

Real-Time Constraint-Free Character Recognition System

Fathallah Nouboud and Réjean Plamondon, Département de Génie Electrique
Ecole Polytechnique de Montréal, C.P. 6079, Succursale "A",
Montréal QC, H3C 3A7

Abstract

In this paper we describe an on-line character recognition system operating in real time (0.07 seconds/character) and with no constraints on the handprinting. The preprocessing step makes it possible to be free of any constraints on either the size of the characters or the speed of the writing. The comparison is performed by a string-comparison processor. Following completion of a learning step, the system can recognize any set of characters defined by the user. The experimental tests show that the system is effective in user-dependent applications. The recognition rate is 96%.

Keywords: On-line, handwriting, character recognition, chain code, dedicated processor, learning, user-dependent.

1. Introduction

A number of problems must be dealt with when designing a character recognition system. Among them, there is the problem of variability. The same character may be written in many different ways by different users or even by the same user, and it is almost impossible to predict all the possible variations of a given character. Many character recognition systems therefore impose constraints on the writing, and these constraints may make users uncomfortable with the process. Other problems are, for example, the slow speed of processing, the use of boxes to separate characters, etc. [1,2,3,8,10].

In the system described here, we use a personal data base for each user, thus the writer does not have to respect any constraint and may use the system with his own natural handprinting style. The use of a specialized

processor for the comparison makes it possible to operate in real time and with a high degree of accuracy.

2. Preprocessing

Characters are digitized with the help of a graphic tablet (PENPAD300, PENCEPT Inc.). The active surface of the tablet is 11" x 11" (27.94 cm x 27.94 cm), the resolution is 1000 points per inch (394 points/cm) and the sampling frequency is 100 points per second.

Segmentation between characters is based on temporal information. The character is considered to be completed when a given time (about 300 ms) has elapsed since the last pen contact with the tablet surface.

A number of preprocessing tasks must be applied to the raw digitized data. These are designed to reduce the amount of information, to eliminate certain imperfections of the trace and to make the process independent of the size of the characters and the speed of the writing. Four preprocessing steps are used in this system.

2.1 Smoothing

The smoothing operation eliminates imperfections caused by the tablet, trembles in the writing, etc. A mobile average filter replaces a point with the average over its neighbours.

The transformation is performed according to the following equation:

$$X_i = (X_{i-3} + 3X_{i-2} + 6X_{i-1} + 7X_i + 6X_{i+1} + 3X_{i+2} + X_{i+3})/27 \quad (1)$$

This equation has been obtained after the convolution:

