A Robust Daily Human Activity Recognition and Prediction System

Md. Zia Uddin
Department of Computer Education
Sungkyunkwan University
Seoul, Republic of Korea
ziauddin@skku.edu

ABSTRACT
In this work, a system is proposed for Human Activity Prediction (HAP) using activity sequence spanning tree constructed from a life-log created via a depth video camera-based daily Human Activity Recognition (HAR) approach using time-sequential augmented Local Binary pattern (LBP) and Enhanced Independent Component (EIC)-based depth silhouette features with Hidden Markov Models (HMMs). Regarding the daily HAR, the augmented local features are extracted first from the collection of the depth silhouettes containing various daily human activities such as walking, sitting, lying, cooking, neutral etc. Using these features, HMMs are used to model the time sequential information and recognize several activities. The depth silhouette-based activity recognition results show superior performance over the traditionally used binary silhouette-based approaches, The depth silhouette-based human activity recognition system can be used to recognize the activities automatically, which can be utilized to create life-log. In this regard, an easy prediction method is also proposed for predicting of activity in next few frames while testing a small video which may consist of multiple activities. It can save recognition time and help us to assign a long-time activity annotation for fast life-log construction. After building a life-log consisting of activity sequences, furthermore, a method for human activity prediction in long video is proposed using activity sequence spanning tree built from the activity sequence database. Based on the consecutive activities recorded in an activity sequence database (i.e. life-log) for a specific period of time of each day over a long time such as a month, the spanning tree can be constructed for the sequences starting with each activity where the leaf nodes contain the frequency of the consecutive activity sequences. Once the tree is constructed, to predict an activity after a sequence of activities, traverse the spanning tree until a path up to the previous node of the leaf nodes is matched with the testing pattern and finally, prediction of the next activity is done based on the highest frequency of the leaf nodes along the matched path. The prediction experiments over computer simulated data shows satisfactory results. The proposed video sensor-based human activity recognition and prediction systems can be utilized for important practical applications such as smart healthcare, proactive computing etc.

C3S2E '14, August 03 - 05 2014, Montreal, QC, Canada
Copyright 2014 ACM 978-1-4503-2712-1/14/08…$15.00.
http://dx.doi.org/10.1145/2641483.2641537

Keywords
LBP, EICA, Depth Silhouette, Spanning Tree.

1. INTRODUCTION
Human Activity Recognition (HAR) from video is getting good attentions nowadays among the researchers of Human Computer Interaction (HCI) [1]-[3]. Besides, an efficient HAR system can be applied in many practical applications such as smart healthcare, automated surveillance etc. The most common methodology for video-based human activity recognition is extracting some features from time sequential activity images and comparing with each trained activity features. Binary silhouettes are very commonly applied in this regard [4]-[9]. Yamato et al. proposed a binary silhouette-based human activity recognition system using mesh features with hidden Markov models (HMMs) [4]. Niu and Abdel-mottaleb applied binary silhouette features for representing view-invariant human activities [5], [6]. Local features such as Local Binary Patterns (LBP) were proposed in some pattern recognition works such as [9], [10]. LBP was introduced initially for the purpose of texture analysis [9] but recently it has been applied for human activity recognition [10] as well. Since LBP is computationally efficient, hence one can come up with strong features by using LBP on the depth images from human activity video. Regarding local human activity feature extraction, there is another famous approach namely Enhanced Independent Component Analysis (EICA) proposed in [11]. Basically, ICA is a blind source separation analysis to decompose an observed signal into a linear combination of some unknown independent signals. Before applying basic ICA, PCA can be used to reduce dimension of total training image data which is known as EICA. Since EICA also extracts useful information to represent human activities in images, it can also be adopted with LBP for strong feature representation. Thus, augmented LBP and EIC features can be adopted from human activity video for efficient HAR. Binary silhouettes are not efficient enough to describe human body properly in the activity videos due to its two-level flat pixel intensity distribution although they have been extensively utilized in many HAR applications. From the binary silhouettes, it is impossible to obtain the difference between the far and near parts of human body in the activity video [11]-[16]. To overcome such limitations of the binary representation of human body, depth information-based whole body silhouette representation for

human activity recognition was proposed in [11]. Thus, depth values can represent the human body better than the binary ones for HAR.

