Daily Human Activity Recognition Using Depth Silhouettes and $\Re$ Transformation for Smart Home

Ahmad Jalal[1], Md. Zia Uddin[2], Jeong Tai Kim[3], and Tae-Seong Kim[1]

[1] Department of Biomedical Engineering, Kyung Hee University,
1 Seochun-ri, Kiheung-eup, Yongin-si, Kyunggi-do, 446-701, Republic of Korea
[2] Department of Electronic Engineering, Inha University,
253 Yonghyun-dong, Nam-gu, Incheon, 402-751, Republic of Korea
[3] Department of Architectural Engineering, Kyung Hee University,
1 Seochun-ri, Kiheung-eup, Yongin-si, Kyunggi-do, 446-701, Republic of Korea
jalalahmad25@yahoo.com, ziauddin@inha.ac.kr,
{jtkim,tskim}@khu.ac.kr

Abstract. We present a human activity recognition (HAR) system for smart homes utilizing depth silhouettes and $\Re$ transformation. Previously, $\Re$ transformation has been applied only on binary silhouettes which provide only the shape information of human activities. In this work, we utilize $\Re$ transformation on depth silhouettes such that the depth information of human body parts can be used in HAR in addition to the shape information. In $\Re$ transformation, 2D directional projection maps are computed through Radon transform, and then 1D feature profiles, that are translation and scaling invariant, are computed through $\Re$ transform. Then, we apply Principle Component Analysis and Linear Discriminant Analysis to extract prominent activity features. Finally, Hidden Markov Models are used to train and recognize daily home activities. Our results show the mean recognition rate of 96.55% over ten typical home activities whereas the same system utilizing binary silhouettes achieves only 85.75%. Our system should be useful as a smart HAR system for smart homes.

Keywords: Human activity recognition, Depth silhouettes, $\Re$ transformation, Smart home.

1 Introduction

Human activity recognition (HAR) has become one of the challenging and active areas of research recently with its essential role for smart homes. General approach of video-based HAR is to extract some significant features from each video frame, and use these features to train a classifier and perform recognition [1]. A HAR system can keep a continuous observation on basic human activities of daily living, allowing various services such as lifecare from physical damage, nursing, rehabilitation, and health assistance to make a more intelligent home environment [1]. In recent years, HAR becomes a key component in a smart home system providing various applications such as home activity monitoring and e-healthcare [2]. Numerous smart home projects (e.g., Microsoft Easyliving project and House_n group at MIT) [1] are currently underway.

B. Abdulrazak et al. (Eds.): ICOST 2011, LNCS 6719, pp. 25–32, 2011.

In the video-based HAR, typically binary silhouettes [3] are used for human activity representation however they produce ambiguity among the same silhouette for different postures of different activities due to their limited pixel value (i.e., 0 or 1, thus within the binary shape of a posture, no information presents). To extract activity shape features, Principle Component Analysis (PCA) and Independent Component Analysis (ICA) have been used [2, 4], producing the spatial features that are global and local respectively. However these features are sensitive to translation and scaling of human body postures which are problems in the silhouette extraction process. To derive the translation and scaling invariant features (thus reducing the burden on the silhouette extraction process in our case), $\Re$ transformation has been proposed. $\Re$ transformation first computes a 2D angular projection map of an activity silhouette via Radon transform, then converts the 2D Radon transformed map into a 1D $\Re$ transform profile. $\Re$ transformation [5] was first introduced to classify objects from images and extract distinct directional features of each binary shape. In [6], Singh et al proposed Radon transform to identify the skeleton representation for the human recognition. In [7], Chen et al. implemented Radon transform for its sensitiveness to angle variation to promote gender recognition using binary silhouettes. In [3], $\Re$ transform was applied on binary silhouettes to describe the spatial information of different human activities. However the binary silhouettes cause limited recognition performance due to a lack of information in the flat pixel intensity, and difficulties to differentiate between the far and near distance of the human body parts. Recently to overcome the limitation of the binary silhouettes, depth based silhouette representation for human activity has been suggested [8-10], since depth silhouettes differentiate the body parts by means of different intensity values.

