Performance Analysis of KNN, SVM and ANN Techniques for Gesture Recognition System

B. P. Pradeep Kumar^{1*} and M. B. Manjunatha²

¹Department of ECE, Jain University, Bangalore - 560069, Karnataka, India; pradi14cta@gmail.com ²Department of ECE, AIT, Tumkur - 572106, Karnataka, India; manju.kari29@gmail.com

Abstract

The Gesture identification or recognition system is an emerging area of a current dynamic research about in PC vision and machine learning concepts. The primary design of the projected method is to develop a framework; it can understand the American Sign Language, which can recognize exact human gestures and practice them to express the information. Gesture signals are captured utilizing Kinect cameras along with a static back ground. The proposed strategy is isolated into 2 sections, significant features extraction stage, and Classification stage. Patterns are calculated using KNN, SVM and finally the ANN; it is a classification phase. It is then subjected to feature extraction and classification. In present work, SVM, KNN, and ANN tools are used in recognizing the gestures of ASL. The gesture recognition system is capable of recognizing only numerical ASL static signs with 97.10% accurateness. The obtained results proved that the SVM had a very good performance compared to other techniques.

Keywords: American Sign Language, ANN, Gesture recognition, KNN, SVM

1. Introduction

Recently, progress in technology is booming in Human-Computer Interaction (HCI) research¹. There are many methods for interaction, where, camera-based gesture recognition is used much to have a natural interaction method². Gesture recognition is a territory of dynamic research in PC, machine level learning¹. Gesture recognition, being a characteristic method for human PC communication, is a region where numerous specialists in the scholarly community and industry are chipping away at various applications to make associations all the more simple, normal and advantageous without wearing any additional devices².

Previous researches used a 2D image, but it lacks the robustness of following environmental changes which make depth based camera research to take a spotlight. Microsoft Kinect sensor gives joint introduction data to the skeletons tracked; that can be easily gestured recognition research³. But, previous research showed an increase in the database makes calculation slow, complex

statistical method is needed. Daily used gestures, such as games, medical system, the interplay between man and robot, hand signal, sign language, and so on. Those gestures have many of data distribution in each class⁴. To make a widely spread hand gesture recognition effectively, we present a method with a simple calculation that users can make easy feed-forward neural networks structure that has a learning algorithm using back-propagation.

A few algorithms exist that can be connected to the undertaking of gaining from hand highlights. There have been changed studies close by gesture recognition going from Hidden Markov Models, Dynamic Time Warping, Support vector machine, Artificial Neural Networks. Disregarding the way that HMM has rich numerical structure and its factual model of succession information that is appropriate in different application field, for example, discourse flag and signal acknowledgment, yet it has inconvenience in the discretization of multi-dimensional information changed over to onedimensional information. DTW is a calculation for measuring time and speed distinction between two arrangements Disregarding the way that has a frailness of information increment to operation increment. SVM is a regulated learning model for gesture acknowledgment, yet its rate and dimension has a point of confinement in learning and testing stage.

Artificial Neural Networks (ANN) is regularly sorted out in layers. NN prepares and upgrades network until the yield and target are coordinated⁵. When the system is trained, recognition and grouping can be utilized as a part of dataset test. It has focal points which require less preparing and conceivable to have non-straight characterization^{6,9}. There is an issue with over fitting and imbalanced information; however, much research has been completed on the issue^{4,7}. ANN is a non-parametric representation that is less demanding and quicker to embrace than different models.



Figure 1. General structure of ANN.

Figure 1 demonstrates the stretched nodes are added to the information nodes. At begin with, we form preparing dataset and test dataset to seen hand signal maintenance in mind the end objective to the rough arrangement. Handle every dissemination highlight of motion direction, which are a piece of the preparation dataset. Here, we utilized a normal purpose of direction and end purpose of direction for appropriation include. Compute every class of signal's illustrative esteem, which is expressible by directions and covariance for handled circulation highlight by utilizing a factual model. Calculated agent value is Ak and covariance is Sk, where k is the class of motion. This factual information is a standard motion point for inexact characterization. Remove dissemination includes B from a hand motion that made out of just motion direction information. Such a strategy was already utilized for the

process the conveyance includes from the preparation dataset. At the estimated characterization, more than once figures Mahalanobis removes between separated Ak and B of a hand signal as every k class. As a signal has a place with certain class, Mahalanobis separation of the class get to be shorter, then make N as an expanded hub, where N is a list of first closest k class motion. Another augmented hub is created with direction extremes.

