# Different Types of Spatiotemporal Classifier and Their Efficiency for Human Gestures

Dr. Vishal Kumar Goar, Dr. Manoj Kuri, Mr. Pawan Tanwar
*Assistant Professor*
*Govt. Engineering College Bikaner (Rajasthan)*
*Email – dr.vishalgoar@gmail.com, kuri.manoj@gmail.com, pktbkn@gmail.com*

***Article Info***
***Volume 82***
***Page Number: 1253 - 1262***
***Publication Issue:***
***January-February 2020***

***Article History***
***Article Received:*** *14 March 2019*
***Revised:*** *27 May 2019*
***Accepted:*** *16 October 2019*
***Publication:*** *07 January 2020*

***Abstract***

As computers have become additional well-liked, the interest in the strategies to act with computers is additionally growing. For the use of the fullest capability of computers, we want to find ways in which to act with computers. There are several new approaches and technologies are developing to bridging the gap between human and computers. Instructing the computers by human gestures may be a natural means of human-computer interaction (HCI), there are several types of research are in progress during this direction. during this work, we tend to summarize a number of the novel spatiotemporal classifier and their potency to affect statistic information that derbies all sorts of human pose and gestural activities to act with a machine. it's quite difficult to acknowledge the communication in sign language recognition (SLR) among deaf and dumb folks for a typical person and to differentiate the suspicious activities in videos.

***Keywords****: 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);*

## I. INTRODUCTION

Communication could be a key issue for performing arts any style of activity which may be categorized by human-to-human, human-to-machine and machine-to-machine interaction [13]. the sort of interaction performed is to work out in videos that could be a difficult task. A video is nothing however composed of frames within the style of spatiotemporal information. In the globe, nearly information belongs to a statistic domain that consists of a big quantity of ambiguous data, exhausting to work out in an unrestraint setting. it's quite difficult to acknowledge the communication in sign language recognition (SLR) among deaf and dumb individuals for a typical person and to tell apart the suspicious activities in videos. a sturdy statistic classifier ensures the performance for such a fancy drawback.

### 1.1 Human-computer interaction

It is a task that is usually helpful whenever we would like a communication to happen between human and a machine, these machines might vary from a little calculator to a fancy mainframe computer or a little measuring system to a fancy CT scanner each machine need some style of human interaction, each recently fictitious machine brings a brand new style of interaction[12]. styles of interaction might vary from a straightforward keypress input or a depression input to voice input or a straightforward intuitive gesture or a fancy dynamic gesture, human gestures has numerous forms and shapes which may be made with the varied combination of various body components. within the next paragraph, I am focusing additional regarding the gestures that person.

TEST Engineering & Management

January - February 2020
ISSN: 0193 - 4120 Page No. 1232 - 1244

### 1.2 Gestures

An acutely aware gesture could be an excellent tool by that we tend to human's square measure able to create nonverbal communication additionally as use it in conjunction with verbal communication. These gestures may be related to any one of the body components or over one connected body components, Gestures square measure the first a part of human communication may be used as a major suggests that for HCI. The add [1] by Karam M., it's shown the comparison between gestures related to totally different body components in natural communication between human to human interactions, the proportion wise gestures related to totally different combination of body components are shown within the table one.1.in the table it's clearly shown that most gestures in human to human communication square measure related to the hands, therefore, hands gestures [19] are most helpful human gesture within the human to laptop interaction. Same gesture in numerous cultures could also be treated as otherwise, therefore, it's exhausting to generalize a gesture for a specific that means except for a machine to require input within the style of gesture is comparatively straightforward during this work it's mentioned a number of the foremost technique for dynamic gesture recognition within the sequent sections.

