A Survey on Hand Gesture Recognition Systems

Ankith A. Prabhu
Research Scholar
CSE Department
SRM Institute of Science and Technology

E. Sasikala
Assistant Professor
CSE Department
SRM Institute of Science and Technology

ABSTRACT
Gesture recognition can be said to be the interpretation of human gestures via mathematical models. Gestures can originate from any bodily motion but this paper focuses exclusively on hand gesture recognition. Hand gesture recognition is referred to as a Perceptual User Interface (PUI). A Perceptual User Interface allows Human Computer Interaction (HCI) without the use of a mouse or a keyboard. Gestures are used primarily to interact with devices without any physical contact with the device. Successful gesture recognition is dependent on the accuracy and efficiency of gesture classification. The gestures are classified using dynamic programming, machine learning or deep learning techniques. In all gesture recognition systems, the relevant input data is collected by a number of sensors. A good gesture recognition system uses this input data to classify the gesture accurately and efficiently. Gesture recognition is deployed in a number of fields like in the medical field where is used to make sign language interpretation devices for the vocally impaired, in virtual gaming and in smart home environments.

General Terms
Gesture recognition; Machine Learning Algorithms; Accelerometers; Dynamic Time Warping.

Keywords
Multi-modal; User Dependent; User Independent; Mixed User.

1. INTRODUCTION
Gesture recognition can be defined as the identification of non-verbal communication by using mathematical, statistical and probabilistic methods. Advances in micro-electrics over the last twenty years have made it possible to mass-produce very accurate and low cost sensors. This has led to the field of gesture recognition to grow at an exponential rate. A modern gesture recognition system must be accurate and quick in classifying gestures with minimum computational time.

Gesture recognition can be classified based on the method used to collect the input data as either vision based like shown in [1] by Chen et al which used cameras or motion based like shown in [2] by Biswas et al which makes use of infrared cameras to track the gesture. Vision based systems require proper lighting, relatively expensive hardware and computationally intensive algorithms while motion based systems use sensors like accelerometers, flex sensors and gyroscopes to collect the input data.

Hand gestures can be divided into two distinct groups as mentioned in [3]. These are static gestures and dynamic gestures. Identification of static gestures requires only knowledge of its start and end points while identification of dynamic gestures requires the entire motion to be captured.

2. SENSORS
Gesture recognition systems use a variety of sensors to collect input data from the gesture being performed. The common sensors used to collect data are accelerometers used in [4], [6], [7], [8] and [10]. Tri-axis accelerometers are commonly used owing to their low cost and low power requirements. In [4] and [6], the accelerometer used was from a wii-remote and a smartwatch.

In addition to accelerometers, other sensors such as contact sensors [10], gyroscopes [12], flex sensors [10] and multichannel electromyography sensors (EMG) [9]. These systems make use of a glove [11] with the sensors embedded in it to collect the gesture data.

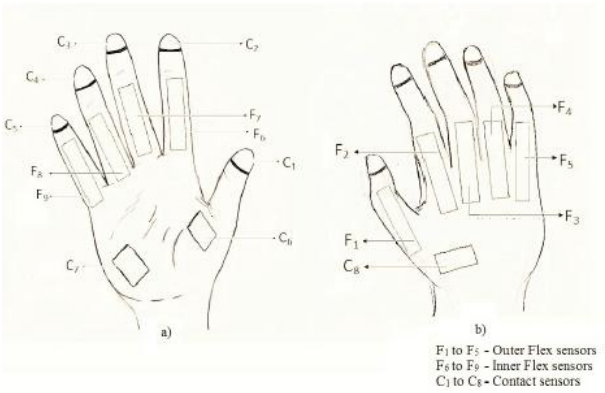

Figure 1. Sensor Glove

Apart from this, [1],[13] use cameras and [2] uses infrared sensors.

3. HAND GESTURES
Hand gesture recognition systems are trained to identify unique and distinguishable gestures accurately. The systems specified in uWave [4] and Sony [5],[6] are designed to classify dynamic gestures exclusively.