If signal holds an extensive number of waving, it might attack different class that can tumble off in acknowledgment execution. Along with a particular objective is to decrease this issue, Gaussian channel bit for smoothing is separated hand direction from motion. We utilize the first subsidiary of the smoothed direction to ascertain quantities of nearby extrema. Also, form extrema as developed hub from the computation. If there is to cover in information appropriation of every class of motion, acknowledgment execution will tumble off⁴. Be that as it may, basic figuring can make adjust in information with creating amplified hub, with surmised grouping information and quantities of extreme.

The next Section in paper is as per the following. To begin with related work is in Segment 2. Segment 3 presents the proposed method along with the implementation techniques, flowchart, system design architecture, etc. In Section 4 we can observe the result obtained in this research work along with the discussion and comparison of all the three algorithms. Finally, the conclusions and future work are presented.

2. Literature Review

There are numerous studies on hand gesture recognition and systems all around displayed.

Author in¹ proposed another strategy for, visionbased acknowledgment of element signs comparing to American Sign Language words. Another strategy is proposed for key casing extraction which is more specific than the current techniques. The edges relating to MCPs of the worldwide direction are taken as the key casings. The technique obliges the spatial transient changeability that may happen when diverse people play out a similar motion. Our technique when contrasted and three other existing strategies have given better execution. The element extraction strategy has ended up being scale invariant and interpretation invariant.

Author in² proposed a programmed motion acknowledgment approach for ISL. Indian gesture-based

communication utilizes both hands to speak to every letter in order. We propose an approach which addresses nearby worldwide equivocalness recognizable proof, between class inconstancies improvements for every hand signal. Hand area is portioned and recognized by YCbCr skin shading model reference. To order every hand pose, multi-class non-direct bolster vector machines (SVM) is utilized, for which an acknowledgment rate of 91.3% is accomplished. The execution of the proposed approach is broke down with surely understood classifiers like SVM, KNN, and DTW. Trial results are contrasted and the routine, and existing calculations to demonstrate the better effectiveness of the proposed approach.

Author in³ built up a communication via gestures acknowledgment framework can greatly affect people with listening to inabilities. Perceiving motions from the ASL with a camera can be troublesome because of multifaceted nature of different motions. The inspiration driving the paper is to build up a way to deal with effectively group motions in the ASL under equivocal conditions from static pictures. A novel approach including the decay of motions into solitary or twofold gave signal has been displayed in this paper. Arranging motion into these subcategories rearranges the procedure of signal acknowledgment in the ASL because of the nearness of lesser number of motions in each subcategory. Different methodologies making utilization of Histogram of Gradients (HOG) highlights and geometric descriptors utilizing KNN and SVM classifiers were attempted on a dataset comprising of pictures of every one of the 26 English letters in order introduces in the ASL under Factor foundation. Hoard highlights when grouped with Support Vector Machine were observed to be the most productive approach bringing about an exactness of 94.23%.

Author in⁴ proposed a system involve recognition of ISL using Leap Motion Sensor. The device extracts significantly characters/attributes from the dataset such that high-accuracy angle between fingers cab is retrieved, and this can be used for robust prediction. Then, System maps all the frames recognized by the sensor to attributes. Feature classifier is used to identify which gesture is being performed. Here, Random Forest (RF) classifier model is used as feature classification. Our method trains an RF to recognize hand gestures for given set of data presented in training dataset. A block-list classifier is used to remove all non-gestures frames, after which accurate gesture frames are recognized by RF to form a meaningful word. This meaningful word is then displayed on LCD. Our proposed method outweighs other state-of-art methods for gesture recognition.

Author in⁵ have used one of the means for acquiring hand Signs like Web Camera, instrumented Glove, Depth Camera, and Kinect sensor. In this paper, an American Sign Language recognition system is developed which recognize using Leap Motion sensor. This sensor overcomes the major issues in real time environment like the background, lightening condition, and occlusion. The leap motion sensor captures the hand gesture and gives finger position in 3D format (X, Y, Z axis values). While performing a sign, the Leap Motion Camera is kept about 10 degrees inclined so that depth information is properly extracted. Although Leap Motion sensor tracks both the hand accurately it can't track non-manual signs which involve other body parts and facial expressions.

Author in⁶ proposed a novel unbroken ASL gesture recognition technique where Possibility Theory (PT) has been applied. Pre-processing and extraction of overlapping frames (start and end point of each gesture) are the major issue which is being covered in this paper using background modeling and noble gradient method. Overlapping frames are helpful for fragmenting a continuous ASL gesture into isolated gestures. These isolated gestures are further processed and classified. During the segmentation process, some of the gesture structures like shape and orientation of hand are deformed. A novel concept of wavelet descriptor has been applied here for extracting correct features of these deformed features and combines with the other two features (orientation and speed). Wavelet descriptor is very generic for finding a moment invariant features from the shape of the hand due to its multiresolution property. 3-dimensional feature vectors (orientation, speed and moment) are parallelly combined and classified using possibility theory. Possibility theory is suited for handling uncertainty as well as the precision present between intermediate frames.