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

| Bodyparts | Percentage | Bodyparts | Percentage |
|---|---|---|---|
| Hand+Head | 7% | Object+Fingers | 4% |
| MultipleHands | 13% | Hand+Finger | 6% |
| Multiplebody | 10% | Hand+Head | 7% |
| Hand | 21% | Finger | 10% |
| Head+Fingers | 2% | Objects | 14% |
| Foot | 2% | Others | 9% |

### *1.3 Gesture Recognition*

The Recognition of human body gesture refers to the method of following the elements of the body the human and their illustration and semantically significant operation. the analysis concerned within the space of hand gesture is to develop such techniques or frameworks that are ready to determine the human gestures whereas taking it as input and perform actions on these gestures so that some device is often controlled by the commands as input to the device. Gesture recognition gives an alternate to the touch-based interaction since the touch-based device doesn't seem to be accepted in several areas thus a vision based mostly approach is provided to spot input by recognition of hand gesture in HCI. we've tried to explain, the chassis gesture as spatiotemporal information, victimization the statistic classifies that organizes the remainder of sections of this work for the abstract description of your time series classifiers.

## II. FSM BASED TECHNIQUE

FSM primarily based technique is being projected for gesture recognition of dynamic hands, this methodology relies on the illustration of finite-state and final of gestures mistreatment video planes that are having some key video objects Planes (VOPs). during this technique, videos are considered as objects for abstraction and ending the frames into segments and thus generating the VOPs. during this technique, they need to be thought of the hands mutually of the video-object (VO). this method selects the key VOPs on the idea of Hausdorff to live off the hard distance then breaking down the entire video clip into the frame that represents the entire video and additionally these frames can represent the gesture related to it [18]. These frames are some specific frames that have some reliable info within the direction of understanding the gesture associated so these frames are thought of because of the keyframes. These VOPs that are thought of because the key VOPs are used because the input for the classification of the gestures and states are being employed for the illustration and identification of those gestures. For knowing the similarity of the shape of the sequence of incoming information and therefore the FSM states are measured by a usually used distance measure known as Hausdorff.

## 2.1. FSM scheme for hand gesture recognition

Fig 2.1 is showing a diagram for various stages concerned in the FSM primarily based system for recognition of hand gestures. Input to the present system contains a sequence of gesture videos known as VOPs for various positions of the hand. Hausdorff huntsman is employed for chase the modification from one frame to consecutive within the incoming sequence of frames of gesture thought of as gesture video, then key VOPs are chosen by mistreatment distance measure of Hausdorff that eliminates the frames that are having same info These frames ar the $64000 inputs for this FSM primarily based technique for recognizing gestures that encompass the entire info of the video thought of as “representative frames”, for every gesture an FSM is being created throughout coaching and therefore 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's thought of because the gesture that is being recognized and given in Fig. 2.2. However, the examination of input to the FSMs is taken into account because of the matching input to each state of the FSM. it's sort of a brute force job however the ART form descriptor is helpful for choosing potential FSMs that ar candidate for recognizing the gesture by examination solely the initial state of FSMs within the vocabulary of gesture, by doing this the technique is in a position to pick out a number of the FSMs that are probable a part of the answer and reject all alternative FSMs that don't seem to be the potential candidate for recognizing gesture and thus this method is in a position to hurries up the popularity method.

