)^2} \quad (4)$$

While, the feature vector size of CDF become 1×5 . Thus, the overall feature vector of body skin joints features used for training/testing each activity become 1×10 , respectively.

2.3 K-mean Clustering for Body Skin Joint Features

These feature vectors are symbolized in the form of codebook which are generated based on k-mean clustering algorithm ($k=32$) [35] and each feature's patch is assigned to it's nearest neighbor in the codebook. Then, each cluster has assigned one element which represents its group and the closest element to the mean of each cluster becomes the best candidate for being representative. In this way, we identify the activity in the training set which is most similar to the test data. However, each activity sequence is a time series of the numerical words. Therefore, these symbol data are generated according to the sequence of each activity and maintained using buffer strategy [36, 37].

2.4 Training/Testing via SOM

The SOM is used as neural network model which can be successfully applied as data mining tool with various applications in image analysis, pattern recognition and computer vision. It provides a way to represent a multidimensional features data into lower dimensional feature vectors. SOM is based on topology preservation properties where each neuron makes a set of patterns which can be activated or rejected [38]. However, neuron acting as best matching (winning) neuron $m(t)$ consists of largest similarity measure (or smallest dissimilarity measure) between all weight vectors $w_j(t)$ and the input vector $x(t)$.

$$m(t) = \arg \min_j \|x(t) - w_j(t)\| \quad (5)$$

While, the weight of the winning neuron and its neighboring units are then updated as

$$\Delta w_j = \gamma (x(t) - w_j(t)) \quad (6)$$

Where γ is the learning rate. Thus, we used SOM engine for both training each activity and recognized the testing input activity sequence by finding the closest prototype. Fig. 6 shows the U-matrix probabilities of the eating meal SOM after training with the map size of 5x5.

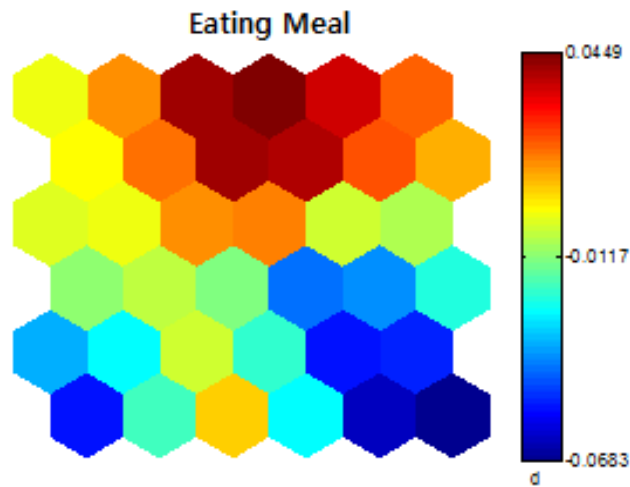


Fig. 6. Eating meal SOM having U-Matrix probabilities after training

Finally, the major contributions of our paper as: 1) To my certain knowledge, it is first time to use depth silhouettes along with the combination of skin detection which is further used for joint information. The joint information are further use for training and testing with SOM to recognize the human activities. 2) We recorded continuous depth dataset from our Kinect sensor and used for tracking and recognizing human activities, which is itself a contribution in the field of activity recognition. 3) Our preprocessing phase, where we used modified flood fill algorithm along with the tracking technique to extract the body silhouettes. This technique used to find variation of pixel intensity values to track bounding box. Moreover, the bounding box is further used for feature extraction of each activity.

3. Experimental Results

In this study, we performed experiments in an indoor environment (i.e., in our lab) having nine different subjects (i.e., gender: male and female, age: 32~48) and recorded our own depth silhouette datasets. The datasets consist of nine different activities which are mostly performed in our daily routine life. These nine activities include: *walking, sit down, exercise, prepare food, stand up, cleaning, watching TV, eating meal and lying down*. The datasets are quite challenging because many of the activities have similar sequential postures to each other especially hands and legs movements. Also, subjects have no restriction to perform the various activities randomly, therefore, the trajectories of their movement make our datasets more challenging.

