# Depth Video-based Human Activity Recognition System Using Translation and Scaling Invariant Features for Life Logging at Smart Home

**Article** *in* IEEE Transactions on Consumer Electronics · July 2018

CITATIONS
7

READS
167

**3 authors**, including:

Ahmad Jalal
Pohang University of Science and Technology
**68** PUBLICATIONS **1,360** CITATIONS

SEE PROFILE

# Depth Video-based Human Activity Recognition System Using Translation and Scaling Invariant Features for Life Logging at Smart Home

A. Jalal, Md. Zia Uddin, and T.-S. Kim, *Member,* IEEE

**Abstract**— *Video-based human activity recognition systems have potential contributions to various applications such as smart homes and healthcare services. In this work, we present a novel depth video-based translation and scaling invariant human activity recognition (HAR) system utilizing R transformation of depth silhouettes. To perform HAR in indoor settings, an invariant HAR method is critical to freely perform activities anywhere in a camera view without translation and scaling problems of human body silhouettes. We obtain such invariant features via R transformation on depth silhouettes. Furthermore, in R transforming depth silhouettes, shape information of human body reflected in depth values is encoded into the features. In R transformation, 2D feature maps are computed first through Radon transform of each depth silhouette followed by computing 1D feature profile through R transform to get the translation and scaling invariant features. Then, we apply Principle Component Analysis (PCA) for dimension reduction and Linear Discriminant Analysis (LDA) to make the features more prominent, compact and robust. Finally, Hidden Markov Models (HMMs) are used to train and recognize different human activities. Our proposed system shows superior recognition rate over the conventional approaches, reaching up to the mean recognition rate of 93.16% for six typical human activities whereas the conventional PC and IC-based depth silhouettes achieved only 74.83% and 86.33% ,while binary silhouettes-based R transformation approach achieved 67.08% respectively. Our experimental results show that the proposed method is robust, reliable, and efficient in recognizing the daily human activities.*

**Index Terms — Depth silhouette images, R transformation, principle component analysis, Hidden Markov Models.**

## I. INTRODUCTION

Recently, advancement in information technologies led to the development of effective and inexpensive consumer depth video cameras which are actively used for interactive consumer-level gaming, motion capture, and gesture recognition devices [1]. These depth cameras technologies produce high quality depth data and could have potential applications in consumer surveillance systems, life-care systems, and smart home systems [1]-[4]. A key function in these applications is to recognize human activities. For instance, in smart home systems, depth video cameras could be used to monitor daily life of residents (i.e., elder, children, disabled people, etc.). By recognizing the daily activities of the residents, homes could be turn into an intelligent living space by making them respond to the needs of residents.

For this reason, human activity recognition (HAR) has become one of the challenging and active areas of research with its considerable interests in smart environments [2]. As HAR systems provide a way of sense to observe basic human activities of daily living, video-based HAR systems are widely used among researches to recognize daily activities for proactive care of health and life of residents [2]. The general approach of video-based HAR involves extraction of some significant features from video images to train classifiers for recognition [3]. Therefore, effective feature extraction and pattern recognition techniques play a significant role in this regard. In general, HAR is a challenging task due to lacks of rigid syntax like gesture or sign language recognition.

Initially, in the video-based HAR, a human activity has been generally represented in a time-series of binary silhouettes derived from RGB cameras [4]-[7]. In [5], Liang and Suter used binary silhouettes to exploit locality preserving projections for dimensionality reduction of moving persons for their action recognition. In [8], Yamato et al. designed a human activity recognition system based on binary silhouettes where mesh feature vector sequences were extracted from the time-sequential images to recognize a number of tennis stokes. However, this approach showed poor recognition rate due to its dependency on the pixel meshes and sensitivity to position displacement. In [9], Uddin et al. utilized Independent Component (IC) features of binary silhouettes to recognize various human activities. One critical limitation of the binary silhouettes is its ambiguity among the same silhouettes for different postures of different activities due to their limited pixel values (i.e., 0 or 1, thus within the binary shape of a posture, no information is present). Due to a lack of information in the flat pixel intensity, it is impossible for binary silhouette to differentiate between the far and near distance of the human body parts. To overcome this limitation of binary silhouettes, depth-based silhouette representation for

This work was supported by the MKE (Ministry of Knowledge Economy), Korea, under the ITRC (Information Technology Research Center) support program supervised by the NIPA (National IT Industry Promotion Agency) (NIPA-2012-(H0301-12-1004). This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No.2012-0000609).

A. Jalal and T.-S. Kim are with the Department of Biomedical Engineering, Kyung Hee University, Seocheon-dong, Giheung-gu, Yongin-si, Gyeonggi-do, 446-701, Republic of Korea. Md. Zia Uddin is with the Department of Electronic Engineering, Inha University, Incheon, 402-751, Republic of Korea (email: ahmjal@yahoo.com, ziauddin@inha.ac.kr, tskim@khu.ac.kr).

Contributed Paper
Manuscript received 05/28/12
Current version published 09/25/12
Electronic version published 09/25/12.

human activity has been suggested lately where the body parts are differentiated by means of different intensity values [10]-[13]. Recent studies showed that with the depth silhouettes, a significantly improved recognition rate could be achieved [11], [12].

As for the activity features, Principle Component (PC) and Independent Component (IC) features of the binary and depth silhouettes have been widely used [9], [14]. However, these spatial features are sensitive to scaling and translation of silhouettes, thus requiring proper scaling and alignment of silhouettes. If body silhouettes are not properly aligned, satisfactory recognition rate may not be achieved. To recognize various subjects having different body characteristics (in their size, height, position, and body structure) performing the same activity, some invariant features to those changes are needed.

Radon transform was originally applied on binary silhouettes [15] to extract distinct shape features of binary object. In [16], Singh et al. used Radon transform to identify the binary skeleton representation of hands and arms movement for human pose recognition. In [17], Chen et al. used Radon transform on binary silhouettes to construct gait templates to recognize the gender through a walking activity. In [18], Boulgouris et al. used Radon transform on binary silhouettes to recognize the human gait. In [19], Li et al. used Radon transform representation of binary silhouettes to recognize human movements. In [20], Tabbone et al. introduced an improved version of Radon transform, R transformation, to recognize the binary shape objects. In [21], Wang et al. used R transformation on binary silhouettes to represent the spatial information of different human activities and they used Gaussian Mixture Model (GMMs) as well as variable number of states to model HMMs for different activities. In [22], Zhang et al. described a method for HAR where key binary frames were extracted from a cycle of video sequences based on R transformation. These previous studies mainly focused on binary silhouettes to recognize different object shapes and activities.