Once a good HAR system is developed, it can be used to recognize different daily human activities such as walking, neutral, lying, sitting, cooking etc. Thus, a computer can act as life-log agent that log what activities a human performed everyday in specific time [17]. There have been a few attempts to log a person's life in different research areas [18]-[20]. Using the human activity log information, many researchers have tried to predict human activities [21]-[23]. In [21], the authors proposed a behavior prediction system for supporting daily lives where the behaviors in daily-life were recorded in an environment with embedded sensors, and the prediction system learned the characteristic patterns that would be followed by the behaviors to be predicted. The prediction system observes behaviors with the sensors and outputs the prediction of the future behaviors based on the some rules. For the experiments, the authors generated 1250 rules to be applied for prediction that really caused a lot of time for prediction. In another work [22], the authors focused on the prediction of the progression of a particular activity on the basis of a 24-hour period to detect unexpected events which could indicate changes in health status. The objective of the authors was to help caregivers by providing better assessments of functional health status and preventative assistance. In [23], data mining association rule-based method was proposed for activity prediction where the frequent activity sequences were mined based on probabilities calculated from the temporal relations of the activities i.e., which activity was followed by which activity at what time. For instance, if anyone takes pills after night meal frequently and suddenly he forgets to take the pill after night meal, it can remind him to take the pills. It is known that human beings habitually repeat same kind of sequential patterns every day and hence, the contiguous activity sequences can be focused for efficient Human Activity Prediction (HAP). Also, to store and search the contiguous activity sequence information, tree-based technique can be considered. Thus, once human activity log for a long period of time is obtained, it can be utilized for real time HAP based on which a computer can assist based on predicted next activity. Therefore, HAP can play a vital role in proactive computing, which is a technology that proactively anticipates peoples' health-related needs and takes whatever action that is appropriate on their behalf.

In this work, a simple activity sequence spanning tree-based HAP is proposed using life-log created through a depth silhouette-based HAR system. As depth silhouettes are more effective than binary silhouettes to represent human activities, depth silhouettes can be employed with HMM for an improved HAR. The depth silhouette-based HAR system can also be employed in real time to recognize different daily activities. Thus, the real time depth silhouette-based HAR system can result in an effective video-based life-log generation of different daily home activities. Besides, for creating the life-log in a fast way, simple prediction method is also proposed for predicting of activity in next few frames in a small video (may consist of multiple activities) for saving time to create life-log i.e., if we have a small testing video consisting of multiple activities, we can split the video in some consecutive time-slides to recognize the slides individually. After recognizing some slides, if we see that the recognized activities are same than we can predict that next activity in next few frames should be the same activity and skip next few frames for recognition. It can be used for video segmentation with small errors. Thus, we can assign a long time activity annotation for life-log in this way which can be used later for activity prediction in long video. The proposed system can be effectively application important applications such as smart home healthcare. Fig. 1 shows a conceptual setup for video-based human activity recognition and prediction for smart room.

Using the daily contiguous activity information from the life-log, activity spanning tree can be built containing the frequency of the contiguous activity sequence. Every node in the tree represents an activity. Furthermore, after recognizing some contiguous activities automatically by a HAR system in real time, let us try to predict the next activity going to be performed by means of the activity spanning tree information. Thus, traversing the tree for matching the contiguous activity sequence can be done for prediction. Along with that matched path in the tree, there may be more than one leaf nodes. Then, considering the leaf node with the highest frequency in the matched path in a tree the next activity going to be performed can be predicted. As real life-log containing the daily activities is available at this moment, the proposed prediction system has been tested by a computer simulated random data and obtained satisfactory results.

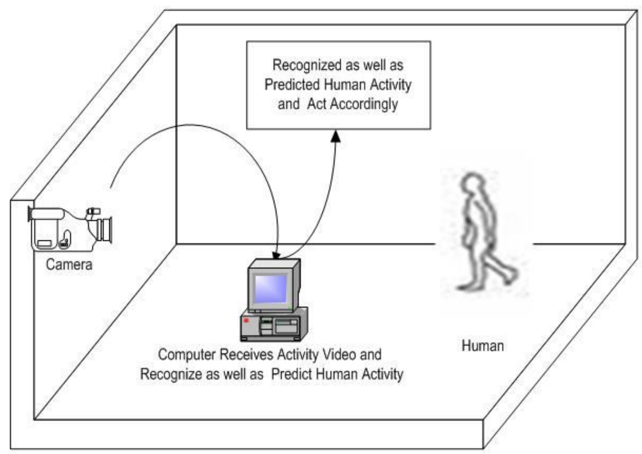


Figure 1. A conceptual setup for a video-based human activity recognition and prediction for smart room.

2. HUMAN ACTIVITY RECOGNITION

Let us start with the processing of depth silhouette information from the time sequential activity video images. Fig. 2 shows the basic architecture of the proposed HAR system to create life-log. Once a life-log is available containing the activities of each day, human activity prediction can be done based on contiguous sequential activity patterns. The RGB and depth images of different activities are acquired by ZCAM^{TM} [24]. The depth video generated by the system indicates the range of each pixel in the scene to the camera as a gray level value such as the shorter ranged pixels have brighter and longer ones contains darker values as shown in Fig. 3(b) where right hand looks brighter than the left hand based on the distance. Fig. 3(a) represents the RGB image corresponding to the depth image shown in Fig. 3(b). The system can provide both the RGB and depth images simultaneously. The binary silhouettes can be obtained from the depth images based on a threshold value. Basically, the binary silhouette representations contain a flat pixel value distribution on all over the human body in the activity video frames. The region of interest (ROI) is extracted from every depth image and generalized to the size of 50x50. Figs 4(a) and 4(b) show ten generalized binary and depth silhouettes respectively from a walking activity. Here, significant differences can be noticed between the pixel intensity distribution of the depth and binary silhouettes. All resized depth silhouettes are collected for further processing of the proposed recognition system.