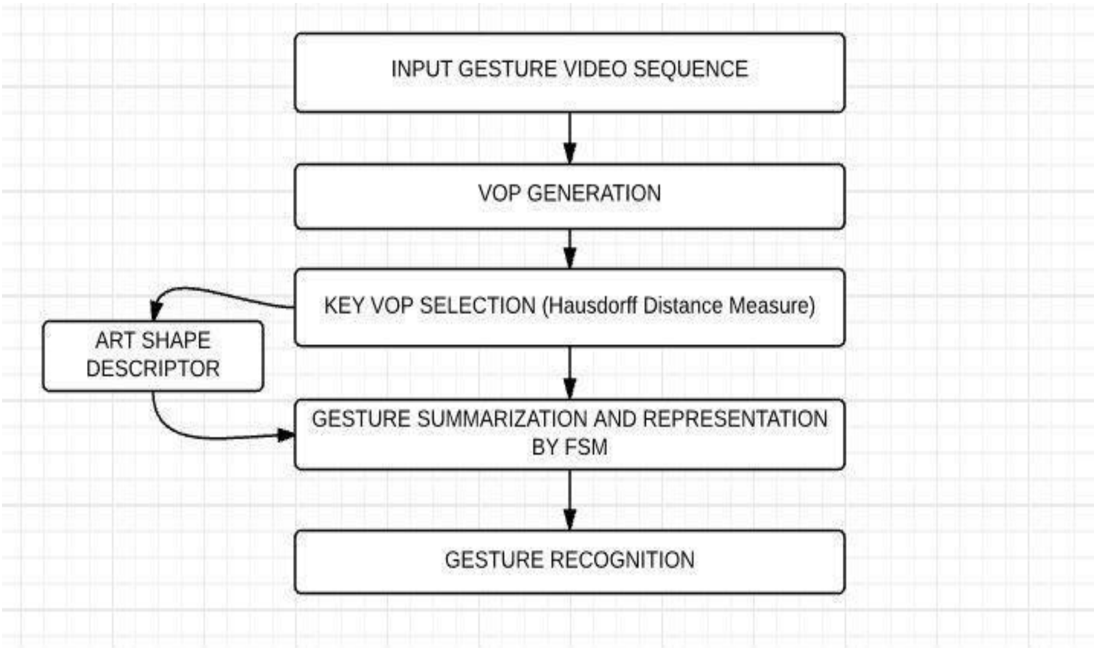


**Figure2.1. Thebasicblock diagramfortheFSM based hand gesturerecognitionsystem**

January - February 2020
ISSN: 0193 - 4120 Page No. 1232 - 1244

*2.1. 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 J to set K is a max-min function, defined as

$$h(J, K) = \max\{\min\{e(j, k)\}\} \quad (1)$$

where j and k are points of sets J and K respectively, and e(j, k) is any metric between these points", for simplicity e(j, k) can be considered as the Euclidian distance measure between point a and point b.

