International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

Tamil Sign Language to Speech Translation

S. Sudha, S. Jothilakshmi[1], R. Rajasoundramani[2]

[1, 2]Department of Computer Science and Engineering, Annamalai University, Chidambaram, Tamilnadu, India

Abstract: ***Sign Language Recognition is one of the most growing fields of research today. Most researches on hand gesture recognition for HCI rely on either Artificial Neural Networks (ANN) or Hidden Markov Model (HMM). There are many effective algorithms for segmentation, classification, pattern matching and recognition. The main goal of this paper is to compare the classifiers for translating Tamil sign language to speech, which will definitely help the researchers to attain an optimal solution. The most important thing in hand gesture recognition system is the input features and the selection of classifiers. To increase the recognition rate and make the recognition system resilient to view-point variations, the concept of shape descriptors from the available feature set is introduced. K-Nearest Neighbor (KNN), Proximal Support Vector Machine (PSVM) and Naïve Bayesian are used as classifiers to recognize static Tamil words. The performance analysis of the proposed approach is presented along with the experimental results. Comparative analysis of these methods with other popular techniques shows that the real time efficiency and robustness are better. Experimental results demonstrate the effectiveness of the proposed work for recognizing efficiency 78% for KNN classifier, 91% for PSVM classifier and 93% for Naïve Bayesian classifier.***

Keywords: Artificial Neural Networks (ANN), Hidden Markov Model (HMM), K-Nearest Neighbor (KNN), Naïve Bayes, Proximal Support Vector (PSVM)

1. Introduction

A sign language is a replacement of speech for hearing and mute people. Because of this reason it has provoked many researchers to work in this field. Sign languages offer a much more structured and constrained research environment than common gestures. Moreover, gesture recognition is a tool for the virtual reality environment with his/her hands. There are different sign languages all over the world. Researchers have contributed to different sign languages like American Sign Language (ASL), British Sign Language (BSL), Taiwanese Sign Language (TSL), etc. The application was extended to several international sign languages including Chinese and Arabic [1, 2]. There have been no such distinct contributions for South Indian Languages by any of the researchers in this area. There may be different regional versions available in a particular language.

However, the sign language is common and applicable to any variant of language. This paper deals with a system which recognizes the Tamil Sign Language and to convert it into speech to help people with such disabilities. The mute person becomes neglected from the society because the normal people neither try to communicate nor try to learn sign language. This makes them to feel isolated and they remain uneducated. This paper targets to break the gap between normal people by introducing a Tamil Sign Language which will enable the user to understand the meaning of the sign without the help of any translator. Sensor-based methods, such as data gloves, can provide accurate measurements of hands and movement. Unfortunately, these methods require extensive calibration; they also restrict the natural movement of hands and are often very expensive. Video-based methods are less intrusive, but new problems arise: locating the hands and segmenting them is a nontrivial task. Recently, depth cameras have become popular at a commodity price. Depth information makes the task of segmenting the hand from the background much easier.

Depth information can be used to improve the segmentation process, as used in [3], [4], [5], [6]. In this paper, static gestures of Tamil are recognized by giving Tamil word as input. The static gestures are recognized using the classifiers K-Nearest Neighbor (KNN), Proximal Support Vector Machine (PSVM) and Naïve Bayesian. The overall objective of this work is to help the hearing impaired people to communicate with the normal people, and replace the conventional language with Tamil Sign Language. This paper is organized in the following manner; Section 2 explains the related work. Section 3 explains the proposed approach. Section 4 deals with a system design for the proposed work with a detailed description of features and classifiers used for recognizing the static gestures. Section 5 explains about the Experiments and results.

2. Related Work

The methodologies used in Sign Language recognition can be categorized into several types based on feature extraction methods, input type and the hardware dependency. Traditionally, there have been three main types of sign language recognition: hand shape classification, isolated sign language recognition, and continuous sign classification [7]. Another application of gesture language is human-computer interaction, which uses hand gestures as input data to a computer through webcam. In HCI, a visual interface is created to provide a natural way of communication between man and machine [8].

Earlier researchers [9] mostly focused on the capture and classification of the gestures of sign language. Researchers have developed several methods for Sign recognition. In [10] edge detection algorithm and boundary tracing are used. Hand gestures are recognized automatically using the shape descriptors. The image of the hand gesture is grabbed and converted into feature vector [11]. The hand gesture input is taken with the help of a data glove and artificial neural networks are used to recognize the gesture [12]. Sara Bilal et al. [13] developed a system for automatic translation of static as well as dynamic gestures of Indian Sign Language.

One prominent approach describes the vision based recognition technique [14] to achieve visual information in the form of feature vector. Hand gestures are represented in terms of hierarchies of multi scale color images [15]. In some systems more than one feature extraction methods and neural networks are implemented to recognize the gestures made by hand [9]. Moreover, other papers have already explored HCRFs [16] and other variants for gesture recognition. Morency et al. used LD-CRFs [17] to perform gesture recognition in continuous image streams, with excellent results. Elmezain et al. [18] also studied CRFs, HCRFs and LD-CRFs in the recognition of alphabet characters and numbers drawn in mid-air using hand trajectories, obtaining 91.52%, 95.28% and 98.05% for each model, respectively.

3. Proposed Approach

The objective discussed in this paper is a vision based translation from Tamil Sign Language to speech. The system deals with images of bare hands which provide an easy interaction with the system. Gestures are of two types. i) Static gesture and ii) Dynamic gesture. Fig.1 shows the block diagram of the Tamil Sign Language Recognition system. The proposed work consists of three stages. First stage is preprocessing, were in the sample images are processed by using the following steps i) resizing ii) gray conversion iii) filtering iv) black and white conversion. Second stage is the feature extraction, which extracts the required feature vectors from the output obtained from the first stage. Features like solidity, eccentricity, perimeter, convex area, Major axis length, Minor axis length, orientation are used to obtain the shape. Third phase is the classification where three different classifiers are used to find better accuracy. The classifiers used are K-Nearest Neighbor (KNN), Proximal Support Vector Machine (PSVM) and Naïve Bayesian.

4. Methodology Used

A vision based analysis is used in this work. Vision based analysis, is based on the way human beings perceive information about their surroundings, it is probably the most difficult to implement in a satisfactory way. Several different approaches have been tested so far.

- One is to build a three-dimensional model of the human hand. The model is matched to images of the hand by one and parameters corresponding to palm orientation and joint angles are estimated. These parameters are then used to perform gesture classification.
- Second one is to capture the image using a camera then extract some features and those features are used as input in a classification algorithm for classification. In this work we have used second method for modeling the system. Images are captured using a camera and the features are extracted and for the extracted feature a classifier is applied to classify the signs. Fig.1 shows the block diagram of Tamil Sign Language Recognition system.

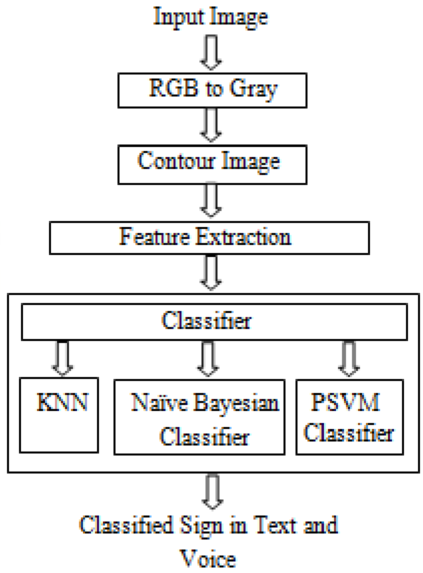

Figure 1: Block Diagram of Tamil Sign Language Recognition System

The system consists of the following stages
- Image Acquisition
- Preprocessing
- Feature extraction
- Classification.

4.1 Image Acquisition

The first stage of any vision system is the image acquisition stage. Static hand gestures and facial gestures were captured using USB connected camera. Each image represents a unique Tamil sign word. The resolution of the grabbed image is large so it is resized to a resolution of 200 into 200, which is given as input to the next stage of the model i.e., preprocessing. The sample images used are shown in Fig. 2.

Figure 2: Collection of Static Images

4.2 Preprocessing

Preprocessing methods use a small neighborhood of a pixel in an input image to get a new brightness value in the output image. It consists of two steps
- Segmentation
- Gaussian filtering

Segmentation is done to convert gray scale images into binary image. The obtained image has some noise. So it is better to filter those noises using Gaussian filtering approach. This approach helps us to obtain a smooth, closed and complete contour of a gesture. The output obtained at this stage is black and white image which uses the steps like RGB to gray conversion, filtering and thresholding. Fig. 3 shows the result of the preprocessed stage.

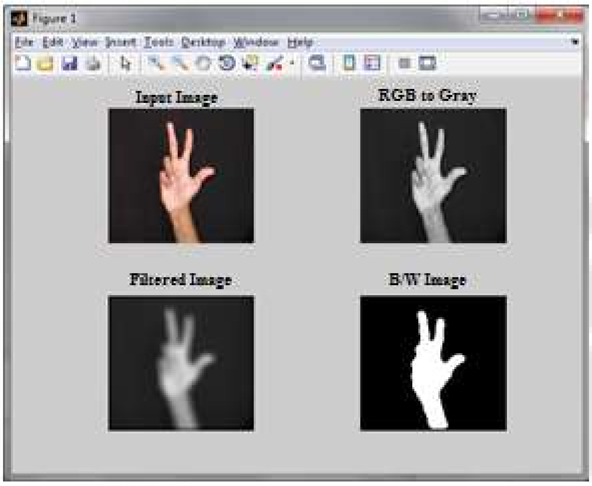

Figure 3: Result of Preprocessing Stage

4.3 Feature extraction

Region-based analysis [10,12] exploits both boundary and interior pixels of an object. The following are the shape descriptors used as features (a) Solidity (b) Eccentricity (c) Perimeter (d) Convex area (e) Major axis length (f) Minor axis length (g) Orientation. These features are described in the following sub sections.

4.3.1 Solidity

A scalar quantity which is defined to be the ratio of area to the convex area of the same object. It is computed as

$$Solidity = \frac{Area}{Convex\ Area}$$

For a solid object or cell, this value is 1.

4.3.2 Eccentricity

A scalar quantity which is defined to be ratio of the major to the minor axis. The value is between 0 and 1. It is given by the equation

$$Eccentricity = \frac{\text{Minor length axis}}{\text{Major length axis}}$$

4.3.3 Perimeter

A scalar quantity which specifies the distance around the boundary of the region. Perimeter is calculated to be the distance between each adjoining pair of pixels around the border of the region. If the image contains discontinuous regions, region props returns unexpected results.

4.3.4 Convex area

Scalar that specifies the number of pixels in 'convex image'. The image is the size of the bounding box of the region. This property is supported only for 2-D input label matrices.

4.3.5 Major axis length

Scalar specifying the length (in pixels) of the major axis of the ellipse that has the same normalized second central moments as the region. This property is supported only for 2-D input label matrices.

4.3.6 Minor axis length

Scalar specifying the length (in pixels) of the minor axis of the ellipse that has the same normalized second central moments as the region. This property is supported only for 2-D input label matrices.

4.3.7 Orientation

Scalar specifying the angle (in degrees ranging from -90 to 90 degrees) between the x-axis and the major axis of the ellipse that has the same second moments as the region. This property is supported only for 2-D input label matrices.

4.4 Classification

Classifier always tries to improve the classification rate by pushing classifiers into an optimized structure [19]. Three different classifiers are used in this work in order to compare and find the best classifier with high accuracy to recognize the static Tamil Sign.

