Real Time Human Activity Recognition System based on Radon Transform

Z.A. Khan
Dept. of Electronics and Radio Engg.
Kyung Hee University, South Korea

W. Sohn
Dept. of Electronics and Radio Engg.
Kyung Hee University, South Korea

ABSTRACT
A real time human activity recognition system based on Radon transform (RT), Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) is presented. RT improves low frequency components and PCA provide global representation of these low frequency components in few eigenvectors. The proposed technique computes radon projections in different directions to obtain directional features of the images from video sequences. PCA is used to reduce the dimensions of radon shape features. LDA is applied on PCA features to provide better class separation. The aim is to develop a proficient recognition system in real time by the combination of local and global features. The dataset consisting of normal and abnormal activities is produced. Artificial Neural Nets (ANN) is used to recognize different human activities in real time. Experimental results show better recognition results for our system as compared to some state of the art methods.

General Terms
Image Processing, Pattern Recognition.

Keywords
Feature Extraction, Radon Transform, PCA, LDA, ANN.

1. INTRODUCTION
Human activity recognition remains a challenging and interesting research area in computer vision due to the applications in intelligent health care, video surveillance, human computer interaction and visual content retrieval systems. Video based real time human activity recognition is a complex and challenging task due to variation in people's appearance, illumination changes and the amount of data generated. The main step of real time human activity recognition system involves person detection, tracking and recognition.

The focus of this research is real time monitoring of elderly people's health and recognizing any abnormal activity e.g. fall recognition. We want to take advantage of recent advances in information technologies and reduce the burden on hospitals and economy. Also, it is proved that people suffering from long term diseases feel much better in their home environment than hospitals.

The human activity and posture recognition have been extensively studied during the past few years. A detailed survey of video based motion and activity recognition systems is discussed in [1, 2]. The projects successfully implemented for abnormal human activity recognition include, PROSAFE project which is a system to detect an accident based on the sensor data and generate an alarm for help [3]. A smart home system for elderly people's health care is implemented by using video and motion sensors to detect and monitor elderly people activities and generate an alert in emergency situation [4]. Most activity and posture recognition research is performed using 2-D approaches [5-9]. Mostly 2-D approaches are implemented for fixed view point. These can be differentiated into, interest point based approach [5, 6], space time based approach [7, 8] and motion template based approach [9]. A method to detect, learn and predict abnormal behavior based on adaptive background subtraction algorithm, appearance based model tracking, and N-ary tree classifier in presented in [10]. A method for human fall detection based on the eigenspace technique, integrated time motion images and support vector machine is discussed in [11].Technique for abnormal human activity recognition based on binary silhouette features using PCA is presented in [12, 13]. A real time system for human activity recognition by utilizing projection histograms of the detected blobs as features for postures classification is presented in [14]. A system called Pfinder to detect real time human detection and tracking by using multiclass statistical model for tracking is presented in [15]. A multisensor video based surveillance and monitoring system called VSAM is used to track people and vehicles is presented in [16]. A real time system called GHOST is developed for the detection of human body and its parts is presented in [17]. Several applications of RT in the field of image processing and computer vision are discussed in detail in [18]. An algorithm for curve detection based on RT using real, simulated data and performance evaluation is presented in [19]. Here, human motion is recognized by using human silhouette as a stick figure with two stage recognition. First, a model-driven approach is used to track human motions. In second stage, neural network classifier is used for motion classification of stick figures into three categories: walking, running and other motions [20].

Due to lack of publically available abnormal activity dataset, we produce the dataset consisting of forward fall, backward fall, chest pain, vomit (abnormal activities) and stand, sit (normal activities).