2.2. Representation of gestures using Finite states

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

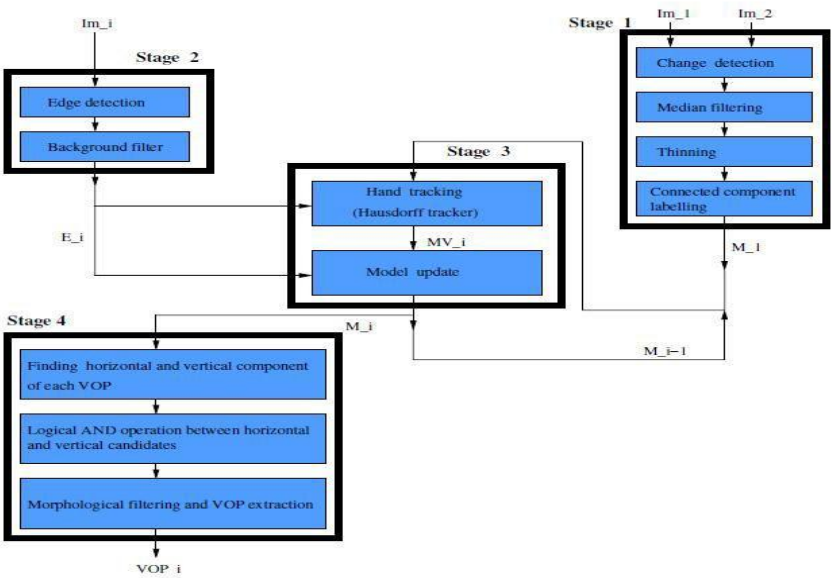


**Figure2.2 VOPgenerationalgorithms(Block**

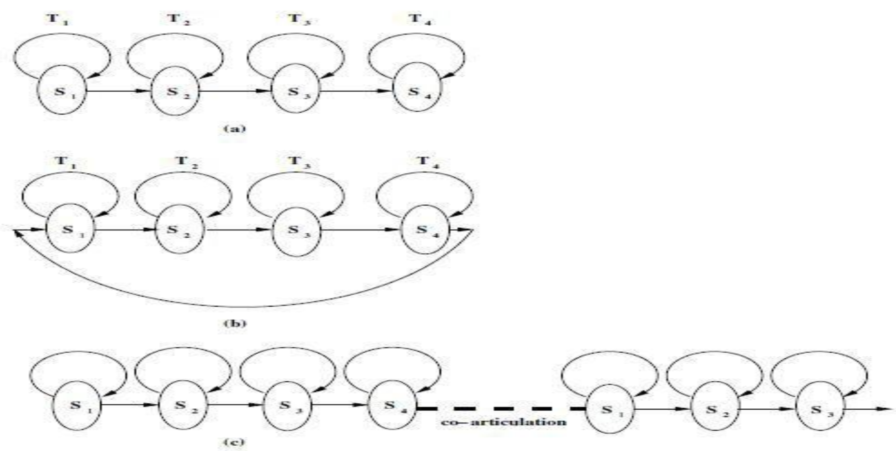


**Figure2.3. (a) FSMrepresentationfora gesture. (b) F.S.M. representation forthesimilargesturewhichisrepeatingmultipletimes.(c)FSM representationif the gestureswhichareconnectedin sequenceoneafteranother.**

January - February 2020
ISSN: 0193 - 4120 Page No. 1253 - 1262

## III. HMMBASEDMODEL

HMM [4] may be a common alternative for the gesture recognition model due to its ability to deal with the segmentation downside, Markoff chains is that the to be thought of for describing HMM, Markoff chain may be an assortment of fastened variety of states rather like a finite state machine with every state transition has some probabilistic worth related to it. the essential design of HMM is projected in [10] and [11]. From a state, there is also several outward arcs area unit attainable with total chance worth one, each outward arc related to associate output image with the restriction that only 1 transition for a specific output, due to this restriction the Markoff chain model is behaving as a settled model. With an equivalent output image, the HMM will have quite one arc, they're nondeterministic, and by watching the output it's considerably uphill to see directly the sequence of the states for specific input. thus this state sequence is hidden in HMM.

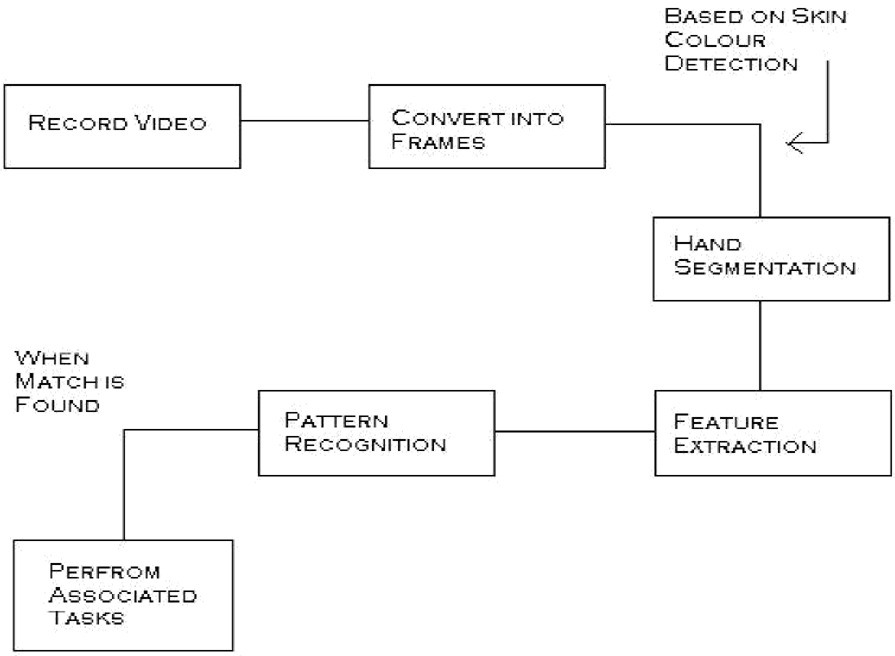


**Figure3.2. HMM based systemblock diagram**

Each of the state transitions 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 an output symbol and the probability corresponding to a particular transition. Various trajectories are represented in the HMM model given by Fig. 3.1. It is found the versatile application of a set of HMMs a particular set of hand gestures, weather forecasting and volcano monitoring [17]. The HMM having the maximum possibility of forwarding indicates the gesture which is most likely of the user.

