ASIAN JOURNAL OF PHARMACEUTICAL AND CLINICAL RESEARCH

INNOVARE ACADEMIC SCIENCES
Knowledge to Innovation

Online - 2455-3891
Print - 0974-2441

Special Issue (April)

Full Proceeding Paper

FIELD-PROGRAMMABLE GATE ARRAY IMPLEMENTATION OF THE DYNAMIC TIME WARPING ALGORITHM FOR SPEECH RECOGNITION

MITALI BHOJWANI, JOHN SAHAYA RANI ALEX*

School of Electronics Engineering, VIT University, Chennai, Tamil Nadu, India. Email: jsranialex@vit.ac.in

Received: 19 January 2017, Revised and Accepted: 20 February 2017

ABSTRACT

Objective of this research is to implement a speech recognition algorithm in smaller form factor device. Speech recognition is an extensively used in mobile and in numerous consumer electronics devices. Dynamic time warping (DTW) method which is based on dynamic programming is chosen to be implemented for speech recognition because of the latest trend in evolving computing power. Implementation of DTW in field-programmable gate array is chosen for its featured flexibility, parallelization and shorter time to market. The above algorithm is implemented using Verilog on Xilinx ISE. The warping cost is less if the similarity is found and is more for dissimilar sequences which is verified in the simulation output. The results indicate that real time implementation of DTW based speech recognition could be done in future.

Keywords: Dynamic time warping, Field-programmable gate array, Speech recognition.

 DOI: http://dx.doi.org/10.22159/ajpcr.2017.v10s1.19753

INTRODUCTION

Dynamic time warping (DTW) is a dynamic programming based algorithm which measures the similarities of the two different time series sequences which may vary in length. Earlier researchers did not use this algorithm because of it computational complexity. However, this algorithm is revisited by researchers because of the current trends in computing power.

In speech recognition, the speech features are extracted by the feature extraction method from the speech signal and the features are treated as time series sequence. Speech feature vectors are compared using DTW for similarity check. Efficient algorithms have been proposed which overcomes the drawbacks of the existing DTW techniques. DTW is used for automatic speech recognition in car environment [1]. Some of the other applications using DTW are bird call recognition, online signature verification, pattern recognition, and database mining [1-4].

2D DTW has been proposed and developed to match the Chinese calligraphic characters [5]. DTW is used in different types of pattern matching application using the threshold DTW [6]. Data mining and clustering application also uses DTW using different algorithm called DTW Barycenter averaging (DBA) with details of the complexity involved in the execution [7]. The shortest warping path is used to find the match through a process of time series classification and proposes weighted DTW (WDTW) algorithm for efficient matching [8]. The comparison has been developed between the hidden Markov model and DTW technique which gives the understanding for the fastest computation for the isolated word recognition [9]. The computational complexity of DTW algorithm is known for efficient performance [10]. The local and the global distance matrix which is used for calculating the A matrix for the shortest path is studied [11].

Most of the DTW algorithms being implemented using Matlab. DTW algorithm is implemented using digital signal processing processors for speech recognition [12]. Various field-programmable gate array (FPGA) implementations of DTW algorithm for applications are explored [13-16].

In this paper, simple digital architecture is designed for DTW algorithm and implemented in FPGA for speech recognition. The comparative study for similarity sequence and dissimilarity sequences are analyzed. This paper is organized as follows, section II describes the block diagram of the feature template extraction, section III describes the DTW block implementation, and section IV describes the simulation results. These are followed by conclusions and future scope in section V.

FEATURE EXTRACTION

The Feature extraction block diagram is shown in Fig. 1. In this, the given incoming speech waveform is pre-emphasized before giving in to feature extraction block.

The feature extraction block provides the selective attributes of the speech signal which are called feature vectors. One such popular feature extraction method is mel-frequency cepstral coefficient (MFCC) method [17]. The block diagram of MFCC feature extraction is shown in Fig. 2. The speech signal is divided into 20 ms length of frames with an overlap of 10 ms. Each framed signal is passed through the mel filter banks which imitates the psychoacoustic model of the ear. The log energies are computed to imitate the dB scale of the ear. Finally, de-correlation is performed using discrete cosine block to remove the initial overlap. The speech features are stored in the read only memory (ROM) for further processing.

