International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 4 Issue 04, April-2015

FPGA Implementation of Sequential Minimal Optimization for MFCC-SVM Based Speech Recognition

Prabha.S
M. Tech VLSI Design
School of Electronics Engineering (SENSE)
VIT University,

Aarthy. M
Assistant professor
School of Electronics Engineering (SENSE)
VelloreVIT University,Vellore

***Abstract*: Support vector machines(SVM) are one of the very efficient supervised machine learning algorithms.Many training algorithms are used to train SVM.This paper presents FPGA implementation of Sequential Minimal Optimization, which is a widely used algorithm for optimization of SVM .In order to obtain better featuresa comparison of performance of svm based on acoustic features extracted from algorithms such as FFT,MFCC,VAD has been done.The feature extraction and classificationis done using MATLAB.SMO algorithm is verified using Modelsim and is implemented in Altium Nanoboard NB3000XN.**

Keywords: SMO, MFCC, SVM, FPGA, VAD

I. INTRODUCTION

Speech is a complex process without clearly distinguished parts.The sounds or phones might vary based on phone context,speaker and style of speech.Hence speech recognition system consists of two basic operations: signal modeling and pattern recognition[1].Signal modeling helps in converting given speech into set of parameter values and Pattern matching helps in matching the parameters to the parameter set in the memory. Feature extraction is a very important process involved in signal modeling.Acoustic features such as energy,power,centroid etc are extracted from the input speech.These features help in differentiating between speech and non-speech signal. Speech recognition system consists of two main blocks as shown in Fig.1.It consists of Feature extraction block to extract acoustic features and a classifier to segregate speech and non-speech features. Based on the input considered for speech recognition,the speech recognition analysis can be classified into spectral and temporal.Speech recognition can also be based on two approaches: pattern recognition approach and acoustic phonetic approach [2].The spectral analysis involves Bank of filter method[3],Linear Predictive Coding(LPC),Cepstral analysis,Mel Frequency Cepstral Coefficients(MFCC).These methods are the most commonly used front end signal processing for speech recognition.In bank of filter approach,the signal is passed through a bank of overlapping band pass filters .In linear Predictive coding,the spectral envelope of speech is represented in compressed form[4].LPC estimates the formants,removes their effects and estimates intensity of the rest of the signal. In cepstral analysis[5],the cepstrum is obtained as a result of taking inverse Fourier transform of logarithm of the speech spectrum. The power cepstrum is very useful in speech recognition. Mel Frequency Cepstral Coefficients (MFCC) is widely used in automatic speech and speaker recognition MFCC,the speech is broken into smaller frames and DFT is computed for the same.The output of DFT is used to find log filter bank energies in which discrete cosine transform is applied to obtain mel cepstrum.The temporal analysis[6] involves methods such as power estimation, fundamental frequency estimation, Cepstrum based pitch determination, whichinvolves analysis of speech waveform directly.

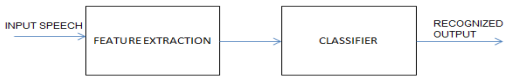

Fig 1.Basic Blocks involved in speech recognition

Hidden Markov Models are widely used for speech recognition.It was developed by L.E.Baum et al [7].It is probability based model and can be easily related to the simple urn problem.In HMM the states are not visible,but the output is visible.S.Iqbal et al[8] have suggested a voice recognition based secure ATM using MFC-HMM along with k-means algorithm. I. Patel et al[9] have suggested a sub band technique to obtain frequency spectral information for MFCC-HMM based speech recognition.Young-kyu Choi et al[10] have analyzed the memory pattern of speech recognizer based on HMM. Aravind et al[11] and J.P.Sendra et al [12],have shown the application of SVM in speech recognition. PhanDinhDuy[13] et al have implemented a speech recognizer based on a FFT controller using FPGA. Cheng-Yuan Chang et al [14] have suggested a speech recognition chip based on FFT. The application of MFCC for speech identification is explained by Haojun Wu[15].S.Verma et al[16] have implemented MFCC-DTW based recognizer for numeric. Chadawan Ittichaichareon et al[17] have suggested a MFCC-maximum likelihood(ML) –SVM based speech recognizer.

The rest of the paper is structured as follows. Section II describes feature extraction techniques and the steps involved in it. Section III describes SVM in detail. Section IV elaborates SMO algorithm.Section V involves the experimental results.

II. FEATURE EXTRACTION

In the speech recognition system, the features given as input plays a very important role. The input given to the classifier should be distinct enough such that better classification results are obtained.Three different types of features were selected as input to the classifier.i.e.FFT,MFCC,VAD-MFCC and the classification performance was analyzed to find which is the better acoustic feature.

A. Fast Fourier Transform(FFT)

Fast Fourier Transform[18] is a fast implementation of Discrete Fourier Transform which is given as in (1)

$$X_i(k) = \sum_{n=1}^{N} xi(n)h(n)\, e^{-j\,2\pi kn\,/_N}$$

$$for\ 1 \le k\ \le K \qquad (1)$$

Where k is the length of DFT.

