I J C T A, 9(41) 2016, pp. 717-726 ISSN: 0974-5572

A SCHEME OF TIME SERIES CLASSIFICATION FOR DYNAMIC HAND GESTURE RECOGNITION

Naresh Kumar[1], Aditi Sharma[2] and Vishal Gaur[3]

Abstract: As computers are becoming more popular, the interest towards the methods to interact with computers is also growing. For the utilization of the fullest capability of the computers we need to discover ways to interact with computers. There are many new approaches and technologies are coming up to bridging the gap between human and computers. Giving instruction to the computers by human gestures is a very natural way of human computer interaction (HCI), there are many researches are ongoing in this direction. In this work we summarize some of the novel spatiotemporal classifier and their efficiency to deal with time series data which derbies all types of human pose and gestural activities to interact with machine. It is quite challenging to recognize the communication in sign language recognition (SLR) among deaf and dumb people for a common person and to distinguish the suspicious activities in videos.

Key Words: *Human Computer Interaction (HCI), Finite State Machine (FSM), Time Delay Neural Network (TDNN), Sign Language Recognition (SLR), Human Action Recognition (HAR), Hidden Markov Model(HMM),Dynamic Time Wrapping(DTW), Derivative Dynamic Time Wrapping(DDTW);*

I. INTRODUCTION

Communication is a key factor for performing any type of activity which can be categorized by human-to-human, human-to-machine and machine-to-machine interaction [13]. The type of interaction performed is to determine in videos is a challenging task. A video is nothing but composed of frames in the form of spatiotemporal data. In real world, almost data belongs to time series domain which consists of huge amount of ambiguous information, hard to determine in unconstraint environment. It is quite challenging to recognize the communication in sign language recognition (SLR) among deaf and dumb people for a common person and to distinguish the suspicious activities in videos. A robust time series classifier ensures the performance for such a complex problem.

1.1 Human computer interaction

It is a task which is always useful whenever we want a communication to happen between human and a machine, these machines may varies from a small calculator to a complex supercomputer or a small thermometer to a complex CT scanner every machine require some form of human interaction, every newly invented machine brings a new form of interaction[12]. Types of interaction may varies from a simple key press input or a mouse click input to a voice input or a simple intuitive gestures or a complex dynamic gesture, human gestures has various forms and shapes which can be produced with the various

[1]*Department of Mathematics, Indian Institute of Technology, Roorkee India, Email- atrindma@iitr.ac.in*
[2]*Department of Computer Science & Engineering, MBM Engineering College,Jai Narain Vyas University,Jodhpur, India Email- aditi11121986@gmail.com*
[3]*Department of Computer Science & Engineering, Government Engineering College,Bikaner,India, Email- vishalg.research@gmail.com*

combination of different body parts. In the next paragraph I am focusing more about the gestures which human being.

1.2 Gestures

A conscious gesture is a very good tool by which we humans are able to make nonverbal communication as well as use it in conjunction with verbal communication. These gestures can be associated with any one of the body parts or more than one related body parts, Gestures are the primary part of human communication can be used as a significant means for HCI. The work in [1] by Karam M., it is shown the comparison between gestures associated with different body parts in natural communication between human to human interactions, the percentage wise gestures associated with different combination of body parts is shown in the table 1.1.in the table it is clearly shown that maximum gestures in human to human communication are associated with the hands so hands gestures [19] are also most useful human gesture in the human to computer interaction. Same gesture in different culture may be treated as differently so it is hard to generalize a gesture for a particular meaning but for a machine to take input in the form of gesture is relatively easy in this work it is discussed some of the major technique for dynamic gesture recognition in the subsequent sections.

