IOSR Journal of Electronicsl and Communication Engineering (IOSR-JECE)
ISSN: 2278-2834-, ISBN: 2278-8735, PP: 07-12
www.iosrjournals.org

A Review On Vision Based Hand Gesture Recognition Approach Using Support Vector Machines.

Mrs.A.R.Patil[1], Dr.S.S.Subbaraman[2]
[1.] (*Asst.Prof., Electronics Dept., DKTE'sTextile and Engineering Institute, Ichalkaranji.,India*
[2.(] *Professor and HOD, E&TC Department, ADCET, Ashta.India)*

Abstract: *The aim of hand gesture recognition researches is to develop such a system that will easily identify gestures, and use them for variety of application ranging from sign language to virtual reality. The discussion on the various approaches for vision based hand gesture recognition based on Support vector Machine (SVM) is given in this paper. SVM is one of the most important machine learning algorithms that analyze data and recognize patterns, used for classification and regression analysis even for the large input space. Brief introduction and working as well as various implementation mechanisms are also discussed in this paper.*

Keywords: *hyperplane, PCA, SIFT, soft margin, SVM, Vision based hand Gesture.*

I. Introduction

A gesture is a form of non-verbal communication in which visible bodily actions communicate particular messages, either in place of speech or together and in parallel with spoken words. Gestures include movement of the hands, face, or other parts of the body. The main purpose of developing such a system lies in the fact that gesture recognition has implemented in motion analysis to machine learning. It also serves many applications from virtual reality to sign language recognition. A lot of research has been carried out in last 2-3 decades on hand gesture recognition approach. These can be roughly divided into two categories, namely, Glove Based Gesture Recognition and Vision Based Gesture Recognition.

Glove based Hand Gesture Recognition hinders the naturalness as cumbersome devices are required to wear. Potentially, any awkwardness in using gloves and other devices can be overcome by using video-based noncontact interaction techniques [1]. This approach suggests computer vision techniques to interpret gestures in a more natural way. This approach is called as vision based approach [2] [3]. *A vision based* approaches uses features extracted from visual appearance of the input image model of the hand, comparing these modeled features with features extracted from input camera(s) or video input. This has garned popularity in last decades.

In this paper, we provide a survey on different aspects of Hand Gesture Recognition with particular emphasis on Support Vector Machine as a classifier. Rest of the paper is organized as follows: Section II will be an outline of different tools used for feature extraction. Section III introduces an overview of SVM. Gesture Recognition using SVM will be explained in Section IV. Discussion and Conclusion are given in Section V.

II. Feature Extraction Techniques

As hand gestures are very rich in shape variation, motion and textures, recognition of hand from video frame is not a straight forward problem. Spatial as well as temporal features can be extracted from the hand gesture inputs. A simple approach is to look for skin colored regions in the image. This is a very popular method [4], [5], with some limitations like it is very sensitive to lighting conditions. [6]. Discrete Cosine Transform (DCT) technique is applied to get the most important data that will be used in gesture classification stage which is the simplest and given better results. L.Bretzner et al. [7] used multi-scale space color features to recognize hand gestures the system shows real-time performance only when no other skin-colored objects exist in the image in background. A.Argyros et.al.in [8] obtained hand contour to recognize hand gestures and then computed the curvature of each point on the contour. Due to noise and unstable illumination in the cluttered background, the segmentation of integrated hand contour had some difficulty.

Features are detected by robust feature detection methods like SIFT [9], its variant principal component analysis (PCA)-SIFT [10] and speeded up robust features (SURF) [11]. In [9], SIFT was used for extracting distinctive invariant features from images that can be invariant to image scale and rotation and partially to illumination. In [10], PCA was used to normalize gradient patch instead of histograms. It turned out that PCA-SIFT-based local descriptors were also distinctive and robust to image deformations. In [11], robust features (SURF) were speeded up, the experiments showed that it was very fast and worked properly. It is possible to optimize the code and speed up the recognition process combining SIFT, PCA-SIFT, and SURF key point detection approach. In [12], authors acquired and preprocessed video image frame from camera, then extracted the normalized moment of inertia features and Hu invariant moments of gestures to constitute feature vector. Histograms of oriented gradients (HOG) were used to describe hand image. Extracted HOG features were projected into low-dimensional subspace using PCA-LDA (Principle Component Analysis and Linear Discriminant Analysis). In [13], C. Messom et al. achieved block based picture information ratio as features Optical Flow and motion estimation algorithms can be also used to extract the features of meaningful hand gestures.