Given N (real or complex) samples $x_0,x_1,..x_{N-1}$, the FFT of these are N complex numbers $y_0,y_1,y_2\ldots.y_{N-1}$ and is given by (2)

$$y_j = \frac{1}{N^{\frac{(1-a)}{2}}} \sum_{k=0}^{N-1} e^{\frac{2\pi i}{N} b_{jk}} X_k \quad for\ j = 0,1,\ldots N \qquad (2)$$

The results of FFT applied to speech and noise signal is shown in Fig 2.

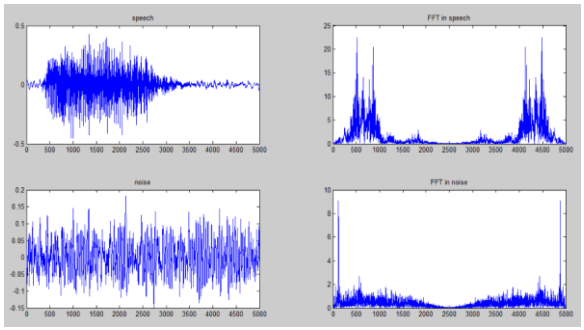

Fig. 2.FFT applied to speech and noise for a sample size of 5000.

B. Voice Activity Detection(VAD)

Voice activity detection [19] can be done with the help of three important features such as short time energy, spectral centroid and zero crossing rate, which distinguishes voice from noise.

*1) Short time Energy:*Short time energy[20] for each frame is computed with the help of the formula given in (3) where $x_i(n)$,n=1,…N is the audio sample of the ith frame of length N.

$$E = 10\ log_{10}\left(\frac{1}{N}\sum_{n=0}^{N-1} x^2(n)\right) \qquad (3)$$

*2) Spectral Centroid:*Spectral centroid C_i of the i-th frame is defined as Centre of gravity of its spectrum.Its also the measure of spectral position with high values corresponding to "brighter" sounds and is computed using the formula in (4)

$$S_c = \frac{\sum_{k=0}^{k-1}(k+1)X_i(k)}{\sum_{k=0}^{k-1} X_i(k)} \qquad (4)$$

Where k = 1,…N and X_i is the DFT coefficient of the ith frame.

*3)Zero crossing rate:*Zero crossing rate[21] is the measure of the number of times the signal changes its sign and is computed using the formula as in (5)

$$ZCR = \frac{1}{2}\sum_{n=1}^{N-1} |sgn(x[n]) - sgn(x[n-1])| \qquad (5)$$

With the help of these features,the threshold values for voice can be detected in the given signal.The results of VAD applied to voice and noise signal is shown in Fig. 3 and Fig. 4 respectively.

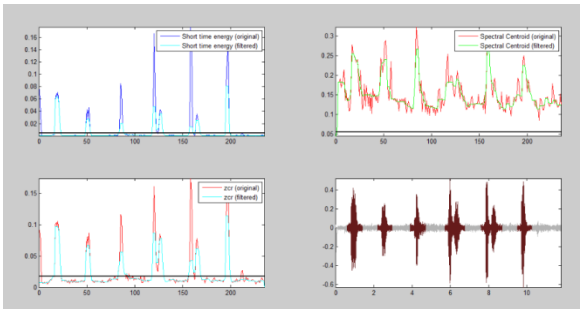

Fig. 3.Voice detected from the input signal after thresholding using three features described in VAD.

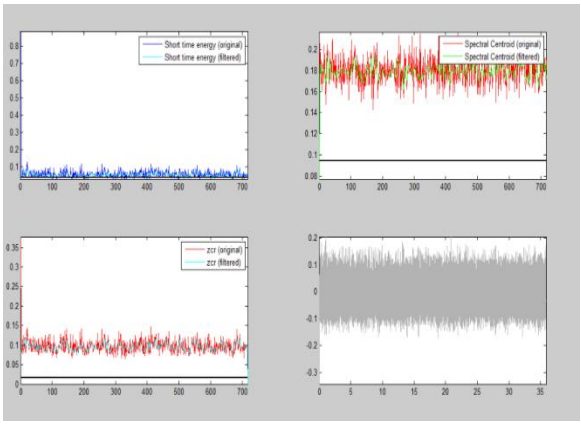

Fig. 4.Voice activity detection applied to noise.

