IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Real Time Action Recognition with Instant Characteristics

[1]Sahar Fatima, [2]Jyoti P. Singh

[1]Rresearch Scholar, [2]Assistant Professor,
[1]Department of Electronics Engineering,
[1]Uma Nath Institute Of Engineering & Technology (V.B.S. Purvanchal University), Jaunpur, India

Abstract: In this research work, we proposed a most effective noble approach for Human action recognition in real-time environments. We recognize several distinct dynamic human activity actions using kinect. A 3D skeleton data is processed from real-time video gesture to sequence of frames and getter skeleton joints (Energy Joints), orientation, rotations of joint angles from random selected number of frames. Thus after extracting the set of frames we implements classification techniques Principal Component Analysis (PCA), Artificial Neural Network (ANN) and Deep Neural Network (DNN) respectively with some variants for classify our all different gesture models. However we conclude that use very less number of frame (10-15%) for train our system efficiently from entire set of gesture frames. After successfully completion of our classification methods we got an excellent overall accuracy 94%, 96% and 98% respectively. We finally observe that our proposed system is more useful then comparing to other existing system, therefore this model is best suitable for real-time application such as in video games for player action/gesture recognition.

I. INTRODUCTION

Since last decades Robotics has been known as a science which creates the Automatic interpretation between human activities and perception, and integrating them by using controls, machines, and electronics with the help of digital system that is computers. Recently Robotics has been used in more areas such as field robotics, service robotics, or human enlarge. The idea that has been proposed in this paper targets to give to robot the skill to execute a task, identifying the human activities and learning with their movements. Many researchers has been demonstrated related work in large scale since the 1050 in these areas such as robotics research and computer vision. The goal of computer vision is to extract the information from a particular scene and design it. The recognition task can be classified in to three stages: feature representation, feature extraction and action classification. This paper aims to present a humanoid robot having the skill of observing, training and representing actions operated by humans for generating new skills. This system has been implemented in such a way that it will distinguish different actions along with grants the permission to robot for regenerating their actions. Being able to identify and recognize human actions is most essential for many applications such as smart homes and assistive robots. Human robot interaction (HRI) has been implemented in the view of real world applications. Human activity recognition is an important functionality in any intelligent system designed to support human daily activities. The measurement of image or camera motion and more on the labeling of the action taking place in the scene. We design a new action feature descriptor Accumulated Motion Energy (AME) is then proposed to perform informative frame selection, which is able to remove noisy frames and reduce computational cost.

II. PREVIOUS WORK

Many research works has been developed till now related to recognizing the actions of human. Human actions can be recognized in the form of skeleton [9], silhouettes [6], and in the form of images [7]. Visual surveillance technique is used for identifying packages of human actions [2]. Many papers have used different techniques to recognize human actions such as hierarchical probabilistic approach [3], multi- modality representation of joints [4], HDP-HMM which is multi-level [5], Eigen-joint based method [8] using NBNN classifier. All the mentioned work is not reliable for real world application. Researchers have been explored different varieties of compact representation for human actions.

1. Real-time Human Action Recognition From Motion Capture Data

In this paper explain that the recognition of the human actions are the most important task in various vision application included the human computer interaction, video surveillance etc. Traditionally recognized human actions or depth video surveillance cameras are used for this purpose. Captured motion provides the accurate motions information of body joint in 3D space. The skeleton joints co-ordinate of user provided by the motion system used to analyze dynamics of action performed. The temporal variances to each joint of skeleton and weighted variance serves as the feature for the classification. These feature can extracted rapidly and the suitable for the real time recognition. They show the performance of proposed method by using correlation based support vector machine on multimodal action detection 3D dataset. This approach independent of duration of action and starting point of action. The weighted time variance for embedding temporal information help to improved classification for the same type of action like as 'stand up', 'sit down' and 'stand up', sit down actions. The propose algorithm evaluated on MO Cap data of multi-model human activity dataset, and show the better performance.