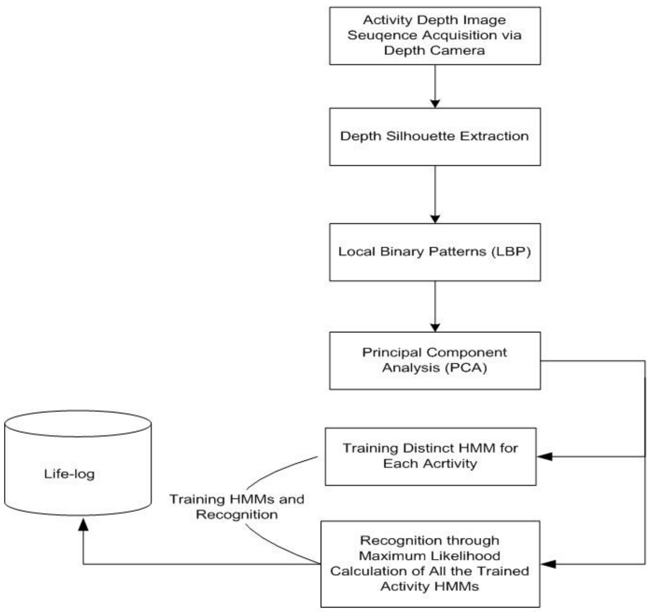


Figure 2. Basic steps for depth silhouette-based real time HAR system to create activity life-log.

For feature extraction, LBP-PCA feature are extracted and then followed by EICA. Finally, both features are augmented. LBP features are local binary patterns based on the intensity values of surrounding pixels of a center pixel. Then, the LBP pattern at the given pixel at (x_i, y_i) can be represented as an ordered set of the binary comparisons as:

$$LBP(x_e, y_e) = \sum_{i=0}^{7} f(g_i - g_e)2^i \tag{1}$$

$$f(l) = \begin{cases} 0, l \le 0 \\ 1, l > 0 \end{cases} \tag{2}$$

where g_e represent the intensity of the given pixel and g_i intensity of the surrounding pixels. Fig. 5 shows a LBP operator used in this work. The image textual feature is basically represented by the histogram of the LBP map of which the i^{th} bin can be defined as follows

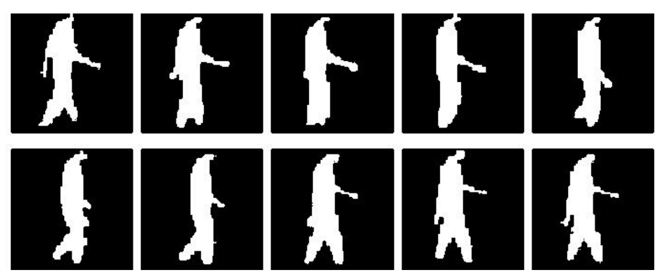
$$Histo_m = \sum_{x,y} I\{LBP(x,y) = i\}, m = 0,1,...n-1 \tag{3}$$

where n is the number of the LBP histogram bins. The histogram of the LBP map is then presented as $Histo = (hist_0, hist_1, ..., hist_{n-1})$. To describe the LBP pattern, an image is often divided into some non-overlapping rectangle regions and then, the histogram is computed for each region. Finally, the whole LBP histogram features are expressed as a concatenated sequence of histograms. After extracting LBP features, PCA is applied on them for dimension reduction.

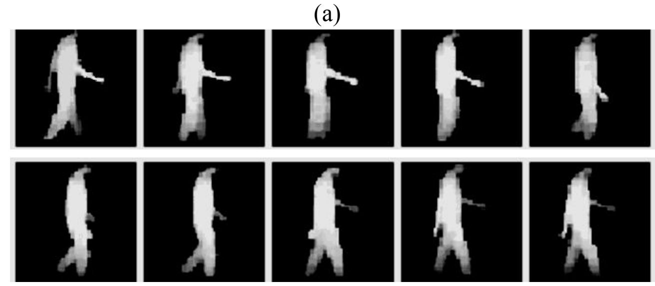
(a)

(b)

Figure 3. A sample (a) RGB and (b) depth frame from a walking activity obtained by the depth camera.



(a)



(b)

Figure 4. Generalized sequences of (a) binary silhouette-based walking, and (b) depth silhouette-based walking.