We compare our body skin joints features approach with the approach using conventional features as PC-R transform [39, 40] where R transformation features computed a 1D features profiles of a depth silhouette which produces a highly compact representation of all daily human activities. However, the collected video clips were split into 85 clips (i.e., 30 for training and 55 for testing) where each clip contained fifteen consecutive frames. During training phase, a total of 30 clips from each activity were used to build the training feature space and the whole training data contained a total of 4,050 depth silhouettes. Each depth silhouette contains body joints features vector with its size of 1x10. During testing, we applied 55 video clips of each activity where SOM map size is equal to 5x5.

3.1 Analysis and Recognition of Continuous Video

To analyze the recorded (continuous) video having mixed activities, a subject performed all nine activities freely and randomly for several hours in a day at a pre-specific path range (i.e., distance range of 1.3m to 3.5m) and recorded in the database.

Fig. 7 shows the recognition of recorded video of human activities against the annotated ground truth having depth silhouettes using body skin joints features approach. There are total of 6,438 frames having all nine activities performed randomly without any instructions. Note that all daily activities show consistent matching between the predicted activity and the ground truth. In the recognition process, feature vectors are extracted per every fifteen frames with an overlap of seven frames.

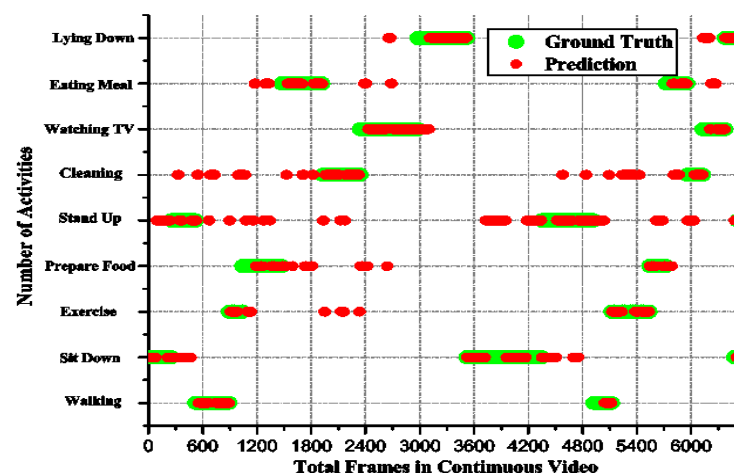


Fig. 7. Recorded video recognition having all nine human activities using body skin joints features approach

3.2 Recognition Results

In this section, the proposed body skin joints system is compared against the conventional system using depth silhouettes. **Table 1** presents the confusion matrix of recognition results using R transformation features. However, the recognition results of stand up, exercise, watching TV, and sit down are 67.0%, 68.0%, 69.50%, and 76.50%, respectively. Their recognition rate is relatively lower than other activities due to closer postures among all these activities.

Table 1. Confusion matrix of the conventional features as R Transformation features based HAR

Activities	WK	SD	EX	PF	SU	CL	WT	EM	LD
WK	79.0	2.0	5.50	8.50	1.0	2.50	1.50	0	0
SD	1.50	76.50	1.0	3.50	11.50	4.50	1.50	0	0
EX	6.50	0	68.0	5.0	3.50	9.0	2.50	5.50	0
PF	1.0	3.50	5.50	82.50	0	4.50	1.50	0	1.50
SU	3.0	14.50	4.0	7.50	67.0	1.0	0	0	3.0
CL	2.50	4.50	0	2.50	3.0	85.50	0	2.0	0
WT	4.0	8.50	6.50	1.50	3.0	1.50	69.50	5.50	0
EM	1.0	9.0	0	3.0	1.50	0	4.0	79.0	2.50
LD	1.50	2.0	0	1.0	0	1.50	3.50	1.0	89.50
Mean Recognition Rate = 77.39%									

WK=Walking, SD=Sit Down, EX=Exercise, PF=Prepare Food, SU=Stand Up, CL=Cleaning, WT=Watching TV, EM=Eating Meal, LD=Lying Down.

Finally, in **Table 2**, the mean recognition rate of our proposed body skin joints features approach proved much higher recognition rate of 89.72% than the conventional R transform features approach as 77.39%

Table 2. Confusion matrix of the proposed features based HAR

Activities	WK	SD	EX	PF	SU	CL	WT	EM	LD
WK	89.50	1.50	3.0	0	2.0	1.50	0	2.50	0
SD	3.50	88.0	2.50	1.0	1.50	0	2.50	1.0	0
EX	3.0	1.50	85.0	4.50	1.0	4.0	0	1.0	0
PF	1.50	0	1.0	93.50	0	1.50	1.50	0	1.0
SU	2.0	4.0	2.50	2.0	84.50	1.0	0	1.50	2.50
CL	2.50	0	1.0	0	1.0	94.0	0	1.50	0
WT	1.0	3.50	1.0	0	1.0	2.0	87.50	3.0	1.0
EM	2.50	4.50	0	0	1.0	0	2.50	89.50	0
LD	0	1.0	0		0	2.0	1.0	0	96.0
Mean Recognition Rate = 89.72%									

Thus, the overall performance among the conventional and proposed approaches showed that the proposed body skin joint features provide stronger features than the R transform features respectively. For experimental pupose, we used standard PC as Intel Pentium IV 2.63GHz having 2GB RAM along with a Kinect Depth camera.

4. Conclusion

In this work, we have presented an effective body skin joints features based HAR system using depth sequences. Our proposed HAR system utilizes a combination of body activity detection, and tracking using effective body skin joints features from the joints points of the skeleton model and modeling, training and activity recognition using SOM. Experimental results showed some promising performance of the proposed HAR technique, achieving the mean recognition rate of 89.72% over the conventional method as 77.39%. Moreover, our system handles subject's body size, self-occlusion, overlapping among people, and hidden body parts prediction which significantly track complex activities and improve recognition rate. We believed that the proposed system is useful for many applications including healthcare system, automatic video surveillance, smart homes and robot learning.

For future work, we have planned to improve the effectiveness of our system especially in case of complex activities, interaction between people, and joints missing by introducing some hybrid HAR system concept. The proposed system is merged together with body parts modeling and recognition [41] system to extract more exact joints positions which make HAR algorithm more effective and robust in the future.

References

- [1] A. Veeraraghavan, A. K. Roy-Chowdhury and R. Chellappa, "Matching shape sequences in video with applications in human movement analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 12, pp. 1896-1909, 2005. [Article \(CrossRef Link\)](#)
- [2] A. Jalal and I. Uddin, "Security architecture for third generation (3g) using GMHS cellular network," in *Proc. ICET 2007, IEEE International Conference on Emerging Technologies*, pp. 74-79. IEEE, 2007. [Article \(CrossRef Link\)](#)
- [3] A. Jalal and M. A. Zeb, "Security and QoS optimization for distributed real time environment," in *Proc. of 7th IEEE International Conference on Computer and Information Technology, 2007*, pp. 369-374. IEEE, 2007. [Article \(CrossRef Link\)](#)
- [4] J. S. Kim, D. H. Yeom and Y. H. Joo, "Fast and robust algorithm of tracking multiple moving objects for intelligent video surveillance systems," *IEEE Transactions on Consumer Electronics*, vol. 57, no. 3, pp. 1165-1170, 2011. [Article \(CrossRef Link\)](#)
- [5] A. Jalal and A. Shahzad, "Multiple facial feature detection using vertex modeling structure," in *Proc. of IEEE Computer Society Conference on Interactive computer aided learning, 2007*. [Article \(CrossRef Link\)](#)
- [6] A. Jalal and Y. A. Rasheed, "Collaboration achievement along with performance maintenance in video streaming," in *Proc. of IEEE Conference on Interactive Computer Aided Learning, 2007*. [Article \(CrossRef Link\)](#)
- [7] V. Megavannan, B. Agarwal and R.V. Babu, "Human action recognition using depth maps," in *Proc. of the international conference on signal processing and communication*, pp. 1-5, 2012. [Article \(CrossRef Link\)](#)
- [8] A. Jalal, MZ. Uddin, J. T. Kim and T. Kim, "Daily Human Activity Recognition Using Depth Silhouettes and R Transformation for Smart Home," in *Proc. of Smart Homes Health Telematics*, pp. 25-32, 2011. [Article \(CrossRef Link\)](#)
- [9] A. Jalal, J. T. Kim and T.-S. Kim, "Human activity recognition using the labeled depth body parts information of depth silhouettes," in *Proc. of the 6th international symposium on Sustainable Healthy Buildings*, pp. 1-8, 2012. [Article \(CrossRef Link\)](#)
- [10] B. Zhang, K. Mei and N. Zheng, "Reconfigurable processor for binary image processing," *IEEE Transaction on Circuits and Systems for Video Technology*, vol. 23, no. 5, pp. 823-831, 2013. [Article \(CrossRef Link\)](#)

- [11] A. Jalal and S. Kim, "Algorithmic implementation and efficiency maintenance of real-time environment using low-bitrate wireless communication," in *Proc. of IEEE workshop on Software technologies for future embedded and ubiquitous systems*, pp.1-6, 2006. [Article \(CrossRef Link\)](#)
- [12] A. Jalal and S. Kim, "A complexity removal in the floating point and rate control phenomenon," in *Proc. of Korea multimedia society*, pp.48-51, 2005. [Article \(CrossRef Link\)](#)
- [13] A. Iosifidis, A. Tefas and I. Pitas, "Activity-based person identification using fuzzy representation and discriminant learning," *IEEE transaction on information forensics and security*, vol. 7, no. 2. pp. 530-542, 2012. [Article \(CrossRef Link\)](#)
- [14] Y. Wang, K. Huang and T. Tan, "Human activity recognition based on R transform," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-8, 2007. [Article \(CrossRef Link\)](#)
- [15] T.-J. Chin, L. Wang, K. Schindler and D. Suter, "Extrapolating learned manifolds for human activity recognition," in *Proc. of IEEE Conference on image processing*, pp. 381-384, 2007. [Article \(CrossRef Link\)](#)
- [16] A. Jalal, S. Kamal and D. Kim, "Depth Map-based Human Activity Tracking and Recognition Using Body Joints Features and self-organized map," in *Proc. of the IEEE International Conference on computing, communication and networking technologies*, 2014. [Article \(CrossRef Link\)](#)
- [17] A. Jalal and Y. Kim, "Dense Depth Maps-based Human Pose Tracking and Recognition in Dynamic Scenes Using Ridge Data," in *Proc. of the IEEE International Conference on Advanced Video and Signal-based Surveillance*, pp. 119-124, 2014. [Article \(CrossRef Link\)](#)
- [18] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman and A. Blake, "Real-time human pose recognition in parts from a single depth image," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1297-1304, 2011. [Article \(CrossRef Link\)](#)
- [19] O. Oreifej and Z. Liu, "HON4D: Histogram of oriented 4D normals for activity recognition from depth sequences," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 716-723, 2013. [Article \(CrossRef Link\)](#)
- [20] 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, pp. 271-279, 2013. [Article \(CrossRef Link\)](#)
- [21] 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 *Proc. of the International Symposium on Sustainable Healthy Buildings*, pp. 91-95, 2012. [Article \(CrossRef Link\)](#)
- [22] M. Karg and A. Kirsch, "Low cost activity recognition using depth cameras and context dependent spatial regions," in *Proc. of international Conference on autonomous agents and multiagent systems*, pp. 1359-1360, 2014. [Article \(CrossRef Link\)](#)
- [23] V. Ganapathi, C. Plagemann, D. Koller and S. Thrun, "Real time motion capture using a single time-of-flight camera," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, pp. 755-762, 2010. [Article \(CrossRef Link\)](#)
- [24] X. Zhao, X. Li, C. Pang and S. Wang, "Human action recognition based on semi-supervised discriminant analysis with global constraints," *Neurocomputing*, vol. 105, pp. 45-50, 2013. [Article \(CrossRef Link\)](#)
- [25] P. Simari, D. Nowrouzezahrai, E. Kalogerakis and K. Singh, "Multi-objective shape segmentation and labeling," in *Proc. International symposium on Geometry Processing*, pp. 1415-1425, 2009. [Article \(CrossRef Link\)](#)
- [26] A. Jalal, S. Lee, J. T. Kim and T. S. Kim, "Human activity recognition via the features of labeled depth body parts," in *Proc. of Smart Homes Health Telematics*, pp. 246-249, 2012. [Article \(CrossRef Link\)](#)
- [27] K. Buys, C. Cagniard, A. Baksheev, T.-D. Laet, J. D. Schutter and C. Pantofaru, "An adaptable system for RGB-D based human body detection and pose estimation," *Journal of visual communication and image representation*, vol. 25, pp. 39-52, 2014. [Article \(CrossRef Link\)](#)