In this paper, we propose a real time human activity recognition system based on RT, PCA and LDA to provide shape based feature extraction and dimensions reduction. ANN is used to train and recognize six human activities: forward fall, backward fall, chest pain, vomit (abnormal activities) and stand, sit (Normal Activities). We used binary shape based features and extracted Region of interest (ROI) from silhouette by adaptive background subtraction from the sequence of images. The testing of our system on six different activities resulted in an average recognition rate of 78.0% as compared to state of the art

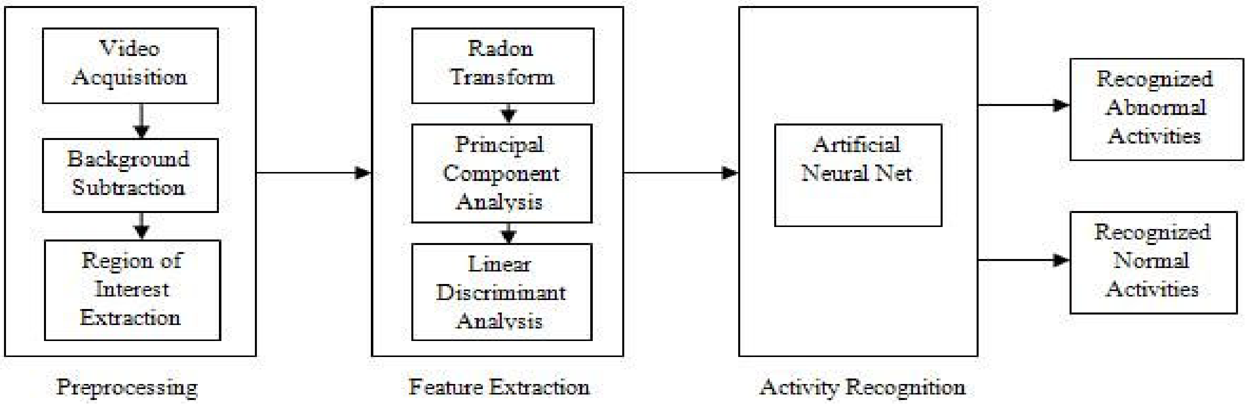

Fig 1: Our proposed system model for human activity recognition.

classifiers: Naive Bayes, AdaBoost and k-NN. This proves the usefulness of our proposed system.

The rest of the paper is organized as follows. In Section 2, the methods of our proposed system model are discussed including preprocessing (Background subtraction, ROI extraction), feature extraction (RT, PCA and LDA) and training and recognition of activities (ANN). Section 3, presents the experimental results for our proposed system model. The conclusion is presented in Section 4.

2. METHODOLOGY

Our recognition system consists of video and image preprocessing, Feature extraction by RT, PCA and LDA, and activity training and recognition by using ANN. Figure 1 shows the overall architecture of our human activity recognition system model.

2.1 Video Preprocessing

For the training purpose we adopt offline approach. Each video is converted to frames. Our research focus is elderly population having healthcare issues, therefore, it is assumed that a single person is performing a single activity at a time in an indoor environment, where background change is not severe. Even than there may be some illumination variations and shadows. Therefore, instead of using static background subtraction algorithm we selected non-parametric background modeling approach used by [21] to adapt to the continuous changing background due to illuminations and shadows. The background subtracted binary image of the moving person is extracted by the probability density function (pdf) as

$$P_r(x_t) = \frac{1}{N}\sum_{i=1}^{N}\prod_{j=1}^{d}\frac{1}{\sqrt{2\pi\sigma_j^2}}e^{-\frac{1}{2}\frac{\left(x_{tj}-x_{ij}\right)^2}{\sigma_j^2}} \qquad (1)$$

The pixel is regarded as a foreground pixel if $P_r(x_t) \leq th$ and a pixel is regarded as a background pixel if $P_r(x_t) > th$. After background subtraction and thresholding to get binary image, an ROI is extracted and is converted to the size of 50 x 50 pixels representing a moving person. To reduce noise and shadow effects after binarization, morphological filtering is performed. Each ROI image is normalized to represent in the form of a row vector such that the dimension of the vector is equal to number of pixels in the image. This means that a row vector has 1x2500 dimensions for each ROI image. Figure 2 shows selected images activity sequences which are converted to binary by using preprocessing and Figure 3 shows different preprocessing steps.

