ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 3, Issue 6, June 2015**

# Generic System for Human Gesture Interaction using Hidden Markov Model

Poonam Sangar, Prof.Vanita Babanne

M.E, Department of Computer Engineering, R. M. D. Sinhgad College of Engineering, Pune, India

Professor, Department of Computer Engineering, R. M. D. Sinhgad College of Engineering, Pune, India

**ABSTRACT**: Human action recognition is motivated by some of the applications like video surveillance, medical field, to interact with dumb and deaf people, human robot interaction etc. A video sequence is given as an input, the task of action recognition is to identify or recognize the most similar action among the action sequences learned by the system. In action recognition system, some pre-processing steps are done for removing the noise. After pre-processing feature extraction is done and extracted features are given as an input to HMM model. HMM model classifies various actions in different classes. Thus a system is capable of recognizing various human actions.

**KEYWORDS**: Distance learning, machine learning,pullback metrics, hidden Markov models, action recognition;

## I. INTRODUCTION

Hidden Markov Models (HMM) have proven to be one of the most widely used tools for learning probabilisticmodels. The Pullback Hidden Markov Model (HMM) is a finite set of states [1]; each state is associated witha probability distribution. Transitions among the state are controlled by a set of probabilities called transitionprobabilities.

To define HMM, we need the following elements:-
1) The number of states of the model.
2) The number of observation symbols in the alphabet, M, if observation are continuous then M is infinite.
3) A set of state transition probabilities.

The Pullback Hidden Markov Model (HMM) is a powerful statistical tool for modeling generative sequences that can be characterized by an underlying process generating an observable sequence.

The probabilistic parameters of a Hidden Markov model [3] in fig1:

X: states; y: possible observations; a: state transition probabilities; b: output probabilities.
In fig 1.{X1, X2, X3} are finite states which are hidden from users.
Output y1, y2, y3, y4 are observable states, which are visible to user.
a12, a23, b11, b12 etc these are probabilities with which they are going to next state or showing output.

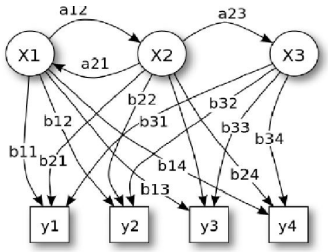


Fig.1:Pullback Hidden Markov Model

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

## II. RELATED WORK

Histogram of Oriented Gradient (HOG) and Histogram of Optical Flow (HOF) descriptors [11].This system is very powerful for classification. The trade-off between accuracy and computational efficiency for the video representation is shown. It is also computationally expensive.

State of the art in action recognition and Bag of space-timefeatures [5] are explained. It does not take into accountthe structure of the action, i.e. does not separate actor andcontext. Also does not allow precise localization.PCAs [12] key advantages are discussed. The PCA method isan unsupervised technique of learning that is mostly suitablefor databases that contain images with no class labels.The paper presented a system able to interpret dynamicand static gestures [3] from a user with the goal of realtimehuman computer interaction. Thus, for hand postureclassification a SVM model was learned from centroiddistance features and for dynamic gesture classification, aHMM model was learned for each gesture and a final averageaccuracy is achieved.

It proposes to learn action subvolumes in a weaklylabelled[13], multiple instance learning (MIL)framework. The resulting action recognition system is suitable for both clipclassification and localization in challenging video datasets. Results demonstrated that the MIL-BoF method achievescomparable performance or improves on the BoF baseline on the most challenging datasets.In this paper, for evaluating the accuracy/efficiency trade-offPCA algorithm [12, 4] is used for feature extraction andHMM [14] is used for classification purpose.