Normally, the feature extraction is done for the set of words that is stored in ROM as dictionary. The templates are matched with the input speech data to find the match using the DTW technique.

DTW TEMPLATE MATCHING

DTW algorithm

$$A_i = \{a_1, a_2, a_3 \ldots\ldots\ldots, a_m\} \quad (1)$$

Then we have the featured spoken word given as:

$$B_j = \{b_1, b_2, b_3 \ldots\ldots\ldots\ldots\ldots, b_n\} \quad (2)$$

Step 1:

The absolute distance between the two elements A_i, B_j is D_{ij}. This result in a local distance matrix of length n*m as given by the following equation:

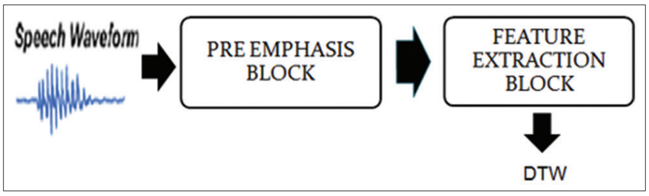

Fig. 1: Block diagram of feature extraction

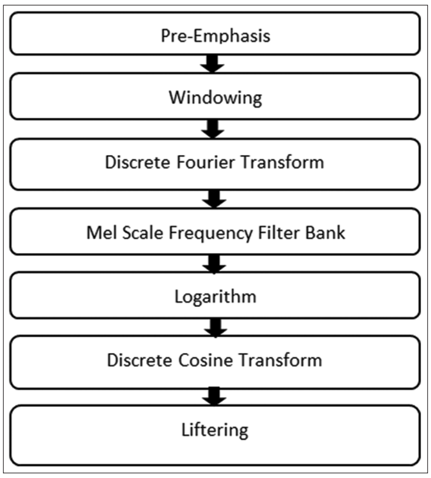

Fig. 2: Mel-frequency cepstral coefficient feature extraction method

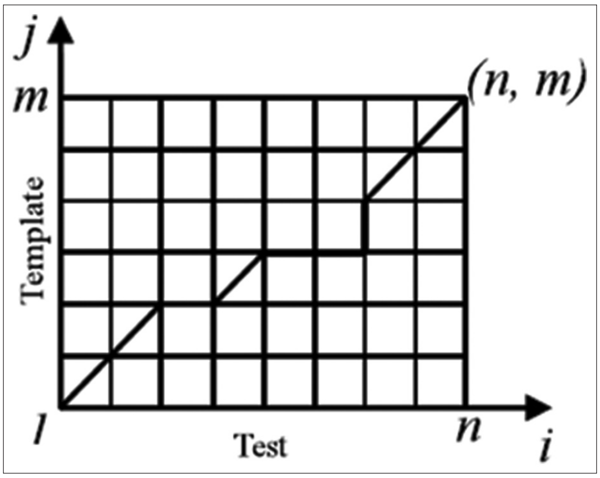

Fig. 3: Illustration of dynamic time warping

$D_{ij}=|a_i-b_j|$ (3)

The above operation results in the matrix of distance m rows and n columns represented by a D matrix. The obtained matrix is stored in a register for next computation.

Step 2:

The global distance matrix can be calculated from the local distance matrix through the following steps:

Starting with the calculation of:

$a(1,1)=D(1,1)$ (4)

Step 3:

Calculating the first row:

$a(i,1)=a(i,1)+d(i,1)$ (5)

Calculating the first column:

$a(1,j)=a(1,j)+d(1,j)$ (6)

The output generated by the above formula is stored in register.

Step 4:

The accumulator matrix of warped data is generated using the formula:

$a_{ij}=D_{ij}+\min(a_{i-1,j-1},a_{i-1,j},a_{i,j-1})$ (7)

In the above Equation 7 for generating the "a" matrix the comparison and the summation operation is done.