2.2 Radon Transform

In this paper, RT is proposed as a feature extractor due to its effectiveness in representing the shape characteristics. RT compact prominent shape and motion features into radon coefficients from 2-D binary images. RT is defined as the linear integration of function $f(x, y)$ of an image over all parallel lines, which corresponds to computing the projection sum of image intensity matrix along the specified direction [18]. It is written as

$$T_{Radon}(\rho,\theta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,\delta\,(x\cos\theta + y\sin\theta - \rho)dxdy\,, \qquad (2)$$

where, $\rho \in [-\infty,\infty]$ is perpendicular distance of the line $\rho = x\cos\theta + y\sin\theta$ from the origin. $\theta \in [0,\pi)$ is an inclination along the line where projections are computed and $\delta(\cdot)$ is the Dirac delta function that gives the projections along the line only. The RT is used to depict the view structure of binary image in 2-D form and to get directional features in the range of 0°–179°[19]. RT for binary image $f(x, y)$ can be defined in discrete form as consisting of summation of pixel intensities taken along the lines at specified directions. It is represented in mathematical form as

$$T_{Radon}(\rho,\theta) = \sum_x\sum_y f(x,y)\delta(x\cos\theta + y\sin\theta - \rho) \qquad (3)$$

The RT is applied to the dataset of 1,200 images from six abnormal activities for each person and it gives the feature vectors with decreased dimensions. We have 1x2500 dimensions for each image which are reduced to 1x180 by RT. The total dimensions for six activities are reduced to 1,200x180 features for each person. Figure 4 shows resized 50x50 ROI binary images along with their individual RT for selected key frame of the six normal and abnormal activities. The change in postures can be represented by the corresponding change in RT signal. Therefore, RT signal for a sequence of activity can be used to explain the particular activity.

