SHORT COMMUNICATION

A MODIFICATION OVER SAKOE AND CHIBA'S DYNAMIC TIME WARPING ALGORITHM FOR ISOLATED WORD RECOGNITION*

K.K. PALIWAL, Anant AGARWAL and Sarvajit S. SINHA

Speech and Digital Systems Group, Tata Institute of Fundamental Research, Homi Bhabha Road, Bombay 400005, India

Abstract. A modification over Sakoe and Chiba's dynamic time warping algorithm for isolated word recognition is proposed. It is shown that this modified algorithm works better without any slope constraint. Also, this algorithm not only consumes less computation time but also improves the word recognition accuracy.

Zusammenfassung. Eine Veränderung von Sakoe und Chiba's 'Dynamic Time warping' Algorithmus für die Erkennung von freistehende Wörter ist vorgestellt. Es wird gezeigt dass das veränderte Algorithmus ohne Abhangszwang bessere Ergebnisse gibt. Es hat sich herausgestellt dass die vereinfachte Methode weniger Rechnenzeit benötigt und dass sie die Erkennungs-schärfe erhöht.

Résumé. Une modification de l'algorithme de programmation dynamique de Sakoe et Chiba pour reconnaissance de mots isolés est proposée. Cet algorithme modifié fonctionne mieux sans contrainte de pente. Il est plus rapide en temps de calcul tout en permettant une amélioration du taux de reconnaissance.

Keywords. Speech, isolated word recognition, dyamic time warping.

1. Introduction

The dynamic time warping (DTW) algorithm provides a procedure to align optimally in time the test and reference patterns and to give the average distance associated with the optimal warping path. Sakoe and Chiba [1] proposed a DTW algorithm for spoken word recognition and showed experimentally its superiority over other algorithms reported in the literature [2-4]. In the present paper, a modification over Sakoe and Chiba's DTW algorithm [1] is proposed. The modified DTW algorithm is applied to a speaker dependent isolated word recognition system. It is

* This paper has been adjudged as the first paper in a Student Paper Contest (1981) conducted by the IEEE India Council.

shown that the modified algorithm works better without the slope constraint. Also, this algorithm consumes less computation time and improves the word recognition accuracy.

2. The modified DTW algorithm

Fig. 1 shows diagrammatically the difference between Sakoe and Chiba's DTW algorithm [1] and the modified DTW algorithm proposed in this paper. The test pattern