C. Mel Frequency Cepstral Coefficients (MFCC)

Mel frequency cepstral coefficients [22] have been widely employed to obtain acoustic features in speech recognition system. MFCC's were originally suggested by Paul Mermelstein[23][24] and Bridle and Brown where they used a set of 19 cepstral coefficients.There are seven main

steps involved in MFCC as shown in Fig. 5. They are described as follows

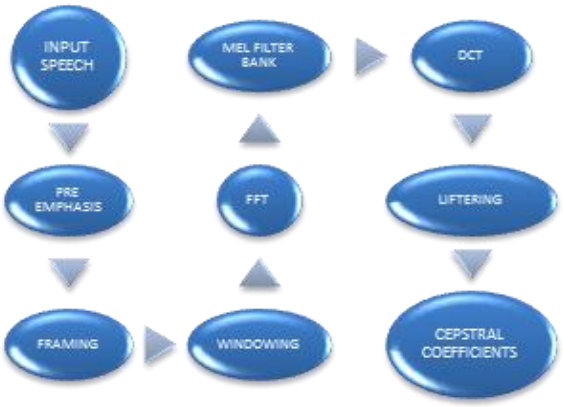

Fig 5. Flow chart of steps involved in MFCC

*1) Pre emphasis:*This is done in order to compensate for rapid decaying of speech. The input signal is passed through a digital filter which emphasizes signal with higher frequencies.This is done using the formula in (6)

$$s'_n = s_n - \propto s_{n-1} \qquad (6)$$

Where $\propto$ is the pre-emphasis filter coefficient and is in the range (0,1)

*2) Framing:*Framing helps in segmenting speech signal into smaller frames such that the range of each frame isbetween 20 to 40ms.Since audio signals vary continuously shorter frames are considered.

*3) Windowing:*Windowingis done in order to minimize disruptions at the start and end of the frame. Generally hamming window (7) is used for this purpose.

$$w(n) = \begin{matrix} 0.54-0.46\cos\left(\frac{2\pi n}{L-1}\right) & for\ 0 \le N \le L-1 \\ 0 & Otherwise \end{matrix} \qquad (7)$$

*4) Fast Fourier Transform- one side:*After applying window to the framed signals,fastfourier transform of the signal is done. FFT converts a signal into magnitudes and phases of various sine and cosine frequencies making up the signal.After using FFT,N/2 signals i.e.half of the FFT output is considered as input for the next step.This is done because the other half of the FFT signal is a mirror image of the first half.

*5) Melfilter bank:*Mels[25] are computed from min to max value by using the formula in (8)

$$Mel2Hz = 700 * \exp\left(\frac{Hz}{1125} - 700\right) \qquad (8)$$

The Mels are converted back to hertz by using (9)

$$Hz2mel = 1125 * \log\left(1 + \frac{Hz}{700}\right) \qquad (9)$$

The cut off frequency which should be less than half the sampling rate is computed using the formula in (10)

$$H_m(k) = \begin{cases} 0 \quad for\ f(k) < fc(m-1)\ and\ f(k) \ge fc(m+1) \\ \dfrac{f(k) - fc(m-1)}{fc(m) - fc(m-1)} \quad for\ fc(m-1) \le f(k) < fc(m) \\ \dfrac{f(k) - fc(m+1)}{fc(m) - fc(m+_1)} \quad for\ fc(m-1) \le f(k) < fc(m) \end{cases} \qquad (10)$$

Where$f_c(m)$: center frequency,

$f(k)$: frequency of the k[th] sample.

Thus the uniformly spaced triangular filters on mel scale between lower and upper frequency limits is applied to magnitude spectrum obtained with the help of FFT in order to produce filter bank energies which is computed using (11).

$$E = log \sum_{n=1}^{N} s_n^2 \qquad (11)$$

Where s_n is the magnitude of nth speech signal and N is total number of input samples.

*6) Discrete Cosine Transform(DCT):*Discrete cosine Transform is applied with the help of the formula as in (12)

$$dct(e[m]) = \sum_{n=0}^{M-1} e[m] cos\left(\frac{\pi n\left(m + \frac{1}{2}\right)}{M}\right)$$

$$for\ 0 \le n \le M \qquad (12)$$

Where e[m] is the output energy of the m[th] filter, M is the number of triangular filters and N is number of cepstral coefficients.This DCT is applied to log of filter bank energies obtained in the previous step.

*7) Liftering:*Filter operating in cepstrum domain is called liftering. Sinusoidal liftering[26] is applied using the formula in (13)

$$c'_n = \left[1 + \frac{L}{2} sin \frac{\pi n}{L}\right] c_n \qquad (13)$$

where L is the lifter parameter.

The coefficients obtained after the DCT step and Liftering is shown in Fig. 6.The difference between the coefficients of voice and noise signal can be observed in the figure.

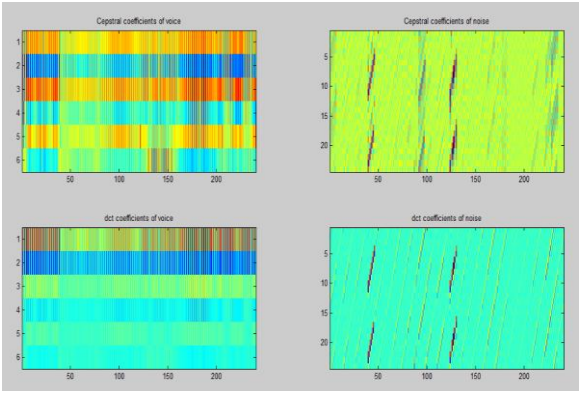

Fig 6.12*20 Coefficients obtained for 6 words after Liftering and DCT.