A Hidden Markov Model, HMM = (S; C; π; J; K), denotes a stochastic process for a time, in terms of the hidden states S, observations C, initial state probabilities π, state transition probabilities J and output probabilities K" [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 the recognition of gesture, recognition of speech, and modeling of protein [16]. Referring to [2] the paper has given an idea about using the HMM model for the Dynamic hand gesture recognition in the static

January - February 2020
ISSN: 0193 - 4120 Page No. 1232 - 1244

background. This model is trying to recognize some finite set of defined gestures and using them as an 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

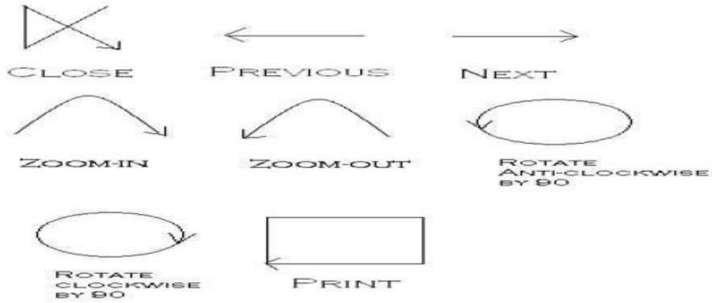


**Figure3.1. VariousTrajectoriesusedin HMM based model**

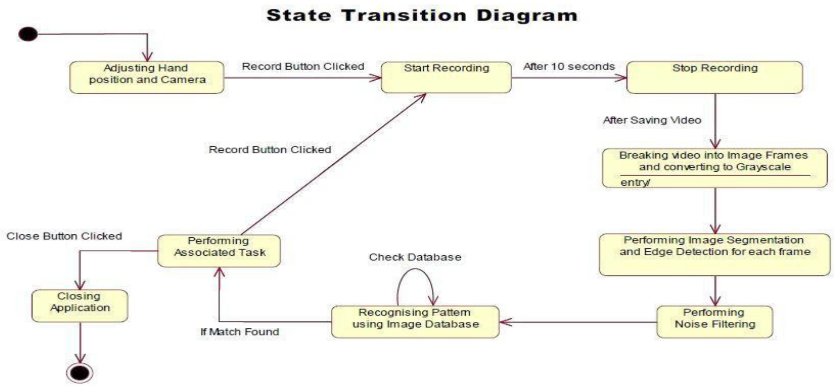


**Figure3.3. HMM systemfollowsthisstatetransition diagram**

## IV. TIMEDELAYNEURALNETWORK

This is another design in Fig.4.1, of the neural network, which is the time-delay neural network (TDNN), to figure on the continual knowledge is that the primary objective of the TDNN. Meng et.al. in [8] demonstrates concerning the sequence facial expressions during which TDNN design [14] is in a position to adapt the network on-line whose primary objective is to handle continuous knowledge and therefore it's advantageous for several real-time applications. The design as shown in figure four.1 has continuous delayed input that is causing input to the neural network. the current or current state within the time-series indicates the required output and also the delayed time-series (which area unit previous values) indicates the input to the neural network [15]. therefore, the past prices of the statistic area unit helpful parameter within the operate for scheming output of neural network that is that the prediction of forthcoming value for the time-series. in theory TDNNs area unit the continuation of a multi-layer perceptron. the idea of TDNNs is time delay which gives the power to every nerve cell to preserve the history of its signals that area unit signal, and therefore the network is in a position to adapt the sequence-patterns [14]. The time delay allows each nerve cell to own access to administer input

TEST Engineering & Management
January - February 2020
ISSN: 0193 - 4120 Page No. 1253 - 1262

at time t similarly as former inputs. Therefore, every nerve cell has some ability to grasp the connection between gift input and also the input values that area unit previous, there can be a specific sequence or arrangement within the signal that area unit taken as input. Also, approximate functions are often derived from the history of signal that is already being sampled by time. For the educational of TDNN normal backpropagation and its variants could also be helpful. TDNN could be a style of the feed-forward network, there are three layers concerned in it: an input layer, output layers, and hidden layers. it's clearly shown within Figure nine that for the input vector x(k) time delays applied that is input to the network. the importance of those time-delay inputs is to supply temporal info concerning the system to the given network. the planning parameter is that the hidden layer activation operates, which could be a tan-sigmoid operate. alternative variance for the planning parameter area unit log- sigmoid and radial basis functions, etc. The coaching rule changes variable parameters like matrices Wx, wy and bias vector b to mimic and generate the I-O mapping of the plant. the subsequent equation offers the output of the TDNN.