In this work, we propose a depth silhouette based home HAR system for smart homes. $\Re$ transformation is applied on the depth silhouettes. PCA is applied to extract features from the $\Re$ transformed profiles of depth silhouettes, and then Linear Discriminant analysis (LDA) is applied to make them more robust. Finally, the features are utilized in Hidden Markov Model (HMM) for recognition of ten daily home activities. Our results show that significant improvement over the systems where only the binary silhouette features are used in recognition. The proposed system could be an essential component of a smart home system for continuous observation of daily human activities.

2 Methodology

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

2.1 Depth Silhouette Preprocessing

To capture depth images of activities, we utilized a ZCAM^{TM} [11]. Activity silhouettes are extracted from depth images and resized to 50×50. Figs. 2 (a) and (b) show some sample of binary and depth silhouette sequential images of a rushing activity.

In Fig. 2, it is clear that binary silhouettes contain limited information due to its flat pixel intensity value (i.e., 0 or 1) distribution over the human body (i.e., only shape information is available) while depth silhouettes show discernable parts in addition to the shape information.

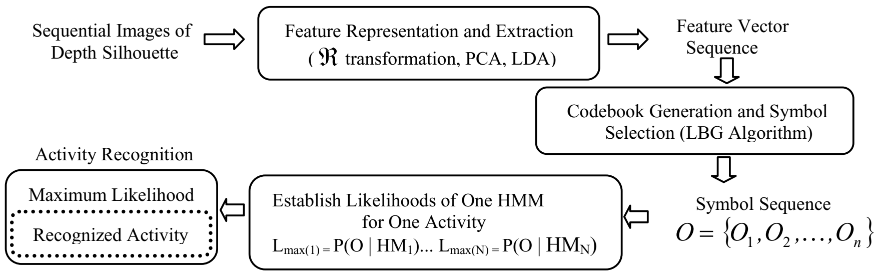


Fig. 1. Overall flow of our proposed human activity recognition system

Fig. 2. Two sequences of (a) binary and (b) depth silhouettes of a rushing activity

2.2 Feature Representation

To derive translation and scaling invariant features from the depth activity silhouettes, $\Re$ transformation including Radon transform is the key technique used in the proposed approach. First, Radon transform computes a 2D projection of a depth silhouette [6] along specified view directions. It is applied on each depth silhouette to establish a mapping between the domain produced by the image coordinate system $f(x, y)$ and the Radon domain indicated as $\Re(\rho, \theta)$.

Let $f(x, y)$ is the depth silhouette, its Radon transform $\Re(\rho, \theta)$ is computed by

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

where δ is the Dirac delta, $\rho \in [-\infty, \infty]$ and $\theta \in [0, \pi]$.

Then $\Re$ transform [3, 5] is used to transform the 2D Radon projection to make a 1D $\Re$ transform profile for every frame.

$$\Re_T(\theta) = \int_{-\infty}^{\infty} \Re^2(\rho, \theta)\, d\rho \tag{2}$$

Basically, $\Re$ transform is the sum of the squared Radon transform values along the specified angle θ. It provides a highly compact shape representation and reflects the time-sequential profiles of each daily home human activity. In addition, these $\Re$ transform profiles are translation and scaling invariant. Fig. 3 shows the general flow of $\Re$ transformation using a set of depth (walking) activity silhouettes.

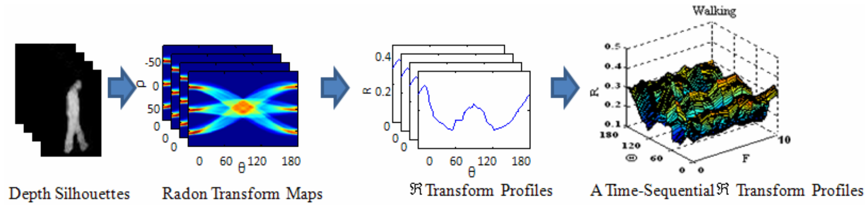


Fig. 3. Overall flow of Radon transform using depth silhouette activities

Fig. 4 shows the whole set of $\Re$ transform time-sequential profiles of all human home activities we recognize in this study: namely, cleaning (vacuuming), cooking, eating, exercise (dumb-bell), lying-down, rushing, sitting-down, standing-up, take-an-object, and walking. These 3D representations of time evolving $\Re$ transform profiles are derived from a video clip of ten consecutive frames F for each activity. It is shown that $\Re$ transform profiles show distinct characteristics of the ten home human activities.