*8) Delta:*The delta coefficients obtained using (14) is also known as differential coefficients and delta-delta coefficients are known as acceleration coefficients.The differential and acceleration coefficients help in getting additional information about the speech i.e. it helps in tracking the variations of MFCC over time.

$$d_t = \frac{\sum_{n=1}^{N} n(c_{t+n} - c_{t-n})}{2\sum_{n=1}^{N} n^2} \quad (14)$$ Similarly delta-delta coefficients can also be calculated by replacing static coefficients by delta coefficients. Thus MFCC gives M*N coefficients where M is number of cepstral coefficients and N is number of triangular filters applied.

III. SUPPORT VECTOR MACHINES (SVM)

Machine learning[27] focuses on prediction based on known properties. In machine learning at first a model is built where the inputs are defined. Based on this model, the machine is trained and the predictions/ decisions are done. Machine learning can be widely classified into three categories: supervised learning, unsupervised learning and semi-supervised learning.Support vector machine(SVM)is an example of supervised learning where the learning algorithms are used to analyze data and recognize patterns. It was originally suggested by V.Vapnik[28].SVMs can be classified into linear svm and nonlinear svm based on the input kernel .When a set of data is given and is to be separated by a single feature, then the hyper plane to be defined is one-dimensional(point) and when the data is to be separated based on 2 features then the hyper plane defined is two-dimensional (line).SVM classifies the data without modeling a probability distribution. A linear kernel in SVM can be classified by good well defined hyperplanes.A two dimensional hyperplane is shown in Fig.7.The hyper plane can be defined as in (15).

$$w.x - b = 1$$
$$w.x - b = -1 \quad (15)$$

Where the training data is given as in (16)

$$D = \{(X_i, y_i) \mid X_i \in R^P, y_i \in \{-1,1\}\}_{i=1}^{n} \quad (16)$$

The distance between the two hyper planes is called the margin and the closest to the hyper planes are the support vectors.Thus the area of interest here is to find maximum margin hyper plane which divides the data having $y_i = 1$ from the data containing $y_i = -1$.

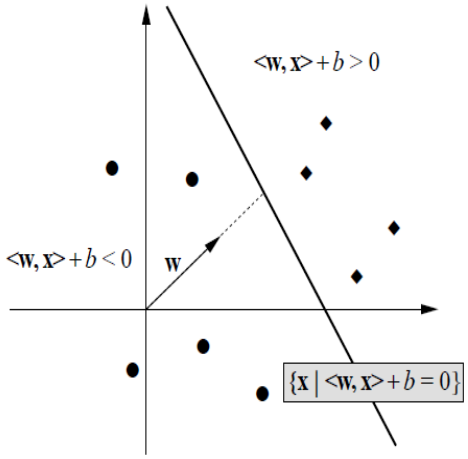

Fig. 7. Two dimensional hyper plane[29]

Hence the optimization problem can be defined as in (17)

$$\max_{\propto} \varphi(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} y_i y_j k(x_i, x_j)\, \alpha_i \alpha_j \quad (17)$$

Which is subject to

$$\sum_{i=1}^{N} y_i \propto_i = 0, \quad 0 \leq \propto_i \leq C, and\ i = 1 \ldots n$$

Where x_i is the training sample ,y denotes the label i.e.+1 or -1,C is the cost parameter and $\propto_I$ represents lagrange multiplier. This Quadratic Programming problem is addressed with the help of Optimization algorithms.

The optimization of svm can be done using Chunking algorithm[30],Osuna theorem[31],Decomposition algorithm[32] and Sequential Minimal Optimization(SMO)[33].In chunking algorithm all the zero multipliers are discarded. The non-zero multipliers are identified and all the examples which violate kkt conditions are considered. In Osuna's theorem the large quadratic programming problem is broken down into smaller series of problems. In Osuna'stheorem,the major drawback was that,the algorithm suggests adding one example and subtracting one example every step.In decomposition algorithm,the strategy used is similar to the one used in active set strategies where the optimization problem is addressed as inequality constratints of simple bounds(Gill et al[34]). SMO was suggested by Platt and the main advantage of this algorithm is that the optimization problem can be solved analytically. Its a special derivative of Osuna's theorem and is similar to Perceptron learning algorithm which finds a linear separator by adjusting weights on misclassified examples.