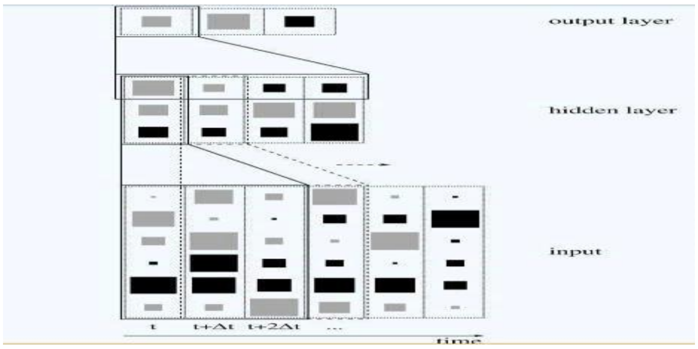


**Figure4.1. Architectureof TDNN**

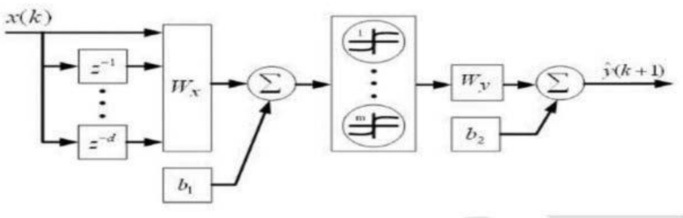


**Figure4.2TDNN feed-forward network**

## V. DYNAMICTIMEWARPING

Finding optimum alignment of two signals has been in use for an extended time. The DTW algorithms [6] square measure ready to establish the gap between two signals by observant every attainable combine of points on the idea of their associated feature values. accumulative distance live matrix is getting used for this purpose that helps locate the trail having the least price. This least costly path represents the warp that minimizes the gap of options between their points that square measure synchronal once attempting to synchronize the two signals, thus it's ideal warp. For shrewd the gap between points in signals, typically signals square measure normalized and smoothened. The usage of DTW is in numerous fields, for instance, data processing, movement recognition, and speech recognition. Enhancing the speed of the algorithms was the foremost common to add the sector of DTW. Eamonn and Pazzani in the year 2001 projected [8] by-product dynamic time warp (DDTW). In DDTW the

January - February 2020
ISSN: 0193 - 4120 Page No. 1232 - 1244

distances square measure being calculated between the primary order derivatives, not for the feature values of the points. but most of the work thought-about solely one- dimensional series. For activity, the standardization and therefore the time alignment by shrewd a metamorphosis (temporal) and for matching of the two signals a DTW rule is employed. DTW is additionally used for the video sequences scrutiny depth options of human joints. every feature is being allotted with weights supported their intra-inter category gesture variation. In DTW a way known as feature coefficient is applicable for recognizing the starting and finish of gestures that are created of knowledge sequences. ordinarily used task of DTW in gesture recognition is to upset gestures having variation in temporal length. In the DTW framework given in Fig.5.1, a gesture pattern set is to be compared with every of the take a look at sequence one by one and therefore the gesture is taken into account as recognized if the value of warp lessor to a given price is exiting within the take a look at the sequence. The movement of the hand is half-track in [7] and Freeman's eight directional code is generated for hand chase and therefore the classification is performed by dynamic time wrapping supported Lowenstein minimum edit distance rule. However, for locating similarity of 2 sequences everyone joints aren't having equal importance. This methodology relies on the timeline of the dynamic operation.