Table 1- Table1.1: Percentage of gesture by Body Parts

Body parts	Percentage	Body parts	Percentage
Hand	**21%**	**Hand + Head**	**7%**
Head+ Fingers	**2%**	**Finger**	**10%**
Foot	**2%**	**Objects**	**14%**
Object + Fingers	**4%**	**Others**	**9%**
Hand + Finger	**6%**	**Multiple Hands**	**13%**
Hand + Head	**7%**	**Multiple body**	**10%**

1.3 Gesture Recognition

The Recognition of human body gesture refers to the process of tracking the parts of body the human and their representation and for semantically meaningful operation. Research involved in the area of hand gesture is to develop such technique or frameworks which are able to identify the human gestures while taking it as input and perform actions on these gestures so that some device can be controlled by the commands as input to the device. Gesture recognition provide an alternate to the touch based interaction, since touch based device are not accepted in many areas so a vision based approach is provided to identify input by recognition of hand gesture in HCI. We have attempted to describe, the human body gesture as a spatiotemporal data, using the time series classifies which organizes rest of sections of this work for the conceptual description of time series classifiers.

II. FSM BASED TECHNIQUE

FSM based technique is being proposed for gesture recognition of hands which are dynamic in nature, this method is based on the representation of finite state and concluding of gestures using video planes which are having some key video objects Planes (VOPs). In this technique videos are considered as objects for abstraction and breaking up the frames into segments and hence generation the VOPs. In this technique they have considered the hands as one of the video-object (VO). This technique selects the key VOPs on the basis of Hausdorff measure of calculating distance and then breaking down the whole video clip into the frame which represents the whole video and also these frames are able to represent the gesture

associated with it [18]. These frames are some specific frames which have some reliable information in the direction of the understanding the gesture associated so these frames are considered as the key frames. These VOPs which are considered as the key vops are used as the input for the classification of the gestures and states are being used for the representation and identification of these gestures. For knowing the shapes similarity of the sequence of incoming data and the FSM states are measured by a commonly used distance measure called Hausdorff.

2.1. FSM scheme for hand gesture recognition

Fig 2.1 is showing diagram for different stages involved in FSM based system for recognition of hand gestures. Input to this system contains sequence of gesture video called VOPs for different positions of the hand. Hausdorff tracker is used for tracking the change from one frame to the next in the incoming sequence of frames of gesture considered as gesture video, and then key VOPs are selected by using distance measure of Hausdorff which eliminates the frames which are having same information These frames are the real inputs for this FSM based technique for recognizing gestures which consist of the whole information of the video considered as "*representative frames"*, for each gesture a FSM is being constructed during training and the recognition is performed by matching the sequence of states for the input with the provided FSMs, if gesture is matched with the any of the FSM then it is considered as the gesture which is being recognized and presented in Fig. 2.2. However, comparing of input to the FSMs is considered as the matching input to every state of the FSM. it looks like a brute force job but the ART shape descriptor is useful for selecting possible FSMs which are candidate for recognizing the gesture by comparing only the initial state of FSMs in the vocabulary of gesture, by doing this the technique is able to select some of the FSMs which are probable part of the solution and reject all other FSMs which are not the potential candidate for recognizing gesture and hence this technique is able to speeds up the recognition process.

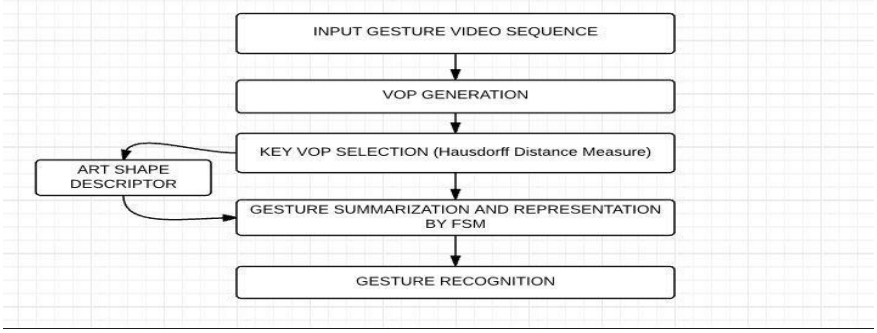

Figure 2.1. The basic block diagram for the FSM based hand gesture recognition system