2. Action Recognition from Motion Capture Data using Meta-cognitive RBF Network Classifier

In this paper action recognition play as an important part in various application, including the smart homes and assistive robotics. They propose an algorithm for the recognition of human action using captured data in motion form. Captured data provides the accurate dimensional position of joints which are constituted in human skeleton and modeled the movement of skeleton joints temporally order to classify the actions. The skeletal joints in each frame of action sequence are represented as 129 dimensional vectors; of each component make a 3D angle by separate joint with a fix point on skeleton. Finally the videos are presented as histogram over codebook obtains from every action sequences. Along this the temporal variances of skeletal joints are used as additional features. The proposed algorithm evaluated on MO Cap data of Multi modal action dataset by using PBL-McRBFN classifier and performs better than state of approaches.

3. Action Recognition Using Local Joints Structure and Histograms of 3D Joints

In this paper present a method for the human action recognition by using local joints structure and 3D joints of histogram. The global features of joints ignore the local structure information's of human body joints, which also essential for the accurate activity recognition. For this problem the proposed joints structure feature as complement and combined global and local feature for the posture description in these method. Linear discriminate analysis method used to reduce feature dimension and the k-mean clustering utilized to generate the code word. Finally code word treated as the discrete symbol for trained hidden markov model which used for action recognition. In future work aimed to explore the other features that's describe spatial relationship between the skeleton joint and try to utilized the other classification technique that's are better in the modeling of temporal information.

4. Human Action Recognition based on Motion Capture Information using Fuzzy Convolution Neural Networks

In this paper the proposed approach for the human action recognition is based on the captured motion information by using fuzzy convolution neural network, and the tracking information of human joint used to compute temporal variations of the displacement between joints at the time of execution of action. The fuzzy membership function design to emphasize discriminative position associated with the every action considered by feature extraction. Temporal variations of the membership value associated by these membership functions considered as feature representation for the action recognition. A convolution neural network capable of the recognized local pattern in input data which trained to human action recognition from the local pattern in feature representation. The future work involve extensive experiment on MOCAP dataset with the predicted human joints information to identify set of features which suitable for the action recognition across multiple dataset.

III. RESEARCH METHODOLOGY

We have to design a new action feature descripted by adopting differences of the skeleton joint in both spatial domain and temporal domains to model dynamics of each joint and configuration of the different joint. In data we have apply principal component analysis (PCA) to finding the eigen joints by reducing the noises and redundancy on joint differences. Thus after extracting the set of frames we implements classification techniques Principal Component Analysis (PCA), Artificial Neural Network (ANN) and Deep Neural Network (DNN) respectively with some variants for classify our all different gesture models. Accordance with the image classification we have to avoid quantization of the frames compute video to class distances and descriptor, instead of the video to video distances in addition most powerful method performed activity recognition by operated whole video sequences. The scope of these work is widely applied many number of the real world application like as human computer interaction, health issues, video surveillances and video search based on contents. The entire work mainly concentrate on video sequences of activities which captured by the RGB cameras.