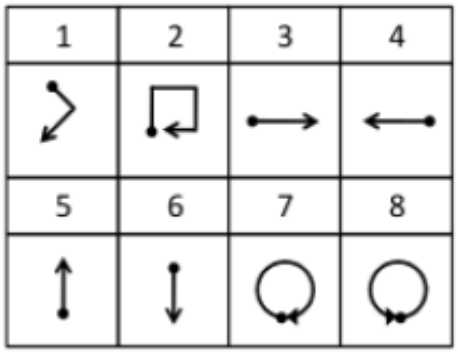

Figure 2. uWave

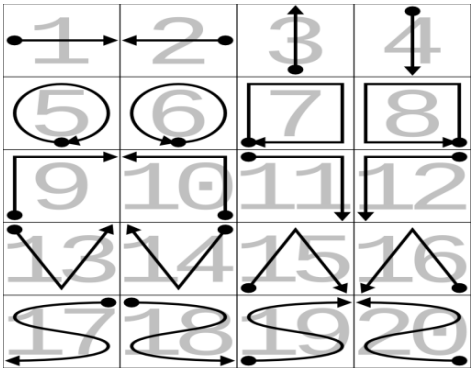

Figure 3. Sony

The uWave and Sony datasets are publicly available free of cost. The uWave dataset (D_u) consists of 8 users who perform 8 different gestures 10 times each every day for 7 days. The Sony dataset (D_S) consists of 8 users who perform 20 different gestures 20 times each.

Table 1. Datasets

Dataset	**Users**	Gestures	**Samples**	**Days**	Total Instances
uWave	8	8	10	7	4480
Sony	8	20	20	1	3200

In addition to the above mentioned gesture recognition systems, the system implemented in [7] correctly classifies American Sign Language (ASL). ASL is a combination of both dynamic and static gestures. J and Z are the only dynamic gestures in ASL, all the other 24 gestures are static gestures. The Hand-Talk [7] dataset consists of all 26 ASL characters performed by 5 individuals.

Figure 4. ASL

4. MODES OF OPERATION

In general, there are 3 distinct modes of operation for gesture recognition systems. These are User Dependent, Mixed User and User Independent. The User Dependent mode ensures that the train-test split is between the gestures of a single user. The Mixed User mode represents the complete set of gestures of all participants. The User Independent mode employs a stratified k-fold cross verification that trains on a number of users and tests on the rest.

5. CLASSIFICATION TECHNIQUES

The field of hand gesture recognition is experiencing rapid growth and as such there are a multitude of different unique methods used for classifying the gestures with new innovative methods being discovered everyday. Here we shall examine some of the common methods used in hand gesture classification.

5.1 Dynamic Time Warping

Dynamic Time Warping (DTW) is a common dynamic programming paradigm used in gesture recognition. It stores a prototypical version of each gesture in a vocabulary or look-up-table. Each instance of this vocabulary is called a template. Each template is a sequence of feature vectors. The incoming gesture is compared to each template to find the closest match. DTW can find the optimal match even for temporally stretched and compressed sequences.

Consider the two temporal sequences to be arranged perpendicular along the sides of a grid, with the known (template) along the side and the unknown (input gesture) along the bottom. It is clearly visible that the 2 temporal sequences are not of the same length. This is possible since even though the sampling rate of the sensors is constant, the speed at which the gesture is performed varies from person to person and gesture to gesture.

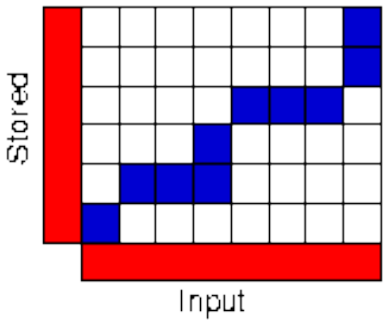

Figure 5. DTW

The best match between two sequences is the path which minimizes the distance between them; this is given in blue in the figure. Here it can be seen that the fourth element of the input matches both the second and third elements of the stored sequence: here a section of the stored sequence has been compressed in the input sequence. Once an overall best path has been found the total distance between the two sequences can be calculated for this stored template. The procedure for computing this overall distance measure is to find all possible routes through the grid and for each one of these compute the overall distance. It should be apparent that for any reasonably sized sequences, the number of possible paths through the grid will be very large. The DTW algorithm is designed to exploit some observations about the likely solution to make the comparison between sequences more efficient.

The hand gesture recognition systems mentioned in [4] and [14] use DTW to match sequences. First the system has to be trained for it to classify the gesture in real-time. This is done in the training mode. During the training mode the templates are defined and the look-up-table is created. During the default mode, the user performs a gesture. This gesture is then compared to all the templates available in the look-up-table to ascertain the optimal match.