In this work, we apply R transformation on depth silhouettes to extract the scale and position invariant features. Then, PCA [23]-[25] is applied to extract features from the R transformed profiles of depth silhouettes. Finally, Linear Discriminant Analysis (LDA) [26] is applied to make them more robust. To recognize various human activities, we use discrete Hidden Markov Models (HMMs) [27], [28] having uniform states for different activities. However, our technique used depth silhouettes which provide much stronger features than binary silhouettes which are not efficient to describe the human body properly in the activity video. Our HAR results show significant improvements over the conventional approaches using PC and IC features of depth silhouettes and binary silhouettes-based R transformation approach. The proposed system could be an essential component of a smart home system for continuous observation of typical human activities.

The rest of the paper is organized as follows. Section II describes the methodology based on the configuration of the proposed human activity recognition system. It provides video-based silhouette features preprocessing. Then, the feature representation and extraction techniques are introduced, followed by activity modeling, training, and recognition using HMM. Section III shows the experimental settings and results utilizing different feature extraction approaches. Finally, conclusions of our work are drawn in Section IV.

## II. METHODOLOGY

Our HAR system consists of depth silhouette extraction, feature representation and extraction (including PCA, or ICA, R transformation), and modeling of HMMs (including codebook generation, training, and recognition). Fig. 1 shows the overall flow of proposed human activity recognition system.

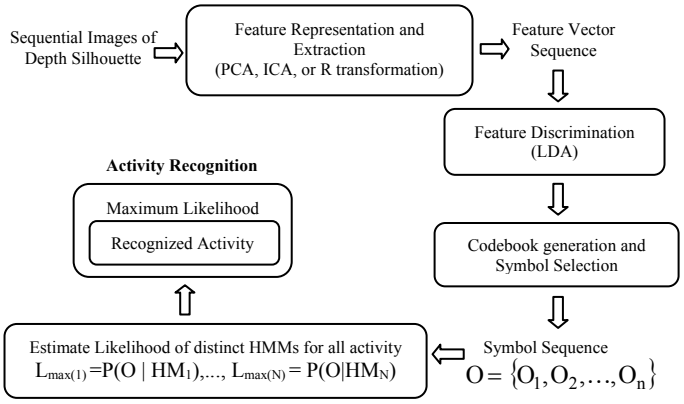



**Fig. 1. Overview flow of proposed human activity recognition system.**

### A. Silhouette Features Preprocessing

To capture depth images of activities, we utilize a depth camera to acquire a sequence of images of six human activities from five human subjects (males of age 25-29) at 30fps. The camera produces RGB images and depth maps. The depth information captured by the camera indicates the strength of each pixel using gray level value.

The Region of Interest (ROI) is extracted from every depth activity images and resized to a matrix of $50 \times 50$ as silhouettes. Considering the comparative study of both (binary and depth) silhouettes in HAR system, we converted the depth silhouettes to binary silhouettes using a simple threshold value. Figs. 2 (a) and (b) show the sequential binary and depth silhouettes of a walking activity. From the silhouette distribution in Fig. 2, it seems to be clear that the binary silhouettes contain limited information due to their flat pixel intensity (i.e., 0 or 1) distribution over the human body (i.e., only shape information is available). On the contrary, in the case of depth silhouettes, every individual component of body parts having far distance contains brighter values and near distance darker ones. Hence, depth silhouettes can represent human body better than binary ones in activity videos.

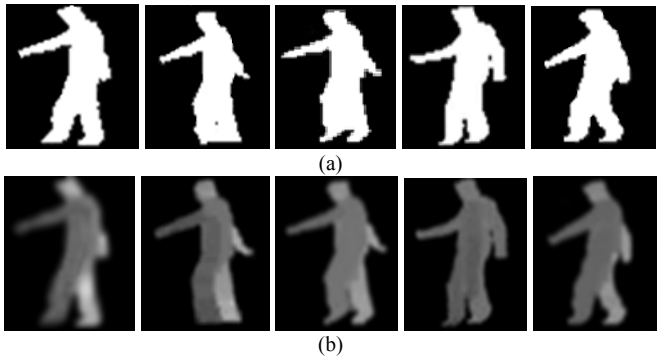


**Fig. 2. Image sequence of (a) binary silhouettes and (b) depth silhouettes of a walking activity.**

During the preprocessing steps, each silhouette is represented as a row vector with the size of the total pixels in the image (i.e., $50 \times 50$ image becomes a vector of $1 \times 2500$). The silhouette vector is then converted to zero mean row vector. Then, feature representation and extraction algorithms are applied on the zero mean vectors. Thus, from $N$ depth silhouette vectors i.e., $(X_1, X_2, \ldots, X_N)$, we calculate the mean vector as $\overline{X} = \left(1/N \sum_{i=1}^{N} X_i\right)$ to subtract it from all silhouette vectors to convert them to zero mean $\overline{\overline{X}} = (X_i - \overline{X})$ where $1 \le i \le N$.

### B. PC and IC Depth Features

Previously, PCA and ICA have been used to extract prominent features for HAR. In general, PCA produces global features of activities [5], [9], [13] whereas ICA produces local features [9], [13]. In this work, we have studied both the features in our experiments to compare them against our proposed R transformation features.

### C. R Transformation

#### 1) Radon Transform

R transformation consists of two processes; namely Radon transform and R transform. Radon transform [15] computes a 2D projection of a depth silhouette along specified view directions. It is applied on a depth silhouette to establish a mapping between the domain produced by the image coordinate system $f(x, y)$ and the Radon domain indicated as $R(\rho, \theta)$. Let $f(x, y)$ is a depth silhouette, its Radon transform $R(\rho, \theta)$ is defined [14] as

$$R(\rho, \theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\delta(x \cos\theta + y \sin\theta - \rho)dx\,dy \qquad (1)$$

where $\delta$ is the Dirac delta function having outputs 1 if the input is 0 and 0 otherwise, $\rho \in [-\infty, \infty]$ and $\theta \in [0, \pi]$.