4.4.1 K-Nearest Neighbor (KNN)

In pattern recognition, the K-Nearest Neighbor algorithm (K-NN) is a method for classifying objects based on closest training examples in the feature space. K-NN is a type of instance-based learning, or lazy learning where the function is only approximated locally and all computation is deferred until classification. Classification (generalization) [20] using an instance-based classifier can be a simple matter of locating the nearest neighbor in instance space and labeling the unknown instance with the same class label as that of the located (known) neighbor. This approach is often referred to as a neighbor classifier. A classifier always tries to improve the classification rate by pushing classifiers into an optimized structure. Classification mainly concentrates on finding the best matching features vector for the new vector among the set of reference features. K-NN [20, 21] is one of the most commonly used methods in sign language recognition systems. It uses feature vectors generated during the training phase to get the K-NN in a dimensional space. The features vector is classified by a majority vote of its neighbors. Neighbors are taken from a set of objects for which the correct classification is known. Euclidean distance measures are used to calculate the difference between the query and the target shape feature vectors and return the number of approximate nearest neighbors.

4.4.2 Proximal SVM Classifier

The proximal SVM also uses a hyper plane $w.x + b = 0$ as the separating surface between positive and negative training examples. But the parameter w and b are determined by solving the following problem

$$\min \frac{1}{2}\left(\|w\|^2 + b^2\right) + C\sum_i \xi_i^2$$

$$s.t. \forall i, y_i\left(w.x_i + b\right) + \xi_i \geq 1,$$

The main difference between standard SVM and proximal SVM is the constraints. Standard SVM employs an inequality constraint whereas proximal SVM employs an equality constraint. We can see that standard SVM only considers points on the wrong side of $w.x_i + b = 1$ and $w.x_i + b = -1$ as training errors. The experiment considered three possible choices of kernel functions; the Gaussian, Polynomial and Sigmoid. For the Gaussian kernel, a coarse-to-fine grid search was conducted in the hyper parameter space in order to find an optimum. For each trained machine, the testing dataset twice was evaluated: at first using the *1-vs-1* voting scheme, then with the DDAG decision. We have annotated the performance of the classifiers, measured in terms of Cohen's kappa ( ), the total number of unique support vectors needed in the voting scheme and the average number of vector evaluations in the DDAG decision path. As linear machines can also be written in a compact form, for linear machines we consider the number of machine evaluations instead of vector evaluations.

4.4.3 Naive Bayesian Classifier

A Naïve Bayesian classifier assigns a new observation to the most probable class, assuming that the features are conditionally independent given the class value [22]. It can outperform more sophisticated classification methods by categorizing incoming objects to their appropriate class. The Naive Bayesian classifiers can handle a random number of independent variables whether continuous or categorical.

The Naive Bayesian classifier is used to justify the objects using new methods to get a maximum. In each image, a measure of properties is taken to determine the sign in different position. They estimate the probability that a sign belongs to each of the target classes that is predetermined. In the training phase, the training set is used to decide how the parameters must be weighted and combined in order to separate the various classes of signs.
It classifies data in two steps:

Training step: Using the training samples, the method estimates the parameters of a probability distribution, assuming features are conditionally independent given the class.

Prediction step: For any unseen test sample, the method computes the posterior probability of that sample belonging to each class. The method then classifies the test sample according to the largest posterior probability.
Bayes theorem used, takes the equation as given in (1) and (2)

$$P(H|X) = P(X|H)P(H)/P(X) \quad \text{........ (1)}$$

It can also be expressed as

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)} \quad \text{......... (2)}$$

Where C is a constant for all classes only need be maximized.

The class-conditional independence assumption greatly simplifies the training step since estimation can be done using one-dimensional class-conditional density for each feature individually. This assumption of class independence allows the Naive Bayesian classifier to better estimate the parameters required for accurate classification while using less training data than many other classifiers. This makes it particularly effective for datasets containing many predictors or features.

5. Experiments and Results

The system for recognizing a set of Tamil sign words using three different classifiers has been developed by using MATLAB R2010a which is processed in a Windows 8 Operating system. MATLAB, which stands for **Mat**rix **Lab**oratory, is a state-of-the-art mathematical software package, which is used extensively in both academia and industry. The proposed work was trained and tested with 41 categories each containing 10 subjects. **Leave-one-out-cross validation** method is used. The Tamil Sign Language Dataset contains 410 samples for each of 41 signs, recorded from 10 different persons. Each sample has a RGB image and a depth image. The sign April, July and August are not used, because these signs have motion and the proposed model only works with static signs. The dataset has variety of background and viewing angles. Due to the variety in the orientation when the signal is performed, signs become strongly similar. Figure 3 shows the most similar signs March, May, June, and January.