2.2. Hausdorff distance

The Hausdorff distance is useful for measuring the similarity between two images having some shapes. Hausdorff distance is the maximum distance (1) of a set to the nearest point in the other set, more formally, Hausdorff distance from set A to set B is a max-min function, defined as

$$h(A,B) = \max\{\min\{d(a,b)\}\} \qquad (1)$$

where a and b are points of sets A and B respectively, and d(a,b) is any metric between these points", for simplicity d(a,b) can be considered as the Euclidian distance measure between point a and point b.

2.3. Representation of gestures using Finite states

The key frame and its duration of key frames can be evaluated as the count of frames of video between present signified frame and the upcoming signified frame. Each state of FSM corresponds to a key frame and a transition in the FSM happens only when the key VOPs shape is similar and the duration criteria also met. Input sequence of the frames provides VOPs for different positions of the hand, generation of VOPs consist of four stages depicted in the Figure 2.2. for hand segmentation and VOP generation.

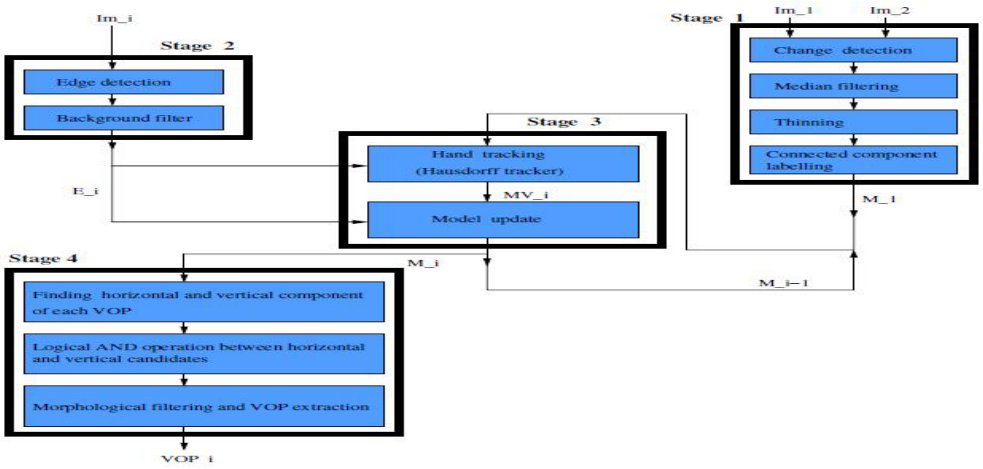

Figure 2.2 VOP generation algorithms (Block

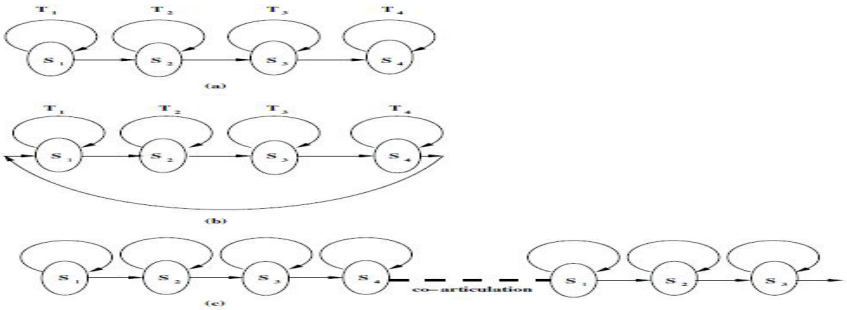

Figure 2.3. (a) FSM representation for a gesture. (b) F.S.M. representation for the similar gesture which is repeating multiple times. (c) FSM representation if the gestures which are connected in sequence one after another.

III. HMM BASED MODEL

HMM [4] is a popular choice for the gesture recognition model because of its ability to deal with the segmentation problem, Markov chains is the to be considered for describing HMM, Markov chain is a collection of fixed number of states just like a finite state machine with each state transition has some probabilistic value associated with it. The basic architecture of HMM is proposed in [10] and [11].From a state there may be many outward arcs are possible with total probability value one, every outward arc associated with an output symbol with restriction that only one transition for a particular output, because of this restriction the markov chain model is behave as a deterministic model.. With the same output symbol, the HMM can have more than one arcs, they are nondeterministic, and by looking at the output it is very much not possible to determine directly the sequence of the states for a particular input. Hence this state sequence is hidden in HMM.

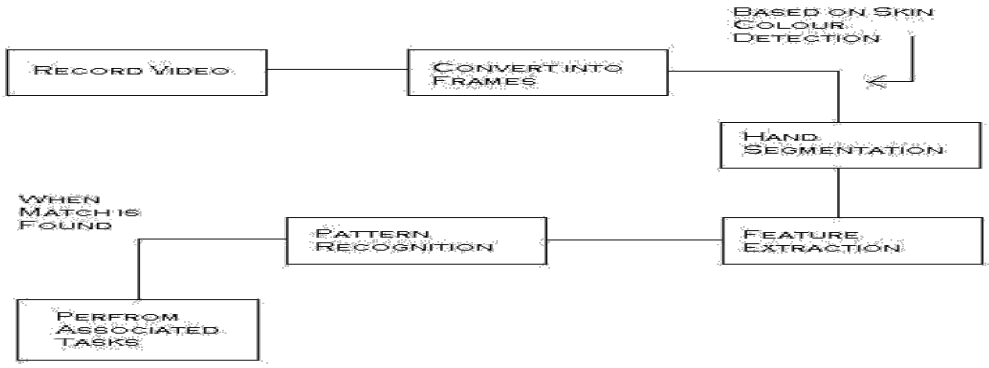

Figure 3.2. HMM based system block diagram

Each of the state transition in HMM. is represented by four parameters, the starting state of that transition, the ending state of the transition Fig. 3.3, the generated symbol which is output symbol and the probability corresponding to a particular transition.Various trajectories are representedin HMM model given by Fig. 3.1. It is found versatile application of a set of HMMs a particular set of hand gestures, wather forcasting and volcano mnitering [17]. The HMM having maximum possibility of forwarding indicates the gesture which is most likely of the user.

A Hidden Markov Model, HMM = (S; C; π; A; B), denotes a stochastic process for time, in terms of the hidden states S, observations C, initial state probabilities π, state transition probabilities A and output probabilities B" [2]. Spatio-temporal variability features of HMM makes it one of the most useful approach to be used in pattern recognition, additionally HMMs can be applied to recognition of gesture, recognition of speech, and modelling of protein [16]. Referring to [2] the paper has given an idea about using HMM model for the Dynamic hand gesture recognition in the static background. This model is trying to recognize some finite set of defined gestures and using them as a input to the system for performing some simple associated tasks. Some of the gestures on which their approach is going to work are given in Figure 3.1. Along with the HMM used in the system, adjacency matrix is also used which is for representing the gestures and also using idea where a principal axis is considered from the center of the hand (usually centroid) to make the gesture standard The block diagram of Figure 3.2 shows the overview of how HMM based model is applicable for such systems.

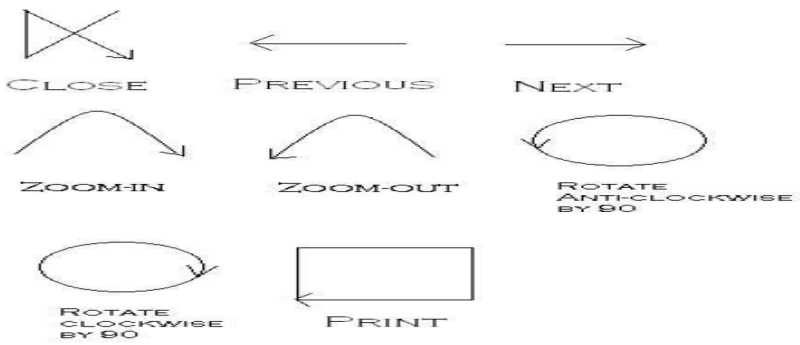

Figure 3.1. Various Trajectories used in HMM based model

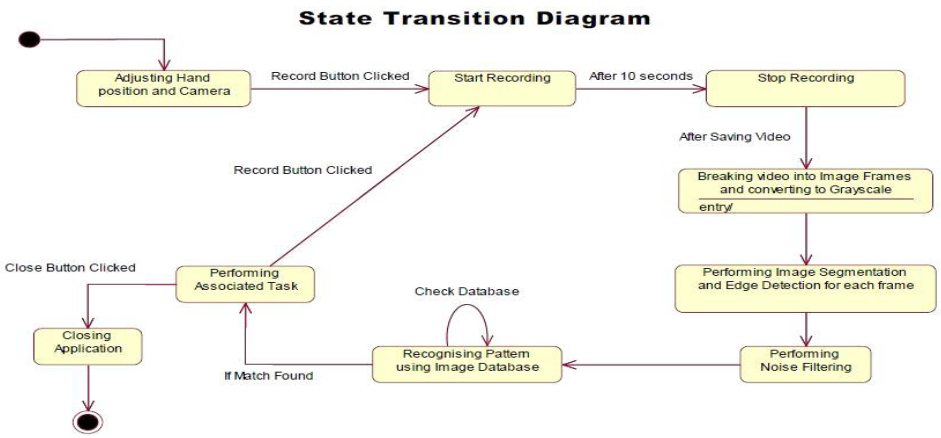

Figure 3.3. HMM system follows this state transition diagram

IV. TIME DELAY NEURAL NETWORK

This is an alternative architecture in Fig.4.1,of neural network is time delay neural network (TDNN), to work on the continuous data is the primary objective of the TDNN. Meng et.al. in [8] demonstrates about the sequence facial expressions in which TDNN architecture [14] is able to adapt the network online whose primary objective is to deal with continuous data and hence it is advantageous for many real time applications. The architecture as shown in figure 4.1 has continuous delayed input which is sending input to the neural network. The present or current state in the time-series indicates the desired output and the delayed time-series (which are previous values) indicates the input to the neural network [15]. Hence the past values of the time series are useful parameter in the function for calculating output of neural network which is the prediction of upcoming value for the time-series. Theoretically TDNNs are the continuation of multi-layer perceptron. The basis of TDNNs is time delay which provide the ability to each and every neuron to preserve the history of its signals which are input signal, and hence the network is able to adapt the sequence-patterns [14]. Time delay enables every neuron to have access to give input at time t as well as former inputs. Therefore, each neuron has some ability to know the relationship between present input and the input values which are previous, there might be a particular sequence or arrangement in the signal which are taken as input. Also, approximate functions can be derived from history of input signal which is already being sampled by time. For the learning of TDNN standard back propagation and its variants may be very useful.TDNN is a type of feed-forward network, there are three layers involved in it: input layer, output layers and hidden layers. It is clearly shown in the Figure 9 that for the input vector x(k) time delays applied which is input to the network. The significance of these time delay inputs is to provide temporal information about the system to the given network. The design parameter is the hidden layer activation function, which is a tan-sigmoid function. Other variance for the design parameter are log-sigmoid and radial basis functions etc. The training algorithm is updating variable parameters like matrices Wx, Wy and bias vector b to mimic and generate the I-O mapping of the plant. The following equation gives the output of the TDNN.

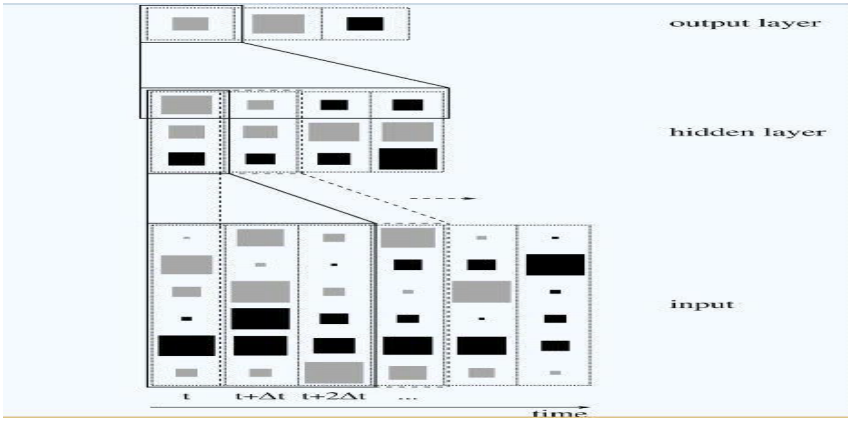

Figure 4.1. Architecture of TDNN

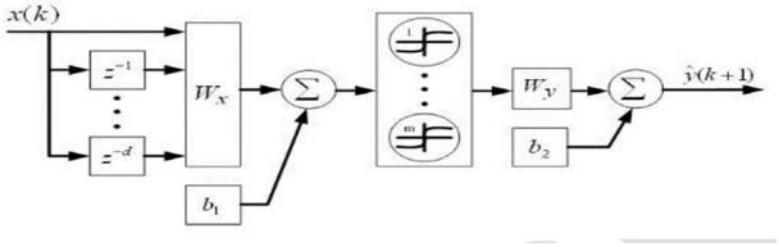

Figure 4.2 TDNN feed-forward network

V. DYNAMIC TIME WARPING

Finding optimal alignment of two signals has been in use for a long time. The DTW algorithms [6] are able to find out the distance between two signals by observing each possible pair of points on the basis of their associated feature values. Cumulative distance measure matrix is being used for this purpose which is useful for finding the path having least cost. This least expensive path represents the warp which minimizes the distance of features between their points which are synchronized when trying to synchronize the two signals, hence it is ideal warp. For calculating the distance between points in signals, generally signals are normalized and smoothed. The usage of DTW is in different fields, for example data mining, movement recognition and speech recognition. Enhancing the speed of the algorithms was the most common work in the field of DTW. Eamonn and Pazzani in year 2001 proposed [8] derivative dynamic time warping (DDTW). In DDTW the distances are being calculated between the first order derivatives, not for the feature values of the points. However most of the work considered only one-dimensional series. For performing the normalization and the time alignment by calculating a transformation (temporal in nature) and for matching of the two signals a DTW algorithm is used. DTW is also used for the video sequences comparing depth features of human joints. Each feature is being assigned with weights based on their intra-inter class gesture variation. In DTW a technique called feature weighting is applicable for recognizing beginning and end of gestures which is made up of data sequences. Commonly used task of DTW in gesture recognition is to deal with gestures having variation in temporal length.In the DTW framework given in Fig.5.1, a gesture pattern set is to be compared with each of the test sequence one by one and the gesture is considered as recognized if the cost of warping

lessor to a given value is exiting in the test sequence. The movement of hand is tracked in [7] and Freeman's eight directional code is generated for hand tracking and the classification is performed by dynamic time wrapping based on Levenshtein minimum edit distance algorithm. However, for finding similarity of two sequences everybody joints are not having equal importance. This method is based on timeline of the dynamic operation.