IV. OPTIMIZATION ALGORITHMS

A. *Sequential Minimal Optimization(SMO)*

SMO addresses the QP problem using Osuna's theorem. The smallest and possible optimization problem is solved at every step. It overcomes the disadvantage of high computational load of other optimization algorithms and is considered to be the best optimization algorithm for SVM.In SMO two Lagrange multipliers are considered at once. The first Lagrange multiplier is used for outer loop and the second lagrange multiplier is used to maximize the size of the step taken during joint optimization.The flow chart for SMO is given in Fig.8.

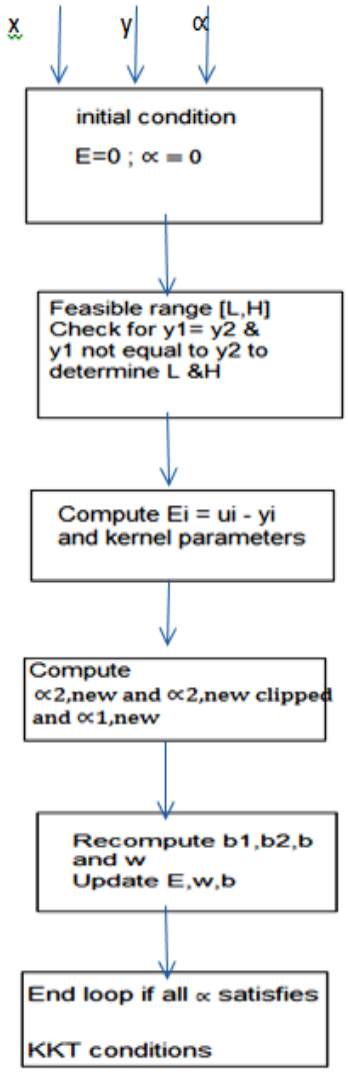

Fig. 8. Flow chart of SMO algorithm

Three conditions are to be satisfied by these Lagrange multipliers which is considered as KKT condition as given in (18)

$$\propto_i = 0 \quad => R_i \geq 0$$

$$0 < \propto_i < C \quad => R_i \approx 0$$

$$\propto_i = C \quad => R_i \leq 0 \qquad (18)$$

Where $R_i = y_i * E_i$

Here linear kernel is considered which is represented as $k_{ii} = x_i * x_i^T$.The second derivative of the objective function ,neta is described as in (19)

$$neta = 2k_{ij} - k_{ii} - k_{jj} \qquad (19)$$

The lagrange multiplier is computed as in (20)

$$\propto_j^{new} = \propto_j^{old} + \frac{y_j(E_j^{old} - E_i^{old})}{neta} \qquad (20)$$

Where the error is computed as $E_i = u_i - y_i$.

The constrained maximum is clipped to the unconstrained maximum by using (21).

$$\propto_j^{new,clipped} = \begin{cases} H, & if \ \propto_j^{new} \geq H \\ \propto_j^{new}, & if \ L < \propto_j^{new} < H \\ L, & if \ \propto_j^{new} \leq L \end{cases} \qquad (21)$$

$$\propto_i^{new} = \propto_i^{old} + y_i y_j (\propto_j^{old} - \propto_j^{new,clipped}) \qquad (22)$$

B.*Decomposition Strategy in SVM^{light}*

SVM^{light} [35] suggested by T.Joachims is used for application of SVM to large domains. It overcomes the disadvantage of other algorithms where long training time is required if the input given is very large.This is accomplished by successive "shrinking" where two strategies are followed i.e. one alpha is always considered to be in upper bound and less support vectors are considered.The optimization of SVM^{light} is nothing but the decomposition strategy involved in Osuna et al. In this algorithm, in each iteration alpha is divided into two categories: set of free variables that are updated in current iteration and set of fixed variables that are temporarily fixed at current value.And when the optimality condition is violated,a set of q variables are considered as free variables and is solved for QP problem.The strategies involved in choosing the working set is different in both SMO and SVM^{light}.In SVM^{light} steepest feasible descent approach is used whereas in SMO heuristic approach is used.SMO accelerates speed of computation for linear SVMs whereas caching concept is used in SVM^{light}.

V. EXPERIMENTAL RESULTS

A. *Feature Extraction*

Performance of two optimization algorithms i.e.SMO of svm and decomposition strategy SVM^{light} are compared. At first 26 alphabets were recorded by 5 different people i.e. 4 male and 1 female voice.In the recorded sample, FFT of 3000 samples (which was the highest for a letter) per word was taken as input to svm. Since this added a lot of non-speech area to the given signal, the input sample size was considered as 1500 samples per letter. Thus from a single recording 26*1500 samples were considered as input to

svm after applying FFT to these alphabets.The results are tabulated in Table 1.