$$A = a_0, a_1, \ldots, a_i, \ldots, a_I$$

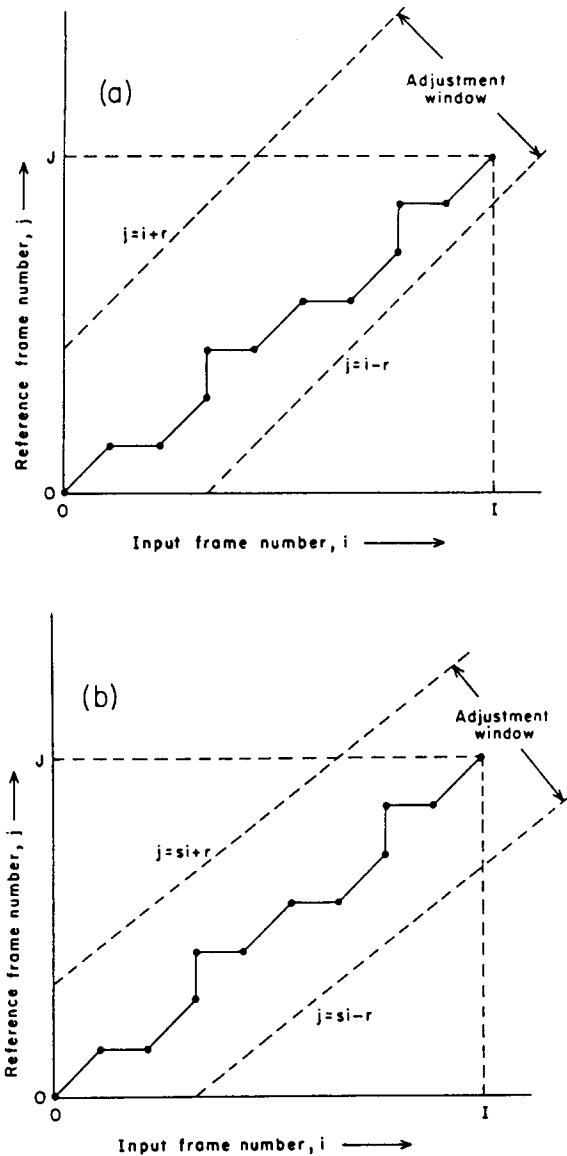and the reference pattern

$$B = b_0, b_1, \ldots, b_j, \ldots, b_J$$

Each spoken digit is represented by a time pattern of feature vectors, each feature vector characterizing a single frame. The test pattern representing the unknown input digit is compared against the reference patterns of all ten digits in the vocabulary and classified as belonging to the digit showing the best match. The symmetric form of DTW algorithm is used to optimally align in time the test and reference patterns and to give average distance associated with the optimal warping path. An L_1 norm between two feature vectors is used for local distance computation. (For a more detailed description of the recognition experiment, see [5].)

The recognition system uses the first ten repetitions of the spoken digits for training and the remaining thirty-nine repetitions for testing. The reference patterns are created for each digit in the vocabulary from the data in the training set by averaging (after dynamic time warping) the patterns of ten repetitions of the same digit. The recognition system is tried out on the data in the test set and performance of the recognition system is evaluated for different values of r, the window length. Sakoe and Chiba's DTW algorithm and the modified DTW algorithm are used with slope constraint conditions of $P = 0$ (i.e., no slope constraint) and $P = 1$. The recognition accuracy as a function of window length is shown in Fig. 2(a) for Sakoe and Chiba's DTW algorithm and in Fig. 2(b) for the modified DTW algorithm. The solid lines correspond here to $P = 0$ and the dashed lines correspond to $P = 1$.

Based on the recognition scores shown in Fig. 2, the following observations can be made:

(1) The recognition accuracy increases at first with r, the window length, attains a maximum and decreases thereafter. The value of window length that gives the highest recognition accuracy is 9 frames (i.e., 90 ms) for Sakoe and Chiba's DTW algorithm and 3 frames (i.e., 30 ms) for the modified DTW algorithm. Since the amount of computation time required by the DTW algorithm to find an optimal warping path is proportional to $(2r + 1)$, the modified DTW algorithm requires
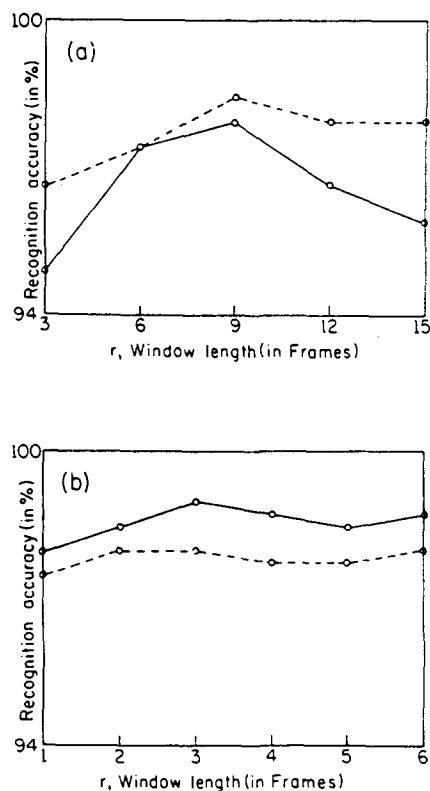

Fig. 2. Recognition accuracy as a function of window length for (a) Sakoe and Chiba's DTW algorithm and (b) the modified DTW algorithm. The solid lines correspond to $P = 0$ and dashed lines to $P = 1$.

computation time which is less than half of the computation time required by Sakoe and Chiba's DTW algorithm.