The computation starts from the 1, 1 position of the matrix. Comparator is used to generate the minimum of the adjacent values.

The output of the comparator is added with the value of city block distance matrix of the same location. The final step is to find the warping path or the shortest path of the mapped values in the matrix which is shown in Fig. 3.

Fig. 4 shows the steps involved in finding a match using DTW. In section II we found feature template extraction process that stores the features of the different words as dictionary in the ROM for matching.

The mod subtracter blocks fetches the feature templates and subtract it with the featured spoken word. The absolute values are then arranged in a matrix which is d matrix.

The D matrix register output is now sent to summation block and adder block so that A matrix can be computed.

For A matrix register we need a comparator so as to obtain the minimum values for calculating global matrix.

The values obtained are now used to find warping path, that is,to find similarity or dissimilarity between the two signals.

OUTPUT AND RESULTS

The dictionary template is taken as. "a" values. Moreover, the input speech template is taken as "b" inputs as shown in Fig. 5.

A_i={8,12,2,6},

B_j={10,14,5,4)

and

A_i={8,12,2,6}.

B_j={10,12,2,4}

The output generated for the absolute difference matrix is as shown in Fig. 6. This gives the matrix computed using the euclidean distance formula explained in section III.

The image in Fig. 7 gives the output of the A matrix register as shown in the block diagram in section III. The output is obtained using the $a_{i,j}$ formula.

The output obtained represents the M*N matrix generated.

Fig. 8 shows the dissimilarity obtained between the speech signals.