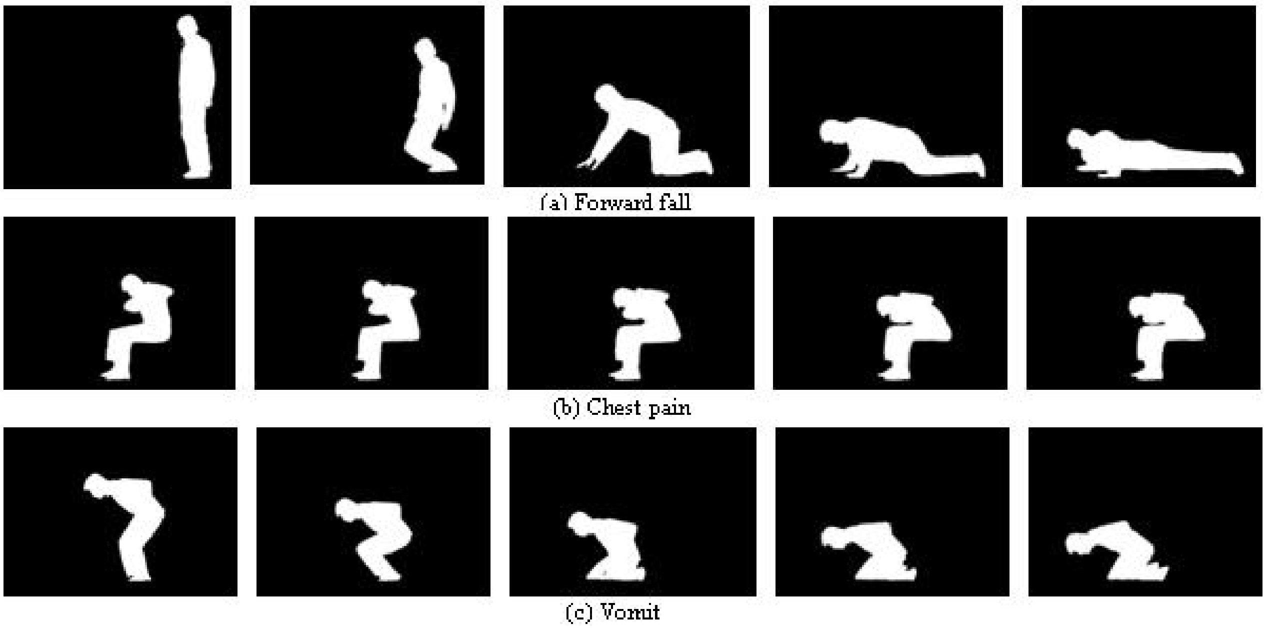

Fig 2: Selected few binary images for (a) Forward fall, (b) Chest pain, (c) Vomit activity sequence.

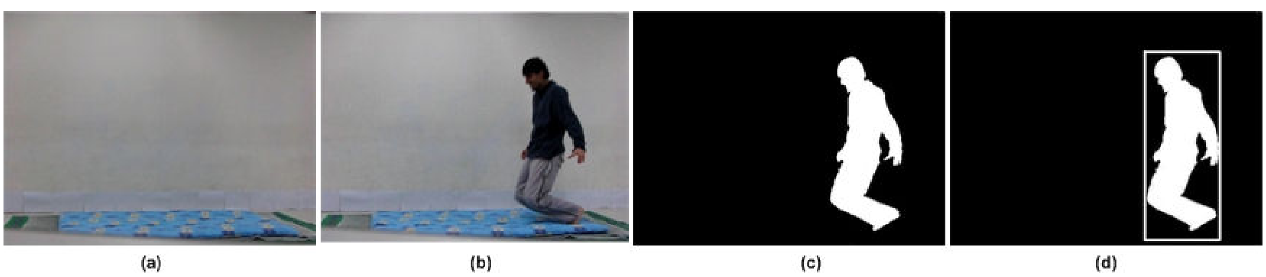

Fig 3: Preprocessing steps (a) Background image, (b) Image from forward fall sequence, (c) Binary image, (d) ROI rectangle.

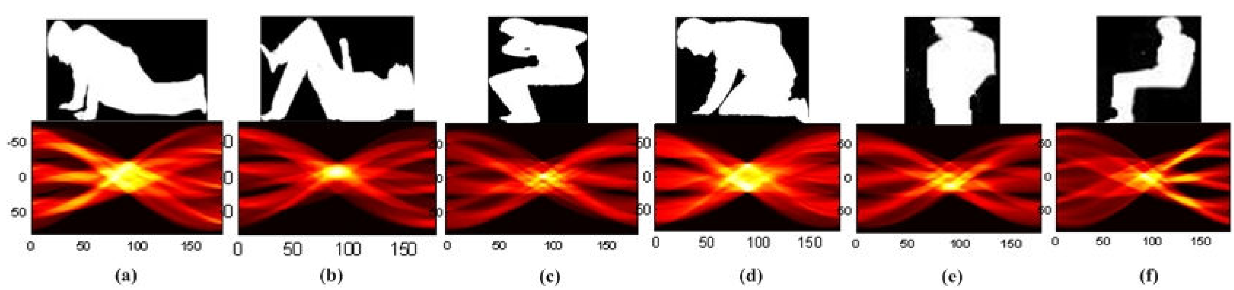

Fig 4: ROI binary images and the corresponding Radon transform representation for (a) Forward fall, (b) Backward fall, (c) Chest pain, (d) Vomit, (e) Stand, (f) Sit activity sequence.

2.3 Principal Component Analysis

PCA method is used to further reduce the dimensions of RT features. PCA reduce the dimensions of input space by projecting the data from a high-dimensional correlated space to low-dimensional uncorrelated space [22]. It computes principal component (PC) vector by calculating eigenvectors of the covariance data matrix C , and performs approximation based on linear transformation on those top eigenvectors that contribute to most variation in the data set. The first PC has the highest variance, second PC has the second highest variance and so on. The covariance data matrix is given as