#### 2) R Transform

Then R transform [15], [16] is used to transform the 2D Radon projection to make a 1D R transform profile.

$$R_T(\theta) = \int_{-\infty}^{\infty} R^2(\rho, \theta)d\rho \qquad (2)$$

Basically, R transform is the sum of the squared Radon transform values along a specified angle $\theta$. The general overview of R transformation along single frame consists of the Radon transform having projection of the image intensity along the radial lines oriented at specific angles which produce a complete Radon transform and 1D R transform profile as shown in Fig. 3.

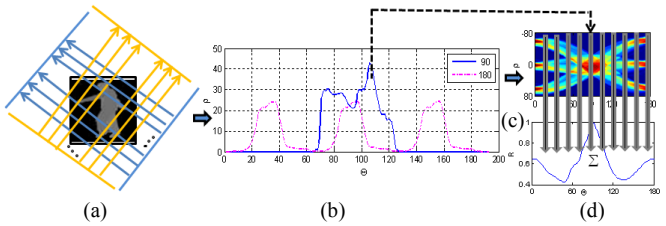


**Fig. 3. Flow of R transformation along a single frame consists of (a) the Radon transform (b) showing two projections at two different angles (90, 180), (c) Radon transform, and (d) 1D R transform profile of a walking silhouette.**

It provides a highly compact shape representation for each depth silhouette. The shape representation of R transform profiles can be further normalized as,

$$R'_T(\theta) = \frac{R_T(\theta)}{max_{\theta}\left(R'_T(\theta)\right)} \qquad (3)$$

#### 3) Invariance of R Transform Features

To account for invariance characteristics, R transform is used to construct invariant R transform profiles of translated and scaled silhouettes. Fig. 4 shows 1D R transform profiles of the original, translated and scaled silhouettes.

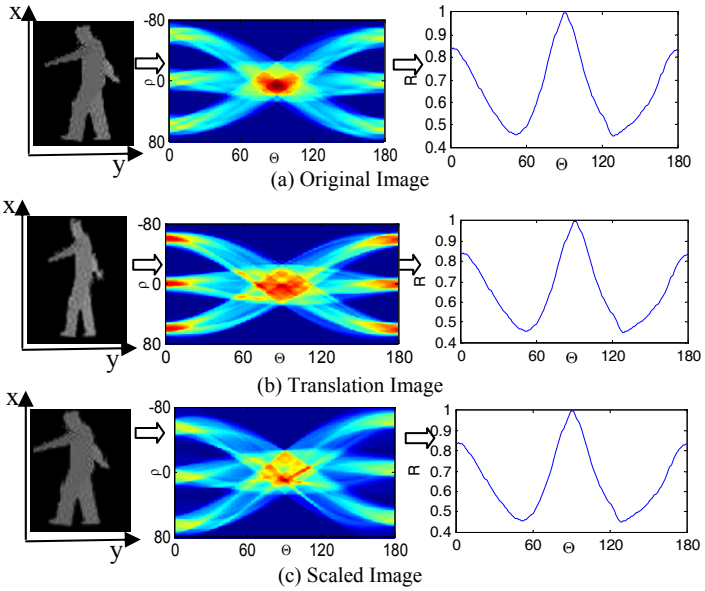


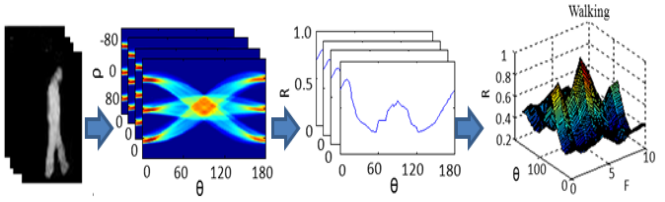
**Fig. 4. R transformation results of a walking silhouette (a) original (b) its translated one (c) and it's scaled respectively. The corresponding R profiles are all identical.**

It clearly indicates that it is invariant with respect to translation and scaling of the depth silhouette. In this regard, R transformation can be a strong candidate to be utilized on

the depth silhouettes to produce translation and scaling invariant features. R transformation reduces a burden of body silhouette alignment due to its invariance properties. Also, it produces a highly compact representation of silhouettes in 1D. Due to these invariant characteristics, image alignment and scaling are not necessary at the preprocessing stage.

*4) R Transformation of an Activity*

From multiple sequential frames of a particular activity, their R transformation features can be obtained as shown in Fig. 5 which shows the general flow of R transformation using a set of depth (walking) activity silhouettes.




(a) (b) (c) (d)

**Fig. 5. Overall flow of R transformation using a series of depth silhouettes of an activity including (a) depth silhouettes (b) Radon transform maps (c) R transform, and (d) A time-sequential R transform profiles respectively.**

Fig. 6 shows the six R transform time-sequential profiles corresponding to our six typical human activities, namely: walking, running, boxing, clapping, sitting-down, and standing-up. These 3D representations of time evolving R transform profiles are derived from a video clip of ten consecutive frames for each activity. The profiles in Fig. 6 seem to effectively characterize and differentiate the six typical human activities. Accordingly, this method establishes distinct feature vectors of each activity sequences by differentiating R transform profiles of each activity shape effectively.

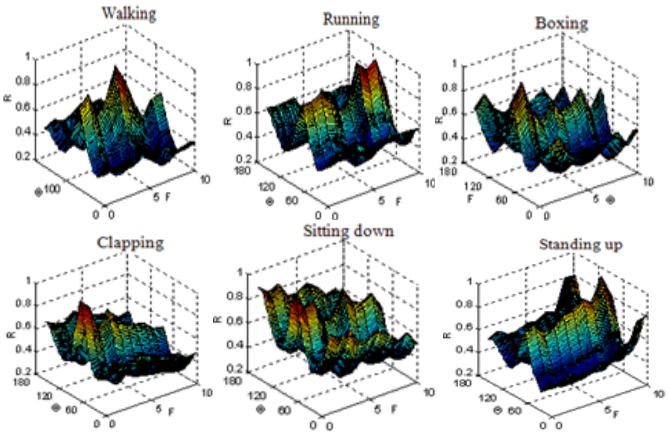



**Fig. 6. Plots of time-sequential R transform profiles from the depth silhouettes of our six human activities.**

Consequently, a large variation in the angles $(\theta)$ of 1D profiles make difference among different human activities due to major moving components of human silhouettes and also the energies (Radon Transform Maps ) of sequential depth silhouettes contain their specific coefficients values which make difference among activities with respect to $(\rho, \theta)$. Thus, R transformation was applied over training data of 1,500 images of human activities and as a result, projecting each silhouette image vector with the size of $1 \times 2{,}500$ on the R transform feature shape is reduced to $1 \times 180$ dimension feature.