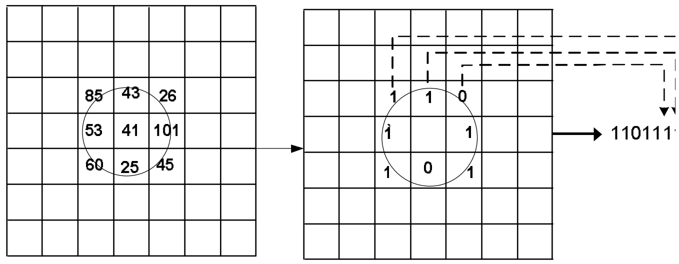


Figure 5. The LBP operator.

Depth silhouette-based HAR system can be applied in real time which can recognize different human activities automatically and save the activity information with time stamps as life-log. In next section, discussion an idea to predict human activity is done using spanning tree containing activity sequences with their frequencies.

After applying LBP-PCA, EICA is applied on the training activity image database. The first step of EICA is to apply PCA on the training depth silhouettes. The PC features are denoted is C. The basic idea of EICA is to represent a set of random observed variables using basis function where the components are statistically independent. If S is collection of basis images and C is collection of input images then the relation between X and S is modeled as

$$C = MS \tag{4}$$

$$C = W^{-1}S \tag{5}$$

$$WC = S \tag{6}$$

where M and W represent linear mixing and unmixing matrix respectively. Thus, EICA finds out the statistically independent basis images, eight of which are shown in Fig. 6.

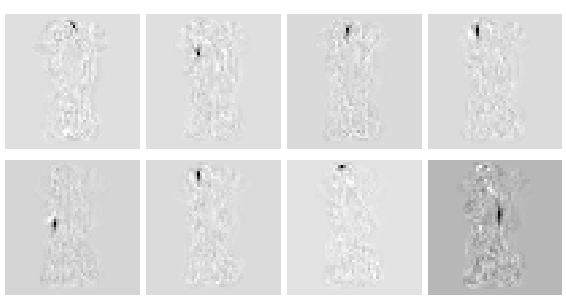


Figure 6. Eight EIC feature basis.

As Hidden Markov Model (HMM) is very efficient to model any time-sequential information [9], [10], [13], for training and recognition, one can apply it on the extracted depth silhouette features of different activities. To train each HMM for each activity, LBG vector quantization algorithm [26] is first performed on the training features from the human activity image sequences to obtain discrete symbols. Those obtained sequential symbols are then trained with HMMs to learn the proper model for each activity. When testing an activity image sequence, after vector quantization process, the obtained observation symbol sequence S is used to determine the proper model by means of highest likelihood computation of all trained q number of HMMs H as

$$HAR\_Decision = \arg\max_{i=1}^{q}((S \mid H_i)). \tag{7}$$

For HAR experiments, the proposed depth silhouette-based activity database was built for five activities (i.e., walking, lying, sitting, cooking, and neutral) where each clip contained variable length consecutive frames. In order to train and test each activity model, 15 and 40 image sequences were applied respectively. The proposed features of binary and depth silhouettes were extracted for the experiments and applied with HMM for training and recognition. The depth silhouette–based HAR approach achieved mean recognition rate of 96.50%, which is superior to binary silhouette-based approach that achieved mean recognition rate of 88.50%.

2.1 HAR Experiments

Since there is no standard database available regarding depth information-based human activities, new database was built for the proposed recognition system. The depth silhouette-based activity database was built for five activities (i.e., walking, lying, sitting, cooking, and neutral) where each clip contained variable length consecutive frames.

Table 1. Experimental results using depth silhouettes with HMM

Features	Activity	Recognition Rate with HMM	Mean
PCA	Walking	85%	85.50
	Lying	87.50	
	Sitting	82.50	
	Cooking	87.50	
	Neutral	85	
LBP	Walking	85	87.50
	Lying	87.50	
	Sitting	90	
	Cooking	87.50	
	Neutral	87.50	
LBP-PCA	Walking	90%	89.50
	Lying	87.50	
	Sitting	92.50	
	Cooking	90	
	Neutral	87.50	
EICA	Walking	92.50	93
	Lying	95	
	Sitting	95	
	Cooking	90	
	Neutral	92.50	
Augmented LBP-PCA and EICA	Walking	97.50	96.50
	Lying	97.50	
	Sitting	95	
	Cooking	97.50	
	Neutral	95	

A total of 15 sequences from each activity were used to build the feature space. In order to train and test each activity model, 15 and 40 image sequences were applied respectively. To apply on the discrete HMMs, the feature vectors were symbolized using the LBG codebook generation algorithm with the size of 32. Thus, the obtained discrete observation sequence from each training or testing video clip was used for training or recognition.

