Dynamic Hand Gesture Recognition for Human Computer interaction; A Comparative Study

Swapnil Athavale[1] ,Mona Deshmukh[2]

[1]Scholar, M.E.I.T, V.E.S.I.T, Chembur, Mumbai

[2]Asst. Professor, MCA Department, M.E.I.T, V.E.S.I.T, Chembur, Mumbai

Abstract:

As computers become more pervasive in society, facilitating natural human–computer interaction (HCI) will have a positive impact on their use. Hence, there has been growing interest in the development of new approaches and technologies for bridging the human–computer barrier. The ultimate aim is to bring HCI to a regime where interactions with computers will be as natural as an interaction between humans, and to this end, incorporating gestures in HCI is an important research area. Gestures have long been considered as an interaction technique that can potentially deliver more natural, creative and intuitive methods for communicating with our computers. The use of hand gestures as a natural interface serves as a motivating force for research in gesture taxonomies. Gesture recognition is done in three main phases i.e. detection, tracking and recognition. The main goal of this study is to provide researchers in the field of gesture based HCI with a summary of progress achieved to date and to help identify areas where further research is needed.

Keywords: Gesture, Detection, Tracking, Hand, Gesture recognition, Human computer interaction, Representations, Recognition, Natural interfaces,

Introduction:

Computer is used by many people either at their work or in their spare-time. Special input and output devices have been designed over the years with the purpose of easing the communication between computers and humans, the two most known are the keyboard and mouse. Every new device can be seen as an attempt to make the computer more intelligent and making humans able to perform more complicated communication with the computer. This has been possible due to the result oriented efforts made by computer professionals for creating successful human computer interfaces. As the complexities of human needs have turned into many folds and continues to grow so, the need for Complex programming ability and intuitiveness are critical attributes of computer programmers to survive in a competitive environment. The computer programmers have been incredibly successful in easing the communication between computers and human. With the emergence of every new product in the market; it attempts to ease the complexity of jobs performed.

Earlier, Computer programmers were avoiding such kind of complex programs as the focus was more on speed than other modifiable features. However, a shift towards a user friendly Environment has driven them to revisit the focus area. The idea is to make computers understand human language and develop a user friendly human computer interfaces (HCI). Making a computer understand speech, facial expressions and human gestures are some steps towards it. Gestures are the non-verbally exchanged information. A person can perform innumerable gestures at a time. Since human gestures are perceived through vision, it is a subject of great interest for computer vision researchers. The project aims to determine human gestures by creating an HCI. Coding of these gestures into machine language demands a complex programming algorithm.

In the present world, the interaction with the computing devices has advanced to such an extent that as humans it has become necessity and we cannot live without it. The technology has become so embedded into our daily lives that we use it to work, shop, communicate and even entertain our self. It has been widely believed that the computing, communication and display technologies progress further, but the existing techniques may become a bottleneck in the effective utilization of the available information flow. To efficiently use them; most computer applications require more and more interaction. For that reason, human-computer interaction (HCI) has been a lively field of research in the last few years. The interaction consists of the direct manipulation of graphic objects such as icons and windows using a pointing device. Even if the invention of keyboard and mouse is a great progress, there are still situations in which these devices are incompatible for HCI.

Gestures:

Gesture acts a medium of communication for non vocal communication in conjunction with or without verbal communication is intended to express meaningful commands. These gestures may be articulated with any of the body parts or with combination of one or many of them. Gestures being major constituent of human communication may serve as an important means for human computer interaction too. Though the significance and meaning associated with different gestures differ very much with cultures having less or invariable or universal meaning for single gesture. For instance different gestures are used for greeting at different geographical separations of the world. For example pointing an extended finger is a common gesture in USA & Europe but it is taken to be as a rude and offensive gesture in Asia. Hence the semantic interpretation of gestures depends strictly on given culture.

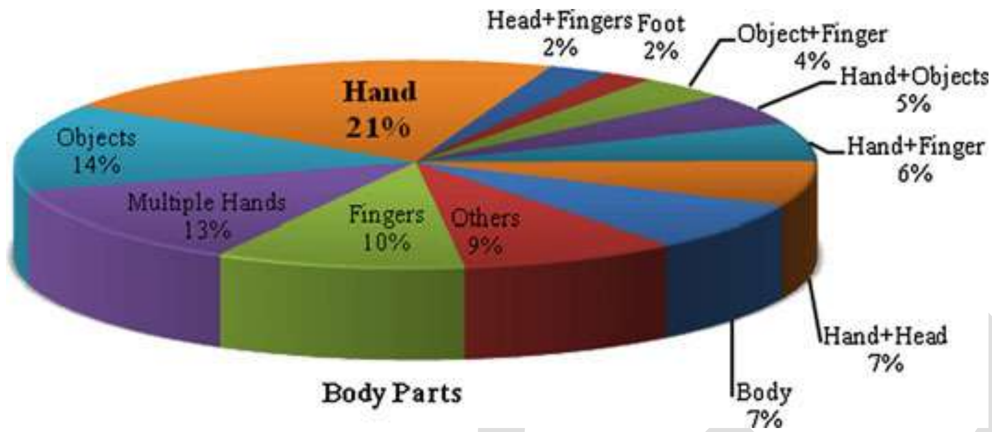
Different Types of Gestures

Types	Meaning	Examples
Symbolic Gesture	Gestures have a single meaning within each culture.	Sign Language, Command Gesture
Deictic Gesture	Gestures direct the listener's attention to specific events or objects in the environment.	Pointing Gestures
Iconic Gesture	Gestures represent meaningful objects or actions.	Predefined Gesture
Pantomimic Gesture	Gestures that depict objects or actions, with or without accompanying speech.	Mimic Gesture

It is hard to settle on a specific useful definition of gestures due to its wide variety of applications and a statement can only specify a particular domain of gestures. Many researchers had tried to define gestures but their actual meaning is still arbitrary.