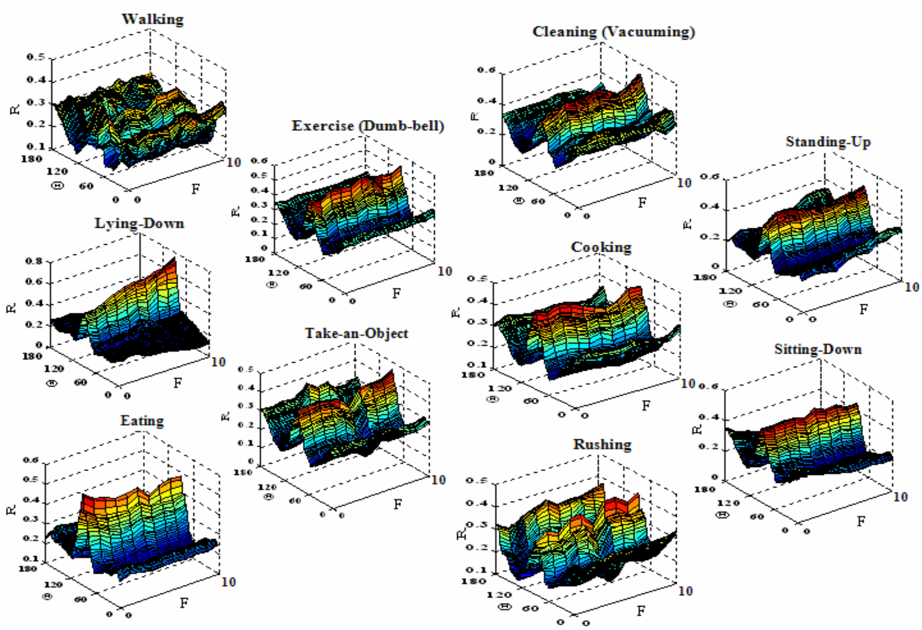
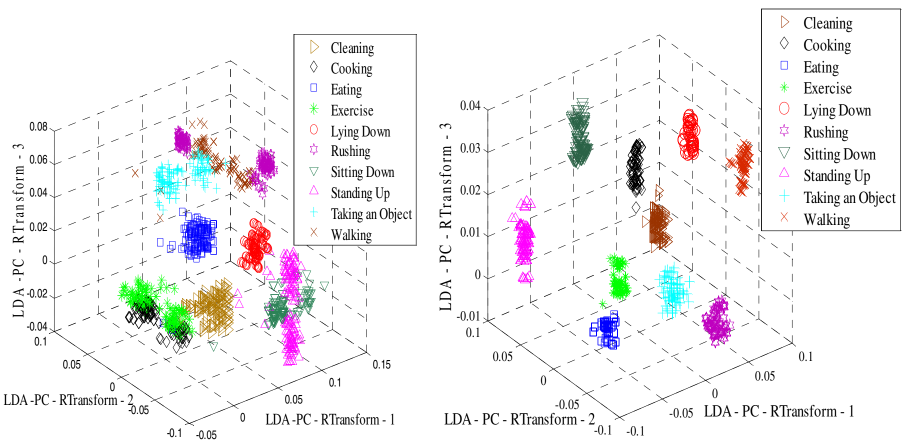


Fig. 4. Plots of time-sequential $\Re$ transform profiles from the depth silhouettes of our ten daily home human activities