*5) Dimension Reduction of R Transformation Profiles Using PCA*

Via R transformation, we obtained the R profiles of 180 dimensions. Now PCA gets used for dimension reduction. Initially, to produce the principal components [29] of the R transformed profiles having maximum variances from the covariance data matrix $M$ , we considered

$$V^T M V = \Lambda \tag{4}$$

where $\Lambda$ is the diagonal matrix of eigenvalues with the size of $e \times e$ and $V$ is the matrix of orthonormal eigenvectors having a size of $t \times e$, where $t$ is the dimension of each silhouette vector and $e$ is the number of principal component.

Thus the principal components (PCs) of the R transformed profiles of the depth silhouettes are expressed as $P_i = \overline{\overline{X}}_i V_e$, where $P_i$ is the PCA projection on the R transformed profiles, and $\overline{\overline{X}}_i$ is the zero mean vector of R transformation profiles. $V_e$ is the leading eigenvectors corresponding to the first top eigenvalues. We considered 100 principal components (PC) of the R transformed profiles after applying PCA over training data of 1,500 images of six human activities and projecting each silhouette vector of R transform with the size of $1 \times 180$ onto the PCA feature space to reduce the dimension feature to $1 \times 100$.

## D. Linear Discriminant Analysis (LDA)

For the purpose of discrimination among the R transformed profiles of different activity classes, LDA acts as an optimal linear discriminant function which maps the input data into the classification space [26]. In addition, LDA utilize class specific information which maximizes the ratio of within and between classes scatter information and obtain maximum discrimination among the features of different classes. Therefore, LDA was applied on the extracted R transform based PC features of depth silhouette human activities to make the PC features more compact and reduce the dimension. The generalized form of within

($S_W$), and between ($S_B$), class scatter matrix [30] is given as:

$$S_B = \sum_{i=1}^{C} D_i (m_1 - m_2)(m_1 - m_2)^T, \quad (5)$$

$$S_W = \sum_{i=1}^{C} \sum_{m \in C_i} (m - m_1)(m - m_1)^T \quad (6)$$

where $D_i$ is the number of vectors in the $i^{th}$ class $C_i$. $c$ is the number of classes which represents the number of activities. $m_2$ represents the mean of all vectors, $m_1$ used as the mean of the class $C_i$, and $m$ is the vector of a specific class. While, the optimal discrimination vector matrix can be considered [26]

$$V^T M V = \Lambda \quad (7)$$

where $F_{LDA}$ is the set of discriminant vectors of $S_W$ and $S_B$ corresponding to the $(c-1)$ largest generalized eigenvalues $\lambda$ which can be obtained as $S_B f_i = \lambda_i S_W f_i$ while the size of $F_{LDA}$ become $t \times v$ where $t \leq v$, as $v$ is the number of elements in a vector. So, the LDA algorithm was applied on our database having six different classes where each vector is 100 dimensional. As a result, the size of the LDA subspace became $5 \times 100$. Thus, to achieve more significant silhouette feature space than PCA or R transform, the LDA algorithm was applied on the R transform based PC silhouette features to produce best discrimination among different classes. The feature vectors using LDA on PC-R features can be represented according to (8) respectively, where $L_{PC\_RTrans}$ indicates the LDA on PC-R features representation for the $i^{th}$ depth silhouettes.

$$L_{PC\_RTrans} = R_T(\theta) P_i F_{LDA} \quad (8)$$

We projected 100 dimensional R transform based PCA representations of 1,500 depth silhouette vectors of six different classes on the LDA subspace, we can further reduce dimensional representation of each vector i.e., $1 \times 5$. Finally, the size of training database containing 1500 vectors becomes $1500 \times 5$ from $1500 \times 2500$.

### *E. Training and Recognition via HMM*

HMM is based on a number of finite states connected by transitions where every state contains transition probability to other state and symbol observation probability. In our study, to model, train, and recognize the human activity, HMM [31] is used since it can deal with the sequential silhouette data with a probabilistic learning process. In a discrete HMM, the feature vectors should be symbolised. We used the Linde, Buzo, and Gray (LBG)'s clustering algorithm [32] to generate a codebook of vectors for HMM. Fig. 7 shows the procedure of codebook generation and symbol selection on the LDA on PC features of the R transform profiles.

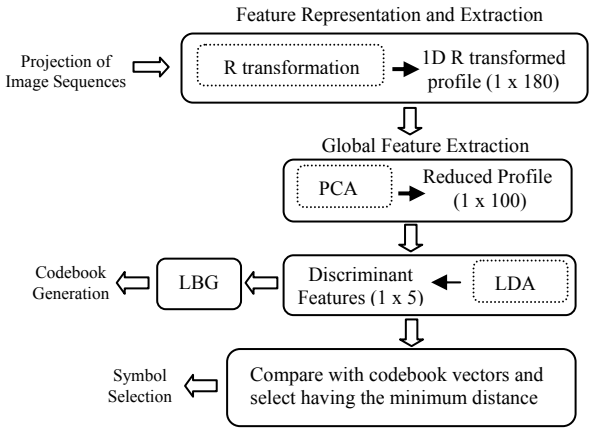



**Fig. 7. Codebook generation and symbol selection.**

Our system used four-state left to right HMMs to model sequential events of human activities. Fig. 8 shows the structure and transition probabilities of a running HMM before and after training with the codebook size of 32.