III. Introduction to SVM

Support Vector Machine (SVM), is one of the best machine learning algorithms, which was proposed in 1990's by Vapnik. SVMs are a set of related supervised learning methods used for classification and regression. Classical machine learning approaches are designed to minimize error on the training dataset and it is called the Empirical Risk Minimization (ERM). Those learning methods follow the ERM principle and neural networks are the most common example of ERM. On the other hand, the SVMs are based on the Structural Risk Minimization (SRM) principle rooted in the statistical learning theory. It gives better generalization abilities (i.e. performances on unseen test data) and SRM is achieved through a minimization of the upper bound (i.e. sum of the training error rate and a term that depends on VC dimension) of the generalization error [14].

Classification is carried out by creating an N-dimensional hyperplane that optimally divides the data into two groups. SVM classifiers are closely related to neural networks. Actually, a SVM classifier model using a sigmoid kernel function is the same as the two-layer, perceptron neural network. In the SVM literature, a predictor variable is known as an attribute, and a transformed attribute that is used to define the hyperplane is known as a feature. The operation of selecting the most appropriate representation is called as feature selection. A group of features that describes one case is known as a vector. Therefore, the objective of SVM modeling is to find the optimal hyperplane that divides clusters of vectors in such a way that cases with one class of the target variable are on one side of the plane and cases with the other classes are on the other side of the plane. The vectors near the hyperplane are the support vectors.

Pattern recognition aims to classify data based on either a priori knowledge or statistical information extracted from raw data, which is a powerful tool in data separation in many disciplines. SVM is a supervised type of machine learning. algorithm in which, given a set of training examples, each marked as belonging to one of the many categories, an SVM training algorithm builds a model that predicts the category of the new example. SVM has the greater ability to generalize the problem, which is the goal in statistical learning.

The basic idea of the SVMs is to construct a hyperplane as the decision plane, which separates the positive (+1) and negative (-1) binary classes with the largest margin, which is related to minimizing the VC dimension of SVM. In a binary classification problem where feature extraction is initially performed, let us label the training data $x_i \in \Re^d$ with a label $y_i = \{-1,+1\}$, for all the training data $i = 1,\ldots\ldots\ l$, where l is the number of data, and d is the dimension of the problem. When the two classes are linearly separable in $\Re^d$, we wish to find a separating hyper plane which gives the smallest generalization error among the infinite number of possible hyperplanes. Such an optimal hyper plane is the one with the maximum margin of separation between the two classes, where the margin is the sum of the distances from the hyper plane to the closest data points of each of the two classes. These closest data points are called Support Vectors (SVs). The middle line on fig.1 represents the optimal separating hyper plane.

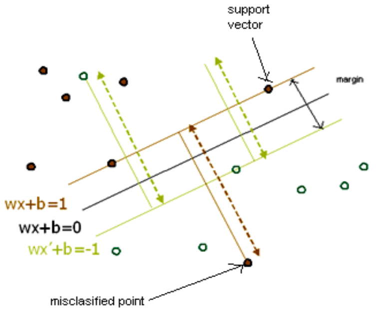

Fig 1.Linear separating hyperplane and Support vectors

A hyper plane is characterized with a *direction vector w (normal to the hyper plane)* and an *offset vector b* that satisfy the equation

$$w.x + b = 0 \quad (1)$$

The separation problem is to determine the hyper plane such that for $wx_i + b \geq 1$ positive examples and $wx_i + b \leq 1$ for negative examples. Since the SVM finds the hyper plane, which has the largest margin, it can be found by minimizing $\frac{1}{2\|w\|^2}$

$$\min_{w,b} \quad \Phi(w) \quad = \frac{\|w\|^2}{2} \qquad (2)$$

The optimal separating hyper plane can thus be found by minimizing equation (2) under the constraint (3) to correctly separate the training data.

$$y_i = (w.x_i + b) - 1 \geq 0, \forall \mathrm{i} \qquad (3)$$

This is a Quadratic Programming (QP) problem for which standard techniques (Lagrange Multipliers, Wolfe dual) can be used [15],[16].

$$f_\lambda : \Re^n \to \{-1,+1\}$$

The set of functions f_λ could be a set of Radial Basis Functions or a multi-layer neural network.

For a given function f_λ the expected risk (the test error) R_λ is the possible average error committed by f_λ *on* the unknown example drawn randomly from the sample distribution *P*

$$R(\lambda) = \int \frac{1}{2} |f_\lambda\ (x) - y| dP(x,y)$$

R_λ is a measure of how well a hypothesiș predicts the correct label y from an input **x**.