$$C = \frac{1}{M} \sum_{n=1}^{M} \left(X_n X_n^T \right) \tag{4}$$

where, $X = \left[X_1, X_2, ..., X_M \right]^T$ represents the high dimensional dataset which has zero mean. The diagonal matrix is represented as

$$D = E^T CE \tag{5}$$

where, E represent an orthonormal eigenvectors matrix which have size of *pxm* and D represents a *mxm* diagonal matrix. The principal components of the shape vector are represented as

$$Y_n = X_n E_m \tag{6}$$

where, Y_n is the projection of nth image by PCA. E_m is the selection of top m eigenvectors of covariance matrix C . The application of PCA on RT features produce reduced feature vectors from 1x180 to 1x50 for each image, and the total of 1200x50 features for six activities. The selected top 50 PCs are shown in Figure 5. It is observed from the figure that most of the variation can be represented by top few PCs.

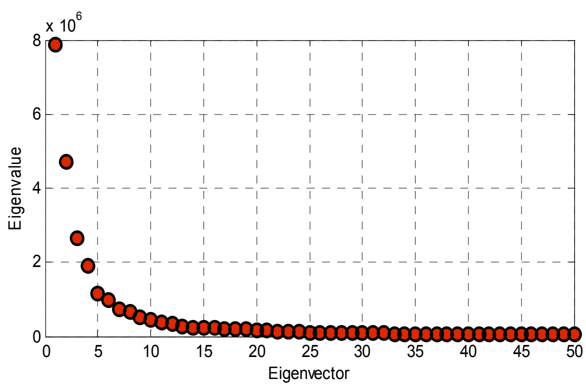

Fig 5: Top 50 Eigenvectors corresponding to the Eigenvectors.

2.4 Linear Discriminant Analysis

The LDA is a supervised dimensionality reduction technique used for pattern recognition and data analysis. LDA produces discrimination among different classes of data by maximizing the separation between different classes and minimizing the dispersion of data from the same class [23]. LDA implementation on Radon-PCA features resulted in a better recognition rate before feeding the features to the classifier.

For data matrix with m feature vectors $\mathbf{x} = \left[\mathbf{x}_1, ..., \mathbf{x}_m \right] \in \Re^n$, where $\mathbf{x}_i$ is the ith data point for $i = 1, ..., m$. The between class S_B , within class S_W scatter matrices are defined as

$$S_B = \frac{1}{n} \sum_{i=1}^{c} n_i \left(\boldsymbol{\mu}_i - \boldsymbol{\mu} \right) \left(\boldsymbol{\mu}_i - \boldsymbol{\mu} \right)^T \tag{7}$$

$$S_W = \frac{1}{n} \sum_{i=1}^{c} \sum_{x \in X_i} \left(\mathbf{x} - \boldsymbol{\mu}_i \right) \left(\mathbf{x} - \boldsymbol{\mu}_i \right)^T \tag{8}$$

where n represents the total data points, n_i is the data points from ith class, $\mathbf{x}$ is a vector of specific class, c is the number of classes, $\mathbf{x}_i$ is the set of data points of ith class. $\boldsymbol{\mu}_i$ is the centroid of ith class, and $\boldsymbol{\mu}$ is the global centroid. The S_W gives the degree of scatter within classes as a summation of the covariance matrices of each class, where as S_B gives the degree of scatter between classes as the covariance matrix of means of each class respectively. The optimal discrimination projection matrix $\mathbf{O}_{opt}$ in the projection space can be computed by solving the optimization problem as

$$\mathbf{O}_{opt} = \arg\max_{\mathbf{O}} \frac{\mathbf{O}^T S_B \mathbf{O}}{\mathbf{O}^T S_W \mathbf{O}} \tag{9}$$

This optimization problem corresponding to $c-1$ largest eigenvalues λ , $\mathbf{O}_{opt}$ can be solved as in [23] by the generalized eigenvalue problem as

$$S_B \mathbf{o}_i = \lambda_i S_W \mathbf{o}_i \tag{10}$$

When S_W is non singular, the problem reduces to eigenvalue problem represented as

$$S_W^{-1} S_B \mathbf{o}_i = \lambda_i \mathbf{o}_i , \tag{11}$$

LDA implementation on PCA results in 1x5 feature vector for each image of a sequence of activity. The 3-D representation of LDA on PCA-Radon features is shown in Figure 6. Forward fall and vomit activities have increased within class differences and decreased between class differences. Within class differences are due to increased variations in each activity sequences and between class differences are due to some similar poses in the two activities respectively.