Five different feature extraction methods (i.e., PCA, LBP, LBP-PCA, EICA, and augmented LBP-PCA as well as EICA) were utilized to evaluate their performances on depth silhouette-based activity recognition. First of all, PCA-based global features were applied with HMM for HAR and it achieved 85.50% mean recognition rate, lowest recognition performance in HAR experiments. Later on, LBP was tried on the depth silhouettes that generated a better recognition performance than PCA (i.e., 87.50%). Then, LBP features were enhanced by PCA and tried for HAR, which achieved mean recognition rate of 89.50%. To obtain better HAR performance, EICA was then applied on the depth silhouettes that achieved 93% mean recognition rate, which shows significant improvement. Finally, the EICA features were augmented with LBP-PCA features that achieved the highest recognition performance than other approaches. Table 1 shows the depth silhouette-based experimental results where augmented LBP-PCA and EICA feature shows superior recognition rate (i.e., 96.50%) over the other methods.

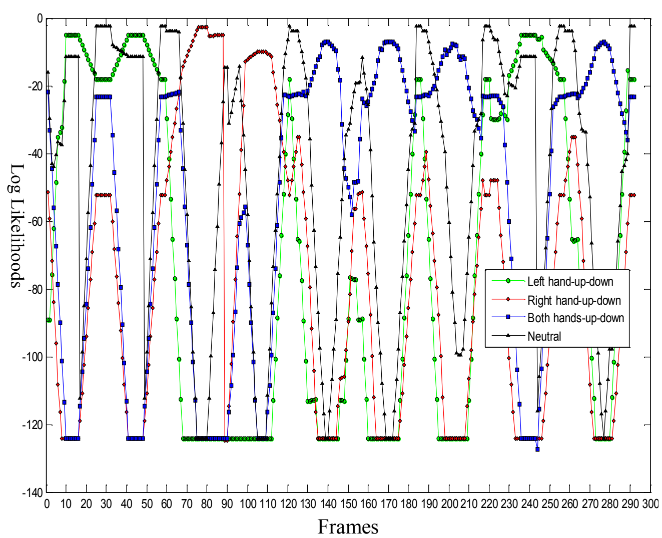


Figure 7. An exemplar Log Likelihood plot of four activities for a ten-length sliding window over 300 frames consisting of multiple activities.

3. LIFE-LOG CREATION

For creating the life-log in a fast way, simple prediction method can be used for predicting of activity in next few frames in a small video (may consist of multiple activities) i.e., if we have a small testing video consisting of multiple activities, we can split the video in some consecutive time-slides to recognize the slides individually. After recognizing some slides, if we see that the recognized activities are same than we can predict that next activity in next few frames should be the same activity and skip next few frames for recognition. It can be used for video segmentation with small errors. Thus, we can assign a long time activity annotation for life-log in this way which can be used later for activity prediction in long video.

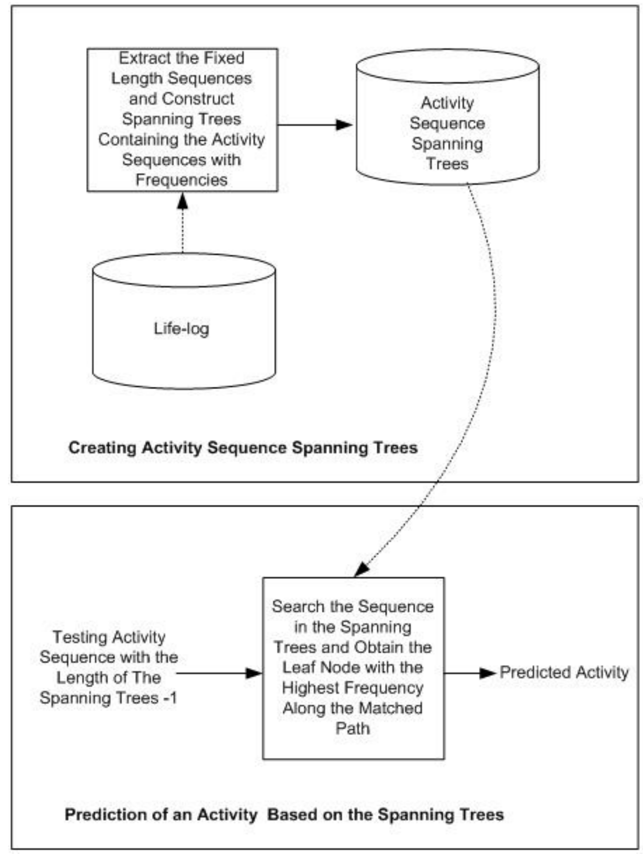


Figure 8. Processes to create activity sequence spanning tree to use them for HAP.

