2011 International Conference on Information and Network Technology
IPCSIT vol.4 (2011) © (2011) IACSIT Press, Singapore

Offline Chinese Character Recognition Using Elastic Matching and Parallel Implementation

Tian Jipeng[1+], G.Hemantha Kumar[2], H.K. Chethan[3]

[1, 2, 3] DoS in Computer Science, University of Mysore, Mysore-570 006, India
[1]Email Address: tianjp1982@gmail.com

Abstract: Computer Information Processing Technology is a very important symbol of today's world. The development of Computer makes human society changing a lot. Chinese language is the most used language in the world. In our experiment we employ Elastic matching algorithm which is a special model of matching. It does not need that the input image is perfectly same as image dataset. Elastic matching use intelligent algorithms to "flexible" deform the input image and sample image. System will first calculate the size, distance and length of input image and sample image, and then match. Parallel model has two kinds of model, which are MIMD and SIMD. In this paper, the parallel programming environment is PVM (Parallel Virtual Machine) which based on message passing. The proposed method is to implement high recognition rate and speed of handwritten Chinese numerals and handwritten Chinese characters Experiment result shows that our proposed approach efficiently and effectively improved recognition speed.

Keywords: Handwritten Chinese Character, High Recognition Speed, Parallel Implementation, Elastic Matching

1. Introduction

In Chinese character recognition, the number of categories is much larger than in alpha numeric recognition.[1] Offline handwritten character recognition is hot issue image pattern recognition. Off-line handwritten input technology is the high-tech areas of computer applications, but off-line handwritten recognition system has not yet been successfully developed. Most of the traditional handwritten character recognition using character strokes and stroke order of the split, which are not offline character recognition methods. Because the same character often change and become a different character, which because of the different strokes and strokes the length. These problems have been extensively studied by a lot of researchers, and very high recognition rates have been reported [1-15]. A survey of various feature extraction methods for character recognition is presented, such as stroke feature method and the feature point method, etc. [1, 8, 9]. A structural-analysis method is known as a powerful approach to the recognition of hand-written Chinese characters [2]. In the method, a character is represented and recognized by a set of structural features (e.g., line segments, strokes). Generally, the data computation was huge, and execute time was unacceptable. Because handwritten Chinese character is not standardized, personalized and structural complicated, therefore feature extraction and character segmentation is more difficult than others character recognition. In this paper, one new algorithm is performed using Elastic matching algorithm and Parallel environment executing. Computer hardware is getting cheaper and higher performance. Using sets of computers to execute recognition experiments is a new tempt.

+ Corresponding author. Tel.: + 91-9916400257 fax: +
E-mail address: tianjp1982@gmail.com

2. Offline Handwritten Chinese Character recognition Parallel Implementation

In text recognition area, in order to evaluate the performance of a recognition system, recognition rate, error rate, rejection rate and credibility of recognition result all will be considered. Recognition rate is the ratio of correctly identified characters number and the total numbers. Error rate is the ratio of wrong identified characters number and the total numbers. Rejection rate is the ratio of unable to identified characters' number and the total numbers. The credibility of recognition result is represented as Bc.

$$Bc = \left(Nc \big/ Mc\right) \times 100\% \qquad (1)$$

2.1 Elastic Matching Algorithm Design

Presently, there are many binary images matching algorithm. For example, image feature information extraction, neural network, etc., these methods are stricter to image quality. Elastic matching algorithm is a special model of matching. It does not need that the input image is perfectly same as image dataset. Elastic matching use intelligent algorithms to "flexible" deform the input image and sample image. System will firstly calculate the size, distance and length of input image and sample image, and then match. Past recognition methods are that input image and sample image is a one to one relationship. It reminds high quality of handwritten characters, and the recognition rate is not high. Base on this question, I consider each Chinese character as one two-dimensional image. Then the problem of character recognition becomes a 2D image matching. Here we do not use the structural features of Chinese character, but an approximate error matching algorithm of elastic matching methods. The method is flexible deforming the input image and sample image, then calculates the distance and size of these two images, and then decides which image from the library corresponds to the input image. The pattern of each handwritten Chinese character is an MxN (20x16) 2D image, and store as a set of MxN. One element in the set corresponds to one image element, to record its gray level as 1 or 0. In this way, two-dimensional image can be represented as a matrix with value of M and N. Then calculate out the distance between input string and sample string. If the value which we have got less than the value we pre-set, that stands for the sample of this character was found.