## III. SYSTEM ARCHITECTURE

The system consist of mainly three phases namely preprocessing followed by feature extraction and thenclassification. Video Input: Input from Web Camera to desired machine is obtained for further processing of recognition. In first phase i.e. preprocessing is required to get the normalized frames. In this YCBCR color conversion is used.
The image which we are getting is in RGB form. RGB image contains drawbacks so instead of this we are using for YCRCB approach. YCRCB approach gives high intensity image and superior for detection of images. By using YCRCB the quality of image can be improved and hence easy for detection purpose. After getting the image from YCRCB filtering is done or noise from the images is removed. For removal of noise Gaussian filter is applied, which blurs the area which is not of interest. So Region of Interest (ROI) is highlighted. Feature Extraction: For Extracting feature existing system uses combination of HoG-HoF [5]and motion boundary histogram (MBH) descriptors locally computed along optical flow trajectories. But in this system for extracting features PCA is used [12]. Features based on size and angular motion is considered to construct the feature vector.

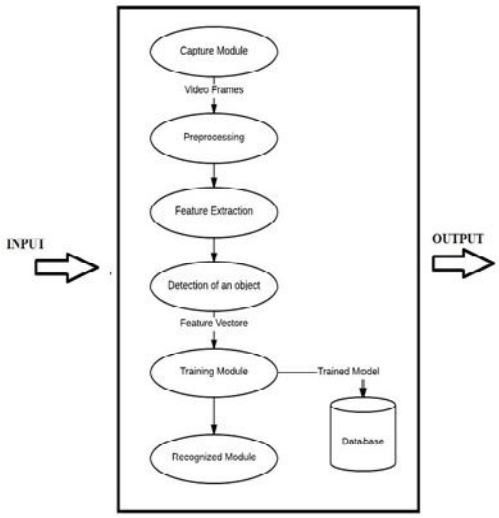


Fig. 2: System architecture

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 3, Issue 6, June 2015**

PCAs key advantages [12] are its low noise sensitivity, the capacity and memory requirements are decreased, and efficiency is increased due to processes taking place in smaller dimensions;

The advantages of PCA are [12]:

1) Lack of redundancy of data given the orthogonal components [15, 16].
2) Reduced complexity in images grouping with the use of PCA [15, 16].
3) Smaller database representation since only the trainee images are stored in the form of their projections on a reduced basis [15].
4) Noise is reduced since the maximum variation basis is chosen and so the small variations in the back-ground are ignored automatically [15].

PCA is used to identify patterns and expressing it in such a way as to highlight their similarities and differences. The patterns are hard to find in data with high dimension, PCA is used for analyzing the data.

## IV. PSEUDO CODE

Steps for PCA:

1) Organize data as an mmatrix, where m is the number of measurement types and n is the number of samples.
2) Subtract off the mean for each measurement type.
3) Calculate covariance matrix.
4) Calculate the eigenvectors and eigen values of the covariance matrix.

- The first step is to obtain a set S with M images S= {T1, T2, T3... Tm}
- Obtain the mean image $\psi$= (T1 + T2 + T3 +..... +Tm)/M.
- Each image differs from the average by $\Phi i = Ti - \psi$ which is called mean centered image.
- Eigen vectors corresponding to the covariance matrix is needed to be calculated

5) Eigen vectors corresponding to the covariance matrix is needed to be calculated

- A covariance matrix is constructed as: $C = A A^T$ where $A = [\Phi 1, \Phi 2... \Phi M]$.
- Eigen vectors corresponding to this covariance matrix is needed to be calculated.
- Eigen vectors Vi of $A^TA$ such that $A^TAX_i = \lambda_i X_i$.
- Eigen vectors corresponding to $A^TA$ can now be easily calculated now with reduced dimensionality where $AX_i$ is the eigen vector and $\lambda_i$is eigen value.

- HMM Algorithm [14]

1) Begin initializesaij,bjk, training sequence V T, Val, z ←0
2) Do z ← z+1
3) Compute a← (z) from a (z-1) and b (z-1)
4) Compute b ← (z) from a (z-1) and b (z-1)
5)aij(z)←aij(z-1)
6)bjk(z) ←bjk(z-1)
7) Until max i,j,k [ aij(z)←aij(z-1) ,bjk(z)←bjk(z-1)] < Val
8)Return aij←aij(z) , bjk←bjk(z)
9)End

## V. MATHEMATICAL MODEL

### *a. SET THEORY:*

- R : be the system
- R={S, A, C, M, P}

Where S ={S: input data| S= {s1, s2…...si}}

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

A = {A: Preprocessing of given input | A= {a1, a2… ai}}.