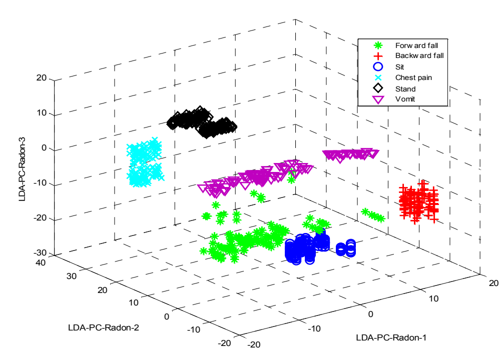

Fig 6: The 3-D plot of LDA features for the six activities.

2.5 Artificial Neural Net

Our goal is to implement an efficient model for real time unconstrained activity tracking and recognition. ANN proved to be effective for this purpose. Artificial Neural network (ANN) is an interesting approach for the study of time-varying data [20]. We selected MLP (Multi layer Perceptron) Neural Network (NN) with feed-forward and back propagation learning algorithm for real time human activity recognition. Back propagation (BP) algorithm is used for the optimization of MLP. A single neuron in MLP can be represented as

$$y_k = \varphi_k \sum_{i=1}^{m} \mathbf{w_i x_i} \tag{12}$$

where, $\mathrm{x_i}$ is data input to NN, $\mathrm{w_i}$ represents synaptic weights between ith neuron of previous layer and kth neuron of the next layer and φ_k represents the activation function. For performance optimization, MLP with different number of layers and neurons is tested. Table 1 shows the MLP training parameters used. A three layer BP network is adopted as shown in Figure 7. There are five neurons in the input layer as features provided by the LDA. Output layer has six neurons, whose output gives the patterns for the six classes of activities. After experimenting with different values, ten neurons are selected in the hidden layer. For an unknown sequence of real time data, we use a sliding window of length 20 frames over a sequence and classify the activity based on the output of NN.

Table 1. MLP training parameters

Number of layers	1 input layer, 1 hidden layer, 1 output layer
Number of neurons in different layers	Input:5, Hidden:10, Output:6
Initial weights	Random between -1 and +1
Activation functions for hidden and output layers	Log sigmoid
Training parameters for learning	Back propagation
Number of iterations	2560
Acceptable mean square error	0.001

3. EXPERIMENTAL RESULTS

In this research, we recognized normal and abnormal human activities in real time. The dataset consists of six activities; four abnormal activities: forward fall, backward fall, chest pain, vomit and two normal activities: stand, sit. We used SONY Cyber-shot HX5V digital camera to capture the activity videos with frame size of 320x240 at 20 fps (frame per second) and stored in AVI format. Experiments are performed with MATLAB version 8.0 on Intel Core2 Duo 3GHz processor, 2GB RAM running Microsoft Windows XP professional operating system.

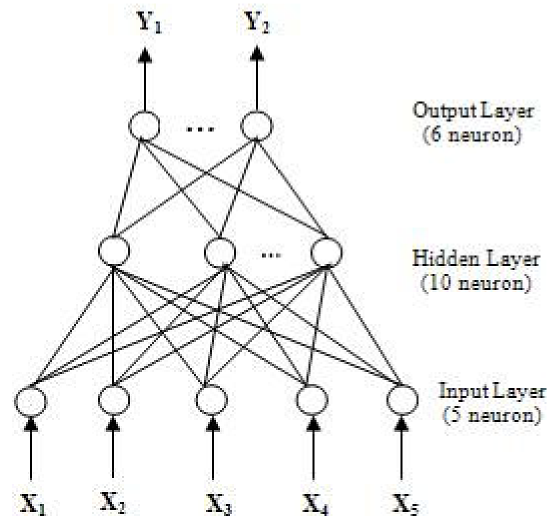

Fig 7: ANN BP network model.

