Analysis and Interpretation of Graphical Documents

17

Bart Lamiroy and Jean-Marc Ogier

Contents

B. Lamiroy (✉)
Université de Lorraine – LORIA, Vandœuvre-lès-Nancy Cedex, France
e-mail: Bart.Lamiroy@loria.fr

J.-M. Ogier
Université de La Rochelle, La Rochelle Cédex 1, France
e-mail: Jean-Marc.Ogier@univ-lr.fr

Abstract

This chapter is dedicated to the analysis and the interpretation of graphical documents and, as such, builds upon many of the topics covered in other parts of this handbook. It will therefore not focus on any of the technical issues related to graphical documents, such as low-level filtering and binarization, primitive extraction and vectorization as developed in ▶Chaps. 4 (Imaging Techniques in Document Analysis Processes) and ▶15 (Graphics Recognition Techniques), or symbol recognition, for instance, as developed in ▶Chap. 16 (An Overview of Symbol Recognition). These tools are put in a broader framework and threaded together in complex pipelines to solve interpretation questions. This chapter provides an overview of how analysis strategies have contributed to constructing these pipelines, how specific domain knowledge is integrated in these analyses, and which interpretation contexts have been contributed to successful approaches.

Keywords

Graphical document interpretation • Heuristics • Image analysis • Interpretation pipeline • Interpretation process • Interpretation systems • Knowledge management • Ontology

Introduction

Graphical Document Analysis

Graphical documents are basically documents containing a significant (if not exclusive) part of line drawings. They are usually considered a separate class between full text documents and photo-realistic documents. Many applications or research areas, related to the interpretation of documents in general, usually consider it good practice to segment composite documents into at least text, line drawings, and photo-realistic subparts, to which more specialized treatments are then applied.

There exists a very rich literature on projects and systems related to graphical document interpretation. Some of them are inspired from concepts which were developed in broader image interpretation problems. From a conceptual point of view, the scientific objectives of these systems consist in trying to implement generic and adaptable strategies, based on knowledge modeling, adaptive interpretation scenarios, etc. Most of the systems are related to projects within organizations managing huge amounts of graphical documents, and for which these documents are operational or organizational assets, the value of which needs to be optimized. Among the organizations/research groups who have presented some realistic systems, it is noteworthy to mention those projects related to the management of network maps and data, such as telephone networks, electric power grids, or waste water networks, for instance. Since efficiently operating these networks raises many management questions and is a major financial issue, many companies have investigated ways of leveraging the use of their digital maps with automated tools that are capable of extracting relevant interpretations.

While there are many references in the state-of-the-art literature to various domains, applications, and categories of documents, one of the most successful approaches concerns the area of electronic diagram interpretation for aircraft, done by Baum et al. at Boeing in the late 1990s. Their goal was to scan paper versions of wiring diagrams and other engineering drawings in order to convert them into a digital and operational maintenance tool that would allow hyper-referenced access to assembly parts, logic diagrams, etc. The illustrations in Fig. 17.2 were reported in [4] and subsequent publications and give an idea of how it was able to cross-reference text-based part numbers and graphical entities.

The main challenge of interpretation systems is to implement flexible and adaptive strategies, being able to produce knowledge compatible with its interpretation domain, and being able to automatically detect and solve semantic ambiguities.

History: Evolution of the Problem

The history of automated document analysis and its related problems finds its origin and initial scope in the corporate business domain. In order to optimize information management in big public or private institutions, which was initially totally paper based, new solutions and document management applications were expected to emerge from focused research. In this context, during the last decades, mainly due to the huge increase and reduction of costs of storage capacities and to the continuous progress in image analysis techniques, there has been a considerable evolution in the various strategies of large institutions when dealing with document analysis problems.

Initially, because of the lack of relevant automatic interpretation systems, most of the organizations decided to only digitize their documents, in order to obtain a representation of data (and therefore, mistakenly, information) which could be easily stored, shared, duplicated, and transmitted and would therefore respond to elementary data management problems. This massive digitizing cannot be considered as part of document analysis per se, but it marks the beginning of an era where automatic document management will progressively attract more and more economic interest.

However, it became rapidly clear that considering digitization only was producing data with too poor a level of information, and most organizations decided it was necessary to improve the level of usability by trying to obtain higher-level interpretations from the digitized data by structuring the contained information through analysis processes. At first, regarding the fact that automatic processes were in their infancy, most of companies and organizations adopted full manual processes for reverse document interpretation. Many low-cost manual digitizing and interpretation projects were started in countries where human labor was cheap. With respect to the definitions and classifications mentioned in the previous section, this is an interesting benchmark, since it fully relies on human document analysis. It remains interesting to have a deeper look into the notions of context, interpretation, and analysis in this particular case. Indeed, the analysis is done by human collaborators who are not necessarily entirely knowledgeable of the