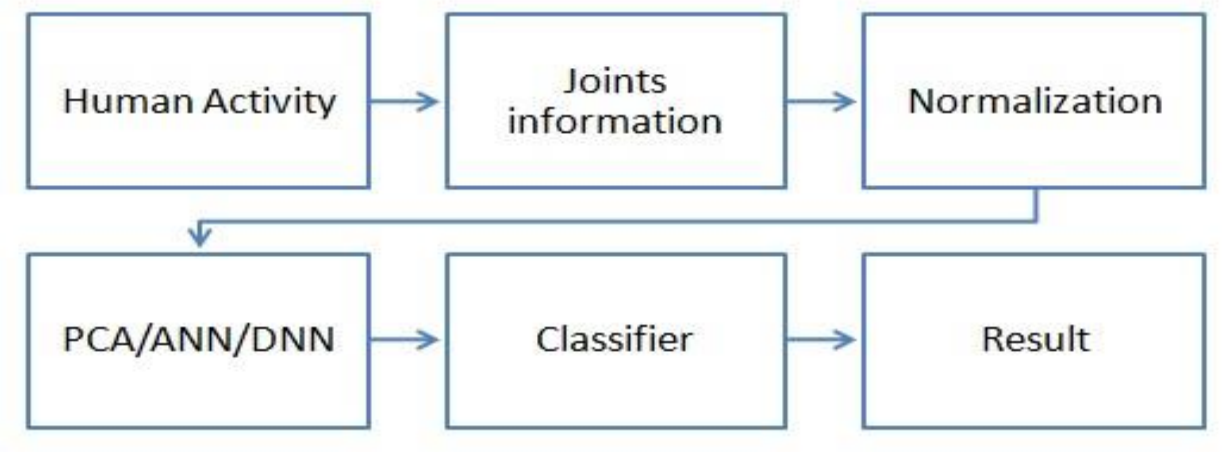


Fig.1.Graphical representation of proposed methodology

1. Human Activity

Human activity recognition is an important functionality in any intelligent system designed to support human daily activities. the measurement of image or camera motion and more on the labeling of the action taking place in the scene.

Moreover the several activity is performed in real-time environment shown in Fig 2.2 where subject person is perform daily routine work such as Brushing teeth, working on computer etc. while our kinect is capturing the activity with help of joint angles information. During our experiment we use all male dataset in kitchen and corridor environments. We are assuming that a kinect is mounted on wall in the front of subject user to capture the activity perform by user. However our kinect start recording the activity of humans from initial starting to until activity is not completed. Thus we have set of video image data for each individual activity. We have currently 25 joint angles for human body.

