

## Arabic Cheque Processing System: Issues and Future Trends

M. Cheriet, Y. Al-Ohali, N.E. Ayat, and C.Y. Suen

### 10.1 Introduction

From the administrative point of view, cheque processing involves all tasks a bank officer may perform to process an incoming cheque for a client. This includes: accessing account numbers, verifying names and signatures on the cheque, verifying the date of the cheque, matching the legal amount with the courtesy amount and verifying the credit of the cheque writer. However, from the technical point of view, cheque processing could involve capturing the cheque image, separating the foreground of the cheque from its background, extracting fields of interest and recognizing each of them.

This work employs theories and methodologies from various fields ranging from Natural language processing, Optical Character Recognition to Banking.

The motivation of the work on Cheque processing is not less than the motivation of the entire research in artificial intelligence, which aims to program the computer to carry out tedious routine processes, freeing time and space for humans to perform tasks that require higher levels of intelligence. A major advantage of such study is that it can be easily adjusted to serve more than 20 different countries (all of them use Arabic as their first language). In addition, legal amounts are widely found in documents other than bank cheques (e.g. business sell/purchase forms). Therefore, this study will be applicable to a wide range of applications. Moreover, similar languages (e.g. Urdu, Farisi) which use the same alphabet can benefit from these studies.

The remaining sections provide a description of datasets available for researchers as well as a detailed description of one system that processes legal amounts and one system dedicated for processing of courtesy amounts.

## 10.2 Datasets

Due to strict banking rules to protect their customers, it is extremely difficult to gain access to real cheques. This led some researchers to perform their research within financial institutions [16]. Such datasets are owned by, controlled and limited to the respected financial institutions. Other researchers like [18], opted to build artificial databases. In [18], about 2600 English cheques, written by 800 writers with pre-set legal amounts, have been collected. Another set of 1900 French cheques from 600 different writers has been collected too. The legal amounts were set for the writers to reflect balanced word distribution.

In the Arabic cheque-processing domain, however, there are real-world datasets that can be used by scientific researchers worldwide. Thus, researchers can test/evaluate their theories and systems in a real world environment. Real data, however, has a number of disadvantages, one must note. First, researchers have no control on the number of samples from each class, as this would be determined by the actual distribution of the classes and by the sampling bias. Another point that some may regard is the exposure of natural carelessness by some people in the society when filling or handling cheques. With this idea in mind, we contributed with a new database that we built in collaboration with Al-Rajhi Bank in Riyadh, Saudi Arabia [2]. These datasets include: legal amount dataset (containing 2499 legal amounts), courtesy amount dataset (containing 2499 courtesy amounts and written in Hindi digits), Arabic sub-word dataset (containing 29,498 sub-words within the domain of legal amount) and Indian digit dataset (containing 15,175 digits). In addition, there is a dataset of complete (original) grey level cheques, which can be used for other research purposes (e.g. date processing). Each of the aforementioned datasets is divided into training and testing sets. As real data may include overlapping parts of consecutive words or digits, such cases are explicitly mentioned by the datasets. Figure 10.1 below shows a sample of our Arabic Cheque database.



Fig. 10.1. A sample of the Arabic cheque data.