Bobick and Wilson have defined gestures as the motion of the body that is intended to communicate with other agents. For a successful communication, a sender and a receiver must have the same set of information for a particular gesture. As per the context of the project, gesture is defined as an expressive movement of body parts which has a particular message, to be communicated precisely between a sender and a receiver.

Researcher Karam in his work reported that hand has been widely used in comparison to other body parts for gesturing as it is a natural form of medium for communication between human to human hence can best suited for human computer interaction also as shown Below.



The *piechart* shows the different body parts or objects identified for gesturing

Kendon has situated these communicative hand movements along a "gesture continuum", defining five different kinds of gestures:

1) *Gesticulation* – spontaneous movements of the hands and arms that accompany speech.
2) *Language-like gestures* – gesticulation that is integrated into a spoken utterance, replacing a particular spoken word or phrase.
3) *Pantomimes* – gestures that depict objects or actions, with or without accompanying speech.
4) *Emblems* – familiar gestures such as "V for victory", "thumbs up", and assorted rude gestures.
5) *Sign languages* – Well defined Linguistic systems, like American Sign Language.

Gesture Recognition:

"Gesture recognition the term collectively refers to the whole process of tracking human gestures to their representation and conversion to semantically meaningful commands."

Research in hand gesture recognition aims to design and development of such systems than can identify explicit human gestures as input and process these gesture representations for device control through mapping of commands as output. Creation and implementation of such efficient and accurate hand gesture recognition systems are aided through two major types of enabling technologies for human computer interaction namely contact based and vision based devices.

Contact based devices employed for hand gesture recognition systems are based on physical interaction of user with the interfacing device i.e. the user needs to be accustomed with the usage of these devices, hence not adaptable to the naïve users. The main disadvantage of contact based devices is the health hazards, which are caused by its devices like mechanical sensor material which raises symptoms of allergy, magnetic devices which raises risk of cancer etc. The contact based can be uncomfortable for user since they require physical contact with the user, still having a verge over the accuracy of recognition and less complexity of implementation goes in favor of these devices.

Restrained by the dependence on experienced users the contact based devices do not provide much acceptability, hence vision based devices have been employed for capturing the inputs for hand gesture recognition in human computer interaction. This set of devices relies on captured video sequence by one or several cameras for interpreting and analyzing the motion. Vision based also uses hand markers for detection of human hand motion and gestures. The hand markers can be further classified as reflective markers which are passive in nature and shines as strobes hit it whereas LED light are active in nature and flashes in sequence. In these systems each camera delivers marker position from its view with a 2D frame which lightens with either strobe lights or normal lights.

Vision based devices though is user friendly but suffer from configuration complexity, implementations and occlusion problems, but are more users friendly and hence more privileged for usage in long run.

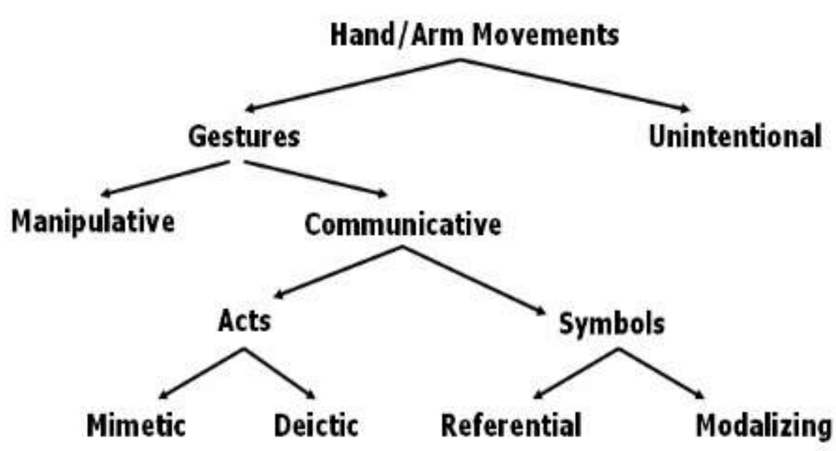
Hand Gesture Recognition:

Hand gesture plays an important part of human communication. Hand gesture has been the most common and natural way for human to interact and communicate with each other. Hand gesture provides expressive means of interactions among people that involves hand postures and dynamic hand movements. A hand posture represents static finger configuration without hand movement, whereas dynamic hand movement consists of a hand gesture with or without finger motion. The ability to detect and recognize the human hand gesture posed many challenges to researchers throughout the decades. Hand gesture recognition studies have gained a lot of attentions for Human-Computer Interaction (HCI).

Hand gestures are classified into two type's static and dynamic gestures.

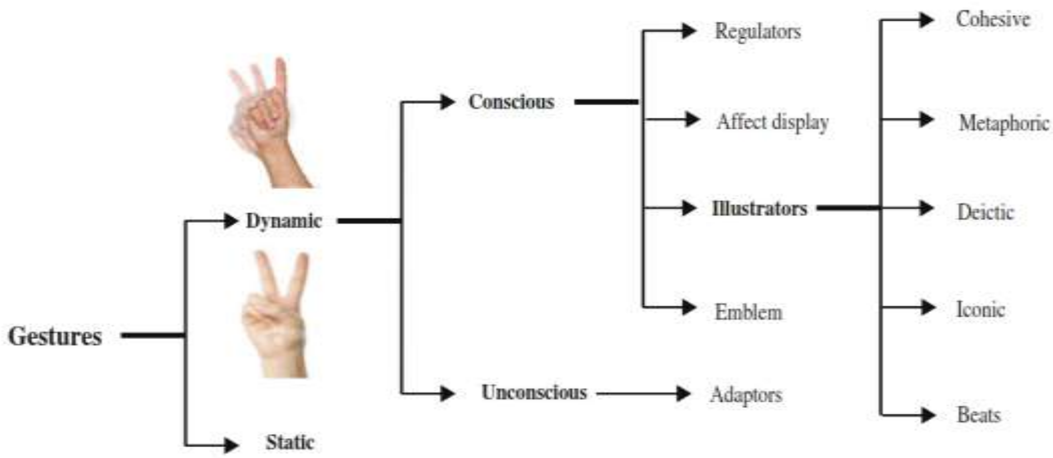
Static hand gestures are defined as orientation and position of hand in the space during an amount of time without any movement and if a movement is there in the aforementioned time duration it is called dynamic gesture.