Figure 5: Similar Gestures

The examples are taken from the same user. It is easy to identify the similarity between these signs, all are represented by a opened fist, and differ only by the thumb position, leading to higher confusion levels. Therefore, these signs are the most difficult to differentiate in the classification task. The accuracy of the system is calculated by taking different number of features into consideration and the comparison chart is shown in Fig. 6

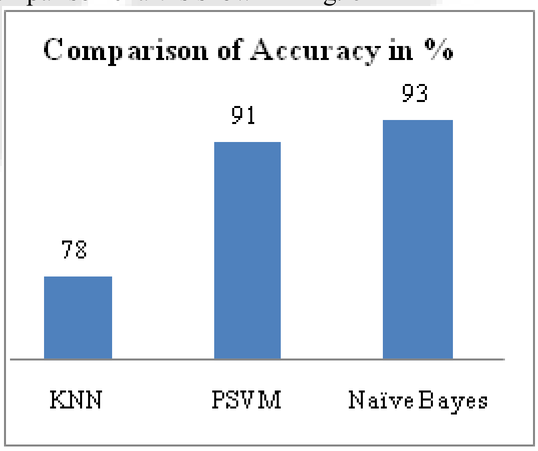

Figure 6: Comparison Chart of Classifiers

If model selection and true error estimates are to be computed simultaneously, the data needs to be divided into three disjoint sets.

Figure 7: Conversion of Sign to Text

Training set: a set of examples used for learning, to fit the parameters of the classifier.
Validation set: a set of examples used to tune the parameters of a classifier.
Test set: a set of examples used only to assess the performance of a fully-trained classifier. The K-Nearest Neighbor algorithm is the simplest of all machine learning algorithms: an object is classified by a majority vote of its neighbors, with the object being assigned to the class most common amongst its k nearest neighbors (k is a positive integer, typically small). If $k = 1$, then the object is simply assigned to the class of its nearest neighbor.

The training process for KNN consists only of storing the feature vectors and class labels of the training samples [27]. One major problem of using this technique is the class with the more frequent training samples would dominate the prediction of the new vector, since they more likely to come up as the neighbor of the new vector due to their large number. k-selection, another important issue which is to be taken into account is how to choose a suitable k for this algorithm. Generally, according to shakhnarovish et.al [23], larger values of k reduce the effect of noise on the classification, but make boundaries between distinct classes. Choosing an appropriate k is essential to make the classification more successful.

The recognition of human gestures and facial expressions in image sequences is an important and challenging problem that enables a host of human computer interaction applications. If a system developed is strong enough for processing the static gestures then it would be the finest system to process the frames obtained while processing the continuous gestures. Since the collected signs were of different shapes, scales and brightness, all the signs could not be perfectly recognized by a single classifier.

The performance evaluation and the comparison of the performance measures of the classifiers are shown in Table.1and Fig. 7 respectively. Though, a maximum of the gestures are recognized to a higher accuracy with three different classifiers namely K-Nearest Neighbor (KNN), Proximal Support Vector Machine (PSVM) and Naïve Bayesian., Naive Bayesian classifier is found to the best classifier with an accuracy rate of 93% where as 91% for PSVM classifier and for K-Nearest Neighbor it is 78%.

Table 1: Performance Evaluation of NB, PSVM and KNN Classifiers

S.No	Performance Measures	NB	PSVM	KNN
1	Classified rate	1.00	0.96	0.95
2	Sensitivity	0.71	0.67	0.57
3	Specificity	0.85	1.00	0.68
4	Error rate	0.34	0.50	0.40
5	Inconclusive rate	0	0	0
6	Positive Predictive value	0.55	0.42	1.00
7	Negative Predictive value	0.92	1.00	0.30
8	Negative likelihood	0.33	0.53	0.63
9	Positive likelihood	2.60	1.86	1.77
10	Prevalence	0.20	0.40	0.20

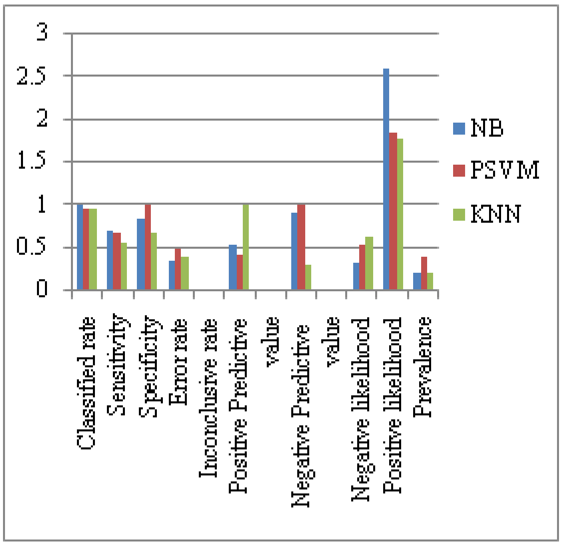

Figure 8: Comparison of Performance Measures of the Classifiers

6. Conclusion and Future Work

Sign language recognition is a wide area of research. An analysis of different classifiers is done in which the Naive bayes approach is proved to be the better for Tamil sign language recognition system. The aim of this work is to develop a Tamil sign language recognition system for deaf-dumb people. In this project, an image processing technique has been presented and designed for recognizing the signs of Tamil language for deaf-dumb persons. In this work more data has been collected and processed. Instead of taking only static hand gestures additionally hand with facial gestures are also taken. So, a large set of data are processed with extracted features called moment descriptors which are classified by using three different classifiers namely K-Nearest Neighbor (KNN), Proximal Support Vector Machine (PSVM) and Naïve Bayesian. The results of the classification technique is evaluated and found that Naïve Bayesian works well with 93% accuracy where as 91% for PSVM classifier and for K-Nearest Neighbor it is 78%. The work presented in this paper recognizes static signs only. In future, the work can be extended to recognize the dynamic signs of Tamil Sign Language. Now, the system deals with images with, uniform background, but it could be made background independent. This research result developed here is a principled technique that will enable the use, not only in sign language or hand gesture recognition but also in other related areas of computer vision.

References

[1] T. Shanableh, K. Assaleh and M. Al-Rousan, "Spatio-Temporal feature extraction techniques for isolated gesture recognition in Arabic sign language," IEEE Trans. Syst., Man, Cybern. B, 37(3), June 2007.

[2] K Assaleh, M Al-Rousan, "Recognition of Arabic Sign Language Alphabet Using Polynomial Classifiers," EURASIP Journal on Applied Signal Processing, vol. 2005, no. 13, pp. 2136-2145, 2005.

[3] Z. Ren, J. Yuan, and Z. Zhang, "Robust hand gesture recognition based on finger-earth mover's distance with a commodity depth camera," in Proceedings of the 19th ACM International Conference on Multimedia. ACM, 2011, pp. 1093–1096.

[4] V. Frati and D. Prattichizzo, "Using Kinect for hand tracking and rendering in wearable haptics," in Proceedings of the IEEE World Haptics Conference (WHC). IEEE, 2011, pp. 317–321.

[5] Y. Li, "Hand gesture recognition using Kinect," in Proceedings of the 3rd IEEE International Conference on Software Engineering and Service Science (ICSESS). IEEE, 2012, pp. 196–199.

[6] Z. Mo and U. Neumann, "Real-time hand pose recognition using lowresolution depth images," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2. IEEE, 2006, pp. 1499–1505.

[7] Sandjaja I., Marcos N., "Sign Language Number Recognition," Fifth International Joint Conference on INC, IMS and IDC, pp.1503-1508, 2009.

[8] I.G.Incertis, J.G.G.Bermejo, and E.Z.Casanova, "Hand Gesture Recognition for Deaf People Interfacing", The 18th International Conference on Pattern Recognition (ICPR), 2006.

[9] Vaishali S Kulkarani, ME Digital Systems, S. D. Lokhande, "Appearance based segmentation of sign language using gesture segmentation", Sinhgad College of Engineering, 2010.