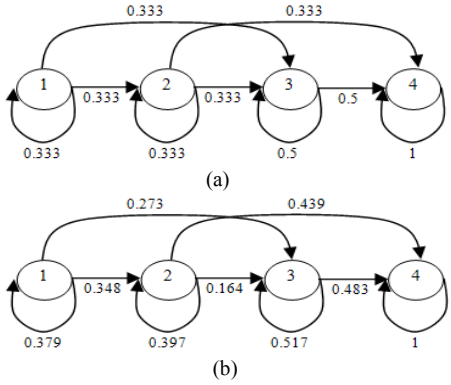



**Fig. 8. Running HMM structure and transition probabilities (a) before and (b) after training.**

To recognize each testing activity, the discrete observation symbol sequence $O = \{O_1, O_2, \ldots, O_n\}$ obtained from the codebook generation and symbol selection are applied on all the trained HMMs and recognized each activity by means of maximum likelihood [33] as

$$L_{max} = arg\ max_{i=1}^{N} \left(P(O|H_i)\right) \quad (9)$$

where $L_{max}$ is the maximum likelihood of sequence observation $O$ on all the distinct HMMs of all activities.

## III. EXPERIMENTAL RESULTS

### *A. Experimental Settings*

To test our system, we built our own depth silhouette database of six most common human activities which recorded multiple test sets from five different subjects. The collected video clips were split into several clips where each clip contained ten consecutive frames. A total of 25 clips from each activity were used to build the training feature space and the whole training data contained a total of 1,500 depth

silhouettes. Initially, each depth silhouette vector with its size of $1 \times 2{,}500$ was transformed via R transformation, producing a 1D profile of $1 \times 180$. PCA reduced the dimension of each vector to $1 \times 100$ and finally LDA was performed on the PC features to build stronger feature space as well as to reduce the dimension further to $1 \times 5$ per each silhouette. In our depth silhouette-based human activity database, we applied 25 video clips for training and 60 video clips for testing each activity.

### *B. Features of PC and IC on Depth Silhouettes*

To compare the performance of the proposed features with other conventional features (i.e., PC and IC features) where the conventional features were both in 2D shape representation while the proposed feature was 1D R transform. Firstly, in the conventional approaches [9] using PCA, PCA was applied on depth silhouettes features to extract the most frequently global features. Fig. 9 shows the five PCA basis images of depth silhouettes using all six typical human activities.

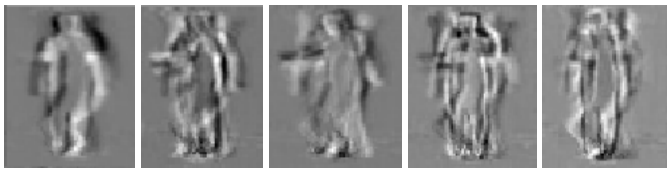


**Fig. 9. Five PCs from the depth silhouettes of all six typical human activities.**

Fig. 10 shows the 3D representation of all the shape silhouettes after applying LDA on the PC features values using depth silhouette. In this plot, using those PC features, some human activities such as boxing and sitting-down were mixed partially and all the features were not seem to be well separated.

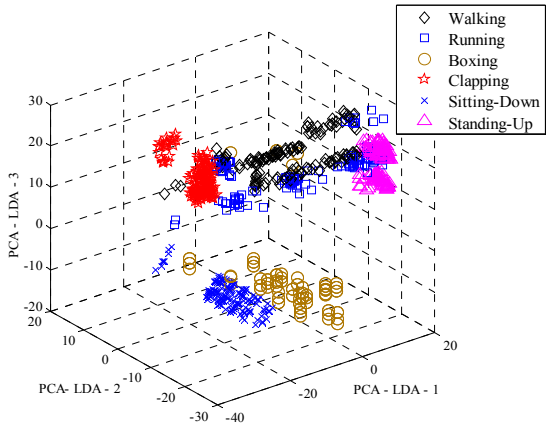



**Fig. 10. 3D plot of LDA on PC features having 1500 depth silhouettes of human activities.**

Fig. 11 shows the IC features. Using ICA [9], IC features of depth images were used to recognize different activities to extract local features. Thus, Fig. 11 shows five ICA basis images having local characteristics of depth silhouette using all six typical human activities.

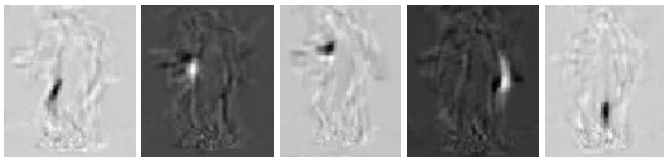


**Fig. 11. Five ICs from the depth silhouettes of all six typical human activities.**

Fig. 12 shows the 3D representation of the depth silhouettes of six different activities after applying on 3 ICs that were chosen based on the top kurtosis values. In the figure, a number of activities were partially separated such as walking, running and boxing as compared to LDA on PC features but a pair of activity such as sitting down and clapping were very close to each other, therefore overlap in the figure among the prototypes of different activities causes low recognition rate.

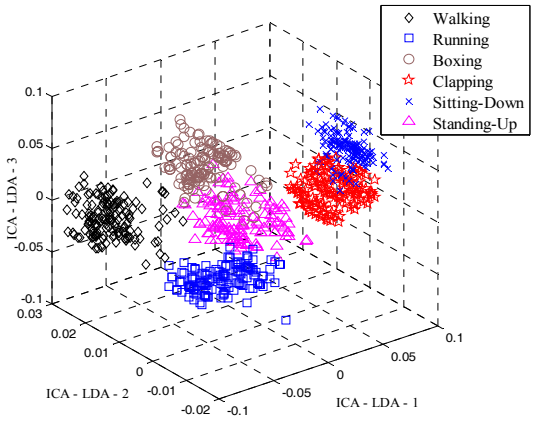



**Fig. 12. 3D plot of LDA on IC features having 1500 depth silhouettes of human activities.**

### *C. R Transform on Binary and Depth Silhouettes*

Furthermore, we continued our study to the proposed R transformation-based feature extraction approach that deals with compact 1D shape representation. While, Fig. 13 shows first three PCs of the binary silhouettes after applying PCA on the R transformed profiles of typical human activities.

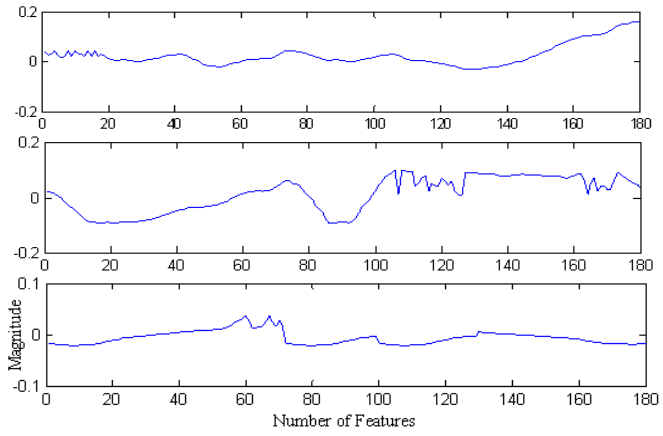



**Fig. 13. Three PCs containing global features of six typical human activities from the binary silhouettes.**

Fig. 14 shows the 3D representation of the binary silhouettes where LDA on the PC-R features from six different activities are not well separated.

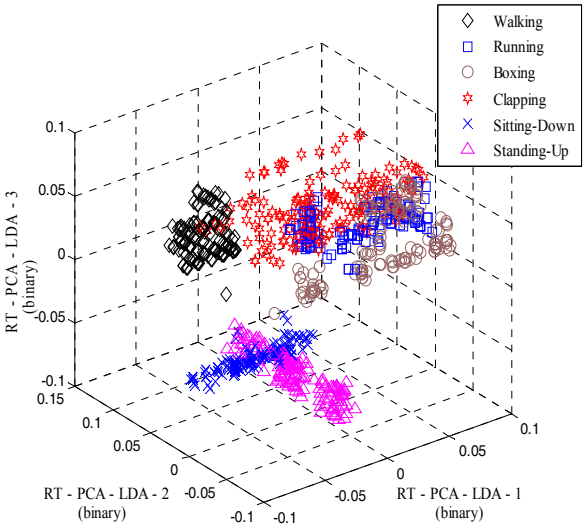



**Fig. 14. 3D plot of LDA on PC-R features having 1500 binary silhouettes of human activities.**

Fig. 15 shows first three PCs of the depth silhouettes after applying PCA on the R transformed profiles of typical human activities. In the figure, 1D profiles show continues variation in magnitude along y-axis at each specific feature.

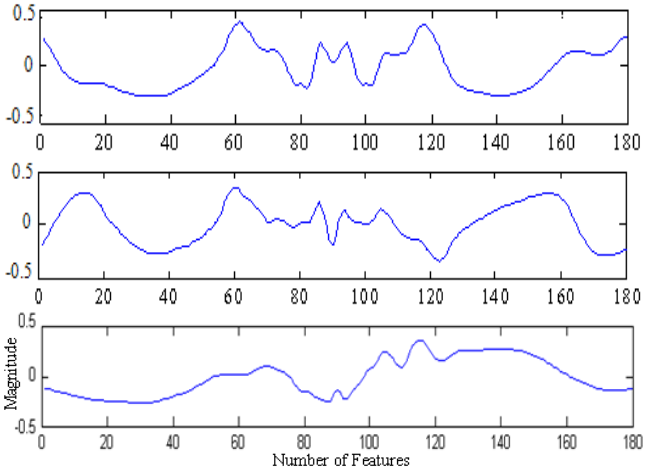



**Fig. 15. Three PCs containing global features of six typical human activities from the depth silhouettes.**

While, Fig. 16 shows the 3D plot of the features after applying the proposed LDA on the PC-R features of the depth silhouette over six human activity classes where 100 PCs were taken. In the figure, the prototypes of all the activities were well separated to each other. Therefore, a comparative overview of 3D plots indicates that the prototypes of all activity classes of proposed method were better separated than the conventional features.

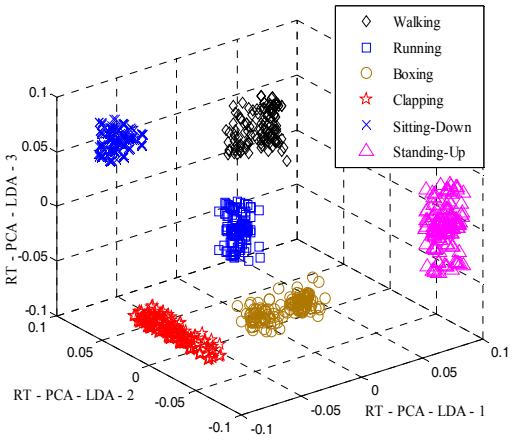



**Fig. 16. 3D plot of LDA on PC-R features having 1500 depth silhouettes of human activities.**

### *D. Recognition Results using PC and IC Depth Silhouettes*

Our approach was evaluated against the conventional approaches using depth silhouettes. As shown in Table 1, the mean recognition rate of LDA on PC feature was 74.83%. And, the results utilizing the IC feature method in Table II shows the improved recognition rate of 86.33% than the results of PC features.

**TABLE I**
**RECOGNITION RESULTS WITH LDA ON THE PC FEATURES APPROACH ON DEPTH SILHOUETTES**

| Activities | WK | RN | BX | CP | SD | SU |
|---|---|---|---|---|---|---|
| **WK** | **82.50** | 12.0 | 0 | 0 | 3.0 | 2.50 |
| **RN** | 8.0 | **83.0** | 2.50 | 0 | 2.0 | 4.50 |
| **BX** | 7.50 | 0 | **68.0** | 16.0 | 0 | 8.50 |
| **CP** | 5.0 | 0 | 19.50 | **71.50** | 0 | 4.0 |
| **SD** | 4.50 | 0 | 5.0 | 0 | **68.50** | 22.0 |
| **SU** | 2.50 | 4.0 | 1.50 | 0 | 16.50 | **75.50** |
| **Mean** | | | **74.83%** | | | |

*WK=Walking; RN=Running; BX=Boxing; CP=Clapping; SD=Sitting Down; SU=Standing Up.