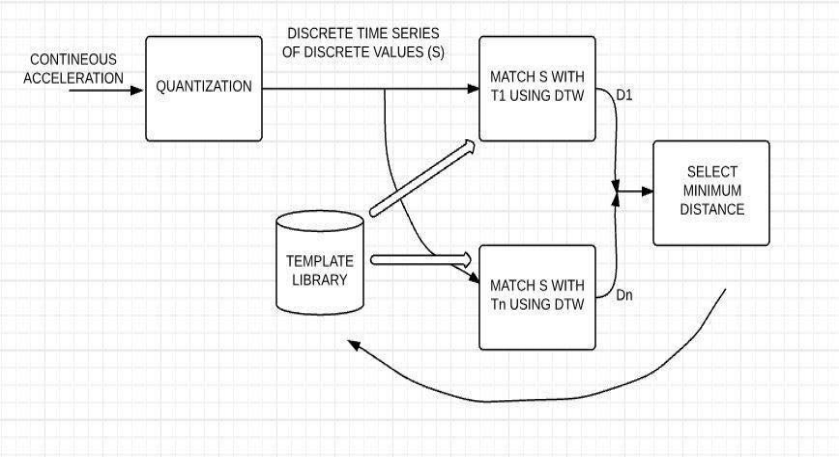

Figure 5.1 Block diagram for DTW

VI. RESULTS

This paper represents an architecture of classification techniques for time series data related to human and its interface to computers. The study in this paper is summarized in table 2 as analytic description of time series classifiers. Its is clearly mentioned that pros and cons of the classification scheme used for particular time series data, vary with the amount of data and environmental issues to capture the datasets for that particular domain.

Table 2. Analytic Description of the classifier for Time Series data

Techniques	Principle	Parameter	Advantages	Disadvantages
Hidden Markov Model	**Without markov chain Generalization is restricted, Set of state transition represents hand positions**	**Pixel in Vision based input**	**Easily Extended to deal with strong TC tasks, Embedded Re-Estimation Possible**	**Large Assumption about the data. Huge no. of Parameters needs to be set. Large Training Data is required.**
Dynamic Time Warping	**Optimal Alignment is found and ideal wrap is obtained based**	**Shape Characteristics**	**Reliable time alignment, Robust to noise, Easy to implement**	**Complexity is quadratic. Distance matrix needs to be defined.**

	on cumulative distance matrix			
Time Delay Neural Network	**Spatial ANN based on time delays giving individual neurons to store history making the system to adapt sequential data,**	**Time Sampled History of Input Signal**	**Faster learning Invariance under time or space translation, Faster execution**	**Lacking robustness. Based on typical patterns of the input is inconsistent.**
Finite State Machine	**Limited or Infinite number of possible states**	**Feature Vector such as trajectory**	**Easy to implement, Efficient predictability, Low Processor Overhead,**	**Not Robust, Rigid Condition for implementation**

VII. CONCLUSION & FUTURE SCOPE

In this paper, four methods are discussed for the recognition of dynamic hand gestures. These methods are Finite state model (FSM), Dynamic time warping (DTW), Hidden markov model (HMM) and Time delay neural network (TDNN). FSM is a simplest technique and very intuitive for hand gesture recognition but it usually involves very lengthy and time consuming operations. HMM based models are preferably used in the robotic control and gives better results in that area. DTW has its own advantage for the continuous gesture recognition whereas the TDNN is mainly used as classifier and identification of hand shape. Different types of applications demands for specific algorithm for recognition purpose. Table 2 represents the key summary of the conceptual scheme for computer vision problems like dynamic gesture recognition, facial expression and other domains that deal time series data. Deep learning approach for classification any sequential activity like sign language recognition, video surmising and image understating sounds as future scope of this work.