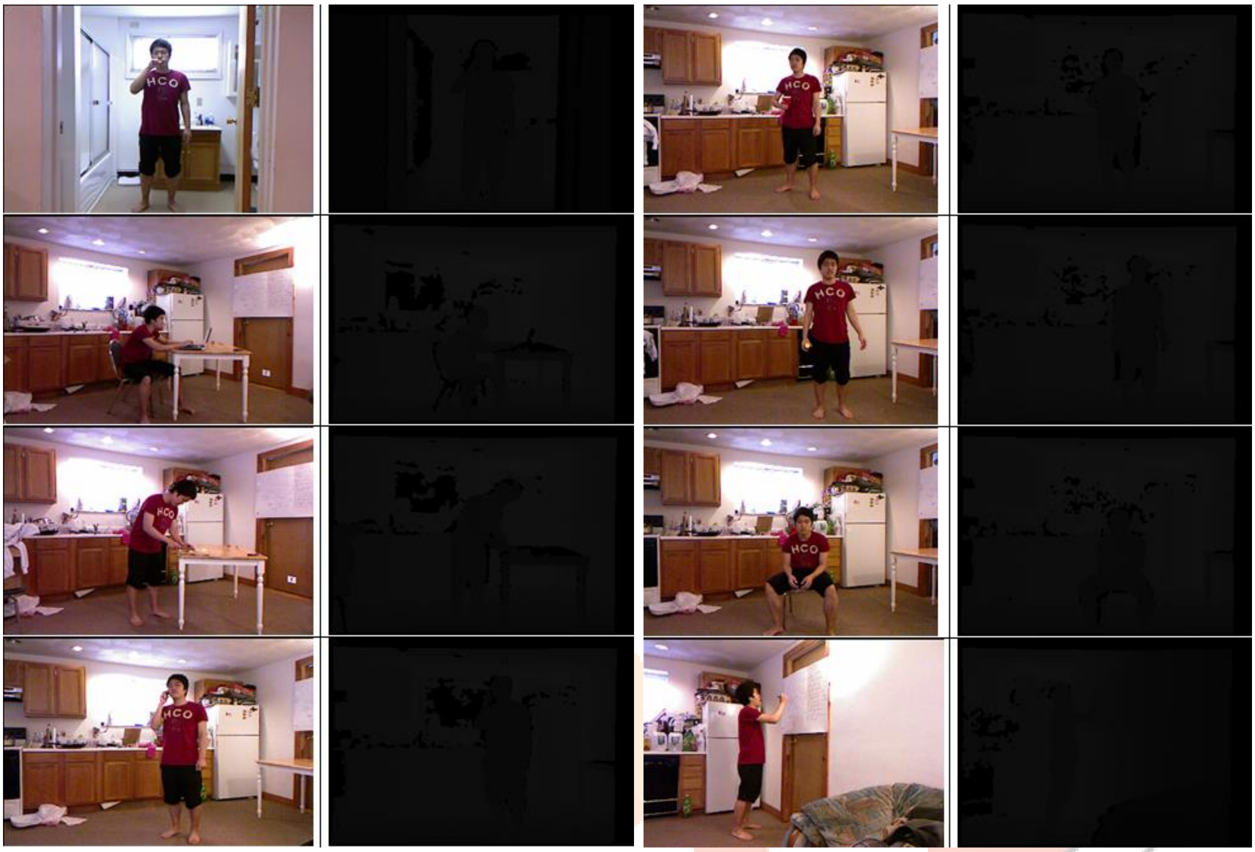


Fig.2.Corresponding activity and depth image of dataset top to down and left to right (1-Brushing teeth. 2-Working on computer. 3- Cooking (chopping). 4- Talking on the phone. 5- Drinking water. 6- Opening pill container. 7- Talking on couch. 8- Writing on whiteboard)

2. Joints Information from kinect

In data total number of joints is 15. Where 11 joints have both the joint orientation value and the joint position value, and 4 have only positions value. The values of orientation and position are in following format

Frame = $P_o(1)$, $P_p(1)$, $P_o(2)$, $P_p(2)$............$P_o(11)$, $P_p(11)$....$P_p(15)$

Po = orientation values of the joints which are 3×3 matrixstored and follows by

=0, 1, 2, 3, 4, 5, 6, 7, 8, C_o

C_o is the Boolean confidence values which are o or 1

3. Differences between joint positions and orientations

The joints positions value are more accurate than the joints angle. The explanation of this point we have to compare two methods of computing a hand position in game. First method is take the position of hand joint in API. The other method is to take torso position and orientation, both shoulder and elbow joints angle typically the result of hand position will different from one returned by API because avatar will have will different lengths than model used in skelton API (specially consider previous mentioned point that skelton allow body segments length to vary the time where avatar model in game have fix length).the position will be match if segment length match exactly same at all time.

4. Image Acquisition

In this section we have to describe that the how static activity image is obtained in the real time. For the acquisition of the activity gesture , Microsoft kinect used. The following steps are in image acquisition:

1.Color image frames extraction: Out of various resolution available we have to obtained only 640x480 resolution images.

2. Depth image frame extraction: It is also obtained at the resolutions of 640x480 images.

3. Skelton data used to track and extract the activity of the user.

5. Background subtraction

For calculation of distance to the pixel from kinect sensor actual value of the pixel are shifted 3 places in depth frame (depth point) to right[22]. The below statements are in C language to show the work.

Depth= depth point << 3

Skelton streams is the most important features of Microsoft kinect. Its provide the position and location of the persons whether they tracked or not. The skeletons which are not tracked are given to zero value returned. Kinect tracked the skelton in two modes:

- Seated mode
- Default mode
- Default mode tracked the all 15 joints and in seated mode the person could be tracked only the upper body part having 10 joint positions is obtained. The most important uses of the depth map is background subtraction from the frames. That pixel which doesn't belonging to region of the intrestare subtracted. So only that pixel remains which have redundant to zero intensity value.

6. Normalization

Normalization is basically a process to change the range of the pixels intensity value. Normalization is also called the histogram stretching or contrast stretching. In general field of data processing such as the image processing, it refers to as the dynamic range expansion. The purpose of the dynamic range expansion in various application usually to bring image or the other type of the signal into the range which are normal or more familiar to sense hence the terms normalization. Often motivation is to achieve consistency in the dynamic range for set of data, images or signal to avoid the mental distractions or fatigue.

Normalization transform the n-dimensional grayscale images with the intensity values in range (min, max) into the new image with the intensity value in range (new min, new max).