Author in⁷ proposed framework having four modules, for example, pre-preparing and hand division, highlight extraction, sign acknowledgment and sign to content and voice change. Division is finished by utilizing picture handling. Distinctive elements are extricated, for example, Eigen qualities and Eigen vectors which are utilized as a part of acknowledgment. The Principle Component Analysis (PCA) calculation was utilized for motion acknowledgment and perceived signal is changed over into content and voice design. The proposed framework minimizes correspondence boundary between hard of hearing idiotic individuals and ordinary individuals.

Author in⁸ propose the technique for programmed include extraction of the pictures of hand. These extricated elements are then used to prepare the Softmax classifier to group them into 20 classes. Five Stacked Denoising Sparse Auto Encoders (DSAE) prepared in unsupervised form are utilized to concentrate highlights from the picture. The proposed design is prepared and tried on a standard dataset¹ which was stretched out by including arbitrary nerves, for example, turn and Gaussian clamor. The execution of the proposed design is 83% which is superior to anything shallow Neural Network prepared on manual hand-built elements called Principal Components which is utilized as a benchmark.

Author in⁹ proposed a visual ASL acknowledgment framework utilizing picture handling, PC vision, and neural system strategies. This approach will change over video of day by day regularly utilized full sentences signal into content and afterward, change it into sound. Distinguishing proof of hand shape from persistent casings will be finished by utilizing an arrangement of picture handling operations. Understanding of signs and comparing significance will be recognized by utilizing Haar Cascade Classifier. At long last showed content would be changed over into discourse utilizing discourse synthesizer.

Author in¹⁰ proposed technique incorporates both neighborhood and also worldwide data of the dynamic sign. An even separate based technique utilizing the seven 3D key focuses (five comparing to every fingertip, one relating to the focal point of the palm and another relating to lower some portion of palm), extricated utilizing Kinect is proposed for neighborhood highlight extraction. Coordinating 3D nearby element has enhanced the execution of the framework as appeared in the outcome. Aside from serving as a guide to the crippled individuals, different uses of the framework likewise, incorporate serving as communication through signing coach, translator furthermore be useful in electronic frameworks that take motion contribution from the clients.

Author in¹¹ propose a half and half framework approach for finger spelling acknowledgment utilizing RGB-D data from Kinect sensor. In a first stage, the hand territory is sectioned from foundation utilizing profundity outline exact hand shape is separated utilizing both profundity information and shading information from Kinect sensor. Propelled by the execution of bit based components, because of its straightforwardness and the capacity to transform any pixel property into fix level elements, we chose to utilize the angle portion descriptor for highlight extraction from profundity pictures. The Scale-Invariant Feature Transform (SIFT) is utilized for portraying the substance of the RGB picture. The execution subjectively assessed on a dataset of genuine pictures of American Sign Language (ASL) hand shapes. Three analyses were performed, utilizing a blend of RGB and profundity data furthermore utilizing just RGB or profundity data independently.

Author in¹² considered a gesture-based communication is a dialect which utilizes hand motions, outward appearances and body developments for correspondence. A gesture based spelling. It is the main correspondence mean for the hard of hearing the idiotic group. In any case, the listening to individuals never attempts to take in the communication via gestures. So the hard of hearing individuals can't interface with the ordinary individuals without a gesture-based communication mediator. The usage of such a framework gives a stage to the collaboration of listening to handicapped individuals world without a translator. The proposed strategy utilizes advanced picture handling strategies and counterfeit neural system for perceiving diverse signs.

Author in¹³ present a strategy for programmed acknowledgment of two gave indications of ISL. The strategy comprises of Recognition. The division is done through Otsu's calculation. In the element extraction stage, shape descriptors, HOG descriptors (Histogram of Oriented Gradients) and SIFT (Scale Invariant Feature Transform) highlight have been melded to process a component vector. The exploratory results give proof of the adequacy of the proposed approach with 93% acknowledgment rate.

3. Proposed Method and Implementation

This section gives the proposed methodology, and the implementation scheme is used to implement the all the three different algorithms for ASL gesture recognition system. Here, the database is created with ASL symbols. These symbols are going to perform the image preprocessing operation. In this operation the image is converted to gray scale and the segmentation will be done to recognize symbol meaning. Form these symbols the specific information is acquired and stored in a particular database. The statistical features are extracted using super pixel method, SURF method and LBP method. These statistical features are stored in a featured set. After getting the test data form database, form the test data the statistical features are extracted and by using three different classification method like KNN, SVM, and ANN schemes are utilized to recognize an image along with the classification processes. Figure 2 shows the system design architecture of the proposed method.



Figure 2. System designs for proposed method.

3.1 Image Pre-Processing

After collecting the database from users, the images were pre-processed. Firstly the RGB pictures were changed over to dim scale picture by rgb2gray work accessible in Matlab environment. It changes over the real nature picture RGB to the dim scale power picture. We utilized first subsidiary Sobel edge identifier technique since it figures angle by utilizing discrete distinction amongst lines and sections of 3×3 neighbors. Sobel technique discovers edges utilizing the Sobel estimation to the subordinate. Feature Extraction

Feature extraction¹ is a form of dimensionality reduction. Input images are too large for processing, so to

process these images in time we diminish the size of input data by feature extraction. Transforming input data into a feature is called feature extraction. Feature extraction is chosen in such a way that image information must be retained. It is often decomposed into feature construction and feature selection. Feature extraction techniques used in this project are direct pixel value and hierarchical centroid.

3.2 Classification

Extracted features are needed as input for classification. Classification techniques are helpful to distinguish the gestures. Present many numbers of classification techniques available¹. Classification is identifying inputs to a set of class by training data set. Classification techniques strategies, by and large, intend to give a sensible response to every single conceivable info and to perform "doubtlessly" coordinating of the data sources, considering their measurable variety.



Figure 3. Flowchart of the proposed method.

Figure 3 shows the flowchart of the proposed method. It gives the implementation step of the projected technique. In present work, SVM, KNN, and ANN tools are used in recognizing the numeral gestures of ASL. The gesture recognition system is capable of recognizing only numerical ASL static signs with 97.10% accurateness.

4. Results and Discussion

We have acquired all videos a recording using kinetic camera and Matlab Image acquisition tool box is used to acquire the images. We have tested distinctive arrangements of the database for these three strategies. Dark background for obtaining the pictures is chosen. The determination of this foundation is because of consistency in the foundation and its pixel values in catching components. Likewise it is useful in erasing foundation so as to removing critical elements. A Sony computerized camera was utilized for picture securing.

The basic document design JPEG was utilized to catch the pictures. There are 5000 pictures for the database. Every picture is 4608*3456 pixel measure and a 5.5 MB estimate roughly. The goal is to make a productive database with sensible size; the pictures were edited to 200×300 RGB pixels, and every picture needs 25 KB space. Figure 4 show ASL Static Signs for creating a standard dataset and it was gathered from 100 clients out of these 100 clients. This image data base was of ASL symbols. Every endorser contributed five specimens of every individual sign. So an aggregate 5000 signs were gathered. At that point database was partitioned into two gatherings, one set was utilized for preparing, and other was utilized for testing.



Figure 4. ASL considered for the proposed framework¹⁴.

Accuracy for ANN, SVM and KNN is also calculated for all ASL gestures.

The neural system administrator takes in a model by a

method for an Encourage forward neural system prepared by a back spread calculation. For trials the taking after parameters (Table 1) were utilized:

Table 1. Precision for dissimilar methods

Gesture	SVM	KNN	ANN
А	97.33%	95.33%	96.33%
E	94.67%	90.67%	92.67%
Ι	96%	92.33%	94.67%
0	94.67%	91.33%	93%
U	98%	93.33%	95.67%
Z	96%	92.33%	95.33%
Κ	95.33%	93.67%	94.67%
М	93.33%	90.33%	91.33%
Q	96.67%	90.03%	94%
G	99.33%	91.58%	96.33%

(Number of Testing Images=200)

Table 1 shows the precision for different techniques like ANN, SVM and KNN methods. Form this table we can observe that the SVM method gives better precisions values compared to other two methods. For the analysis purposes, we took 200 images in a database. Similarly, Table 2 shows the parameters used with the ANN methods. Here, we took 2 different data set. The parameters like learning rate, momentum rate, training cycle and hidden layers in a network are analyzed for ANN methods.

Table 2. Parameters utilized with the ANN

Parameters	Values		
Learning Rate	Data Set 1	Data Set 2	
Momentum Rate	0.11	0.39	
Training Cycles	1000	1000	
Hidden Layers in the Network	1	1	



Figure 5. Comparison of KNN, SVM, and ANN algorithms regarding Accuracy.