C ={C: Feature Extraction | C= {c1, c2… ci}}.

M = {M: Classification | M= {m1, m2… mi}}.

P = {P: Action recognition}.

- R is the system which contains five elements as follows.
- S represents the input given to the system. The input is in the form of video. From these video, frames are grabbed and given to next step i.e. preprocessing "A".
- A is the preprocessing step which removes the noise and blurs the region which is not of interest.
- After this step features are extracted and represented by "C". Each frames grabbed contains number of features as c1, c2….ci. Each grabbed frame may contain more than one feature. So there is one to many relationships between them.
- Then extracted features are classified into different classes represented by m1, m2…mi; from extracted features there is one to one relationship on M.
- Based on these classes action are recognized and given as an output.

*b. Mapping Diagrams:*

- V= {f1, f2, ….. , fi}
- F1= Frame grabbed from video; contains many features.
- C ={c1,c2…} These are the features of the frames
- Single frame may contain more than one feature, so there is one tomany relations between them.

Then these features are classified in different classes M={m1,m2…mi}, it may happen that more than one feature may be included in one class, so there is many to one relation.
After this, actionis recognized.

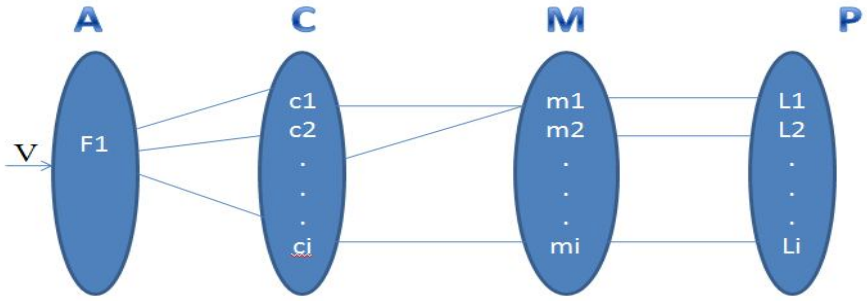


Fig.3: Mapping Diagram

## VI. RESULTS

The experimental methodology is divided into two parts. First one is for static face expression and second is for dynamic gesture recognition. In static face gesture recognition, the image is given as an input which passes through three stages namely pre-processing stage, feature extraction using PCA and Learning HMM model. In pre-processing, noise removal, Selection of Region of interest (ROI), color conversion and largest connected region between face is detected. After this feature extraction using PCA is done, which extracts feature like eyes, lips. Then through HMM it will learn the action based on the trained data and then expression of face is detected.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

For dynamic gesture recognition, the region of interest (ROI) is selected then various features from frames are extracted. Features selected are height, width, angular motion etc and after this HMM will learn and if any moment is done it is captured and result is displayed. The analysis is done on the basis of detection rate and estimation time. The estimation time for the system to detect the action is approximately in 8.4 sec. Thus average accuracy is achieved and the system with its detection and estimation time is given in fig 4.