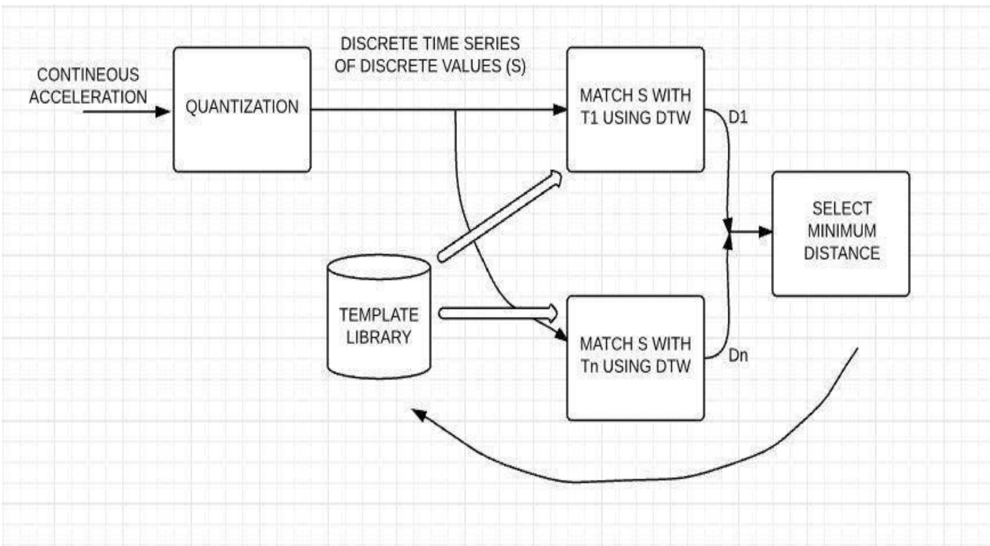


**Figure5.1Blockdiagram forDTW**

## VI. RESULTS

This paper represents an architecture of classification techniques for time series data related to humans and their interface to computers. The study in this paper is summarized in table 2 as an analytic description of time series classifiers. It is mentioned that the 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.

**Table2. Analytic Description of theclassifier forTimeSeriesdata**

| Techniques | Principle | Parameter | Advantages | Disadvantages |
|---|---|---|---|---|
| **HiddenMarkovModel** | **Without Markov, chain Generalization is restricted, set of state transition represents hand positions** | **The pixel in Vision-based input** | **Easily Extended to deal with strong TC tasks, Embedded Re-Estimation Possible** | **Easily Extended to deal with strong TC tasks, Embedded Re-Estimation Possible** |

January - February 2020
ISSN: 0193 - 4120 Page No. 1253 - 1262

| | | | | |
|---|---|---|---|---|
| **DynamicTime Warping** | **Optimal Alignment is found and the ideal wrap is obtained based on the cumulative distance matrix** | **Optimal Alignment is found and the ideal wrap is obtained based on the cumulative distance matrix** | **Optimal Alignment is found and the ideal wrap is obtained based on the cumulative distance matrix** | **Optimal Alignment is found and the ideal wrap is obtained based on the cumulative distance matrix** |
| **Time Delay Neural Network** | **Spatial ANN based on time delays giving individual neurons to store history making the system to adaptsequential 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 In finite number of possible states** | **Feature Vector such as trajectory** | **Easy to implement, Efficient predict ability, Low Processor Overhead,** | **Not Robust, Rigid Condition for implementation** |

## VII. CONCLUSION&FUTURESCOPE

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 the 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 give better results in that area. DTW has its advantage for continuous gesture recognition whereas the TDNN is mainly used as classifier and identification of hand shape. Different types of applications demand a specific algorithm for recognition purposes. 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 with 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 gesturesinhumancomputerinteractions.

[2] Bhuyan,M.K.(2012).FSM-basedrecognitionofdynamichandgesturesviagesturesummarizationusingkeyvideoobjectplanes. InternationalJournalof ComputerandCommunicationEngineering, 6, 248-259.

[3] Bansal,M.,Saxena,S.,Desale,D.,&Jadhav,D.(2011).Dynamicgesturerecognitionusinghiddenmarkovmodelinstaticbackground.IJCSI.