The uWave system implemented in [4] implements some template adaptation techniques to improve classification accuracy. It is able to achieve an accuracy of 98.6% when operated on user dependent mode and 75.4% in user

independent mode. The system implemented in [14] shows accuracies of 99.2% in user dependent mode and 96.4% in user independent mode. Although DTW is a fairly simple and straightforward technique to classify hand gestures, it is very computationally intensive requiring an enormous amount of memory to find the optimal match.

5.2 Deep learning

Gestures can be classified using artificial neural networks. The systems proposed in [7] and [12] use neural networks to classify gestures.

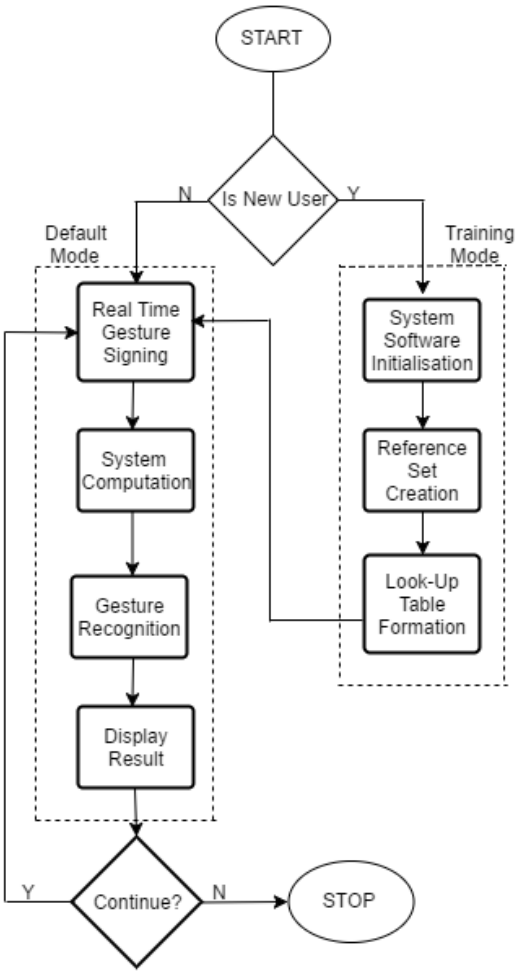

Figure 6. DTW based System Operation

In general, neural networks are a group of computational units called neurons that are interconnected through weighted connections. Each neuron has an associated activation function and the neuron only produces output if its value is more than a certain threshold value. The neural network is divided into 3 layers. The layers being 1 input layer, 1 output layer and the hidden layers. The single input layer takes the input from the sensors, the output layer indicates what class the gesture belongs to and the hidden layers are where the computation takes place.

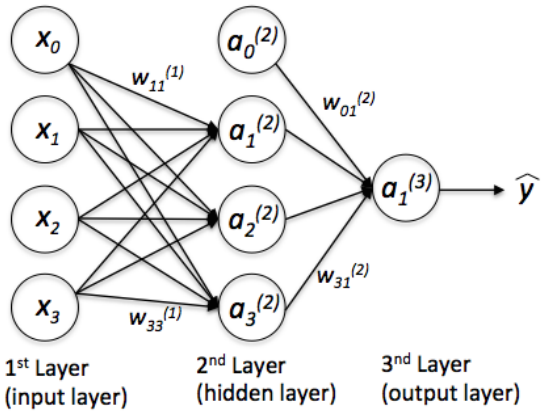

Figure 7. Feedforward Network

Initially the system has to be trained using the training data for it to be able to classify the gesture in real-time. During training, the weights are randomly chosen and the input is given to the system. Labeled or training data is used for this purpose. Labeled data is any input data for which the output is known beforehand. Thus the observed output generated is compared to the expected output and the weights are adjusted to minimize error. This is repeated multiple times till the difference between the expected output and actual output is minimal. This method is called back-propagation.

After the training phase, the system can be used to classify gestures in real time. The systems implemented in [7] and [12] make use of an array of sensors including accelerometer, flex sensor, contact sensor and electromyography sensor to collect data from the gesture being performed. These systems use gloves fitted with these sensors that transform hand and finger movement into real-time data for the gesture recognition system. This data is fed into the neural network and the output is observed from the output layer. The systems mentioned in [7] and [12] achieve accuracies of 95% and 92.57% in user independent mode.

