Part D

Processing of Non-textual Information

Until now, we have concentrated on documents conveying information through text. But if we come back to the general definition of a document we gave at the beginning of ▸Part A (Introduction, Background, Fundamentals), it is evident that people can also use other kinds of representations to record facts and information for the purpose of communicating them in a legible way to other people. ▸Part D (Processing of Non-textual Information) therefore extends the scope we have had so far of processing and recognizing images of textual documents to other kinds of documents currently printed on paper, archived, scanned, and used in human communication.

Of course, we could have addressed the general problem of documents containing pictures, paintings, photographs, and so on. But although a picture is worth a thousand words, it is still of another class. Documents can of course contain pictures, and a lot of efforts are devoted to natural image analysis, to retrieving information from photographs, to retrieving the right picture from a large set – the well-known problem of content-based retrieval. But covering all these techniques, questions, problems, and achievements would probably need one or two thousand pages more and is actually better documented in other handbooks. We therefore will basically restrict ourselves to our view of a document as a man-made symbolic representation of facts, and in this part we will address the specific questions arising from the analysis of documents where the information is only or mainly conveyed by *graphical representations*.

Although basic techniques for document image processing or page segmentation are still of great use for graphics-rich documents, they lead to specific segmentation problems, the two most prominent being the separation between text and graphics, and the conversion of graphics from the raster representation provided by digital imaging to a vector representation more suitable for many analysis and recognition tasks. These specific techniques are presented by Josep Lladós and Marçal Rusiñol in ▸Chap. 15 (Graphics Recognition Techniques).

We have already studied in great detail the methods used for recognizing characters. In graphics-rich documents, the "alphabet" can be much larger, as hundreds of different *symbols* may be used. Salvatore Tabbone and Oriol Ramos Terrades

give us an overview, in ▸Chap. 16 (An Overview of Symbol Recognition), of the main differences between character recognition and graphical symbol recognition. In addition to the large number of possible symbols, some of them being very close to each other, specific problems arise from the varying size of the symbols and from the fact that they are often intimately connected to lines and other graphics, which leads to the famous dilemma that you need to segment to correctly recognize, but you need to recognize to correctly segment.

But it is not enough to master basic segmentation and symbol recognition methods. Graphical documents rely heavily on a *context*; an engineering drawing is meant to be understandable by engineers and technical workers within the field it belongs to; a map relays information about features, locations, positions of various items, and is only understandable in close connection to an explicit or an implicit legend. In ▸Chap. 17 (Analysis and Interpretation of Graphical Documents), Bart Lamiroy and Jean-Marc Ogier lead us through the numerous issues raised by the need to take this context into account in graphical document analysis systems.

The three next chapters go into the details of state-of-the-art solutions for three classes of problems which arise very often, where the graphical part of textual documents must absolutely be analyzed and understood for the recognition process to be successful as a whole.

Anastasios Kesidis and Dimosthenis Karatzas explore in ▸Chap. 18 (Logo and Trademark Recognition) the important question of *logo and trademark recognition*; typically, this is a graphics recognition problem which often comes up in the midst of plain textual document analysis, as a logo or a trademark can be present in many documents, and it is necessary to correctly recognize them.

Tables and forms are also a very common way to organize information in plain documents. Again, although they are made of a lot of text, they also have a graphical framework whose analysis is vital for correct recognition and understanding of the document. Bertrand Coüasnon and Aurélie Lemaitre address this issue in ▸Chap. 19 (Recognition of Tables and Forms).

Finally, in ▸Chap. 20 (Processing Mathematical Notation), Dorothea Blostein and Richard Zanibbi explore the intricate question of *mathematical notation*, which at first glance appears to be close to impossible, as math can be very complex. But luckily, they also obey a strict and well-mastered syntax, so that math recognition systems become a very good illustration of the power of using the context expressed as syntactical rules to decipher complex formulas.