**TABLE II**
**RECOGNITION RESULTS WITH LDA ON THE IC FEATURES APPROACH ON DEPTH SILHOUETTES**

| Activities | WK | RN | BX | CP | SD | SU |
|---|---|---|---|---|---|---|
| **WK** | **89.0** | 6.50 | 0 | 0 | 0 | 4.50 |
| **RN** | 1.50 | **98.50** | 0 | 0 | 0 | 0 |
| **BX** | 3.0 | 0 | **83.50** | 13.50 | 0 | 0 |
| **CP** | 1.50 | 2.50 | 15.50 | **80.50** | 0 | 0 |
| **SD** | 3.0 | 1.0 | 0 | 0 | **81.50** | 14.50 |
| **SU** | 3.50 | 0 | 2.50 | 0 | 9.0 | **85.0** |
| **Mean** | | | **86.33%** | | | |

### *E. Recognition Results using R Transform on Binary and Depth Silhouettes*

The proposed approach was applied on both the binary and depth silhouettes where the recognition results of the experiments reflect the highest recognition rate using depth silhouettes based LDA on the PC-R features approach. While, in Table III, the mean recognition rate using the binary silhouettes based LDA on the PC-R features approach was

overall 67.08%, the lowest recognition rate. Finally, Table IV shows the recognition results of our proposed LDA on the PC-R features approach proved superior recognition rate of 93.16% over all other conventional methods as well as the binary silhouettes.

**TABLE III**
**RECOGNITION RESULTS WITH LDA ON THE PC-R FEATURES APPROACH ON BINARY SILHOUETTES**

| Activities | WK | RN | BX | CP | SD | SU |
|---|---|---|---|---|---|---|
| **WK** | **81.0** | 15.50 | 0 | 0 | 0 | 3.50 |
| **RN** | 11.50 | **84.50** | 0 | 0 | 2.50 | 1.50 |
| **BX** | 5.0 | 0 | **54.0** | 29.50 | 0 | 11.50 |
| **CP** | 7.50 | 0 | 25.0 | **61.50** | 0 | 6.0 |
| **SD** | 0 | 6.50 | 0 | 3.50 | **57.50** | 32.50 |
| **SU** | 7.50 | 4.50 | 0 | 0 | 24.0 | **64.0** |
| **Mean** | | | **67.08%** | | | |

**TABLE IV**
**RECOGNITION RESULTS WITH LDA ON THE PC-R FEATURES APPROACH ON DEPTH SILHOUETTES**

| Activities | WK | RN | BX | CP | SD | SU |
|---|---|---|---|---|---|---|
| **WK** | **95.50** | 4.50 | 0 | 0 | 0 | 0 |
| **RN** | 2.0 | **98.0** | 0 | 0 | 0 | 0 |
| **BX** | 0 | 0 | **93.50** | 6.50 | 0 | 0 |
| **CP** | 3.50 | 0 | 9.50 | **87.0** | 0 | 0 |
| **SD** | 2.50 | 0 | 0 | 0 | **90.50** | 7.0 |
| **SU** | 0 | 0 | 0 | 0 | 5.50 | **94.50** |
| **Mean** | | | **93.16%** | | | |

However, the comparison of overall experimental results showed that LDA on the PC-R features produce the superior recognition rate overall other methods due to their limitations such as difference of subject's size and position changes.

## IV. CONCLUSION

In this paper, we have presented a depth silhouette and R transformation-based translation and scaling invariant HAR system in conjunction with HMM. These invariant feature characteristics are provided by R transformation to produce time sequential feature profiles which have been applied on HMM to model and recognize typical human activities. In addition, our proposed method should be able to handle different sizes and positions of subjects. We have provided the performance of our proposed method on sequential datasets of six typical human activities recognition problems. In our results, the use of depth silhouettes and R transformation has improved the recognition rate up to 93.16% over the conventional systems where the LDA on PC features and LDA on IC features based on the depth silhouettes achieved 74.83% and 86.33%, while binary silhouettes-based R transformation approach achieved 67.08% respectively. Through our experiment using routine daily activities with specific real data, we demonstrated that our algorithm can provide improved recognition rate using depth camera of any other complex activities having unrestricted datasets. The proposed system could be used in any consumer's daily activities monitoring systems to examine the activities of residents including the elderly and disable people at smart home.

## REFERENCES


[1] S. Y. Kim, S. B. Lee, and Y. S. Ho, "Three-Dimensional Natural Video System based on Layered Representation of Depth Maps," *IEEE Transactions on Consumer Electronics,* vol. 52, no.3, pp. 1035-1042, 2006.

[2] M. Chan, D. Esteve, C. Escriba and E. Campo, "A review of smart home-present state and future challenges," *Computer Methods and Programs in Biomedicine*, vol. 91, pp. 55-81, 2008.

[3] T.-Y. Chung, S. Sull and C.-S. Kim, "Frame loss concealment for stereoscopic video plus depth sequences," *IEEE Transactions on Consumer Electronics,* vol. 57, pp. 1336-1344, 2011.

[4] W. Lao, J. Han and P.H.N. De With, "Automatic video-based human motion analyzer for consumer surveillance system," *IEEE Transactions on Consumer Electronics,* vol. 55, pp. 591-598, 2009.

[5] L. Wang and D. Suter, "Learning and matching of dynamic shape manifolds for human action recognition," *IEEE Transactions on Image Processing*, vol. 16, pp. 1646-1661, 2007.

[6] I. Haritaoglu, D. Harwood and L. S. Davis, "$W^4$: Real-time surveillance of people and their activities," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 809-830, 2000.

[7] P. Turaga, R. Chellappa, V. Subrahmanian and O. Udrea, "Machine Recognition of Human Activities: A Survey," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, pp. 1473-1488, 2008.

[8] J. Yamato, J. Ohya and K. Ishii, "Recognizing human action in time-sequential images using hidden markov model," *In: Proceedings of the IEEE 6th International conference on Computer Vision and Pattern Recognition*, pp. 379-385, 1992.

[9] M. Z. Uddin, J. J. Lee and T.-S Kim, "Independent shape component-based human activity recognition via Hidden Markov Model," *Applied Intelligence*, vol. 33, pp. 193-206, 2010.

[10] I. Y. Jang and K. H. Lee, "Depth Video based Human Model Reconstruction resolving Self-Occlusion," *IEEE Transactions on Consumer Electronics,* vol. 56, pp. 1933-1941, 2010.