2.2 Parallel Recognition Program Design and the Parallel Architecture

Parallel model has two kinds of model, which are MIMD and SIMD. MIMD (Multiple Instruction stream, Multiple Data stream) means different processors may be executing different instructions on different pieces of data at same time. It has two different classes shown in Fig. 1, SPMD (Single Program, Multiple Data) which means Tasks are split up and run simultaneously on multiple processors with different input in order to obtain results faster, and MPMD(Multiple Processor, Multiple Data) which means sets of processors simultaneously execute different instruction sequences with different sets of data. Programs of SPMD have same frame and same program code, but different data stream. Programs of MPMD have different frame and different program code. Presently, there are several kinds of parallel systems, such as SMP (Symmetric Multi-processor System), MPP (Massively Parallel Processor System), and Cluster (Computer Cluster). Cluster, SMP, MPP and Distribute systems are four overlapping parallel architecture concepts.

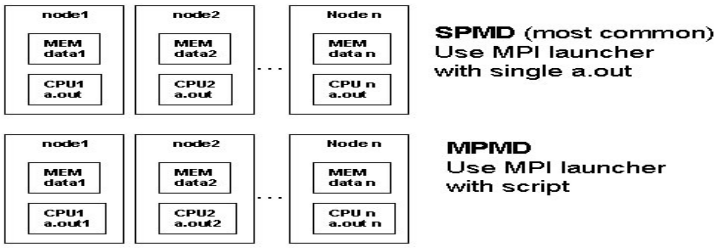

Fig. 1: SPMD and MPMD

2.3 Parallel Programming Model and Environment

Parallel programming is defined as a given algorithm to construct a parallel program. Parallel programming model has implicit mode and explicit mode. In this paper, the parallel programming environment is PVM (Parallel Virtual Machine) which bases on message passing. PVM includes many tools.

It has six specialties.

- Strong versatility, suitable for TCP/IP network environments, but also suitable for MPP.
- Program source code is a few MB only, size of the system is small
- Almost all parallel computer vendors support PVM, and wide range of applications
- PVM is open source software, copy right and modify are authorized
- High maturity, there are development and debugging packages available
- Number of standard digital software has been ported into PVM

3. Algorithm of Parallel Implementation

The basic idea of this algorithm is dividing the sample database into several sub-databases. If there are number of Pn processors, then divides sample data library into number of Pn sub- database to each processor. When a Chinese character recognition task was given to the system, firstly it will recognize in its own system by using own sub-database, then finalize the smallest elastic distance from those Pn values.

3.1 In the Parallel Programming Environment, the Signification of Parallel Computing Implementation is:

- Pn stands for number of processor, Pn = H, t stands process ID
- Pattern stands Chinese character image, which is waiting for recognition (a two-dimensional array)
- ElasiticDist is an array, which records the distance between Pattern and each sample data.
- AvergElasticDist stands the average distance between Patten and sample data.
- MinDist stands the minimum distance between Patten and sample data.
- GroupNumber stands the character from original sample data, which was recognized from Pattern.

3.2 The Parallel Algorithm Process:

(1) Spread Pattern to each sub-schedule

(2) Select H samples data from the image library current group t, sent to first process, and make current group t plus 1

(3) Repeat step 2 Pn times, send to 2nd, 3rd… Pn processor.

(4) Waiting to receive Pn return values

(5) Select the minimum value from Pn, if it is less than MinDist, evaluate it to MinDist, and GroupNumber equals to its correspond group number

(6) If the testing of sample data in the database have been completed, then the entire program is over, return the value of GroupNumber. Otherwise, repeat the procedure (2).

3.3 Sub-Schedule Process:

(1) Receive the sample character data Pattern from main process

(2) Accept a set of samples and extract character data

(3) Calculate the elastic distance between sample character data and each character data from this group, get the average value

(4) Return the average value

The algorithm is a master-slave mode (See in Fig. 2). The main process response sending sample character data to each sub-schedule process, and collects each return value, then process the return value to get the recognition result. Finally identify the character from the sample data library.