6. CONCLUSION

It is evident from the various systems studied that the field of hand gesture recognition is evolving at a rapid rate with numerous unique methods being employed to accurately classify gestures. The field of gesture recognition is moving towards more robust and hi-tech implementations mainly involving the use of computer vision together with deep reinforcement learning in smart gesture recognition systems.

Table 2. System Comparisons

System	Training Data	User Dependant Efficiency	User Independent Efficiency	Technique
[4]	4480	93.5%	75.4%	DTW
[14]	3200	99.2%	96.4%	DTW
[7]	130	95%	95%	Neural Network
[12]	1540	85%	91%	Neural Network

7. ACKNOWLEDGMENTS

The authors would like to acknowledge SRM Institute of Science and technology for its support of the research carried out for this survey by making labs and workspace available to the authors.

8. REFERENCES

[1] Q. Chen, N. D. Georganas and E. M. Petriu, "Real-time Vision based Hand Gesture Recognition Using Haar-like Features,"2007IEEE Instrumentation & Measurement Technology Conference IMTC 2007.

[2] K. K. Biswas and S. K. Basu, "Gesture recognition using Microsoft Kinect," The 5th International Conference on Automation, Robotics and Applications, Wellington, 2011, pp. 100-103.

[3] S. Mitra and T. Acharya, "Gesture Recognition: A Survey," in IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 37, no. 3, pp. 311-324, May 2007.

[4] Jiayang Liu, Zhen Wang, Lin Zhong, Jehan Wickramasuriya, and Venu Vasudevan, "uWave: Accelerometer-based personalized gesture recognition and its applications," in *IEEE Int. Conf. Pervasive Computing and Communication (PerCom)*, March 2009.

[5] SmartWatch Gestures Dataset, Technologies of Vision, Fondazione Bruno Kessler, [Online] Available: https://tev.fbk.eu/technologies/smartwatch-gestures-dataset.

[6] Gabriele Costante, Lorenzo Porzi, Oswald Lanz, Paolo Valigi, Elisa Ricci, Personalizing a Smartwatch-based Gesture Interface With Transfer Learning, 22nd European Signal Processing Conference (EUSIPCO), 2014.

[7] Anetha K and Rejina Parvin J. (2014, Jul). Hand Talk-A Sign Language Recognition Based On Accelerometer and SEMG Data. IJIRCCE. [online]. 2(3), pp.206-215.

[8] Pylvänäinen T. (2005) Accelerometer Based Gesture Recognition Using Continuous HMMs. In: Marques J.S., Pérez de la Blanca N., Pina P. (eds) Pattern Recognition and Image Analysis. IbPRIA 2005. Lecture Notes in Computer Science, vol 3522. Springer, Berlin, Heidelberg.

[9] X. Zhang, X. Chen, Y. Li, V. Lantz, K. Wang, and J. Yang, "A framework for hand gesture recognition based on accelerometer and emg sensors," IEEE Trans. Syst., Man, Cybern. A, vol. 41, no. 6, pp. 1064–1076, 2011.

[10] H. Sekar, R. Rajashekar, G. Srinivasan, P. Suresh and V. Vijayaraghavan, "Low-cost intelligent static gesture recognition system," 2016 Annual IEEE Systems Conference (SysCon), Orlando, FL, 2016, pp. 1-6.

[11] T. Chouhan, A. Panse, A. K. Voona and S. M. Sameer, "Smart glove with gesture recognition ability for the hearing and speech impaired," 2014 IEEE Global Humanitarian Technology Conference - South Asia Satellite (GHTC-SAS), Trivandrum, 2014, pp. 105-110.

[12] Lefebvre G., Berlemont S., Mamalet F., Garcia C. (2013) BLSTM-RNN Based 3D Gesture Classification. In: Mladenov V., Koprinkova-Hristova P., Palm G., Villa A.E.P., Appollini B., Kasabov N. (eds) Artificial Neural Networks and Machine Learning – ICANN 2013. ICANN 2013. Lecture Notes in Computer Science, vol 8131. Springer, Berlin, Heidelberg.

[13] Starner T., Weaver J., Pentland A., "Real-Time American Sign Language Recognition Using Desk and Wearable Computer Based Video", IEEE TPAMI, IEEE CS, Vol 20, Issue 12, Dec 1998, pp. 1371-1375.

[14] Hussain S.M.A., Harun-ur Rashid A.B.M., User Independent Hand Gesture Recognition by Accelerated DTW, IEEE/OSA/IAPR International Conference on Informatics, Electronics & Vision, Proceedings, Dhaka, Bangladesh2012.