[4] Xiaojuan,W.,&Zijian,Z.(2005,May).DynamicgesturetrackrecognitionbasedonHMM.InProceedingsof2005IEEE InternationalWorkshop on VLSI Design and Video Technology, 2005. (pp. 169-174). IEEE.

[5] Modler,P.,&Myatt,T.(2008,October).Recognitionofseparatehandgesturesbytime-delayneuralnetworksbasedonmulti-statespectralimagepatternsfromcyclichandmovements.InSystems,ManandCybernetics,2008.SMC 2008. IEEE International Conferenceon(pp. 1539-1544). IEEE.

[6] Hussain,S.M.A.,&Rashid,A.H.U.(2012,May).Userindependenthandgesturerecognitionbyacceleratеddtw.

[7] In Informatics,Electronics&Vision(ICIEV), 2012 International Conferenceon(pp. 1033-1037). IEEE.

[8] Sreekanth,N.S.,&Narayanan,N.K.(2017).DynamicGestureRecognition—AMachineVisionBasedApproach.In Proceedings oftheInternationalConferenceonSignal,Networks, Computing,andSystems(pp.105-115).SpringerIndia.

[9] Keogh, E. J., &Pazzani,M.J.(2001, April). DerivativeDynamicTimeWarping. InSdm(Vol. 1, pp. 5-7).

[10] Meng,H.,Bianchi-Berthouze,N.,Deng,Y.,Cheng,J.,&Cosmas,J.P.(2016).Time-delayneuralnetworkforcontinuousemotionaldimension prediction fromfacial expressionsequences.IEEEtransactions oncybernetics,46(4),916-929.

[11] Eddy,S.R. (1996). Hidden markovmodels.Currentopinioninstructural biology,6(3), 361-365.

[12] Rabiner,L., &Juang,B. (1986).Anintroductionto hiddenMarkov models.ieeeasspmagazine,3(1), 4-16.

[13] Hibbeln,M.,Jenkins,J.L.,Schneider,C.,Valacich,J.S.,&Weinmann,M.(2016).Howisyouruserfeeling?

[14] Inferringemotion through human-computerinteraction devices.Management InformationSystemsQuarterly.

[15] Collazos,C.A.,Ortega,M.,Granollers,A.,Rusu,C.,&Gutierrez,F.L.(2016).Human-ComputerInteractioninIbero-America: Academic,Research,and Professional Issues.ITProfessional,18(2), 8-11.

[16] Meng,H.,Bianchi-Berthouze,N.,Deng,Y.,Cheng,J.,&Cosmas,J.P.(2016).Time-delayneuralnetworkforcontinuousemotionaldimension prediction fromfacial expressionsequences.IEEEtransactions oncybernetics,46(4),916-929.

[17] Petitjean,F.,Forestier,G.,Webb,G.I.,Nicholson,A.E.,Chen,Y.,&Keogh,E.(2016).Fasterandmoreaccurateclassification of timeseriesbyexploitinganoveldynamictimewarpingaveragingalgorithm.KnowledgeandInformation Systems,47(1), 1-26.

[18] Sivakumar,B. (2017). StochasticTimeSeriesMethods.In Chaos in Hydrology (pp. 63-110).SpringerNetherlands.[17]Whoriskey,K.,Auger-Méthé,M.,Albertsen,C.M.,Whoriskey,F.G.,Binder,T.R.,Krueger,C.C.,&Flemming,J.

[19] M.(2016).AHiddenMarkovMovementModelforrapidlyidentifyingbehavioralstatesfromanimaltracks.arXivpreprint arXiv:1612.06921.

[20] Rokade,R.S.,&Doye,D.D.(2016).Signrecognition usingkeyframeselection.InternationalJournalofSignalandImagingSystemsEngineering,9(4-5), 320-332.

[21] Martínez-Camarena,M.,Oramas,J.,Montagud-Climent,M.,&Tuytelaars,T.(2016).ReasoningaboutBody-PartsRelationsfor Sign LanguageRecognition.arXiv preprint arXiv:1607.06356.