TABLE 1.COMPARISON OF CLASSIFICATION RESULTS OF FFT FOR 26 ALPHABET UTTERANCES OF 5 DIFFERENT VOICES

INPUT	SUCCESS RATE OF CLASSIFICATION	
	SMO(%)	SVM^{light}(%)
ALPHABETS FFT (3000)	90.384	92.31
ALPHABETS FFT (1500)	94.2308	98.08

In order to obtain better features an analysis on number of cepstral coefficients to be taken was done.Six command words i.e.Start, Stop, Right, Left, Forward and Backward were recorded by 4 people(2 male and 2 female voice).Different set of coefficients were given as input to svm.A set of 39*N coefficients where N is the number of triangular filters were given by concatenating cepstral coefficients and energy along with their delta and delta-delta derivatives. Similarly a set of 36*N(cepstral coefficients along with first and second derivative) ,13*N(cepstral coefficient along with energy coefficient) was considered as input to svm and the corresponding results were compared and tabulated in Table 2 and Table 3.Similarly the VAD was applied to the input signal and the silent part was removed in the input speech.

TABLE 2.CLASSIFICATION RESULTS OF 4 DIFFERENT INPUT SAMPLE SIZES FOR 6 WORDS ('START','STOP' ,'RIGHT','LEFT','FORWARD','BACKWARD') OF 4 DIFFERENT VOICES.THE SPEECH WAS COMPARED AGAINST NON-SPEECH

INPUT	SUCCESS RATE OF CLASSIFICATION	
	SMO(%)	SVM^{light}(%)
MFCC 36*120	92.778	86.66
MFCC 39*120	93.889	88.89
MFCC 12*120	90.553	83.33
MFCC 13*120	88.889	66.67

TABLE 3.CLASSIFICATION RESULTS OF 4 DIFFERENT INPUT SAMPLE SIZES FOR 6 WORDS ('START','STOP' ,'RIGHT','LEFT','FORWARD','BACKWARD') OF 4 DIFFERENT VOICES.THE SPEECH WAS COMPARED AGAINST LOUD NOISE

INPUT	SUCCESS RATE OF CLASSIFICATION	
	SMO(%)	SVM^{light}(%)
MFCC 36*120	72.22	57.14
MFCC 39*120	77.77	61.11
MFCC 12*120	66.11	44.44
MFCC 13*120	64.44	61.11

The comparison of efficiency of the proposed method and already existing algorithm is tabulated in Table .5.MFCC was computed to this data after removing the silent part and was given to the svm.

TABLE 4.CLASSIFICATION RESULTS OF VAD-MFCC

	SMO (%)	SVM^{light}(%)
VAD_MFCC	97.5	87.5

TABLE 5.COMPARISON OF EFFICIENCY

	SMO (%)
VAD_MFCC	97.5
[36]	95

The flow chart of algorithm for better speech recognition is shown in Fig. 9.

Fig. 9.Algorithm for better speech recognition

B.*Hardware Implementation of SMO*

SMO algorithm can be divided into three main modules[37]. The first module is where the parameters are initialized and kkt condition is verified. In the second module the lagrange multipliers are computed and updated and in the third module the svm parameters are updated.The state machine involved in SMO algorithm is shown in Fig.10 where N1 goes high when interrupt signal is not generated in pre-process step.N3 is set high if all $\propto$ satisfy kkt conditions and if not N2 is set high and the loop is not terminated.

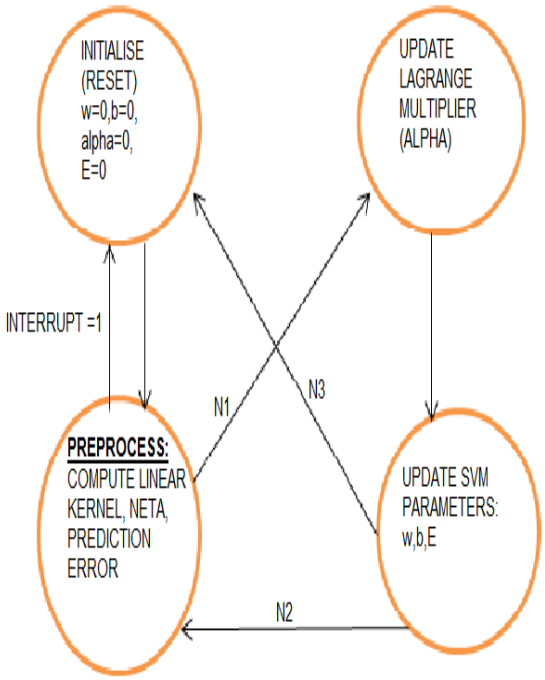

Fig. 10.State machine of SMO algorithm

As given in the [37] the VLSI architecture of SMO was considered breaking the whole algorithm into three modules. The Verilog modules were done with the help of Modelsim.Considering the pseudocode of SMO,RTL design was done using Verilog in Modelsim and this code was dumped into the Altium Nanoboard NB3000XN using Altium designer tool.14.3.The error signals R1 and R2 were found to vary when conditions such as different values for the labels, infeasible range of L and H were observed.These variation in the error signals were studied using the LEDs in the Altium NanoBoard. The number of resources used in the design is tabulated in Table 6. and the output of the algorithm is shown in Fig . 11.