First of all, a sliding window can be run over a small video consisting of some activities and the likelihood of each activity for each window is measured to get the appropriate activity performed in that window. For instance, ten-length sliding window has been run and the likelihood of each activity is measured in 300 frames long testing sequence consists of three activities: namely walking, lying, and cooking. The likelihood result is plotted in the Fig. 7. Let us consider that a neutral activity (i.e., when the subject's movement is very little) is in between two consecutive activities. Thus, an HMM is also trained with neutral activity as each activity used for training and recognition here starts and ends with neutral activity. Consequently, at the beginning or end of each activity, the neutral activity likelihood becomes highest and for the rest of the cases, the maximum likelihoods are according to the corresponding activities. We can use this idea to segment the activities (although it is hard to obtain the exact boundary of activities in the testing sequence) and predict some frames based on the likelihoods without prior knowing the activity area in the testing sequence. Thus, the obtained activity sequence is Neutral → Walking→ Neutral → Walking→ Neutral → Lying→ Neutral → Lying→ Neutral → Cooking→ Neutral → Cooking→ Neutral → Cooking → Neutral → Left Hand-Up-Down→ Neutral → Cooking. Thus, it is possible to predict few future frames based on the likelihood comparison of current window slide. When any activity likelihood becomes highest for few slides, we can predict that definitely next few frames consist of the same activity. Thus, we can skip next few frames and continue the sliding window likelihood calculation after that.

Table 2 A sample activity sequence database for six days

Day	Activity Sequence
1	CCCCSSSSCCCCSSS
2	SSCCCCCLLLLLLL
3	SSSLLLLLLCCCCWW
4	NNNNNWWWCCCCLLL
5	CCCCCWWWWWCCCCC
6	WWWWWWWWWWWWSSS

4. HUMAN ACTIVITY PREDICTION USING ACTIVITY DATABASE

It is obvious that to start the prediction process of next human activity, a life-log of human activities for a specific period of every day for one month or two months is required. Each day's time-line is divided by a fixed number of minutes such as three minutes and tagged with what activities are performed in that period. Based, on the tagged information, spanning tree containing activity sequences can be built where each node represents an activity and the leaf nodes are used to predict using the upper level nodes. Fig. 8 shows the basic processes related to create activity sequence spanning tree to use them for HAP.

To show how the proposed algorithm works, an example database is used as Table 2 where the letters represent the activities such as W for walking, L for lying, S for sitting, N for neutral, and C for cooking. Now, execute a window through each row of the database and build spanning tree where the leaf node stores the frequency of patterns. Fig. 9 shows a length-5 sliding window approach to scan the activity database shown in table 2. Fig. 10 shows the activity spanning tree respectively.
As the tree is of five-length, activities in consecutive four times can be recognized and then try to predict the fifth activity to be performed in next time using the activity spanning tree.

```
CCCCSSSSCCCCSSS
CCCCS
 CCCSS
       ...
       ...
       ...
              CCSSS
SSCCCCCLLLLLLL
SSCCC
 SCCCC
       ...
       ...
       ...
              LLLLL
SSSLLLLLLCCCCWW
SSSLL
 SSLLL
...
...
...
```

Figure 9. Length-5 sliding window to scan the database in table 2.