During training each activity sequence is performed 10 times by 3 different persons. After preprocessing, video frames are converted to binary ROI image sequences and 20 key frames are selected to represent an activity sequence. RT, PCA and LDA are used for feature extraction. RT projects the data on 180 dimensions and provided 1x180 features for each image. PCA is applied to further reduce the dimensions to 1x50 and LDA implemented on Radon-PCA features produced 1x5 feature vector for each image of the activity sequence. After feature extraction and dimensions reduction, we used ANN for training. LDA features are fed to the ANN for training and activity recognition. The features obtained after LDA are used to train the corresponding ANNs for the six classes of activities.

For recognition, we slide a window of length 20 frames over a sequence of frame features with a frame difference of 5 and no overlapping between the frames and recognize the activity which is represented by the sequence in the sliding window. The difference between the frames is suggested to increase the efficiency of our system working in real time. Key frames are manually selected for each activity sequence during training. By carefully looking at the pattern of frames for each activity sequence, we selected the size of window as 20 frames. The skip or difference of 5 frames is used between the consecutive frames to represent an activity sequence by a reduced set of frames. Our proposed system model using ANN classifier resulted in an average recognition rate of 78.0% for six activities in real time. We have tested a variety of classifiers including Naive Bayes, AdaBoost and k-NN on our system. TABLE 2 shows the recognition result for Naive Bayes, TABLE 3 shows the recognition result for AdaBoost, TABLE 4 shows the recognition result for k-NN and TABLE 5 shows the recognition result for MLP ANN classifier. The results show that the use of ANN classifier gives the highest recognition rate, therefore we have selected MLP ANN classifier for our proposed system model.

Table 2. Confusion matrix for normal and abnormal activities using Naive Bayes

Activity	Normal	Abnormal
Normal	61.50%	38.50%
Abnormal	31.0%	69.0%
Mean Rec. Rate	65.25%	

Table 3. Confusion matrix for normal and abnormal activities using AdaBoost

Activity	Normal	Abnormal
Normal	57.0%	43.0%
Abnormal	21.50%	78.50%
Mean Rec. Rate	67.75%	

Table 4. Confusion matrix for normal and abnormal activities using k-NN

Activity	Normal	Abnormal
Normal	62.0%	38.0%
Abnormal	18.50%	81.50%
Mean Rec. Rate	71.75%	

Table 5. Confusion matrix for normal and abnormal activities using MLP neural network

Activity	Normal	Abnormal
Normal	67.50%	32.50%
Abnormal	11.50%	88.50%
Mean Rec. Rate	78.0%	

4. CONCLUSION

In this research, a novel real time feature representation and activity recognition system for human activity recognition is proposed. The key frames selected to represent a sequence of activity, significantly reduced the computational complexity. The feature extraction and dimension reduction algorithms: RT, PCA and LDA implementation on the binary shape based ROI features resulted in algorithm optimization. This results in increased efficiency for our human activity recognition system. The recognition process is performed by using Naive Bayes, AdaBoost and k-NN and neural network classifiers. The ANN classifier shows highest recognition results (78%) for real time human activity recognition. The proposed approach can be applied in real life applications such as elderly healthcare monitoring.

5. REFERENCES

[1] Moeslund, T. B., Hilton, A., and Kruger, V., "A survey of advances in vision-based human motion capture and analysis," *Computer Vision and Image Understanding*, vol. 104, pp. 90-126, 2006.

[2] Nevatia, G. R., and Cohen, I., "Event Detection and Analysis from Video Streams" *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol.23, Aug. 2001.

[3] Chan, M., Campo, E., and Estève, D. "PROSAFE, a multisensory remote monitoring system for the elderly or the handicapped," in Proc. *1st Intern. Conf. On Smart homes & health Telematics, Independent living for persons with disabilities and elderly people*, Paris, pp. 89-95, 2003.