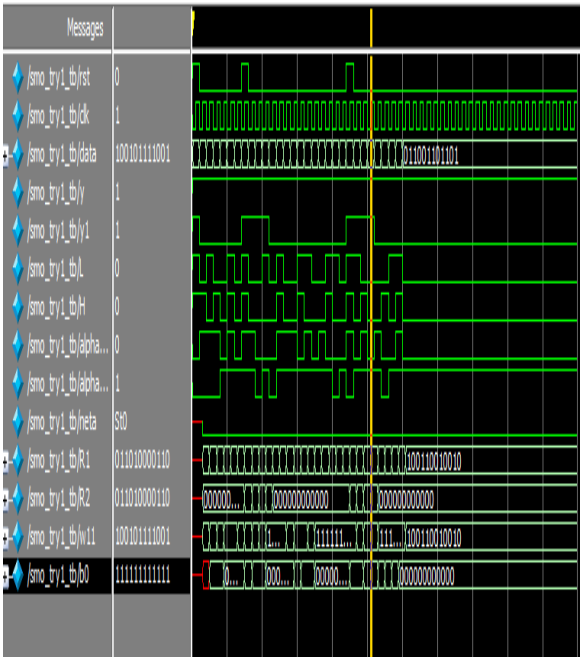

Fig 11.Verilog output for SMO algorithm

TABLE 6.RESOURCE UTILISATION OF ALTIUM NANOBOARD 3000 XN.

FPGA RESOURCES	Number used	Maximum available	Used rate
4-INPUT LUT	198	22528	1%
BUFGMUXs	1	24	4%
I/O PINS	30	502	5%
MUTLT	32	32	100%
SLICE FLIP FLOPS	227	22528	1%
SLICES WITH RELATED LOGIC	199	199	100%
SLICES WITH UNRELATED LOGIC	0	199	0%
SLICES	199	11264	1%
TOTAL 4-INPUT LUT	237	22528	1%

VI. CONCLUSION

This project involved two phases. The first phase involved choosing better features for speech recognition and the second phase was to implement SMO in FPGA. A wide comparison of sample size of input signal along with optimization algorithms and feature extraction methods were done.It was observed that the classification results for alphabets were better when 1500 sample size per word was used after applying FFT to the input,when decomposition strategy of SVM^{light} was used for optimization.With respect to words,better classification results were obtained in general for SMO optimization especially when the input sample size was 39*120.The algorithm suggested (VAD-MFCC-SVM) yielded better results compared to previous speech recognition algorithms.Since SMO optimization yielded better results,Verilog implementation of this algorithm was done and the same was tested using FPGA.This analysis helped in understanding better algorithms for speech recognition along with a deeper understanding of optimization algorithms involved in support vector machines.

ACKNOWLEDGMENT

The author would like to express profound gratitude to Mr. Suriya Prakash.J ,Scientist (CEERI) for his valuable suggestions and constant encouragement throughout this project, without which this work wouldn't have been possible.

REFERENCES

(1) D.Reynolds and R.C.Rose, "Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models",IEEE Transactions on Speech and Audio Processing,vol 3,No 1,January 1995.
(2) L.R.Rabiner,"A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition",Proceedings of the IEEE,Vol.77,No 2,February,1989.
(3) Lawrence Rabiner and Biing-Hwang Juang,"Fundamentals of Speech Recognition",published by PTR Prentice-Hall,1993.
(4) Deng,Li,DouglasO'Shaughnessy,"Speech Processing: A dynamic and optimization-oriented approach",pp.41-48,ISBN 0-8247-4040-8.
(5) A.Michael Noll and M.R.Schroeder,"Short-time Cepstrum Pitch Detection",Journal of the Acoustical Society of America,Vol. 36,No. 2,pp. 296-302.
(6) C.R.Rashmi,"Review of Algorithms and Applications in Speech Recognition System",International Journal of Computer Science and Information Technologies,Vol 5,2014.
(7) L.E.Baum and T.Petrie,"Statistical Influence for Probabilistic Functions of Finite State Markov Chains",The Annals of Mathematical Statistics,37(6):1554-1563,1963.
(8) ShumailaIqbal, TahiraMehboob, Malik, "Voice Recognition using HMM with MFCC for secure ATM",IJCS Vol.8,Issue 6 Nov 2011.
(9) I.Patel and SrinivasRao,"Speech Recognition using HMM with MFCC-An analysis using frequency specral Decomposition technique",Signal and Image Processing:An International Journal,vol.1,No.2,December 2010.
(10) Young- KyuChoi,KisunYou,JungwookChoi,WonyongSung,"A real time FPGA based 20000 word speech recognizer with optimized DRAM Access",IEEE Transactions on Circuits and Systems,Vol 57,No 8,August 2010.
(11) AravingGanapathiraju, Jonathan E .Hamaker ,Joseph Picone,"Applications of support vector machines to speech recognition",IEEE Transactions on Signal Processing,Vol 52,No.8.August 2004.
(12) J.P.Sendra,D.M.Iglesias,F.diaz-de-Marla,"Support vector machines for continuous speech recognition",14th European Signal Processing Conference, Florence, Italy, September 2006.
(13) P.D.Duy,V.D.Lung,N.Q.D.Trang,N.C.Toan,"Speech recognition on robot controller implemented on FPGA",Journal of Automation and Control Engineering, Vol.1,No.3,September 2013.
(14) Cheng-Yuan Chang,Ching-FaChen,Shing-Tai Pan,Xu-Yu Li,"The speech recognition chip implementation on FPGA",2nd International Conference on Mechanical and Electronics Engineering(ICMEE),2010.

(15) Haojun Wu,YongWang,JiwuHuang,"Identification of Electronic Disguised Voices",IEEE Trans on Information Forensics and Security,Vol 9,No 3,March 2014.
(16) SahilVerma,TarunGulati,RohitLamba,"Recognizing voice for numeric using MFCC and DTW",International Journal of Application or Innovation in Engineering and Management(IJAIEM),Vol 2,Issue 5,May 2013.
(17) Chadawan Ittichaichareon,SiwatSuksri and ThaweesakYingthawornsuk ,"Speech recognition using MFCC" , International Conference on Computer Graphics,Simulation and Modelling,Thailand,July 2012.
(18) Chris Lomont,"The Fast Fourier Transform",Jan 2010.
(19) T.Giannakopoulos,"A method for silence removal and segmentation of speech signals,implemented in MATLAB".
(20) A. Pikrakis, T. Giannakopoulos, and S. Theodoridis, "Anoverview of speech/music discrimination techniques in thecontext of audio recordings," vol. 120, pp. 81–102, 2008.
(21) "Short Term Time Domain Processing of Speech manual",Sakshat virtual Lab,IIT Guwahati.
(22) Y.Lavner and DimaRuinskly,"A decision-tree-based Algorithm for speech/Music classification and segmentation",EURASIP Journal on Audio,Speech and Music Processing,June,2009.
(23) P.Mermelstein,"Distance measures for speech recognition,psychological and instrumental",in Pattern Recognition and Artificial Intelligence, C.H. Chen, Ed ,pp.374-388.Academic,New York.
(24) S.B.Davis and P.Mermelstein(1980),"Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken sentences", in IEEE Transactions on Acoustics,Speech,and Signal Processing,28(4),pp. 357-366.
(25) "Mel Frequency Cepstral Coefficient(MFCC) tutorial",Practical cryptography.
(26) DebalinaGhosh et al"A Comparative Study of Performance of FPGA Based Mel Filter Bank & Bark filter Bank " ,Cornell University Library.
(27) Ron Kohavi,Foster Provost, "Glossary of terms", Machine Learning ,vol 30,271-274,1998.
(28) V.Vapnik,"The Nature of Statistical Learning Theory",Berlin,Germany,Springer-Verlag,1995.
(29) V.Vapnik,"Estimation of Dependencies Based on Emprical Data",Springer-Verlag,1982.
(30) Jason Weston,"Support Vector Machine and Statistical Learning Theory Tutorial".
(31) Edgar E.Osuna,Robert Freund and Federico Girosi, "Support Vector Machines: Training and Applications",A.I.Memo No1602,C.B.C.I. Paper No.144,March 1994.
(32) C.C.Chang, C.W.Hsu and C.J.Lin , "The analysis of decomposition methods for support vector machines",IEEETrans.NeuralNetworks,vol 11,no.4,pp. 1003-1008,Jul. 2000.
(33) J.C.Platt,"Fast training of support vector machines using sequential minimal optimization", in Advances in kernel Methods of Support Vector Machine, B. Schölkopf, C. Burges, and A. Smola, Eds.Cambridge,MA:MIT Press,1998.
(34) P.E.Gill,W.Murrayand M.H.Wright,"PracticalOptimization",Academic Press,1981.
(35) T. Joachims, "Making large-scale SVM learning practical," in Advancesin Kernel Methods—Support Vector Learning, B. Schölkopf,C. J. C. Burges, and A. J. Smola, Eds. Cambridge, MA: MIT Press,1998.
(36) S.M.Kamruzzam,A.N.M>R>Karim,Md.SaifIslam,Md.EmdadulHaque,"Speaker Identification using MFCC-Domain Support vector Machine".
(37) Ta-Wen Kuan,Jhing-FaWang,Jia-ChingWang,Po-Chuan Lin and Gaung-HuiGu ,"VLSI Design of an SVM Learning Core on Sequential Minimal Optimization Algorithm",IEEE Transactions on Very Large Scale Integration(VLSI) systems,Vol 20,No 4, April 2012.