This approximation is called empirical risk *Remp* (training error)

$$Remp(\lambda) = \frac{1}{l} \sum_{i=1}^{l} |f_\lambda\ (xi) - yi|$$

Where *l* is the number of samples. The mathematical function used for the transformation is known as the **kernel** function. Following are some kernel functions:

• Linear • Polynomial • Radial basis function (RBF) • Sigmoid

A linear kernel function is recommended when linear separation of the data is straightforward. In other cases, one of the other functions should be used. You will need to experiment with the different functions to obtain the best model in each case, as they each use different algorithms and parameters.

3.1 How SVM Works?

The general equation of a plane in n-dimensions is **w .**F**x=***b* where **x** is a *n*x1 vector. Of all the points on the plane, one has minimum distance *dmin* perpendicular from hyperplane to the origin.

$$\boldsymbol{d_{min}} = \frac{|\boldsymbol{b}|}{\|\boldsymbol{w}\|}$$

Training patterns(**x1,y**1)_,. . . ,(_**xl,**_*yl)*_are given where **xi** is a *d*-dimensional vector and

y_i =1 if x_i is in group 1

y_i =-1 if x_i is in group 2

If the data are linearly separable, then there exist a d-dimensional vector **w** and a scalar *b* such that

$$(w.x_i - b) \geq 1, \text{if yi} = 1$$
$$(w.x_i - b) \leq 1, \text{if yi} = -1$$

3.2 Multiclass SVM

The basic SVMs are for two-class problem. However it should be extended to multiclass to classify into more than two classes [15],[16]. There are two basic strategies for solving *q*-class problems with SVMs.

3.2.1 Multi-class SVMs: One to Others

Take the training samples with the same label as one class and the others as the other class, and then it becomes a two-class problem.

3.2.2 Multi-class SVMs: Pair wise SVMs

In the pair wise approach, *q*2machines are trained for *q*-class problem [17]. The pair wise classifiers are arranged in trees, where each tree node represents an SVM. Regarding the training effort, the one-to-others approach is preferable, since only *q* SVMs have to be trained compared to *q*2 SVMs in the pair wise approach. However, at
runtime both strategies require the evaluation of *q*-1 SVMs . Recent experiments on people recognition show similar classification performances for the two strategies. Fig 2 gives a brief idea of implementing multiclass SVM using Binary classifiers either in top-down approach or in bottom-up approach.

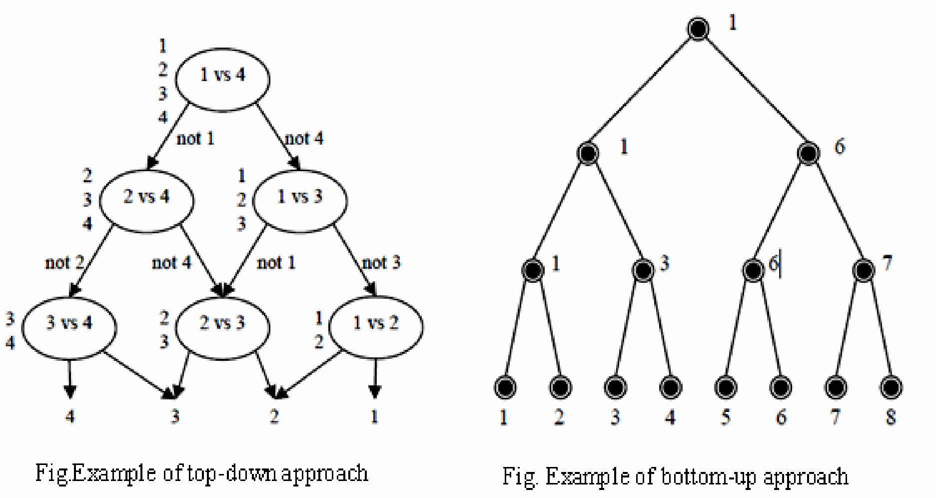

Fig.2 Examples of multiclass SVM

IV. Hand Gesture Recognition using Support Vector Machine.

Various approaches have been utilized to deal with gesture recognition problem ranging from soft computing approaches to statistical models based on Hidden Markov Model HMM [18], and Finite state Machine FSM [19]. Soft computing tools generally include ANN [20] [21], fuzzy Logic sets [22] and Genetic Algorithms GAs [23]. In this paper we focus on the SVMs.

Since different hand gestures are required to classify (and not only two), there is a need to combine several binary SVMs in a *Classifier Tournament* structure as shown in Fig.3. For every pair of classes, training to a binary classifier is required to classify between them. Computed image features in the preprocessing stage will be entered into all the binary classifiers. Each binary classifier returns a class number representing a gesture. The gesture that has the most votes of classifiers wins and that will be the result of the multiclass SVM.