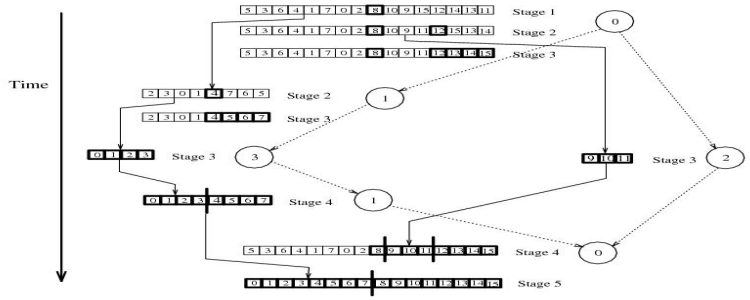

Fig. 2: PVM 4 Node working float chart

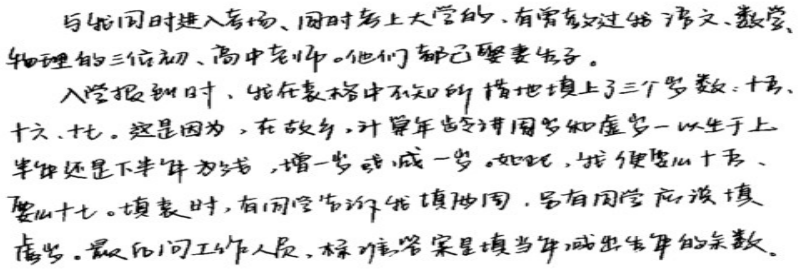

Fig. 3: Random selected character dataset

4. Experimental Result Analysis

Experiments were performed in two different parallel environments. One is Cluster which is composed by a number of PC. Each PC was installed Windows NT system. The other is Dawn-1000 computer system. The experiment result is shown in TABLE 1.

	1 Node machine	2 Node machine	4 Node machine	8 Node machine
Main Process running time	2152	1074	626	432
Main Process communication time	2152	1074	624	430
Sub-procedure average running time	2134	917	436.5	228.7
Sub-procedure average communication time	18	157	186	169.3
Accelerate rate	0.74	1.48	2.53	3.69
The cost of Parallel computing	2152	2148	2502	4512
efficiency	0.74	0.74	0.63	0.36

TABLE 1

Experiment results show that the communication time does not significantly decrease along with the node machine increasing, and sub-schedule communication time increased. After analysis, the parallel algorithm is not perfect, which caused this result. The synchronization requirements are too high. Main process will send the second data group, only after has received all the return values from sub-schedules. This makes some earlier finished sub-schedule waste time on waiting for next round data computing.

5. Conclusion

A Chinese character recognition system with elastic matching and parallel implementation is described in this paper. Two experimental environments were conducted, and two datasets were tested with this method. Extensive experiments are conducted to process the efficiency of the proposed method. This method greatly improves the efficiency of notes and forms data analysis, and brings a new information evolution

method. Furthermore, we have to think how to reduce the waiting time of sub-schedule is the problem that we face.

6. Reference

[1] Xiaofan Lin, Adaptive confidence transform based classifier conbination for Chinese charactrer recongtion, *Pattern Recongtion* Letter *19*(1998)975-988

[2] T.H. Hilderbrand, W. Liu, Optical recognition of Chinese characters:advancessince1980,*PatternRecognition26(2)*

[3] Guo Baolan, Zhang Cailu, Summarize of Optical Recognition Technology Development, *Computer World Journal ,* 1992, (10)

[4] Liu Qingxiang, Xiong Jie, Research and Application on Handwritten Chinese Character Recognition System, *High Education University Journal,* 2006.8

[5] Monroe R T.Kompanek A. Melton P.Gartan D.Arechitectural Styles,Design *Patterns,and Objects.IEEE Software*,1997-01

[6] Cheng LinLiu, In Jung Kim, jin H Kim, Model-based stroke extraction and matching for handwriting Chinese character recognition. *Pattern Recognition, 2001,* (34) 2339-2352

[7] Y Mizukami, a handwritten Chinese character recognition system using hierarchical displacement extraction based on directional features. *Pattern Recognition Letters, 1998*, (19): 595-604

[8] Liu Changpin, Chinese character recognition technology status and prospect, *Chinese Info Assoc Beijing, 2001*, 108-110

[9] Fu Qiang, Handwritten Chinese character recognition using MQDF classfier, *J Tsinghua Univ (Sci & Tech), 2008*, Vol 48, No. 10, 1609-1612

[10] R.S. Mitra, Elastic, maximal matching, *Pattern Recognition, volume 24, Issue 8, 1991*, Pages 747-753

[11] Seiichi Uchida, Eigen-deformations for elastic maching based handwritten character recognition, *Pattern Recognition, Volume 36, Issue 9, September 2003,* Pages 2031-2040

[12] C.H. Leung, Recognition of handwritten Chinese character by elstaic matching, *Image and Vision Computing, volume* 16*, issue 14, December 1998,* Page 979-988

[13] Mehran Moshfeghi, Elastic matching of mutimodality medical images, CVGIP: Graphical models and image processing, Volume 53, issue 3, May 1991, Pages 271-282

[14] G.W. Stewart, A parallel implementation of the QR-algorithm, *Parallel Computing, Volume 5, issues 1-2, july 1987*, Pages 187-196

[15] Fabrizio Pagano, Parallel implementation of associative memories for image classification, *Parallel Computing, Volume 19, issue 6, june 1993,* Pages 667-684