2.3 PCA and LDA for Silhouette Feature Extraction

After computing the $\Re$ transformation profiles, PCA [4] is used to extract the dominant features. Thus the principal components (PCs) of the $\Re$ 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 $\Re$ transformed profiles, and $\overline{\overline{X}}_i$ is the zero mean vector of $\Re$ transformation profiles. V_e is the leading eigenvectors corresponding to the first top eigenvalues. Finally, LDA [12] finds a set of discriminant vectors F_{LDA} that maximizes the ratio of the determinant of the between S_B and within S_W class scatter matrix as $F_{LDA} = \left|F^T S_B F\right| / \left|F^T S_W F\right|$. Thus, the feature vectors using LDA on PC- $\Re$ features can be represented as $L_{PC\_RTrans} = \Re_T(\theta) P_i F_{LDA}$ respectively where L_{PC_RTrans} indicates the LDA on PC- $\Re$ features representation for the i^{th} depth silhouette image.
Figs. 5(a) and (b) show the 3D plots of the features after applying the proposed feature extraction methods on the binary and depth silhouettes of our home activities. In Fig. 5 (a), a pair of activities such as (cooking and exercise) and (walking and rushing) are very close to each other. While in Fig. 5 (b), all activities are well separated among each other.



(a) (b)

Fig. 5. Plot of the LDA on PC- $\Re$ features on (a) the 1,500 binary silhouettes, and (b) 1,500 depth silhouettes of our home activities

2.4 Activity 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, for modeling, training, and recognizing the home activity, HMM [12] 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 symbolized. We used the Linde, Buzo, and Gray (LBG)'s clustering algorithm [14] to generate a codebook of vectors for HMM.

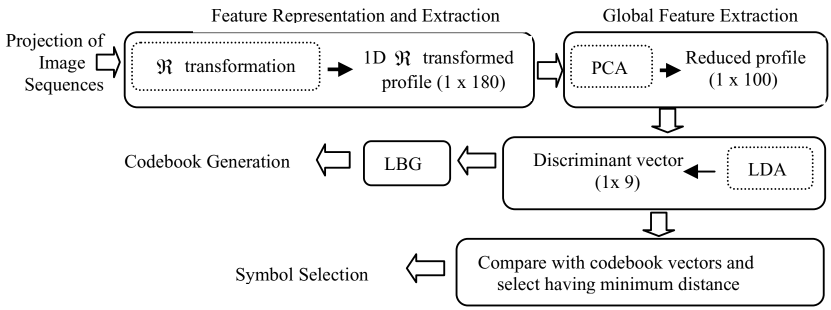


Fig. 6. Codebook generation and symbol selection

Fig. 6 shows the procedure of codebook generation and symbol selection on the LDA on PC features of the ℜ transformed silhouettes. Fig. 7 shows the structure and transition probabilities of a walking HMM before and after training with the codebook size of 32 [13].

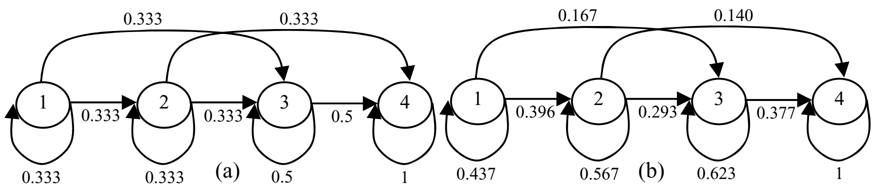


Fig. 7. Walking HMM structure and transition probabilities (a) before and (b) after training

3 Experimental Results

To test our system, we built our own depth silhouette database of the ten daily home activities and recorded multiple test sets from three different subjects. The collected video clips were split into several clips where each clip contained ten consecutive frames. A total of 15 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 1×180. Further reduction in dimension was performed by PCA producing the feature of 1×100, and finally LDA was performed ending up with a feature vector of 1×9. In our home activity model, we applied 15 video clips of each activity in training and 45 video clips in testing of depth silhouette images respectively. Table 1 shows the recognition results from the binary and depth activity silhouettes respectively utilizing two different features sets where the proposed LDA on PC-ℜ feature approach on the depth silhouette shows significantly superior mean recognition of 96.55% over that of the binary silhouettes (i.e., 85.75%).