Linear normalization technique to scale each elements in the range [-1, +1]. The benefit of the normalization is to reduce the intra – class variation of same action performed by the different subjects. The linear normalization of the gray scale image performed by the formula,

$$I_n = (I - I_{min}) \frac{(new \max - new \min)}{\max - min} + \text{new min}$$

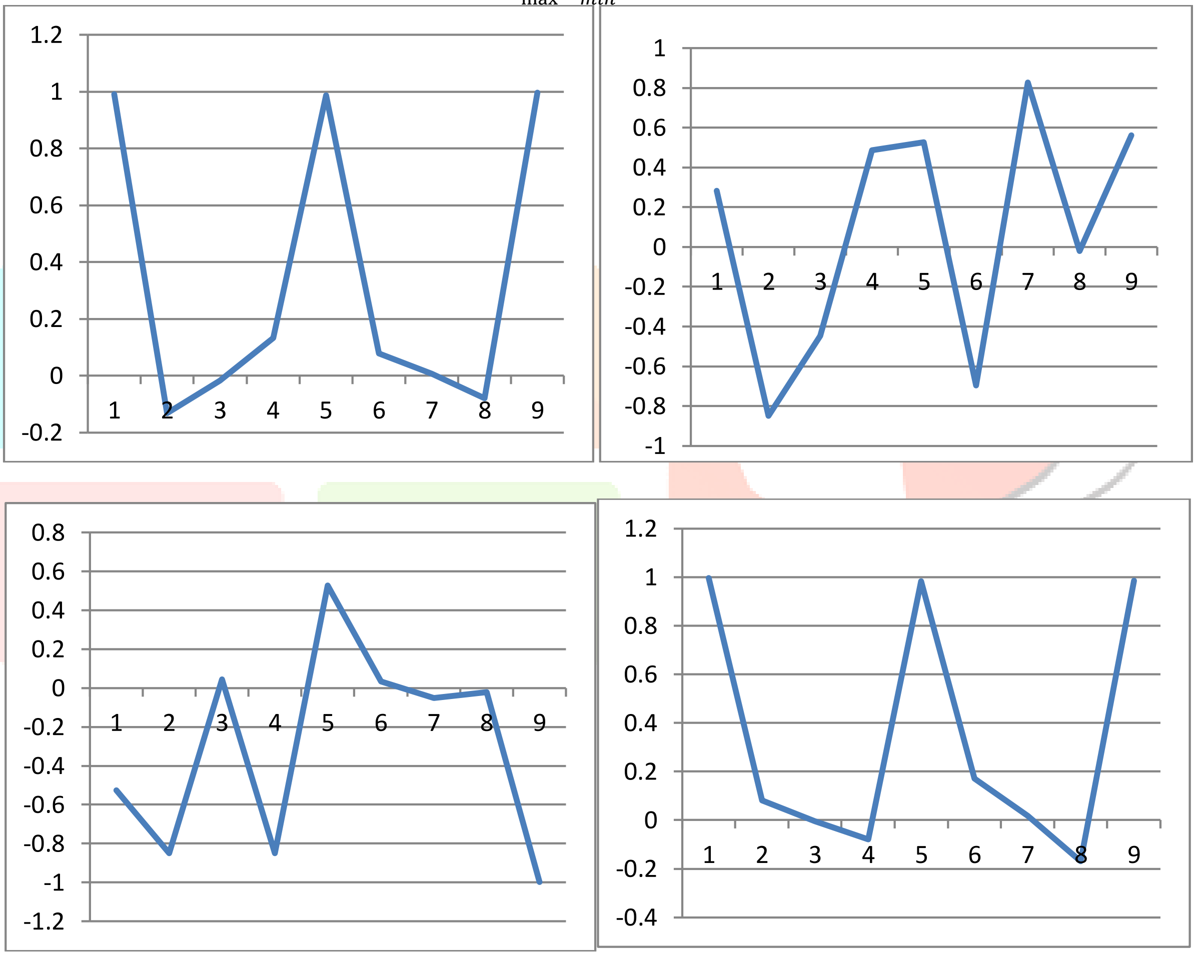
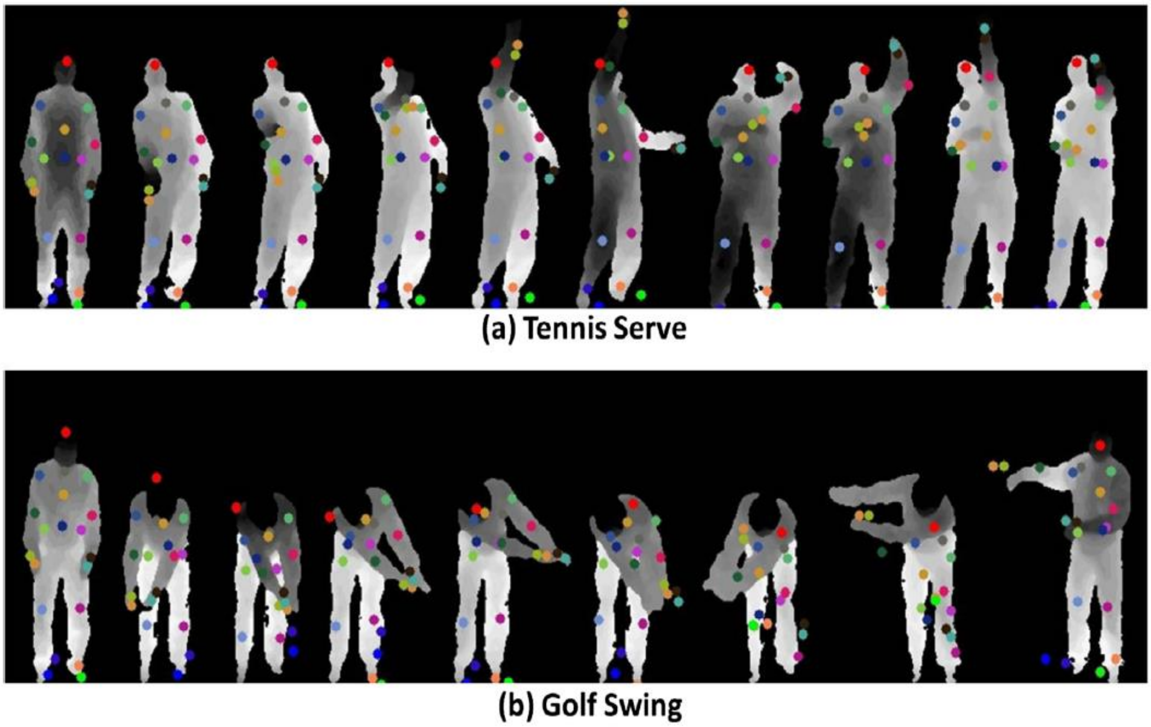


