Adaptive Recognition System of the *Il-Pa* Stenographic Character Images by Using Line Scan Method and BEP

Sangkeun Kim, Sungoh Lee and Gwitae Park

Department of Electrical Engineering , Korea University Sungbuk Ku , Seoul, Korea
(Tel : 82-2-910-4639; Fax : 82-2-910-4609 ; E-mail: sangkeun@korea.ac.kr)

Abstract

In this paper, we would study the applicability of neural networks to the recognition process of Korean stenographic character image, applying the classification function, which is the greatest merit of those of neural networks applied to the various parts so far, to the stenographic character recognition, relatively simple classification work.

Korean stenographic recognition algorithms, which recognize the characters by using some methods, have a quantitative problem that despite the simplicity of the structure, a lot of basic characters are impossible to classify into a type. They also have qualitative one that it is not easy to classify characters for the delicacy of the character forms. Even though this is the result of experiment under the limited environment of the basic characters, this shows the possibility that the stenographic characters can be recognized effectively by neural network system. In this system, we got 90.86% recognition rate as an average.

1. Introduction

As we are living in the information age in which much information is deluged, we need to develop the ability to process much information quickly. Here, the media which are able to record, keep, and process every information are essential to the information age, however media held in common by both human and machine are not numerous so much. So operative work input by keyboard may be an obstacle to promote informationalization, but it must be essential work. In selecting the way held in common by both human and machine, if it is obviously impossible to make human read disks or tapes which are exclusive for machine, it will be far easy and appropriate to make machine read printing media exclusive for human. In this meaning, it is very important task to make a computer recognize character automatically.

Also neural networks, analyzing the structure of human brain as an human intellect scientific approach, making clear the mechanism of that process and making a computer which has a structure like it is appeared as calculation paradigm and researched for over the last 40 years to imitate the human ability image recognition parts[1]. Neural network has powerful classification function and recognition ability in addition that it has the simplicity in structure, and it is expected to present a new way as a parallel data preprocessing system of "all-at-once" way in the part where the existing computer of *Von Neuman* pattern does not fulfill its function [2],[3],[4].

So recently the research carrying out character recognition by neural networks with powerful pattern recognition ability and fault tolerance is in progress actively. Human is able to recognize character easily but if performing this by using the present computer system, it will be very complicated exorbitantly, therefore the research area of character recognition like this is a typical example to make good use of a merit of neural network. The good results of research by neural networks with an object to English characters, figures, Japanese *KANA* characters were not only published abroad but also in the practicalization state and it was reported to achieve high recognized rate over 98% by neural networks in printed *Hangul* Korean character recognition in Korea[1].

2. The Purpose of Research and Expected Effect

Recently the research to the development in the device of character recognition to process many documents automatically according to the information socialization is in progress actively. Also the neural networks which have a powerful function of pattern classification as a model for an artificial realization of human brain, overcoming the limit in the structure of the present computer are in the limelight[5].

Clustering or pattern recognition is defined as classifying data sets into each class according to the degree of similarity. One of the pattern recognition is the character recognition, whose methods are a template matching method, a statistical method and a structural method[6]. However a template matching method has a problem in selecting a special template for the variety of character form and a statistical method in the recognition of characters whose space between consonants and vowels is delicate, and a structural method in forming structural rule of character recognition. Recently the way by using neural networks model in character recognition is used a lot as a way of solving these problems[1].

Among some researches by neural networks, various methods in the *Hangul* Korean character recognition are developed as mentioned above, and it is almost in the practicalization state, but it is necessary to develop many algorithms for improvement of recognized rate in case of hand written style recognition. Also there are some researches in progress in the application part for the necessity, which are car number recognition, zip code recognition and etc., but there is not any research in the stenographic character recognition except some cases[7],[8].

In this paper we would study the applicability of neural networks to the recognition process of stenographic character image, applying the classification function, which is the greatest merit of those of neural networks applied to the various parts so far, to the stenographic character recognition, relatively simple classification work.

Korean stenographic recognition algorithms, which recognize the characters by using some methods, have a quantitative problem that despite the simplicity of the structure a lot of basic characters are impossible to classify into a type. They also have qualitative one that it is not easy to classify characters for the delicacy of the character forms. In this paper, it performs a recognition of basic 126 characters and after preprocessing to the stenographic character input first, it performs learning, extracting 104 DC component and inputting them to the neural networks[9]. The character learning