Dynamic hand gestures include gestures involving body parts like waving of hand while static hand gestures include single formation without movement like jamming the thumb and forefinger to form the "ok" symbol is a static pose which represents static gesture. Dynamic hand gestures done intentionally for communication are called conscious dynamic gestures, whereas unintentionally (unawareness) done gesture carried out during causal physical activity is known as unconscious dynamic gesture. 35% of human communication consists of verbal communication and 65 % is non verbal gesture based communication.



From Figure, gestures can be divided into manipulative and communicative. Manipulative gestures are the ones used to act on objects in an environment (object movement, rotation, etc.). Communicative gestures, on the other hand, have an inherent communicational purpose.

The use of hand gestures provides an attractive and natural alternative to these cumbersome interface devices for human computer interaction. Using hands as a device can help people communicate with computers in a more intuitive way. When we interact with other people, our hand movements play an important role and the information they convey is very rich in many ways. We use our hands for pointing at a person or at an object, conveying information about space, shape and temporal characteristics. We constantly use our hands to interact with objects: move them, modify them, and transform them.



A gesture is scientifically categorized into two distinctive categories: dynamic and static. A dynamic gesture is intended to change over a period of time whereas a static gesture is observed at the spurt of time. A waving hand means goodbye is an example of dynamic gesture and the stop sign is an example of static gesture. To understand a full message, it is necessary to interpret all the static and dynamic gestures over a period of time. This complex process is called gesture recognition.

Gesture can be categorized into five types i.e. emblems, affect displays, regulators, adaptors and illustrators. Emblematic gestures also referred as emblem or quotable gestures are direct translation of short verbal communication like waving hand for good bye or nodding for assurance. The quotable gestures are specifically culture specific. Gestures conveying emotion or intensions are called affect displays.

The affect displays are generally universal less dependent on culture. Gestures controlling interaction are called regulators. Gestures like headshaking, quickly moving one's leg that enables the release of body tension are called adaptors. Adaptors are generally habit of communicators that are not used intentionally during a communication. Illustrator gestures emphasize the key point in speech to depict the communications pronouncing statements. Being emphasized by the communicators pronouncing statements these gestures are inherently dependent on communicators thought process and speech. These gesticulations could further be categorized into five sub category namely beats, deictic gestures, iconic gestures, metaphoric gestures and cohesive gestures.

- ✓ Beats are short and quick, rhythmics and after repetitive gestures.
- ✓ Concrete pointing to real location object or person and abstract pointing to abstract location or period of time are called deictic gestures.
- ✓ Hand movements depicting figural representation or actions for example moving hand upward with wiggling gingers to depict tree climbing are called iconic gestures.
- ✓ Abstractions are depicted by metaphoric gestures.
- ✓ Thematically related but temporally separated gestures are called cohesive gestures. The temporal separation of these thematically related gestures is due to interruption of current communicator by any other communicator.

The two major categories of hand gesture representation are 3D model based methods and appearance based methods.

- The 3D model based hand gesture recognition has different techniques for gesture representation namely 3D textured volumetric, 3D geometric model and 3D skeleton model.

The 3D model based hand gesture representation defines 3D spatial description of human hand for representation with temporal aspect being handled by automation. This automation divides the temporal characteristics of hand gesture into three phases (McNeill 1992) i.e. the preparation or prestroke phase, the nucleus or stroke phase and the retraction or poststroke phase in which every phase corresponds to one or many transitions of spatial states of the 3D human model. One or many cameras focus on the real target and compute parameters spatially matching the real target and then follow its motion during the recognition process in 3D model. Thus the 3D model has an advantage that it updates the model parameters while checking the matches of transition in temporal model, leading to precise hand gesture recognition and representation, though making it computationally intensive with requirement of dedicated hardware. There are also many methods that combine silhouette extraction with 3D model projection fitting by finding target self oriented. Generally three kinds of model are generally used.

- ✓ 3D textured kinematic/volumetric model contains very high details of human body skeleton and skin surface information.
- ✓ 3D geometric models are less precise than 3D textures kinematic/volumetric models with respect to skin information but contains essential skeleton information.

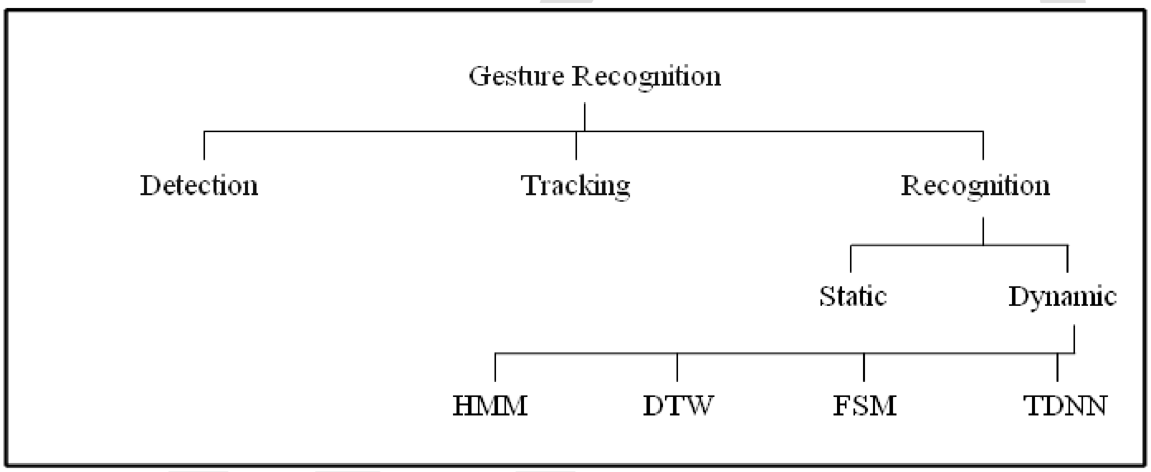
Appearance based hand gesture representation include color based model, silhouette geometry model, deformable gabarit model and motion based model. Appearance based hand gesture representation methods are though broadly classified into two major subcategories i.e. 2D static model based methods and motion based methods, each sub category is having further variants. The commonly used 2D models include:

- ✓ Color based model uses body markers to track the motion of body or body part. As Bretzner et al. (2002) proposed hand gesture recognition employing multi-scale color features, hierarchal models and particle filtering.
- ✓ Silhouette geometry based models include several geometric properties of the silhouette such as perimeter, convexity, surface, bounding box/ellipse, elongation, rectangularity, centroid and orientation. The geometric properties of the bounding box of the hand skin were used to recognize hand gestures.
- ✓ Deformable gabarit based models: they are generally based on deformable active contours (i.e. snake parameterized with motion and their variants. used snakes for the analysis of gestures and actions in technical talks for video indexing.
- ✓ Motion based models are used for recognition of an object or its motion based on the motion of object in an image sequence. Local motion histogram was introduced which uses an Adaboost framework for learning action models.

Gesture Recognition Classification:

Hand movements are thus a mean of non-verbal communication, ranging from simple actions (pointing at objects for example) to more complex ones (such as expressing feelings or communicating with others). In this sense, gestures are not only an ornament of spoken language, but are essential components of the language generation process itself. A gesture can be defined as a physical movement of the hands, arms, face and body with the intent to convey information or meaning. In particular, recognizing hand gestures for interaction can help in achieving the ease and naturalness desired for human computer interaction. Users generally use hand gestures for expression of their feelings and notifications of their thoughts.

Most of the complete hand interactive mechanisms that act as a building block for hand gesture recognition system are comprised of three fundamental phases: detection, tracking and recognition.



Detection:

The primary step in hand gesture recognition systems is the detection of hands and the segmentation of the corresponding image regions. This segmentation is crucial because it isolates the task-relevant data from the image background, before passing them to the subsequent tracking and recognition stages.

Tracking:

If the detection method is fast enough to operate at image acquisition frame rate, it can be used for tracking as well. However, tracking hands is notoriously difficult since they can move very fast and their appearance can change vastly within a few frames. Tracking can be defined as the frame-to-frame correspondence of the segmented hand regions or features towards understanding the observed hand movements. The importance of robust tracking is twofold. First, it provides the inter-frame linking of hand/finger appearances, giving rise to trajectories of features in time. These trajectories convey essential information regarding the gesture and might be used either in a raw form (e.g. in certain control applications like virtual drawing the tracked hand trajectory directly guides the drawing operation) or after further analysis (e.g. recognition of a certain type of hand gesture). Second, in model-based methods, tracking also provides a way to maintain estimates of model parameters variables and features that are not directly observable at a certain moment in time.

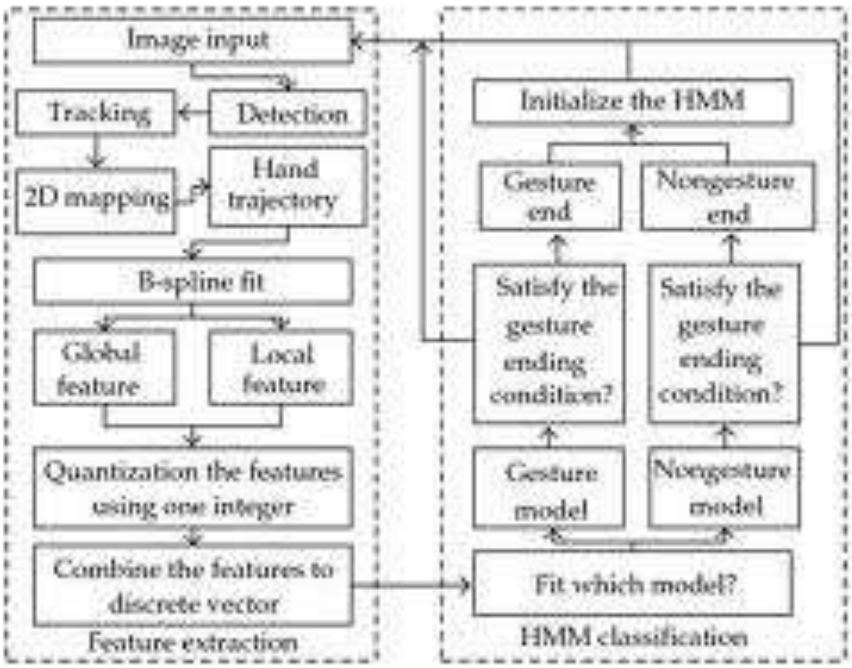
Recognition:

The overall goal of hand gesture recognition is the interpretation of the semantics that the hand(s) location, posture, or gesture conveys. Hand gesture recognition techniques can be further classified under static and dynamic gestures. To detect static gestures (i.e. postures), a general classifier or a template-matcher can be used.

A dynamic hand gesture is then considered as a path between an initial state and a final state. The main limitation of the approaches based on automata is that the gesture model must be modified when a new gesture needs to be recognized. Moreover, the computational complexity of such approaches is generally huge Vision based hand gesture recognition for human computer interaction since it is proportional to the number of gestures to be recognized which is not the case for methods based on other techniques.