[4] Skubic, M., Alexander, G., Popescu, M., Rantz, M., Keller, J.: A smart home application to eldercare: Current status and lessons learned: Technol. Health Care;17(3):183–201, 2009.

[5] Ke, Y., Sukthankar, R., Hebert, M., "Efficient Visual Event Detection using Volumetric Features", *ICCV*, pp.166-173, 2005.

[6] Laptev, I., Lindeberg, T., "Space-time interest points", *ICCV*, pp. 432-439, 2003.

[7] Blank, M., Gorelick, L., Shechtman, E., Irani, M, Basri, R., "Actions as space-time shapes", *ICCV*, pp. 1395-1402, 2005.

[8] Yilmaz, A., Shah, M., "Actions sketch: a novel action representation", *CVPR*, pp. 984-989, 2005.

[9] Bobick, A.F., Davis, J.W., "The recognition of human movement using temporal templates", *PAMI* 23(3), pp. 257-267, 2001.

[10] Duque, D., Santos, H., Cortez, P., "The OBSERVER: An Intelligent and Automated Video Surveillance System," Lecture Notes in Computer Science, Springer Series, vol. 4142, pp. 898-909, 2006

[11] Foroughi, H., Yazdi, H.S., Pourreza, H., and Javidi, M. "An eigenspace-based approach for human fall detection using integrated time motion image and multi-class support vector machine," *4th International Conference on Intelligent Computer Communication and Processing*, pp. 83-90 , 2008.

[12] Niu, F., Abdel-Mottaleb, M. "View Invariant Human Activity Recognition Based on Shape and Motion Features", in proc. *IEEE Symposium on Multimedia software engineering*, pp. 546-556, 2004.

[13] Niu, F., Abdel-Mottaleb, M. "HMM-based segmentation and recognition of human activities from video sequences", in proc. *IEEE international conference on multimedia & expo*, pp 804–807, 2005.

[14] Haritaoglu, I., Harwood D. and Davis, L. S., "W4: Realtime surveillance of people and their actions", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):809–830, August 2000.

[15] Wren, C.R., Azarbayejani, A., Darrell, T., Pentland, A.P., "Pfinder: real-time tracking of human body", *IEEE*

Transactions Pattern on Analysis and Machine Intelligence, 19(7):780 – 785, 1997.

[16] Collins, R.T., Lipton A.J, Kanade, T., Fujiyoshi, H., Duggins, D., Tsin, Y., Tolliver, D., Enomoto, N., Hasegawa O., "A system for video surveillance and monitoring: VSAM", Technical report CMU-RI-TR-00-12, Robotics Institute, Carnegie Mellon University, Pottsburgh PA, 2000.

[17] Haritaoglu, I., Harwood, D., Davis, L.S., "Ghost: A Human Body Part Labeling System Using Silhouettes", *14th International Conference on Face and Gesture Recognition*, April 1998.

[18] Deans, S.R., Applications of the Radon Transform. New York: Wiley Interscience Publications, 1983.

[19] Kobasyar, M., Rusyn, B. "The Radon transform application for accurate and efficient curve," *International Conference on Modern Problems of Radio Engineering, Telecommunications and Computer Science*, pp.223-224, 2004.

[20] Guo, Y., Xu, G., and Tsuji, S., "Understanding human motion patterns", in Proc. *Intl. Conf. on Pattern Recognition*, pp. 325-329, 1994.

[21] Elgammal, A., Harwood, D., and Davis, L. "Non-parametric model for background subtraction," *6th European Conference on Computer Vision*, Dublin, Ireland, pp. 751-767, 2000.

[22] Jolliffe, I.T., "Principal component analysis," Springer Series in Statistics, 2nd ed., Springer, 2002.

[23] Fukunaga, K., Introduction to Statistical Pattern Recognition, 2nd ed., Acamemic Press Professional, 1990.