Table 1. Recognition results of different feature extraction approaches using both binary silhouettes and depth silhouettes

Approaches	Home human Activities	Recognition Rate(Binary)	Recognition Rate (Depth)	Mean (Binary)	Mean (Depth)
PC-$\Re$ features	Cleaning	72.50%	87.50%	72.40%	88.65%
	Cooking	51.50	79.0		
	Eating-Drink	84.0	100		
	Exercise	57.50	84.50		
	Lying Down	100	100		
	Rushing	66.50	86.50		
	Sit Down	63.0	84.0		
	Stand Up	81.50	90.50		
	Take Object	72.0	83.50		
	Walking	75.50	91.0		
LDA on PC-$\Re$ features	Cleaning	85.50	94.50	85.75	96.55
	Cooking	75.50	91.0		
	Eating-Drink	97.50	100		
	Exercise	78.0	96.0		
	Lying Down	100	100		
	Rushing	82.0	96.0		
	Sit Down	80.50	95.50		
	Stand Up	88.50	92.50		
	Take Object	82.50	100		
	Walking	87.50	100		

4 Conclusions

We have presented a depth silhouette and $\Re$ transformation based HAR system for smart homes. In our results, the use of depth silhouettes and $\Re$ transformation improves the recognition rate up to 96.55% over the conventional systems where the PC-$\Re$ features and LDA on PC-$\Re$ features based on the binary silhouettes achieved only 72.40% and 85.75% respectively. The proposed system can be implemented at smart home to recognize daily activities of residents.

Acknowledgement

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (No. 2010-0001860).

References

1. Chan, M., Esteve, D., Escriba, C., Campo, E.: A review of smart homes-Present state and future challenges. Computer Methods and Programs in Biomedicine 91, 55–81 (2008)
2. Uddin, M.Z., Lee, J.J., Kim, T.-S.: Independent shape component-based human activity recognition via Hidden Markov Model. Journal of Applied Intelligence, 193–206 (2009)
3. Wang, Y., Huang, K., Tan, T.: Human Activity Recognition Based on $\Re$ Transform. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1--8 (2007)

4. Han, J., Bhanu, B.: Human Activity Recognition in Thermal Infrared Imagery. In: Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), vol. 3, pp. 17–25 (2005)
5. Tabbone, S., Wendling, L., Salmon, J.-P.: A new shape descriptor defined on the Radon transform. Computer Vision and Image Understanding 102, 42–51 (2006)
6. Singh, M., Mandal, M., Basu, A.: Pose recognition using the Radon transform. In: 48th Midwest Symposium on Circuits and Systems, vol. 2, pp. 1091–1094 (2005)
7. Chen, L., Wang, Y., Wang, Y., Zhang, D.: Gender recognition from gait using radon transform and relevant component analysis. In: Huang, D.-S., Jo, K.-H., Lee, H.-H., Kang, H.-J., Bevilacqua, V. (eds.) ICIC 2009. LNCS, vol. 5754, pp. 92–101. Springer, Heidelberg (2009)
8. Salinas, R.M., Carnicer, R.M., Cuevas, F.J.M., Poyato, A.C.: Depth silhouettes for gesture recognition. Pattern Recognition Letter, PRL (29), 319–329 (2008)
9. Uddin, M.Z., Truc, P.T.H., Lee, J.J., Kim, T.-S.: Human Activity Recognition Using Independent Component Features from Depth Images. In: Proc. of the 5th International Conference on Ubiquitous Healthcare, pp. 181–183 (2008)
10. Frueh, C., Zakhor, A.: Capturing 2½ D Depth and Texture of Time-Varying Scenes Using Structured Infrared Light. In: 5th International Conference on 3-D Digital Imaging and Modeling, pp. 318--325 (2005)
11. http://www.primesense.com
12. Aradhya, V.N.M., Kumar, G.H., Noushath, S.: Fisher Linear Discriminant Analysis and Connectionist Model for Efficient Image Recognition. Studies in Computational Intelligence 83, 269–278 (2008) ISSN 1860-9503
13. Rabiner, L.R.: A tutorial on Hidden Markov Models and selected applications in speech recognition. Proc. of the IEEE 77(2), 257–286 (1989)
14. Linde, Y., Buzo, A., Gray, R.M.: An Algorithm for Vector Quantizer Design. IEEE Transactions on Communication 28, 84–95 (1980)