Hidden Markov model:

HMM were introduced in the mid 1990s, and quickly became the recognition method of choice, due to its implicit solution to the segmentation problem. In describing hidden Markov models (Ramage 2007) it is convenient first to consider Markov chains. Markov chains are simply finite-state automata in which each state transition arc has an associated probability value; the probability values of the arcs leaving a single state sum to one. Markov chains impose the restriction on the finite-state automaton that a state can have only one transition arc with a given output; a restriction that makes Markov chains deterministic. A hidden markov model (HMM) can be considered a generalization of a Markov chain without this Markov chain restriction (Charniak 1993). Since HMMs can have more than one arc with the same output symbol, they are nondeterministic, and it is impossible to directly determine the state sequence for a set of inputs simply by looking at the output (hence the “hidden” in “hidden Markov model”). More formally, a HMM is defined as a set of states of which one state is the initial state, a set of output symbols, and a set of state transitions. Each state transition is represented by the state from which the transition starts, the state to which transition moves, the output symbol generated, and the probability that the transition is taken (Charniak 1993). In the context of hand gesture recognition, each state could represent a set of possible hand positions. The state transitions represent the probability that a certain hand position transitions into another; the corresponding output symbol represents a specific posture and sequences of output symbols represent a hand gesture. One then uses a group of HMMs, one for each gesture, and runs a sequence of input data through each HMM. The input data, derived from pixels in a vision-based solution can be represented in many different ways, the Vision based hand gesture recognition for human computer interaction most common by feature vectors (Staner and Pentland 1995a). The HMM with the highest forward probability determines the users’ most likely gesture. An HMM can also be used for hand posture recognition.



HMM is a dual-stochastic process model indicated by parameters, which is used to describe the probability statistic performance of stochastic processes and it is evolved from Markov chain.

According to the state observation probability density continuity, HMM can be divided into continuous hidden Markov model (CHMM) semi-continuous Hidden Markov Model (SCHMM) Discrete Hidden Markov Model (DHMM).

DHMM: Signal as a discrete concept of measurement, first observation to quantify the characteristics of the signal after carrying out identification, this method is characterized by training speed and recognition speed, but the recognition rate is relatively low.

CHMM: Observation features as a continuous probability density function, each state has a set of different probability density function. This method is characterized by relatively high recognition rate, however, identification of training speed and slow, real-time are bad.

SCHMM: Observation vector is a continuous probability density, but the sequence of all samples of the entire state share a group of probability density. DHMM overcome the disadvantage of low accuracy to identify, also reduces the complexity of calculating the CHMM, take into account the real-time and recognition rate.

A Hidden Markov Model, HMM = (S; C; π; A; B) , represents a stochastic process in time, in terms of (hidden) states S, (visible) observations C, initial state probabilities π, state transition probabilities A and output probabilities B.

HMM is one of the best approaches used in pattern recognition as it has the ability
to overcome the problems of spatio-temporal variabilities. In addition, HMMs have been successfully applied to gesture recognition, speech recognition and protein modeling. Introduction of HMMs makes the recognition-based segmentation more powerful because segmentation and recognition are optimized simultaneously during recognition with HMMs.

In a *hidden* Markov model, the state is not directly visible, but output, dependent on the state, is visible. The adjective 'hidden' refers to the state sequence through which the model passes, not to the parameters of the model HMM has three fundamental problems: evaluation, decoding and training.

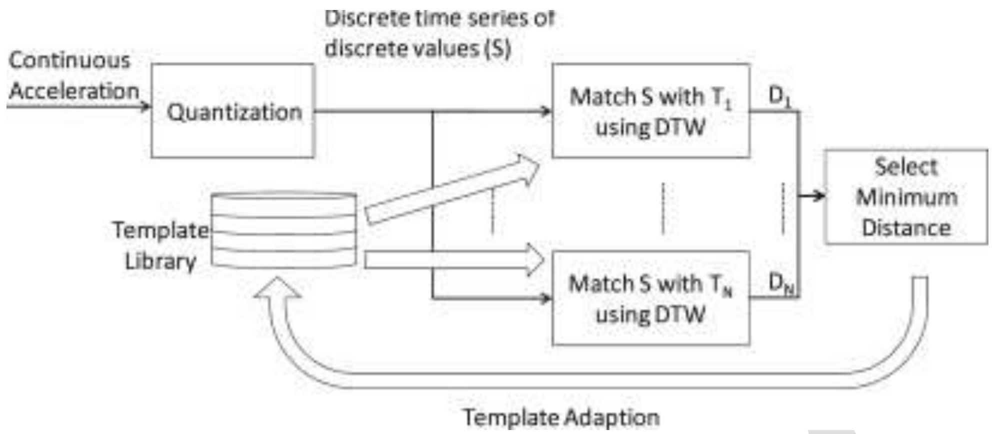
1. Evaluation: The problem is to calculate an output observable symbol sequence or vector O given an HMM parameter set λ. The following problem is solved by using
Forward-Backward algorithm.
2. Decoding: The problem is to determine an optimal state sequence which is associated with the given observable symbol sequence or vector O by a given HMM parameter set λ. The following problem is solved by using Viterbi algorithm.
3. Training: The problem is to maximize the output probability of generating an observable symbol sequence or vector. The following problem is solved by using BaumWelch algorithm.

HMM has three topologies Fully Connected (i.e. Ergodic model) where any state can be reached LR such that each state can go back to itself or to the following states and LRB model in which each state can go back to itself or the next state only. We choose left-right banded model as the HMM topology, because it is good for modelling-order constrained time-series and its properties also change over time in sequence and the number of states are decided on the basis of complexity of a gesture.

One main concern of gesture recognition is how to segment some meaningful gestures from the continuous sequence of hand motion. This is considered as a highly difficult process for two major problems, which arise in real-time gesture recognition system to extract meaningful gestures from continuous gesture . The first problem is the segmentation (spotting) that means how to determine when gesture starts and when it ends in hand motion trajectory. The second problem is caused by the fact that the same gesture varies in shape, trajectory and duration, even for the same person. To overcome these problems, HMM is used in our system because it is capable of modeling spatiotemporal time series of gestures effectively and can handle non-gesture patterns. Moreover, NN and DTW hardly represent the non-gesture patterns. Lee proposed an ergodic model based on adaptive threshold to spot the start and the end points of input patterns, and also classify the meaningful gestures by combining all states from all trained gesture models using HMM.