Figure 5 shows the performance analysis of 3 different methods for accuracy, Figure 6 shows the performance analysis of 3 different methods for sensitivity, Figure 7 shows the performance analysis of 3 different methods for precision, and Figure 8 and 9 shows the performance analysis of 3 different methods for sensitivity FNR and FDR parameters. Form this graphs we can observe that the SVM method gives better gesture recognition rate as well as it is very speedily detecting the ASL symbols.



Figure 6. Comparison of KNN, SVM, and ANN algorithms regarding Sensitivity.



Figure 7. Comparison of KNN, SVM and ANN algorithms regarding Precision.



Figure 8. Comparison of KNN, SVM, and ANN algorithms regarding FNR.



Figure 9. Comparison of KNN, SVM and ANN algorithms regarding FDR.

5. Conclusion

Hand gesture recognition is a difficult concern, and the present work is just a little stride towards accomplishing the outcomes required in the field. This paper introduced a performance analysis of the 3 different techniques like KNN, ANN and SVM schemes for hand gesture recognitions system. The outcomes accomplished recognizing hand signals in the two datasets, empower to presume that the dataset 1 qualities result better close by motion distinguishing proof as well as SVM had a decent execution for the current issue. The outcomes acquired additionally demonstrated that the component choice and information arrangement stage is a critical one, particularly with low determination pictures, which is the situation of profundity pictures should be with Kinect camera. The future work will be worried about the investigation of various hand highlight determination connected to hand signal acknowledgment, commotion lessening in the profundity pictures gained, and the presentation of machine learning calculations given a blend of strategies.

6. References

- 1. Geetha M, Aswathi PV. Dynamic gesture recognition of Indian sign language considering local motion of hand using spatial location of Key Maximum Curvature Points. IEEE Recent Advances in Intelligent Computational Systems (RAICS); Trivandrum. 2013. p. 86–91.
- Rekha J, Bhattacharya J, Majumder S. Shape, texture and local movement hand gesture features for Indian sign language recognition. 3rd International Conference on Trendz in Information Sciences and Computing (TISC2011); Chennai. 2011. p. 30–5.
- Singh A, Arora S, Shukla P, Mittal A. Indian sign language gesture classification as single or double handed gestures. 2015 3rd International Conference on Image Information Processing (ICIIP); Waknaghat. 2015. p. 378–81.
- Chavan P, Ghorpade T, Padiya P. Indian sign language to forecast text using leap motion sensor and RF classifier. 2016 Symposium on Colossal Data Analysis and Networking (CDAN); Indore. 2016. p. 1–5.
- Mapari RB, Kharat G. Real time human pose recognition using leap motion sensor. 2015 IEEE International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN); Kolkata. 2015. p. 323–8.

- Baranwal N, Tripathi K, Nandi GC. Possibility theory based continuous Indian sign language gesture recognition. TEN-CON 2015, IEEE Region 10 Conference; Macao. 2015. p. 1–5.
- Sawant SN, Kumbhar MS. Real time sign language recognition using PCA. 2014 International Conference on, Advanced Communication Control and Computing Technologies (ICACCCT); Ramanathapuram. 2014. p. 1412–5.
- Kumar V, Nandi GC, Kala R. Static hand gesture recognition using stacked denoising sparse auto encoders. 2014 Seventh International Conference on Contemporary Computing (IC3); Noida. 2014. p. 99–104.
- 9. Dabre K, Dholay S. Machine learning model for sign language interpretation using webcam images. 2014 International Conference on Circuits, Systems, Communication and Information Technology Applications (CSCITA); Mumbai. 2014. p. 317–21.
- Geetha M, Manjusha C, Unnikrishnan P, Harikrishnan R. A vision based dynamic gesture recognition of Indian sign language on Kinect based depth images. 2013 International Conference on Emerging Trends in Communication, Control, Signal Processing and Computing Applications (C2SP-CA); Bangalore. 2013. p. 1–7.
- Rodríguez KO, Chávez GC. Finger spelling recognition from RGB-D information using kernel descriptor. 2013 26th Conference on Graphics, Patterns and Images; Arequipa. 2013. p. 1–7.
- Adithya V, Vinod PR, Gopalakrishnan U. Artificial neural network based method for Indian sign language recognition. 2013 IEEE Conference on Information and Communication Technologies (ICT); JeJu Island. 2013. p. 1080–5.
- Agrawal SC, Jalal AS, Bhatnagar C. Recognition of Indian sign language using feature fusion. 2012 4th International Conference on Intelligent Human Computer Interaction (IHCI); Kharagpur. 2012. p. 1–5.
- 14. Kumar BPP, Manjunath MB. A hybrid gesture recognition method for American sign language. Indian Journal of Science and Technology. 2017 Jan; 10(1):1–12.