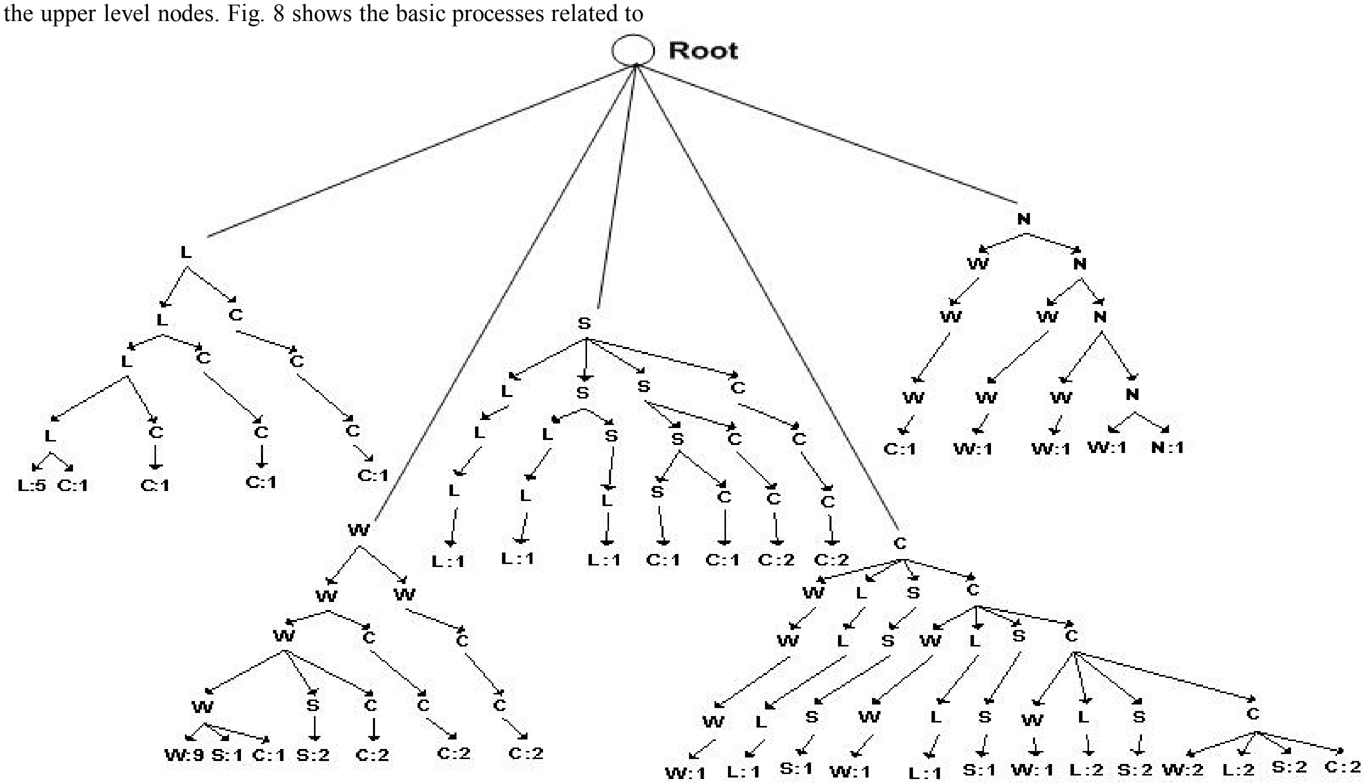


Figure 10. A Five-length spanning-tree.

Thus, start matching the 4-length consecutive tree pattern in the tree and if there is any match, one can find the leaf nodes in that path and their frequencies. If there is only one node, there will be a good possibility of that activity for being next activity. If there are many, one can notice that which one contains the highest frequency and then predict that activity as the next activity. Besides, a threshold can be applied on the highest frequency i.e., the highest frequency should be greater than the threshold percentage of the total frequency of all the leaf nodes. For instance, four consecutive activities as WWWW are recognized and now prediction of the next activity to be performed can be done. One can traverse the tree for the pattern and get a match in W sub-tree. But in that path, there are three leaf nodes as W with frequency of 9, S with frequency of 1 and C with frequency of 1. As frequency of W is much higher than that of others, next activity can be predicted as W (i.e., walking). Another instance is CCCC where one can see that in that path, all the leaf nodes - have the same frequency. So, the next activity cannot be predicted here and hence, continue recognition for the next time slot without prediction. Later on, prediction can tried again using the spanning tree.

As any actual activity database containing life-log is not available here, the experiments were performed using computer simulation. So, an activity life-log database was created by means of a program in Matlab 7.4.0 platform that randomly generates the data sequences. Two random numbers were applied here to generate the datasets. First one was to generate the random activity among five activities and the second one to determine the number of repetitions of the most recent activity. A total number of 60 activity sequences for 60 days are available in the database where each sequence was 2048 in length and consists of five distinct items (i.e., W, L, S, C, and E) to represent different activities. Using these 60 days data sequences the length-5 spanning tree was created for each activity. Later on, to test the proposed prediction approach, another datasets were created which consists of 6210 test patterns of length-5 to predict 5th activity. For every test pattern, matching a path of length-4 in the spanning tree were tried and based on the frequencies of the leaf nodes in the matched path, 5th activity was predicted with which 5th activity of the test pattern is compared to verify the prediction. Finally, using the proposed approach, the obtained prediction rate was 91.31%. Besides, during prediction, a threshold was applied to check the frequency of the leaf node containing the maximum frequency is greater than or equal to 80% of total frequency of the leaf nodes along the matched path or not. When any leaf node in that path satisfied the condition, prediction was continued, otherwise no predication for that sequence and the next testing sequence was used for prediction.

5. CONCLUSION

In this paper, an activity spanning tree-based novel approach has been proposed for human activity prediction based on the life-log created through human activity recognition using depth silhouette features with HMM. Here, the LBP and EIC-based local features of the depth silhouettes have been investigated and obtained superior recognition results over the traditional HAR approaches. Besides, the proposed efficient HAR approach can be successfully implemented in real time to create activity life-logs. A fast way of creating life-log has also been presented here. The spanning tree-based activity prediction is done based on computer simulation. Utilizing the proposed HAP approach, good prediction results were obtained using computer simulated activity sequence datasets. Hopefully, it can work well on actual life-log in different smart environments such as smart homes.

ACKNOWLEDGMENTS

This research was supported by the faculty research grant of Sungkyunkwan University, Republic of Korea.

REFERENCES

[1] Robertson N. and Reid I. 2006. A General Method for Human Activity Recognition in Video. *Computer Vision and Image Understanding*. 104, 2. 232 - 248.

[2] Kang H., Lee C.W., and Jung, K. 2004. Recognition-based gesture spotting in video games. *Pattern Recognition Letters*. 25, 1701-1714.