Dynamic time warping:

It has long been used to find the optimal alignment of two signals. The DTW algorithm calculates the distance between each possible pair of points out of two signals in terms of their associated feature values. It uses these distances to calculate a cumulative distance matrix and finds the least expensive path through this matrix. This path represents the ideal warp—the synchronization of the two signals which causes the feature distance between their synchronized points to be minimized. Usually, the signals are normalized and smoothed before the distances between points are calculated. DTW has been used in various fields, such as speech recognition, data mining, and movement recognition. Previous work in the field of DTW mainly focused on speeding up the algorithm, the complexity of which is quadratic in the length of the series. Examples are applying constraints to DTW, approximation of the algorithm and lower bounding techniques. Eamonn and Pazzani (2001) proposed a form of DTW called Derivative DTW (DDTW). Here, the distances calculated are not between the feature values of the points, but between their associated first order derivatives. In this way, synchronization is based on shape characteristics (slopes, peaks) rather than simple values. Most work, however, only considered one-dimensional series.



A dynamic time warping algorithm is used to perform the time alignment and normalization by computing a temporal transformation allowing the two signals to be matched.

Depth features from human joints are compared through video sequences using Dynamic Time Warping, and weights are assigned to features based on inter-intra class gesture variability.

Feature Weighting in Dynamic Time Warping is then applied for recognizing begin-end of gestures in data sequences Dynamic Time Warping (DTW) is commonly used in gesture recognition tasks in order to tackle the temporal length variability of gestures. In the DTW framework, a set of gesture patterns are compared one by one to a maybe infinite test sequence and a query gesture category is recognized if a warping cost inferior to a given value is found within the test sequence.

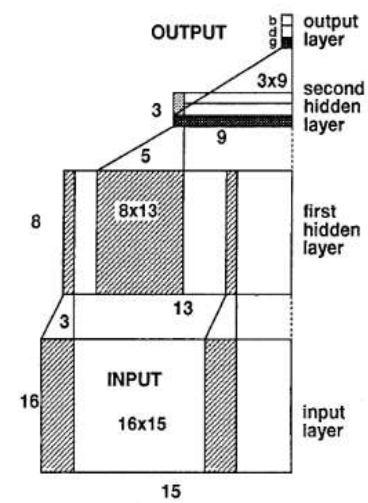
Dynamic time warping (DTW) is a template matching algorithm. To recognize a gesture, DTW warps a time sequence of joint positions to reference time sequences and produces a similarity value.

However, all body joints are not equally important in computing the similarity of two sequences.

DTW method is a timeline based on the dynamic operation of technology; it applies to those with time variable problem. Although it in some small semantic mission was a success the next, but in order to achieve changes in the scope of variables, required a very large number of template matching to calculate the volume too much; but also for the definition of the mode of identification is also not ideal.

Time delay neural networks:

Time delay neural network (**TDNN**) is an alternative neural network architecture whose primary purpose is to work on continuous data. The advantage of this architecture is to adapt the network online and hence helpful in many real time applications.

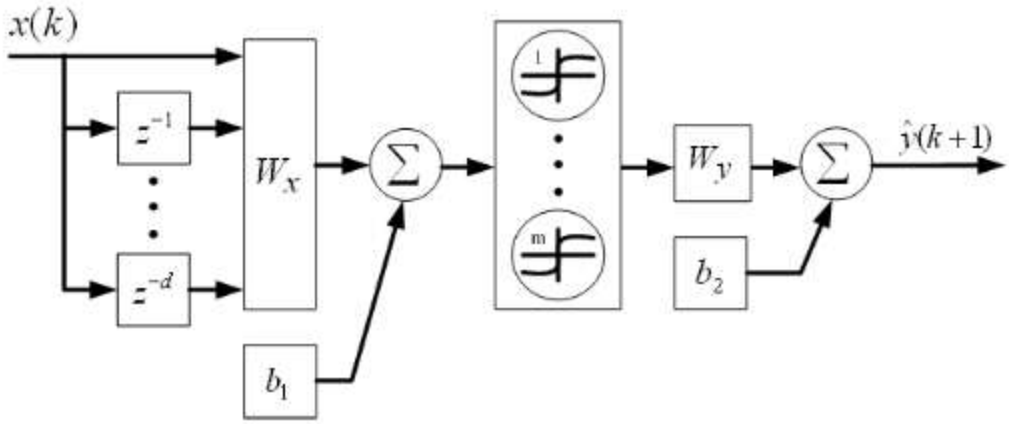


The architecture has a continuous input that is delayed and sent as an input to the neural network. The desired output of the network is the present state of the time series and inputs to the neural network are the delayed time series (past values). Hence, the output of the neural network is the predicted next value in the time series which is computed as the function of the past values of the time series.

These are special artificial neural networks which focus on working with continuous data making the architecture adaptable to online networks hence advantageous to real time applications. Theoretically, time delay neural networks are also considered as an extension of multi-layer perceptron. TDNN is based on time delays which gives individual neurons the ability to store the history of their input signals. Therefore the network can adapt to sequence of patterns. Due to the concept of time delay, each neuron has access not only to present input at time t but also to the inputs at time. Therefore each neuron can detect relationship between the current and former input values which might be a typical pattern in the input signal. Also, the network is able to approximate functions that are derived from time sampled history of input signal. Learning of typical TDNN can be accomplished by standard back propagation as well as its variants.

