International Journal of Innovative
Computing, Information and Control
Volume **5**, Number **11(B)**, November **2009**

ICIC International ©2009 ISSN 1349-4198
pp. **4033–4040**

A STATISTICAL APPROACH FOR HANDWRITTEN CHARACTER RECOGNITION USING BAYESIAN FILTER

Nozomu Araki, Yasuo Konishi and Hiroyuki Ishigaki

Graduate School of Engineering
University of Hyogo
2167 Shosha, Himeji-shi, Hyogo, 671-2201, Japan
araki@eng.u-hyogo.ac.jp

Received September 2008; revised February 2009

Abstract. *This paper proposes a new handwritten character recognition method using a statistical technique based on Bayes' theorem. This technique is the so-called Bayesian filter. The Bayesian filter was originally intended for spam mail recognition, and is known to be very effective in spite its simplicity. The objective of this study is to verify the applicability of this approach to handwritten character recognition.*
Keywords: Bayesian filter, Handwritten character recognition, Image processing

1. **Introduction.** There has been a lot of research on handwritten character recognition in recent years, resulting in a number of proposed pattern recognition techniques. One such method uses Mahalanobis generalized distance of a feature vector calculated from a character image [1, 2]. Techniques using a machine-learning approach such as a neural network (NN) and a support vector machine (SVM) are also well known [3, 4, 5]. However, there are problems with these algorithms. One is that a large number of learning data are required to achieve a high recognition rate, and they are very complex.

However, a new pattern recognition approach using statistical processing based on Bayes' theorem was proposed by Graham for distinguishing spam (junk) mail [6, 7]. This approach comprises the following steps. First, we calculate the appearance probability of words included in spam mails. Then, we calculate the probability that an arbitrary mail is spam using Bayes' combining probability. This approach is the so-called Bayesian filter. The Bayesian filter's algorithm is very simple because it only determines the appearance count of a word. Despite this, it has a very high recognition performance. Furthermore, it is possible to respond to a change in the pattern of a spam mail by learning words sequentially. For these reasons, the Bayesian filter has attracted attention in recent years.

Based on this background, this paper proposes a new handwritten character recognition method using the Bayesian filter algorithm. Here, we use a handwritten character image that is read by a scanner. The method consists of three steps: preprocessing, learning and recognition step. In the preprocessing step, there are many effective approaches to extracting image features [8, 9]. However, our objective was to verify the applicability of the Bayesian filter algorithm to handwritten character recognition. Thus, we simply carried out threshold processing for the scanner image and normalized the image size. Then, to obtain the appearance probability of a black pixel in each pixel of a certain character image, we determined the appearance count of a black pixel of learning data in the learning step. Finally, we carried out the character recognition to calculate the probability that the given character image was the same as a learning character by Bayesian filter using the appearance probability. We then verified the effectiveness of the proposed method by an experiment that using two Japanese characters written by four examinees.