[3] Chen F.S., Fu C.M., and Huang, C.L. 2005. Hand gesture recognition using a real-time tracking method and Hidden Markov Models. *Image and Vision Computing*. 21, 745-758.

[4] Yamato J., Ohya, J. and Ishii, K. 1992. Recognizing human action in time-sequential images using hidden markov model. *in Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*. 379-385.

[5] Niu F. and Abdel-Mottaleb M. 2004. View-invariant human activity recognition based on shape and motion Features. *in Proceedings of the IEEE Sixth International Symposium on Multimedia Software Engineerin.*. 546-556.

[6] Niu F. and Abdel-Mottaleb M. 2005. HMM-based segmentation and recognition of human activities from video sequences. *in Proceedings of IEEE International Conference on Multimedia & Expo*. 804-807.

[7] Agarwal A. and Triggs B. 2006. Recovering 3D human pose from monocular images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 28, 44-58.

[8] Howe N. R. 2007. Silhouette lookup for monocular 3D pose tracking. *Image and Vision Computing*. 25, 331-341.

[9] Zhao G. and Pietikainen M. 2007. Dynamic Texture Recognition using Local Binary Patterns with An Application to Facial Expressions. *IEEE Trans. Pattern Anal. Mach. Intell*, 29, 6, 915-928.

[10] Uddin M. Z., Kim D. H., Kim J. T., and Kim T.-S. 2012. An Indoor Human Activity Recognition System for Smart Home Using Local Binary Pattern Features with Hidden Markov Models. *Indoor and Built Environment*, DOI:10.1177/1420326X12469734.

[11] Uddin M. Z. , Lee J. J., and Kim T.-S. 2008. Human Activity Recognition Using Independent Component Features from Depth Images. *in Proceedings of the 5th International Conference on Ubiquitous Healthcare.* 181-183.

[12] Kale A., Sundaresan A., Rajagopalan A., Cuntoor N., Chowdhury A. R., Kruger V., and Chellappa R. 2004. Identification of humans using gait. *IEEE Transactions on Image Processing.* 13, 1163-1173.

[13] Lam T. H. W., Lee R. S. T., and Zhang D. 2007. Human gait recognition by the fusion of motion and static spatio-temporal templates. *Pattern Recognition*, 40, 2563-2573.

[14] Munoz-Salinas R., Medina-Carnicer R., Madrid-Cuevas F. J., and Carmona-Poyato A. 2008. Depth silhouettes for

gesture recognition. *Pattern Recognition Letters*. 29, 319-329, 2008.

[15] Liu X. and Fujimura K. 2004. Hand gesture recognition using depth data. *in Proceedings of Sixth IEEE International Conference on Automatic Face and Gesture Recognition*. 529-534.

[16] Park J.-C., Kim S.-M., and Lee K.-H. 2006. 3D mesh construction from depth images with occlusion. *Lecture Notes in Computer Science*. 4261, 770-778.

[17] Hori T. and Aizawa K. 2003. Context-based video retrieval system for the life-log applications. *In Proceedings of the 5th ACM SIGMM international workshop on Multimedia information retrieval*. 31-38.

[18] Mann S. 1998. wearcam' (the wearable camera):personal imaging system for long-term use in wearable tether less computer-mediated reality and personal photo/video graphic memory prosthesis. *In Proceedings of IEEE ISWC98*, 124–131.

[19] Aizawa K., Ishijima K., and Shiina M. 2001. Summarizing wearable video. *In Proceedings of IEE ICIP,* 398–401.

[20] Shaikh M. A. M., Molla M.K.I., Hirose K. 2008. Automatic Life-Logging: A novel approach to sense real-world activities by environmental sound cues and common sense. *in Proceedings of 11th International Conference on Computer and Information Technology*. 294-299.

[21] Mori T., Takada A., Noguchi H., Harada T., and Sato T. 2005. Behavior Prediction Based on Daily-Life Record Database in Distributed Sensing Space. *in Proceedings of the International Conference on Intelligent Robots and Systems*. 1703 - 1709.

[22] Cook D.J., Youngblood M., Heierman E.O., Gopalratnam K., Rao S., Litvin A., Khawaja F. 2003. MavHome: an agent-based smart home. *in Proceedings of the First IEEE International Conference on Pervasive Computing and Communications*. 521 - 524.

[23] Jakkula V. R. and Cook D. J. 2001. Using temporal relations in smart environment data for activity prediction. *in Proceedings of the 24th International Conference on Machine Learning,* 2007.

[24] Iddan G. J. and Yahav G. 2001. 3D imaging in the studio (and elsewhere…). *in Proceedings of SPIE*. 4298, 48-55.

[25] Belhumeur P. N., Hespanha J. P., and Kriegman D. J. . Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 19, 7, 711–720.

[26] Linde Y., Buzo A., and Gray R. 1980. An algorithm for vector quantizer design. *IEEE Transaction on Communications*. 28, 1, 84-94.