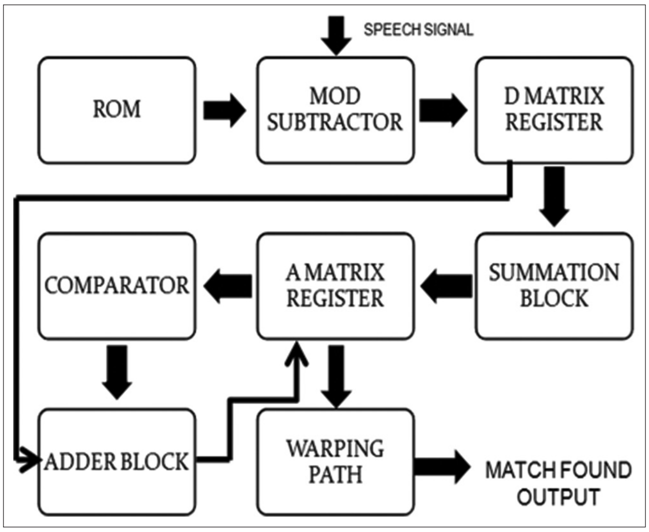

Fig. 4: Design diagram of dynamic time warping algorithm

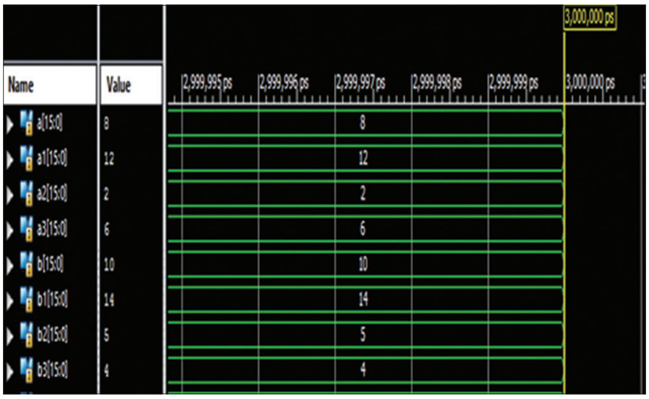

Fig. 5: Input matrix

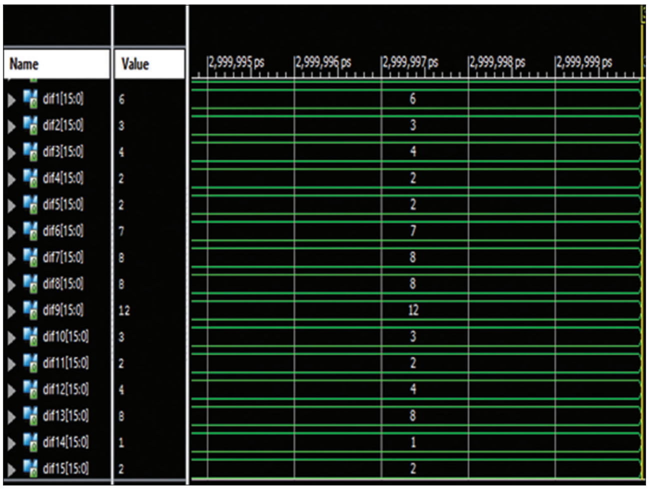

Fig. 6: Local distance "d" matrix output

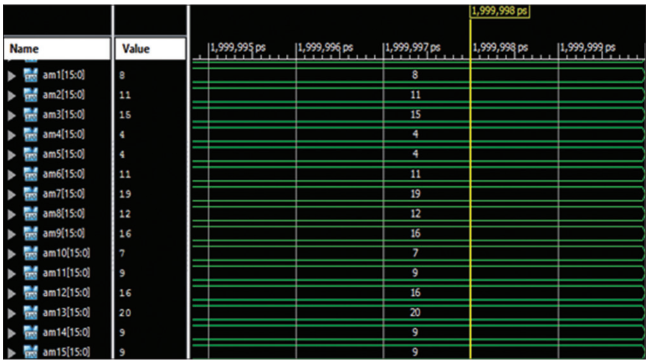

Fig. 7: Global distance "A" matrix

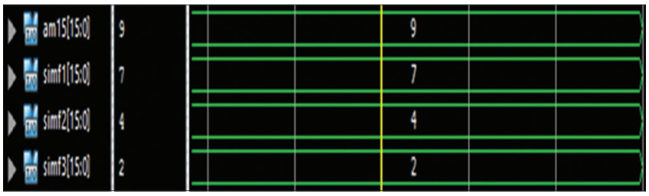

Fig. 8: Dis-similarity results of 2 time series sequence

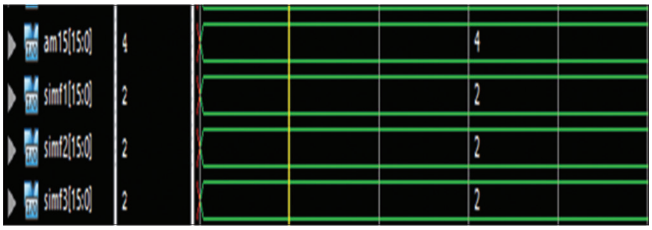

Fig. 9: Similarity

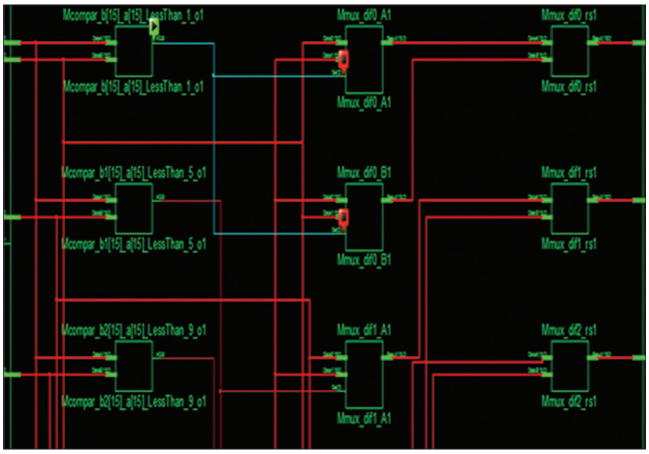

Fig. 10: Schematic overview of dynamic time warping algorithm

Device Utilization Summary (estimated values)			
Logic Utilization	Used	Available	Utilization
Number of Slices	813	2448	33%
Number of 4 input LUTs	1564	4896	31%
Number of bonded IOBs	401	66	607%
Number of GCLKs	1	24	4%

Fig. 11: Design summary

Fig. 9 shows the similarity obtained between the speech signals.