[11] R. Munoz-Salinas, R. Madina-Carnicer, F. J. Madrid-Cuevas and A. Carmona-Poyato, "Depth silhouettes for gesture recognition," *Pattern Recognition Letters*, pp. 319-329, 2008.

[12] X. Liu and K. Fujimura, "Hand gesture recognition using depth data," *In: Proceedings of the IEEE international conference on Automatic Face and Gesture Recognition*, pp. 529-534, 2004.

[13] S. H. Lee and S. Sharma, "Real-time Disparity estimation algorithm for Stereo Camera Systems," *IEEE Transactions on Consumer Electronics*, vol. 57, pp. 1018-1026, 2011.

[14] J. Han and B. Bhanu, "Human Activity Recognition in Thermal Infrared Imagery," *in Proceedings of the IEEE Computer Society on Computer Vision and Pattern Recognition* , pp. 17-25, 2005.

[15] S. R. Deans, "The Radon transform and some of its application," *Wiley Interscience Publication,* New York, 1983.

[16] M. Singh, M. Mandal and A. Basu, "Pose recognition using the Radon transform," *in Proceedings of the 48th Midwest Symposium on Circuit and Systems*, vol. 2, pp. 1091-1094, 2005.

[17] L. Chen, Y. H. Wang, Y. Wang and D. Zhang, "Gender Recognition from Gait using Radon transform and Relevant Component Analysis," *Lecture of Computer Science 5754*, pp. 92-101, 2009.

[18] N. V. Boulgouris and Z. X. Chi, "Gait Recognition Using Radon Transform and Linear Discriminant Analysis," *IEEE Transactions on Image Processing*, vol. 16, pp. 731-740, 2007.

[19] J. Li, S. K. Zhou and R. Chellappa, "Appearance Modeling Using a Geometric Transform," *IEEE Transactions on Image Processing*, vol. 16, pp. 889-902, 2009.

[20] S. Tabbone, L. Wendling and J. P. Salmon, "A new shape descriptor defined on the Radon transform," *Computer Vision and Image Understanding*, vol. 102, pp. 42-51, 2006.

[21] Y. Wang, K. Huang and T. Tan, "Human Activity Recognition based on R transform," *in Proceedings of the IEEE conference on Computer Vision and Pattern Recognition,* pp. 1-8, 2007.

[22] H. Zhang, Z. Liu, H. Zhao and G. Cheng, "Recognizing Human Activities by Key frame in video sequences," *Journal of Software*, vol. 5, pp. 818-825, 2010.

[23] F. I. Bashir, A. A. Khokhar and D. Schonfeld, "Object Trajectory-based Activity Classification and Recognition using Hidden Markov

Models," *IEEE Transactions on Image Processing*, vol. 16, pp. 1912-1919, 2007.

[24] X. He and P. Niyogi, "Locality preserving projections," *In: Advances in Neural Information Processing Systems,* MIT Press, 2003.

[25] A. M. Martinez and A. C. Kak, "PCA versus LDA," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 23, pp. 228-233, 2001.

[26] S. Ji and J. Ye, "Generalized Linear Discriminant Analysis: A Unified Framework and Efficient Model Selection," *IEEE Transactions on Neural Networks*, vol. 19, pp. 1768-1782, 2008.

[27] O. Rohlik, P. Mautner, V. Matousek and J. Kempf, "HMM based handwritten text Recognition using biometrical data acquisition Pen," *In: proceeding of IEEE International symposium on Computational Intelligence in Robotics and Automation*, pp. 950-953, 2003.

[28] H-K. Lee and J. H. Kim, "An HMM-based threshold model approach for gesture recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 21, pp. 961-973, 1999.

[29] B. A. Draper, K. Baek, MS. Barlett and J. R. Beveridge, "Recognizing faces with PCA and ICA," *Computer Vision and Image Understanding*, vol. 91, pp. 115-137, 2003.

[30] A. Eleyan and H. Demirel, "PCA and LDA based face recognition using feed forward neural network classifier," *in Proceedings of multimedia contents representation, classification and security,* pp. 199-206, 2006.

[31] H. Othman and T. Aboulnasr, "A Separable low complexity 2D HMM with application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 25, pp. 1229-1238, 2003.

[32] Y. Linde, A. Buzo and R. Gray, "An Algorithm for Vector Quantizer Design," *IEEE Transactions on Communications*, vol. 28, no. 1, pp. 84-94, 1980.

[33] L. R. Rabiner, "A Tutorial on Hidden Markov Modes and selected application in speech recognition," *in Proceedings of IEEE*, vol. 77, pp. 257-286, 1989.

## BIOGRAPHIES

**A. 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 is currently working toward 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.

**Md. Zia Uddin** received his Ph.D. degree in Biomedical Engineering from Kyung Hee University, Republic of Korea in 2011. He is currently working as a faculty member in the Department of Electronic Engineering, Inha University, Republic of Korea. Dr. Zia's research interest includes pattern recognition, image processing, computer vision, and machine learning.

**T.-S. Kim** received the B.S. degree in Biomedical Engineering from the University of Southern California (USC) in 1991, M.S. degrees in Biomedical and Electrical Engineering from USC in 1993 and 1998 respectively, and Ph.D. in Biomedical Engineering from USC in 1999. After his postdoctoral work in cognitive sciences at the University of California, Irvine in 2000, he joined the Alfred E. Mann Institute for Biomedical Engineering and Dept. of Biomedical Engineering at USC as a Research Scientist and Research Assistant Professor. In 2004, he moved to Kyung Hee University in Korea where he is currently an Associate Professor in the Biomedical Engineering Department. His research interests have spanned various areas of biomedical imaging including Magnetic Resonance Imaging (MRI), functional MRI, E/MEG imaging, DT-MRI, transmission ultrasonic CT, and Magnetic Resonance Electrical Impedance Imaging. Lately, he has started research work in proactive computing at the u-Lifecare Research Center where he serves as Vice Director. Dr. Kim has published more than 80 peer reviewed papers and 150 proceedings, 6 international book chapters, and holds 10 international and domestic patents. He is a member of IEEE from 2005-2012, KOSOMBE, and Tau Beta Pi, and listed in Who's Who in the World '09-'12 and Who's Who in Science and Engineering '11-'12.