

Bank Cheque Data Mining: Integrated Cheque Recognition Technologies

Nikolai Gorski

20.1 Introduction

Recently, one can observe new trends in bank cheque processing. Earlier, the customers (who are banks and other financial institutions) were mainly interested in automation of reading the cheque amount. Nowadays, there are more and more demands on deeper analysis of cheque content. They originated from general political and economical situation in the world: banks try to prevent money laundering, reduce losses from fraud cheques or even detect potential terrorists among their clients.

When processing a cheque, a bank is interested to read automatically as much information from the document as possible. Besides the cheque amount, this can include the date of cheque issue, the beneficiary name, the payer's address and signature, code line(s), etc. This poses new tasks for developers of document analysis systems. They should be able to process and understand a cheque as an integral document with many loosely structured information fields, some of which are mandatory and others are optional.

Many papers on bank cheque processing have been published recently [2, 4, 6–15]. Most of them are devoted to amount recognition as the most actual task for cheque industry. In this paper we also touch it, but mainly concentrate attention on new-coming tasks and describe approaches to their solution by the example of A2iA CheckReader recognition system.

Section 20.2 outlines cheque processing task definitions and demands to their automation. Section 20.3 describes in detail recognition technologies for one of the most important cheque fields: payee name (including results post-processing with user-defined dictionaries). Location, extraction and understanding of this field on both cursive handwritten and machine-printed documents are discussed. Section 20.4 briefly presents recognition of other cheque fields, such as date, payee address and name, and a code

line. It also discusses experimental and exploitation results of the A2iA CheckReader system achieved on bank cheques originated from different countries.

20.2 Challenges of the Cheque Processing Industry

20.2.1 What to Read?

In many countries, the bank cheque is one of the most popular ways of payment. The financial institutions receive daily hundreds of millions of paper cheques and other payment documents, from which miscellaneous data should be read, converted into electronic form and send on for further analysis. Reading and keying these data to a great extent remains a manual job performed by human operators. Replacing them by automatic reading systems is the greatest challenge of cheque processing industry. An important question is what should be read from a bank cheque?

Traditionally, banks were interested in three pieces of information: amount to be paid, account to be debited with this amount and account to be credited.

In most countries, information of debit account is coded in a special field (so called *code line*) pre-printed on each cheque with magnetic ink – see Figure 20.1. In the USA, the use of code lines on cheques was introduced in 1956, being the first step towards automated cheque processing. Special MICR scanners can read magnetic code lines with high accuracy, so the information they contain (e.g. payer account number, issue bank, etc.) need not be manually keyed for a great majority of cheques.

The credit account number is normally known from the payee – a person or an organization who deposits the cheque at his bank. The payee is identified by his name indicated in the cheque (Figure 20.1), while his account number is frequently present in a separate document – deposit ticket – associated with the cheque. The credit account number also comes in electronic form keyed by a clerk of payee's bank at the moment of depositing.

Thus, the first information to be automatically read from the cheque is the amount to be paid. Normally, it exists in two forms: literal (legal amount) and numeric (courtesy amount), both being subject to read and cross-validate each other for higher reliability. These amounts can be hand-written or machine-printed, typical fraction of machine-printed variants being around 20% of the total number of items in the document flow.

Other important information fields of a cheque are the date and the payer's signature. In some countries (e.g. Canada, Ireland, Italy), a cheque becomes payable only from its date, which might be far in future from the current date. In these countries, reading the cheque date is a mandatory operation, which should be automated together with amount reading. As for the payer's signature, in practice, banks do verify it only on cheques