MEB-SVM (Minimum Enclosing Ball-SVM) was implemented on the large scale data classification by [24]. Comparing with others that algorithm proved efficient with less computation time. Some tricks such as mean shift and Fourier descriptor were implemented in that paper. These methods enhanced the hand gesture recognition's efficiency. The main algorithm MEB-SVM shows powerful while deal with multi-class classification.

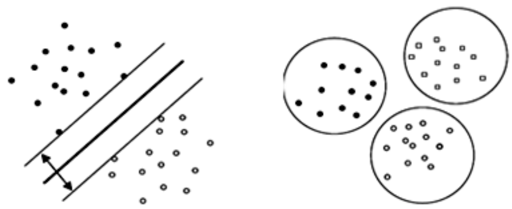

Figure:3 a. Typical SVM model b.MEB model

A multiclass SVM is addressed by [25].Multiclass SVM training and testing were performed using the library for SVM described in [26]. This library supports multiclass classification and uses a *one-against-one* (OAO) approach for multiclass classification in SVM [27]. For the M-class problems (M being greater than 2), the OAO approach creates M (M-1)/2 two-class classifiers, using all the binary pair-wise combinations of the M classes. Each classifier was trained using the samples of the first class as positive examples and the samples of the second class as negative examples. To combine these classifiers, the Max Wins method was used to find the resultant class by selecting the class voted by the majority of the classifiers [28].Fig.4 shows some of the images used by author from Sebastien Marcel database [29].

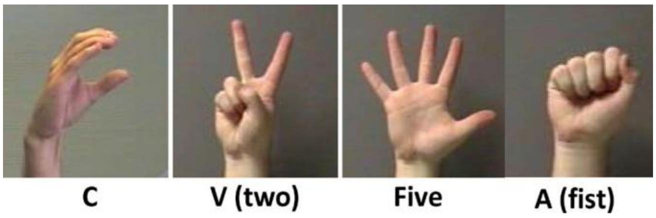

Figure 4: **Hand postures used in Marcel database images. [courtesy[25]]**

[30] tested two different SVM configurations. In the first configuration, a single multi-class SVM whose output classes were equal to the cells of the grid was used. In the second configuration, two multi-class SVMs were used. The first SVM was trained to recognize the row of the pointed cell whereas the second one was trained to recognize the column. Thus the number of the output classes of the two SVMs was equal to the number of rows and columns of the grid, respectively. Since the SVM is a binary classifier, it should be extended for an *m*-class problem in hand gesture recognition. The authors used the so called one against one approach, which was a pair wise method and needs to train SVM classifiers. In addition, another two distance measures, i.e., Euclidean distance and Cosine similarity distance, were also computed to make a comparison with the SVM method

Various issues related to SVM were discussed in [31] like *Optimal Separating Hyperplane, Linearly non-separable case, Multi-Class Learning, Nonlinear Support Vector Machines etc.* This was evaluated using dot product, but it can be computationally large. Thus, it is the choice of kernel **k** that is important, so that large quadratic computation can be reduced. The fig .6 shows a mapping of the nonlinear input space to a linear feature space. There are various kernels to choose from, for example Polynomial learning machines, Radial-basis function networks and Two-layer perceptrons. The choice of the kernel is theoretically of interest since it determines the feature space in which to work. (Fig.5).

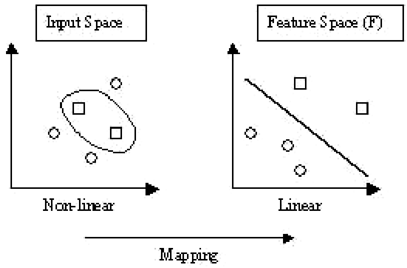

Figure 5: Mapping of Non-linear input feature space to linear feature space.

Hybrid shape motion sequence kernel is one of the innovative kernel functions. The classifying procedure was a classification of an offline video showing hand gestures or classification of a real-time online camera video showing hand gestures. In both cases the video was broken down to frames and the frames converted to images. The images were then preprocessed and the relevant features were extracted. These features were passed onto the multiclass SVM and a hand gestures classification was made [32].

V. Conclusion and Discussion:

Hand gesture recognition is truly an interdisciplinary research area having wide range of applications. In this paper we provided some tools to extract the features. Low level features like contour, skin detection, HOG, DCT, and PCA will prove good when they are used for constant background. To make a flexible system SURF, SIFT, improvements on PCA or hybridization will also provide good results. We emphasized on the Support Vector Machine which can be used as a classifier. SVMs are binary class classifiers and it was first applied for verification or 2 class classification problems. But SVMs had been used for multi-class classification problems since one to others and pair wise bottom-up, top-down multi-class classification methods were developed. Most of applications using SVMs showed SVMs-based problem solving approach is better than other methods. Even SVM is a newer technique; it has been applied to a wide range of machine learning tasks and used to generate much possible learning architecture with proper selection of kernel functions. If some limitations related with the choice of kernels, training speed and size are solved, it can be applied to more real-life classification problems.