- [28] 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(7), pp. 11735-11759, 2014. [Article \(CrossRef Link\)](#)
- [29] R. P. Otero. "Induction of the effects of actions by monotonic methods," in *Proc. of 13th international conference on Inductive logic programming*, pp. 299-310, 2003. [Article \(CrossRef Link\)](#)
- [30] A. Jalal, Y. Kim and D. Kim, "Ridge body parts features for human pose estimation and recognition from RGB-D video data," in *Proc. of the IEEE International Conference on computing, communication and networking technologies*, 2014. [Article \(CrossRef Link\)](#)
- [31] A. Jalal and S. Kamal, "Real-Time Life Logging via a Depth Silhouette-based Human Activity Recognition System for Smart Home Services," in *Proc. of the IEEE International Conference on Advanced Video and Signal-based Surveillance*, pp. 74-80, 2014. [Article \(CrossRef Link\)](#)
- [32] W. Yang, Z. Song and X. Wu, "Histogram of silhouette direction code: An efficient HOG-based descriptor for accurate human detection," in *Proc. of the IEEE International Conference on Robotics and biomimetics*, pp. 330-335, 2012. [Article \(CrossRef Link\)](#)
- [33] 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. [Article \(CrossRef Link\)](#)
- [34] A. Jalal and S. Kim, "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. [Article \(CrossRef Link\)](#)
- [35] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman and A. Y. Wu, "An efficient k-means clustering algorithm: analysis and implementation," *IEEE Transaction on pattern analysis and machine intelligence*, vol. 24, no. 7, pp. 881-892, 2002. [Article \(CrossRef Link\)](#)
- [36] A. Jalal, S. Kim and B. J. Yun, "Assembled algorithm in the realtime h.263 codec for advanced performance," in *Proc. of 7th International Workshop on Enterprise networking and Computing in Healthcare Industry, 2005. HEALTHCOM 2005*, pp. 295-298. IEEE, 2005. [Article \(CrossRef Link\)](#)
- [37] 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. [Article \(CrossRef Link\)](#)
- [38] T. Kohonen, "The self-organizing map," in *Proc. of the IEEE*, vol. 78, no. 9, pp. 1464-1480, 1990. [Article \(CrossRef Link\)](#)
- [39] 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, pp. 184-190, 2011. [Article \(CrossRef Link\)](#)
- [40] Ahmad Jalal, Md. Zia 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 Transaction on Consumer Electronics*, ISSN: 0098-3063, vol. 58, Issue.3, pp. 863-871, 2012. [Article \(CrossRef Link\)](#)
- [41] J. Oberg, K. Eguro, R. Bittner and A. Forin, "Random decision tree body part recognition using FPGAS," in *Proc. of international conference on Field Programmable Logic and Applications*, pp. 330-337, 2012. [Article \(CrossRef Link\)](#)



Adnan Farooq received his B.S degree in Computer Engineering from COMSATS Institute of Science and Technology, Abbottabad, Pakistan and M.S. degree in Biomedical Engineering from Kyung Hee University, Republic of Korea. His research interest includes Image Processing, Computer vision.



Ahmad Jalal received his B.S. degree in Computer Science from Iqra University, Peshawar, Pakistan and 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 received his B.S. degree in Software Engineering from City University of Science and Information Technology, Peshawar, Pakistan and M.S. degree in Computer Engineering from Mid Sweden University, Sweden. He is currently enrolled as Ph.D. candidate in the Department of Radio and Electronics Engineering at Kyung Hee University, Republic of Korea. His research interest includes advanced wireless communication, image and signal processing.