Fig.3.Normalized data joints from top to down and left to right(head, left shoulder, right shoulder, left hip)

For example if we have to extract 20 skeletal joints in the each frame then the F_{norm} have the dimension of 2970. As the skelton joints have already high level information's recover from the depth maps, these large dimensions may be redundant and include noises. Therefore we have to apply principal component analysis(PCA) to reduce the redundancy and noises in centralized F_{norm}. the final representation are eigen joints which are action descriptor of each frames. We are observing that the most of eigen values are covered by first few leading eigenvector


(a) Tennis Serve

(b) Golf Swing

Fig.4. Sampled sequences of depth maps and skeleton joints in actions of (a) Tennis Serve and (b) Golf Swing. Each depth map includes 20 joints. The joints of each body part are encoded in corresponding colors

7. Classification

We have to use Principal component analysis for the feature extraction from the kinect data set and different classification technique like Euclidean, negative and Manhattan as a classifier and neural network technique.

Principal Component Analysis

In the language of information theory, we want to extract the relevant information in a face image, encode it as effectively as possible, and compare one face encoding with a database of models encoded similarly. A simple approach to extracting the information contained in an image of a face is to somehow capture the variation in a collection of face images, independent of any judgment of features

- Most common form of factor analysis
- The new variables/dimensions
  - Are linear combinations of the original ones
  - Are uncorrelated with one another
- Orthogonal in original dimension space
  - Capture as much of the original variance in the
- data as possible
  - Are called Principal Components

IV. Result and Analysis

In my research, work proposed a activity recognition technique which are based on the Eigen joint captured by the Microsoft kinect camera for different distance based classifier technique. three distances are used in my work and compared the result by the other classification techniques. We have to see that the 10-12 % frames are more than enough to recognize the activity with the best accuracy from the activity video. In other computer vision systems, mostly the human activity recognitions are highly depends on context. The activities are different in every situation. Besides, what composes an activity is a part of the perception. It would be interesting to debate we have a wide range of scenarios can define a common set of action or not. Any recognition problem is finding the key to strong representative facility Pattern sets.