References

[1] Chun Zhu and Weihua Sheng, "Wearable Sensor-Based Hand Gesture and Daily Activity Recognition for Robot-Assisted Living", IEEE transactions on systems, man and cybernetics-part a: systems and humans, vol.41, no.3, May 2011.

[2] Vladimir I. Pavlovic, Rajeev Sharma and Thomas S. Huang, "Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review" IEEE Transactions on Pattern Analysis and Machine intelligence, vol.19, no.7, July1997.

[3] Sushmita Mitra, and Tinku Acharya, "Gesture Recognition: A Survey", IEEE Transactions on Systems, Man and Cybernetics-part c: applications and reviews, vol.37, no.3 May 2007.

[4] R. Oka, "Spotting Method for Classification of Real World Data", the Computer Journal, vol. 41, no. 8, pp. 559-565, July 1998.

[5] H. Yoon, J. Soh, Y. Bae, and H. Yang, "Hand Gesture RecognitionUsing Combined Features of Location, Angle and Velocity," Pattern Recognition, vol. 34, no. 7, pp. 1491-1501, July 2001

[6] Paulraj M P1, Hazry Desa1 ,et.al., "Feature Extraction Based On Centroid And DCTE for Sign Language Recognition", Proceedings of the International Conference on Digital Factory, ICDF 2008, August 11-13, 2008 USA. 222

[7] L. Bretzner, I. Laptev, T. Lindeberg, "Hand gesture recognition using multiscale color features, hieracrchichal models and particle filtering", in Proceedings of Int. Conf. on Automatic face and Gesture recognition, Washington D.C., May 2002.

[8] Argyros, M. Lourakis, "Vision-Based Interpretation of Hand Gestures for Remote Control of a Computer Mouse", Proc. of the workshop on Computer Human Interaction, 2006, pp. 40-51.

[9] L. Yun and Z. Peng, "An automatic hand gesture recognition system based on Viola-Jones method and SVMs", in Proc. 2nd Int. Workshop Comput. Sci. Eng., 2009, pp. 72–76.

[10] Y. Ren and C. Gu, "Real-time hand gesture recognition based on vision," in Proceeding Edutainment, 2010, pp. 468–475.

[11] C. Messom and A. Barczak, "Fast and efficient rotated haar-like features using rotated integral images," in Proc. ACRA, 2006.

[12] Ayan Chaki, Pragya Jain,Rohit Kumar Gupta, "Information Measure Ratio Based Real Time Approach for Hand Region Segmentation With a Focus on Gesture Recognition", Innovation Lab, TATA Consultancy Services Ltd.,Kolkata, India

[13] Zhou Hang. Ruan Qiuqi , "VisuaI-Gesture Recognition with Color Segmentation and Support Vector Machines" ICSP'O4 Proceedings.

[14] Hyeran Byun, Seong-Whan Lee ," Applications of Support Vector Machines for Pattern Recognition: A Survey," Proceedings of Int. Conference on Data Mining and Knowledge Discovery, 2(2),pp. 121-167, (1998).

[15] S. Bernhard, C.C. Burges, and A.J. Smola, Pairwise classification and support vector machines, The MIT Press, C. Massachusetts, London England, pp. 255-268, Jan. 1999

[16] Pontil and A. Verri. Support vector machines for 3-D object recognition,IEEE Trans. on Pattern Analysis and Machine Intelligence, pp. 637-646, (1998).

[17] C. Nakajima, M. Pontil, and T. Poggio, People recognition and pose estimation in image sequences, , In Proceedings of IEEE Int. Joint Conference on Neural Net- works, vol. 4, pp. 189-194, (2000).

[18] Mahmoud Elmezain, Ayoub Al-Hamadi, Jorg Appenrodt, and Bernd Michaelis. (2009). "A Hidden Markov Model-Based Isolated and Meaningful Hand Gesture Recognition,"International Journal of Electrical and Electronics Engineering, pp. 156-163.

[19] Verma R., Dev A. (2009, Dec.). " Vision based hand gesture recognition using finite state machines and fuzzy logic", IEEE International Conference on Ultra Modern Telecommunications & Workshops, (ICUMT '09), pp. 1-6, Petersburg,