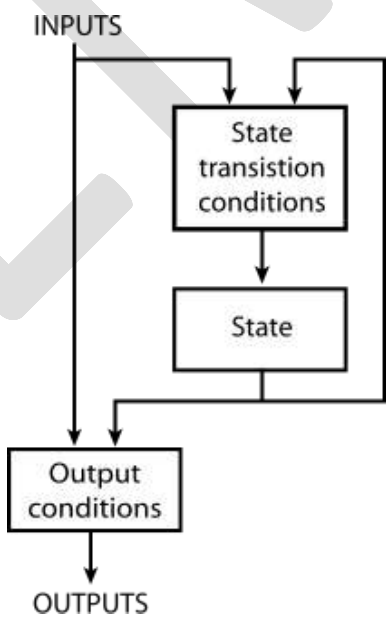
The TDNN is a feed-forward network consisting of three layers: input, hidden, and output layers. Figure illustrates a TDNN. Note that the time delays are applied to the input vector x(k) and presented as inputs to the network. These time-delayed inputs provide the network with temporal information about the system being identified. The hidden layer activation function is a design parameter, which in this study was selected to be the tan-sigmoid function. Other commonly used options are radial basis functions, log-sigmoid, etc. The matrices Wx,Wy, and the bias vectors b are the variable parameters that are updated following a particular training algorithm in order to reproduce or mimic the input-output mapping of the plant. The output of the TDNN is given by.

$$\hat{y}(k+1) = f\left(W_x \times \left[x(k) \quad x(k-1) \quad \cdots \quad x(k-d)\right]^T + b_1\right) \times W_y + b_2$$



Finite state machine:

A finite state machine is one that has a limited or finite number of possible states (an infinite state machine can be conceived but is not practical). A finite state machine can be used both as a development tool for approaching and solving problems and as a formal way of describing the solution for later developers and system maintainers. There are a number of ways to show state machines, from simple tables through graphically animated illustrations. Usually, the training of the model is done off-line, using many possible examples of each gesture as training data, and the parameters (criteria or characteristics) of each state in the FSM are derived. The recognition of hand gestures can be performed online using the trained FSM. When input data (feature vectors such as trajectories) are supplied to the gesture recognizer, the latter decides whether to stay at the current state of the FSM or jump to the next state based on the parameters of the input data. If it reaches a final state, we say that a gesture has been recognized.

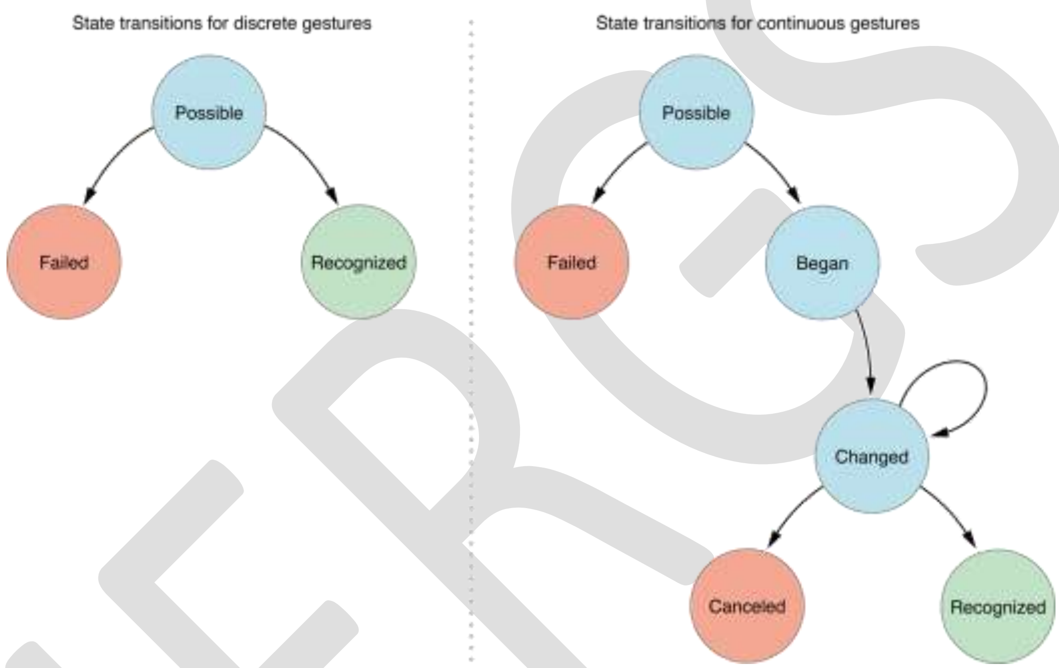


Compared to the HMM based systems, while the number of states and the structure of the HMM must be predefined, in our proposed approach, a gesture model is available immediately. The statistical nature of an HMM precludes a rapid training phase. To train an HMM, well-aligned data segments are required, whereas in the FSM representation the training data is segmented and aligned

ISSN 2091-2730

simultaneously to produce a gesture model. Another advantage of using FSM is that it can handle gestures with different lengths/states. The only one input to the FSM is the spatio-temporal variance, which produces the recognizer after some training sessions.

FSM is more robust to spatio-temporal variability of incoming gesture sequence,
where FSM adapts quickly to accommodate this variability during the training phase. Unlike other template matching algorithms, where an image sequence is first converted into a static shape pattern, and then compares it to prestored action prototypes during recognition, recognizer only compares the selected key frames of the incoming video sequence with the existing key frames in FSMs. The key frame based shape comparison greatly enhances the recognition speed.



Comparison between different Gesture Recognition Techniques:

Technique	Principle	Parameter	Advantages	Disadvantage
Hidden Markov Model	Generalization of a markov chain without markov chain restriction. Set of status, transitions represent the set of possible hand positions	Pixels in vision-based input	Easily extended to deal with strong TC tasks. Embedded re-estimation in possible easy to understand	Large assumption about the data. Huge no. Of parameters needs to be set. Training data required is large.