[10] Dengsheng Zhang and Guojun Lu, "Review of shape representation and description techniques", Pattern Recognition, Vol. 37, 1–19, 2004.

[11] Yasushi Hamada, Nobutaka Shimada and Yoshiaki Shirai, "Hand shape estimation for complex background images", Osaka University, Japan, 2004

[12] Liviu Vladutu, "Shape Recognition for Irish Sign Language Understanding", International Conference on Simulation, Modeling and Optimization, Vol. 4, 242-247, 2009.

[13] Sara Bilal1, Rini Akmeliawati2, Momoh Jimoh El Salami ,Amir A. Shafie, "Vision-based Hand Posture Detection and Recognition for Sign Language-A study", 2011 4th International Conference on Mechatronics (ICOM), 17-19 May 2011, Kuala Lumpur, Malaysia.

[14] Thomas Coogan , George Awad, Junwei Han and Alistair Sutherland, "Real time hand gesture recognition including hand segmentation and tracking", ISVC 2006 - 2nd International Symposium on Visual Computing, 6-8 November 2006, Lake Tahoe, NV, USA. ISBN 978-3-540-48628-2. Geetha M et al. / International Journal on Computer Science and Engineering (IJCSE) ISSN

[15] Lars Bretzner, Ivan Laptev, Toney Lindberg, On Hand gesture recognition using multiscale color features hierarchical models and partial filtering, CVAP Laboratory, Department of numerical analysis and Computer Science, Sweden,2002'

[16] Wang, S., Quattoni, A., Morency, L.-P., Demirdjian, D.: Hidden Conditional Random Fields for Gesture Recognition. In: IEEE Conf. on Computer Vision and Pattern Recogni-tion, 2006. CVPR '06, Washington, DC, USA, pp.1521-1527 (2006)

[17] Morency, L.-P., Quattoni, A., Darrell, T.: Latent-Dynamic Discriminative Models for Con-tinuous Gesture Recognition. In: IEEE Conf. on Computer Vision and Pattern Recogni-tion, 2007. CVPR '07, pp.1-8 (2007)

[18] Elmezain, M., Al-Hamadi, A., Michaelis, B.: Discriminative Models-Based Hand Gesture Recognition. In : International Conference on Machine Vision, Los Alamitos, CA, USA, pp.123-127 (2009) .

[19] Danesh, A., B. Moshiri and O. Fatemi, "Improve text classification accuracy based on classifier fusion methods", Proceeding of the 10th International Conference on Information Fusion, July 9-12, IEEE Computer Society, USA July 9-12, 2007.

[20] M. Krishnaveni and V. Radha, "Performance evaluation of statistical classifiers using Indian Sign Language datasets", International Journal of Computer Science, Engineering and Applications, Vol.1 (5), 2011.

[21] Nadia R, Albelwi and Yasser M. Alginahi, "Real-time Arabic Sign Language (ASL) recognition", International Conference on Computers and Information Technology, 2012.

[22] Yong qiang TANG and Hao Helen ZHANG, "Multiclass Proximal Support Vector Machines", Journal of Computational and Graphical Statistics, Vol. 15 (2), 339–355, 2006.

[23] Glenn Fung and Olvi L. Mangasarian, "Multicategory Proximal Support Vector Machine Classifiers", Machine Learning, Vol. 59, 77–97, 2005.

[24] T. Hanumantha Reddya, K. Karibasappaa, A.Damodaramb, "Proximal SVM for Face Recognition", Journal of Computational Linguistics, Vol. 1(1), 1-4, 2011.

[25] Olvi L. Mangasarian and Edward W. Wild, " Multisurface Proximal Support Vector Machine Classification via Generalized Eigen values", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 28 (1), 69-74, 2006.

[26] Shakhnarrovish, Darrell and Indyk, Nearest Neighbor Methods in learning and vision,MIT-press, 2005.

[27] Prof. Hetal Bhavsar, Dr. Amit Ganatra, "Variations of Support Vector Machine classification Technique: A survey", International Journal of Advanced Computer Research, Vol. 2 (6), 2012.