(2) For Sakoe and Chiba's DTW algorithm, the recognition accuracy obtained with the slope constraint $P = 1$ is more than that obtained without any slope constraint (i.e., $P = 0$), thus indicating the necessity of introducing a slope constraint in this case. Sakoe and Chiba [1] also obtained similar results. These results can be explained as follows. Sakoe and Chiba's DTW algorithm, requiring relatively large window length, provides too much of the warping function flexibility. This leads to poor discrimination between different digits. So for improving the recognition results, it becomes necessary to introduce a slope constraint to limit the warping function flexibility.

(3) For the modified DTW algorithm, the recognition accuracy obtained with slope constraint $P = 1$ is less than that obtained without the slope constraint (i.e., $P = 0$). This can be explained as follows. The modified DTW algorithm requires relatively small window length and thus limits itself the warping function flexibility. So, it is not necessary to have an additional slope constraint for limiting the warping function flexibility.

(4) The modified DTW algorithm gives the optimal recognition accuracy of 99.0% with $P = 0$ and $r = 3$ frames, while Sakoe and Chiba's DTW algorithm gives the optimal recognition accuracy of 98.5% with $P = 1$ and $r = 9$ frames. Thus, the modified DTW algorithm gives better recognition accuracy than that given by Sakoe and Chiba's DTW algorithm.

The recognition experiment is repeated with reference patterns generated as the miniav centres of the clusters formed from the multiple repetitions of each digit in the training set. (The miniav centre of the cluster is obtained as the pattern whose average distance to all other patterns in the cluster is minimum). The results obtained with these reference patterns are similar to those described above for the average reference patterns. For example, the recognition accuracies with these reference patterns by using Sakoe and Chiba's DTW algorithm are 95.6% with $P = 0$ and $r = 9$ frames and 96.9% with $P = 1$ and $r = 9$ frames. Using the modified DTW algorithm, these recognition accuracies are 98.7% with $P = 0$ and $r = 3$ frames and 97.2% with $P = 1$ and $r = 3$ frames. Thus, we see here also that the modified DTW algorithm gives better results without any slope constraint, while Sakoe and Chiba's DTW algorithm works better with the slope constraint. The window length used by the modified DTW algorithm is less than that used by Sakoe and Chiba's DTW algorithm. Thus, the modified DTW algorithm is computationally more efficient than Sakoe and Chiba's DTW algorithm. Also, the recognition accuracy obtained by the modified DTW algorithm is more than that obtained by Sakoe and Chiba's DTW algorithm.

It should be noted here that these recognition results are based on limited data used for experimentation. However, these results demonstrate clearly the advantages of the modified DTW algorithm over Sakoe and Chiba's DTW algorithm. Tsuruta et al. [6] have used Sakoe and Chiba's DTW algorithm for real-time implementation of the word recognition system DP-100. Since the modified DTW algorithm is computationally more efficient than Sakoe and Chiba's DTW algorithm, it will be more suitable for real-time implementation of an isolated word recognition system.

4. Conclusion

A modification over Sakoe and Chiba's DTW algorithm is proposed. The modified algorithm is applied to a speaker dependent isolated word recogniton system. It is shown that the modified DTW algorithm works better without imposing any slope constraint on the warping function. Also, the modified algorithm requires less computation time and gives better recognition accuracy.

Acknowledgement

References

[1] H. Sakoe and S. Chiba, "Dynamic programming optimization for spoken word recognition", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. ASSP-26, No. 1, Feb. 1978, pp. 43–49.
[2] V.M. Velichko and N.G. Zagoruyko, "Automatic recognition of 200 words", *Int. J. Man-Machine Studies*, Vol. 2, June 1970, pp. 223–234.
[3] G. White and R. Neely, "Speech recognition experiments with linear prediction, bandpass filtering, and dynamic programming", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. ASSP-24, No. 2, April 1976, pp. 183–188.
[4] F. Itakura, "Minimum prediction residual principle applied to speech recognition", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. ASSP-23, No. 1, Feb. 1975, pp. 67–72.

[5] K.K. Paliwal, A. Agarwal and S.S. Sinha, "Automatic recognition of spoken (Hindi) digits using DP time warping", *Indian J. Technology*, Vol. 20, No. 1, Jan. 1982, pp. 10–14.

[6] S. Tsuruta, H. Sakoe and S. Chiba, "Dp-100 connected speech recognition system", paper presented at INTEL-COM 1979, Feb. 1979.