Dynamic Time Warping	Optimal alignment of features is found & ideal warp is obtained based on cumulative distance matrix	Shape characteristics	Reliable time alignment. Robust to noise. Easy to implement	Complexity is quadratic. Distance matrix need to be defined. DTW to strong TC tasks not achieved.
Time Delay Neural Network	Special artificial neural network based on time delays giving individual neurons ability to store history making the system adapt to sequence of patterns.	Time sampled history of input signal	Faster learning. Invariance under time or space translation. Faster execution	Lacking robustness. Based on typical patterns of input that is inconsistent
Finite State Machine	Limited or infinite no of possible states	Feature vector such as trajectories	Easy to implement. Efficient predictability. Low processor overhead	Not robust. Ridged condition of implementation

Applications areas:

Application	Principle	Substitute	Purpose
Information Visualization	Interact data with more natural way	Mouse and keyboard	Scientific visualization. Audio-video presentation.
Desktop application	Interact with computer generated environment in more natural way	Mouse and keyboard	Wearable computers
Sign language	Automatic translation	Camera based devices	Communication system for disabled
Robotics	Control and program robots	Mouse and keyboard	Mobile robots. Multi DoF robots. Programming by demonstrations.
Virtual reality	Provide realism to man-made environment	Data glove	Realistic manipulation of virtual object.
Ubiquitous computing	Use signals to transfer data between different devices	Remote contact devices	Smart room, interaction
Games	Interactions with player's hands and body positions	Mouse and keyboard	Feels on field effect

Conclusion:

In today's digitized world, processing speeds have increased dramatically, with computers being advanced to the levels where they can assist humans in complex tasks. Yet, input technologies seem to cause a major bottleneck in performing some of the tasks, under-utilizing the available resources and restricting the expressiveness of application use. Hand Gesture recognition comes to rescue here. Computer Vision methods for hand gesture interfaces must surpass current performance in terms of robustness and speed to achieve interactivity and usability.

The importance of gesture recognition lies in building efficient human–machine interaction. Its applications range from sign language recognition through medical rehabilitation to virtual reality. Constructing an efficient hand gesture recognition system is an important aspect for easily interaction between human and machine. In this work we provided a comparative study on various gesture recognition systems with emphasis on detection, tracking and recognition phases which are essential for gesture detection and extraction.

Over the last decade numerous methods for gesture taxonomies and representations have been evaluated for the core technologies proposed in the gesture recognition systems. However the evaluations are not dependent on the standard methods in some organized format but have been done on the basis of more usage in the gesture recognition systems. Hence the analysis of the detailed survey presented in the paper states the fact that the appearance based gesture representations are more preferred than the 3D based gesture representations in the hand gesture recognition systems. Though there are vast amount of information and research publications available in both the techniques but due to complexity of implementation the 3D model based representations are less preferred.

In this paper various methods are discussed for gesture recognition, these methods include Time Delay Neural Network, HMM, Dynamic Time Warping, Finite State Model. For dynamic gestures, HMM tools are perfect and efficient especially for robot control. TDNNs are used as classifier and for capturing hand shape. The selection of specific algorithm for recognition depends on the application needed. FSM is a simple to implement but lengthy in operation. DTW is advantageous in continuous recognition applications.

REFERENCES:

[1] Prateem Chakraborty, Prashant Sarawgi, Ankit Mehrotra, Gaurav Agarwal, Ratika Pradhan "Hand Gesture Recognition: A Comparative Study" Proc. The International MultiConference of Engineers and Computer Scientist 2008 Vol. 1

[2] Ginu Thomas "A review of Various Hand Gesture Recognition Techniques" Proc. VSRD-IJEECE, Vol. 1(7)2011

[3] Shah Muhammed Abid Hussain and A.B.M. Harun-ur Rashid "User Independent hand gesture recognition by Accelerated DTW" IEEE/OSA/IAPR 2012

[4] Frederic Bettens, Todor Todoroff "REAL-TIME DTW-BASED GESTURE RECOGNITION EXTERNAL OBJECT FOR MAX/MSP AND PUREDATA" Proc. SMC 2009

[5] M.K.Bhuyan "FSM Based recognition of dynamic hand gesture" Proc. International Journal of Computer and Communication Engineering 6 2012

[6] Rohit Verma, Ankit Dev "Vision Based Hand Gesture Recognition using Finite State Machine" Proc. IEEE 2009

[7] Wenjun Tan, Chengdong Wu, Shuying Zhao, Shuo Chen "Recognition of Dynamic Hand Gesture Based on SCHMM Model" Proc. IEEE 2009

[10] Rajat Shrivastava "A Hidden Markov Model based Dynamic Hand Gesture Recognition System using OpenCV" Proc. IEEE 2012

[11] Wu Xiaojuan, Zhao Zijian "DYNAMIC GESTURE TRACK RECOGNITION BASED ON HMM" Proc. IEEE 2009

[12] Ruiduo Yang and Sudeep Sarkar "Gesture Recognition using Hidden Markov Models from Fragmented Observations" Proc. IEEE2009

[13] Noor A. Ibraheem , Rafiqul Z. Khan "Gesture Recognition using Hidden Markov Models from Fragmented Observations" Proc. IEEE 2012

[14] Paul Modler and Tony Myatt "Recognition of Separate Hand Gestures by Time delay Neural Networks Based on Multi-state Spectral Image Patterns from Cyclic Hand movement" Proc. IEEE 2008

[15] Senin P (2008) Dynamic time warping algorithm review, technical report. http://csdl.ics.hawaii.edu/ techreports/08-04/08-04.pdf

[16] Triesch J, Von der Malsburg C (1998) A gesture interface for human-robot-interaction. In: Proceedings of the international conference on automatic face and gesture recognition (FG). IEEE, Nara, Japan, pp 546–551

[17]Tseng KT, Huang WF, Wu CH (2006) Vision-based finger guessing game in human machine interaction. In: IEEE international conference on robotics and biomimetics, pp 619–624

[18] Chen Q, Georganas ND, Petriu M (2008) Hand gesture recognition using Haar-like features and a stochastic context-free grammar. IEEE Trans Instrum Meas 57(8):1562–1571

[19] Cheng J, Xie X, Bian W, Tao D (2012) Feature fusion for 3D hand gesture recognition by learning a shared hidden space. Pattern Recogn Lett 33:476–484