Fig. 10 shows the schematic view of the DTW algorithm which has difference output and comparator output and finally showing the warping path.

Fig. 11 gives detailed information of the design summary of the program.

The obtained output gives a detailed way the steps are performed with the execution of each step explained in the previous sections.

CONCLUSION AND FUTURE SCOPE

Recently, the DTW algorithm is used revisited because of the current trends in increased computing power. DTW algorithm is used in pattern recognition such as in biometric, bird call, speech recognition, and data mining. Since most of the pattern recognition employed in small form factor devices, for example, a mobile phone, an embedded device instead of a personal computer. From the literature, it can be found that the pattern recognition is very application specific. In this research work, the implementation of the DTW algorithm is done FPGA–Virtex6 ML506. The time series values are taken, and matrix has been generated, and the warping path is calculated using the basics of the DTW. The effective way of computing the distance has been done. The coefficients of speech signals can be obtained using MFCC method and similarity can be obtained using DTW algorithm.

REFERENCES

1. Syu D, Syu S, Ruan S, Huang Y, Yang C. FPGA Implementation of Automatic Speech Recognition System in a Car; 2015. p. 485-6.
2. López-García M, Ramos-Lara R, Miguel-Hurtado O, Cantó-Navarro E. Embedded system for biometric online signature verification. IEEE Trans Ind Inform 2014;10(1):491-501.
3. Liu H, Bergmann NW. An FPGA Softcore Based Implementation of a Bird Call Recognition System for Sensor Networks; 2010. p. 1-6.
4. Xinxing J, Xu S. Procedia engineering speech recognition based on efficient DTW algorithm and its DSP implementation. Procedia Eng 2012;29:832-6.
5. Zhang X, Zhuang Y. Dynamic time warping for Chinese calligraphic character matching and recognition. Elsevier J 2012;33(16):2262-9.
6. Adwan S, Arof H. On improving dynamic time warping for pattern matching. Elsevier J 2012;45:1609-20.
7. Petitjean F, Ketterlin A, Gancarski P. A global averaging method for dynamic time warping with applications to clustering. Pattern Recognit 2011;44:678-93.
8. Jeong YS, Jeong MK, Omitaomu OK. Weighted dynamic time warping for time series classification. Pattern Recognit 2011;44:2231-40.
9. Barbon S Jr, Guido RC, Vieira LS. Wavelet based dynamic time warping. Elsevier J 2007;J Comput Appl Math 2009;227(2):271-87.
10. Alex JS, Venkatesan N. Spoken utterance detection using dynamic time warping method along with a hashing technique. Int J Eng Technol 2014;6(2):1100-8.
11. Alex JS, Shivkumar TS, Venkatesan N. Adapted DTW joint with wavelet transform for isolated digit recognition. ARPN J Eng Appl Sci 2015;10(1):1-7.
12. Xihao S, Miyanaga Y. Dynamic time warping for speech recognition with training part to reduce the computation. IEEE J 2013;Signals, Circuits and Systems (ISSCS), 2013 International Symposium on, 11-12 July; 2013.
13. Suneja K, Bansal M. Hardware design of dynamic time warping algorithm based on FPGA in verilog. Int J Adv Res Electron Commun Eng2015;4(2):165-8.
14. Tai JS, Li KF. Dynamic Time Warping Algorithm : A Hardware Realization in VHDL; 2013. p. 31-4.
15. Ogawa M, Shibata T. A delay-encoding-logic array processor for dynamic-programming matching of data sequences. IEEE J Solid-State Circuits 2005;40(7):1578-82.
16. Ahem ME. A dynamic-time-warp integrated circuit for a 1000-word speech recognition system. IEEE J Solid-State Circuits 1987;22(1):3-14.
17. Alex J, Venkatesan N. Modified MFCC methods based on KL-transform and power law for robust. J Theor Appl Inf Technol 2014;67(2):527-32.