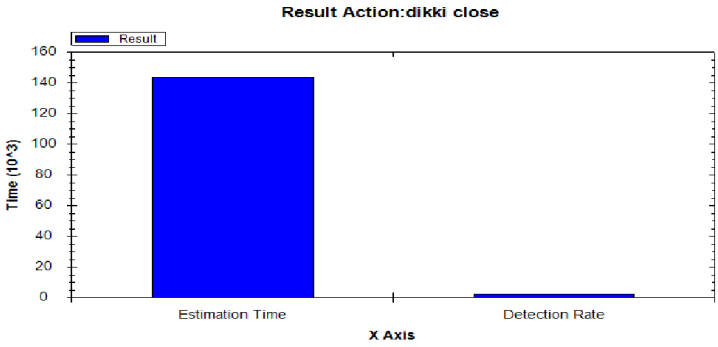


Fig 4: Result Action

## VII. CONCLUSION

This project gives work flow for analysis and classification of dynamic models. In this paper a system is presented that can accurately recognize different human actions and analyzed the performance of human action recognition based on HMM distance. Preprocessing of video is also most important for utilization of system. Preprocessing of video makes classification accurate. For classification of objectsHMM distance is used, which classify objects and recognize them and put them in proper class.In this paper,preprocessing is done first where noise is removed and the better image quality is presented. Then it is followed by feature Extraction in which various features from object is extracted and clusters are formed. These features areclassified into groups or classes using HMM distance and various actions are recognized. Thus this work gives better result for recognition of various actions performed by human.

## REFERENCES

[1] Fabio Cuzzolin and Michael SapienzaLearning Pullback HMM Distances,IEEE Transactions on pattern analysis and machine intelligence, VOL. 36, NO. 7, JULY 2014.
[2] S. Ali, A. Basharat, and M. Shah, Chaotic Invariants for Human Action Recognition, Proc. IEEE 11th Intl Conf. Computer Vision (ICCV 07), 2007.
[3] Paulo Trigueiros, Fernando Ribeiro and Luis Paulo Reis, Generic System for Human-Computer Gesture Interaction, (ICARSC 14), 2014.
[4] Wei Qua, XiaoleiHuangb, and YuanyuanJiac,Segmentation, Noisy Medical Images Using PCA Model Based Particle Filtering,.
[5] I. Laptev Learning Realistic Human Actions from Movies, Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR 08), 2008.
[6] S. Fine, Y. Singer, and N. Tishby, The Hierarchical Hidden Markov Model: Analysis and Applications, Machine Learning, vol. 32, no. 1,pp. 41-62, 1998.
[7] A. Galata, N. Johnson, and D. Hogg, Learning Variable-Length Markov Models of Behavior, Computer Vision and Image Understanding (CVIU), vol. 81, no. 3, pp. 398-413, 2001.
[8] T.S. Jaakkola and D. Haussler, Exploiting Generative Models in Discriminative Classifiers, Proc. Conf. Advances in Neural Information Processing Systems II (NIPS 99), pp. 487-493, 1999.
[9] J.D. Lafferty and G. Lebanon, Diffusion Kernels on Statistical Manifolds, Machine Learning Research, vol. 6, pp. 129-163, 2005.
[10] A. Gupta, P. Srinivasan, J. Shi, and L.S. Davis, Understanding Videos, Constructing Plots Learning a Visually Grounded Storyline Model from Annotated Videos, Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR 09), pp. 2012-2019, 2009.
[11] J.R.R. Uijlings, I.C. Duta, N. Rostamzadeh, N. Sebe, Realtime Video Classification using Dense HOF/HOG, ICMR 14 April 01 - 04 2014.
[12] SasanKaramizadeh, Shahidan M. Abdullah, Azizah A. Manaf, MazdakZamani, AlirezaHooman, An Overview of Principal Component Analysis, Journal of Signal and Information Processing, 2013, 4, 173-175.
[13] Michael Sapienza, Fabio Cuzzolin, Philip H.S. Torr, Learning discriminative space-time actions from weakly labelled videos, DISCRIMINATIVE SPACE-TIME SUBVOLUMES.
[14] Richard O. Duda, Peter E. Hart, David G. Stork, Pattern classification II edition of Wiley-India publication.
[15] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min and W. Worek, Overview of the Face Recognition Grand Challenge, in Computer vision and pattern recognition, 2005. CVPR 2005, IEEE Computer Society Conference on, 2005, pp. 947-954.
[16] D. SrinivasuluAsadi, Ch. DV SubbaRao and V. Saikrishna, A Comparative Study of Face Recognition with Principal Component Analysis and Cross-Correlation Technique, International Journal of Computer Applications Vol. 10, 2010