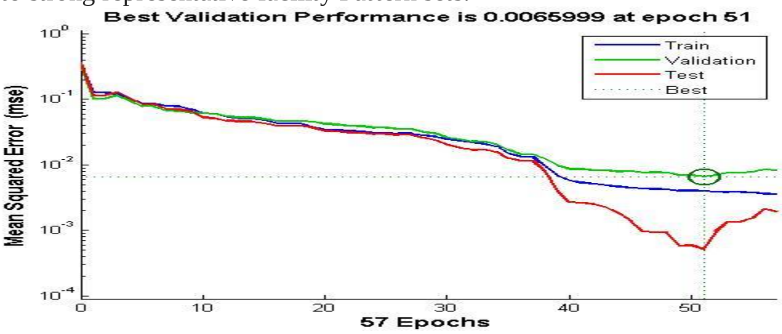


Fig.5.Performance Analyses with mean square error(MSE)

The results of proposed system have been demonstrated considering the different parameters such as performance and percentage accuracy of classification. We have also calculated the errors in terms of mean square error (MSE) of training samples of dataset.

V. REFERENCES

[1] Cedras, Claudette, and Mubarak Shah. "Motion-based recognition a survey."Image and Vision Computing 13.2 (1995): 129-155.

[2]Ye, Juan, Simon Dobson, and Susan McKeever. "Situation identification techniques in pervasive computing: A review." Pervasive and mobile computing8.1 (2012): 36-66.

[3]Augustyniak, Piotr, et al. "Seamless tracing of human behavior using complementary wearable and house-embedded sensors." Sensors 14.5 (2014): 7831-7856.

[4] Liu, An-An, et al. "Coupled hidden conditional random fields for RGB-D human action recognition." Signal Processing 112 (2015): 74-82.

[5]Raman, Natraj, and Stephen J. Maybank."Action classification using a discriminative multilevel HDP-HMM." Neurocomputing 154 (2015): 149-161.

[6] Foggia, Pasquale, GennaroPercannella, and Mario Vento."Graph matching and learning in pattern recognition in the last 10 years." International Journal of Pattern Recognition and Artificial Intelligence 28.01 (2014): 1450001.

[7]Li, Y. F., Jianwei Zhang, and Wanliang Wang. "Active sensor planning for multiview vision tasks"Vol. 1. Heidelberg: Springer, 2008.

[8]Xiaodong Yang; YingLiTian, "EigenJoints-based action recognition using Naïve-Bayes-Nearest-Neighbor," Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on , vol., no., pp.14,19, 16-21 June 2012.

[9] Xia, Lu, Chia-Chih Chen, and J. K. Aggarwal. "View invariant human action recognition using histograms of 3d joints." Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on.IEEE, 2012.

[10] Vantigodi, S.; Babu, R.V., "Real-time human action recognition from motion capture data," Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), 2013 Fourth National Conference on , vol., no., pp.1,4, 18-21 Dec. 2013

[11] Vantigodi, Suraj, and VenkateshBabuRadhakrishnan. "Action recognition from motion capture data using meta-cognitive rbf network classifier." Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), 2014 IEEE Ninth International Conference on. IEEE, 2014.

[12] Andrew Ng. Cs229 machine learning autumn 2013. http://cs229. stanford.edu. Accessed: 2014-06-20.

[13] J. Sun, X. Wu, S. Yan, L. Cheong, T. Chua, J. Li, "Hierarchical spatio-temporal context modeling for action recognition", in: Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 2004–2011.

[14] A. Bobick, J. Davis, "The recognition of human movement using temporal templates", IEEE Trans. Pattern Anal. Mach. Intell. 23 (3) (2001) 257–267.

[15] K. Schindler, L. Gool, Action snippets: "how many frames does human action recognition require?" in: Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2008, pp. 1–8.