REFERENCES

[1] Karam, M. (2005). A taxonomy of gestures in human computer interactions.
[2] Bhuyan, M. K. (2012). FSM-based recognition of dynamic hand gestures via gesture summarization using key video object planes. International Journal of Computer and Communication Engineering, 6, 248-259.
[3] Bansal, M., Saxena, S., Desale, D., & Jadhav, D. (2011). Dynamic gesture recognition using hidden markov model in static background. IJCSI.
[4] Xiaojuan, W., & Zijian, Z. (2005, May). Dynamic gesture track recognition based on HMM. In Proceedings of 2005 IEEE International Workshop on VLSI Design and Video Technology, 2005. (pp. 169-174). IEEE.
[5] Modler, P., & Myatt, T. (2008, October). Recognition of separate hand gestures by time-delay neural networks based on multi-state spectral image patterns from cyclic hand movements. In Systems, Man and Cybernetics, 2008. SMC 2008. IEEE International Conference on (pp. 1539-1544). IEEE.
[6] Hussain, S. M. A., & Rashid, A. H. U. (2012, May). User independent hand gesture recognition by accelerated dtw.

In Informatics, Electronics & Vision (ICIEV), 2012 International Conference on (pp. 1033-1037). IEEE.
[7] Sreekanth, N. S., & Narayanan, N. K. (2017). Dynamic Gesture Recognition—A Machine Vision Based Approach. In Proceedings of the International Conference on Signal, Networks, Computing, and Systems (pp. 105-115). Springer India.
[8] Keogh, E. J., & Pazzani, M. J. (2001, April). Derivative Dynamic Time Warping. In Sdm (Vol. 1, pp. 5-7).
[9] Meng, H., Bianchi-Berthouze, N., Deng, Y., Cheng, J., & Cosmas, J. P. (2016). Time-delay neural network for continuous emotional dimension prediction from facial expression sequences. IEEE transactions on cybernetics, 46(4), 916-929.
[10] Eddy, S. R. (1996). Hidden markov models. Current opinion in structural biology, 6(3), 361-365.
[11] Rabiner, L., & Juang, B. (1986). An introduction to hidden Markov models. ieee assp magazine, 3(1), 4-16.
[12] Hibbeln, M., Jenkins, J. L., Schneider, C., Valacich, J. S., & Weinmann, M. (2016). How is your user feeling? Inferring emotion through human-computer interaction devices. Management Information Systems Quarterly.
[13] Collazos, C. A., Ortega, M., Granollers, A., Rusu, C., & Gutierrez, F. L. (2016). Human-Computer Interaction in Ibero-America: Academic, Research, and Professional Issues. IT Professional, 18(2), 8-11.
[14] Meng, H., Bianchi-Berthouze, N., Deng, Y., Cheng, J., & Cosmas, J. P. (2016). Time-delay neural network for continuous emotional dimension prediction from facial expression sequences. IEEE transactions on cybernetics, 46(4), 916-929.
[15] Petitjean, F., Forestier, G., Webb, G. I., Nicholson, A. E., Chen, Y., & Keogh, E. (2016). Faster and more accurate classification of time series by exploiting a novel dynamic time warping averaging algorithm. Knowledge and Information Systems, 47(1), 1-26.
[16] Sivakumar, B. (2017). Stochastic Time Series Methods. In Chaos in Hydrology (pp. 63-110). Springer Netherlands.
[17] Whoriskey, K., Auger-Méthé, M., Albertsen, C. M., Whoriskey, F. G., Binder, T. R., Krueger, C. C., & Flemming, J. M. (2016). A Hidden Markov Movement Model for rapidly identifying behavioral states from animal tracks. arXiv preprint arXiv:1612.06921.
[18] Rokade, R. S., & Doye, D. D. (2016). Sign recognition using key frame selection. International Journal of Signal and Imaging Systems Engineering, 9(4-5), 320-332.
[19] Martínez-Camarena, M., Oramas, J., Montagud-Climent, M., & Tuytelaars, T. (2016). Reasoning about Body-Parts Relations for Sign Language Recognition. arXiv preprint arXiv:1607.06356.