interpretation context, since they are externally hired to process the documents and provide the interpretation. This required the context to be documented and a quality process to be instituted in order to make sure that the produced interpretations were conforming to the expected context. However, facing prohibitive costs of the manual interpretation resources and considering quality problems related to human digitizing, many of them tried to implement interactive processes, coupling reliable image-processing tools and human correction interfaces. Then, during the 1990s full automatic interpretation systems were presented in the literature, based on complex approaches, integrating sophisticated strategies. The complexity of these approaches essentially came from the inability of the developers to fully capture the interpretation context, who therefore resorted to compensating this (often unconsciously) by embedding them into the algorithms themselves. Because of this, the produced analysis programs were often very focused on specific interpretation contexts, without offering any satisfactory hints to whether they could be adapted to other contexts without significant re-engineering. Considering that fully automatic systems that would also be generic and usable over a wide range of interpretation context were considered as quite Utopian by some authors, and given the fact that this seemed to be confirmed by the observation of the state of the art, a significant paradigm shift appeared suggesting to develop alternatives to full interpretation systems by only partially interpreting document contents and by using "spotting techniques" integrating "contextual information." Generally speaking, the *information spotting* problem can be defined as the location of a set of regions of interest from a document image, which are likely to contain an instance of a certain queried object without explicitly recognizing it. The most famous applications of this kind of concepts are word spotting on manuscript documents on the one hand and symbol spotting in the context of graphical documents on the other hand. One of the main applications for information spotting methods is its use in large collections of documents. In a sense, this is a very pragmatic answer to the previously mentioned inability to capture interpretation context. Spotting techniques implicitly admit that full interpretation seems to be out of reach and focuses either on very generic ("low-level") analyses or on very broad interpretation contexts, leaving it to a later stage process step to combine this partially complete information to achieve full interpretation.

This explains why more recent literature does much less focus on complete interpretation systems. Figure 17.1 illustrates these different alternatives, as well as corresponding milestones during the last five decades.

Structure of This Chapter

As will become gradually clear through the development of this chapter, there is no precise or generally adopted approach to graphical document analysis. Furthermore, there is a significant overlap with the more general Machine Perception domain, and it will sometimes be necessary to digress and illustrate some approaches that are appropriate in that area.
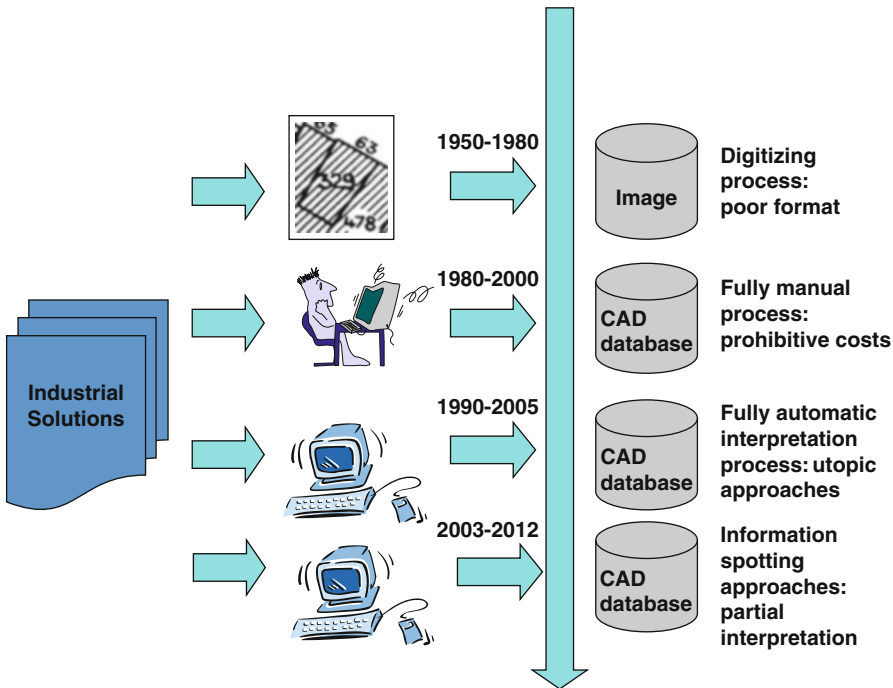
Fig. 17.1 Main periods of the graphic interpretation history

However, it is possible to identify broad categories of approaches to graphical document analysis. These are highlighted in section "Structure of This Chapter." Section "Examples of Analysis Approaches and Interpretation Knowledge Representations" will give a quick state-of-the-art overview of some of the most significant approaches for each category.

State of the Art and Classification of Graphical Document Analysis

Overview of Graphical Document Analysis Problems

The most generally adopted approach to graphical document analysis is related to the document reverse production process. This essentially means that the analysis aims at extracting information from a 2D representation. The principle of document production and the corresponding reverse interpretation is represented in Fig. 17.3.

As a consequence, a classical analysis process includes several stages, which try to reconstruct higher-level information from a 2D representation, on the basis of a progressive analysis, going from low-level information i.e., graphical primitives) to
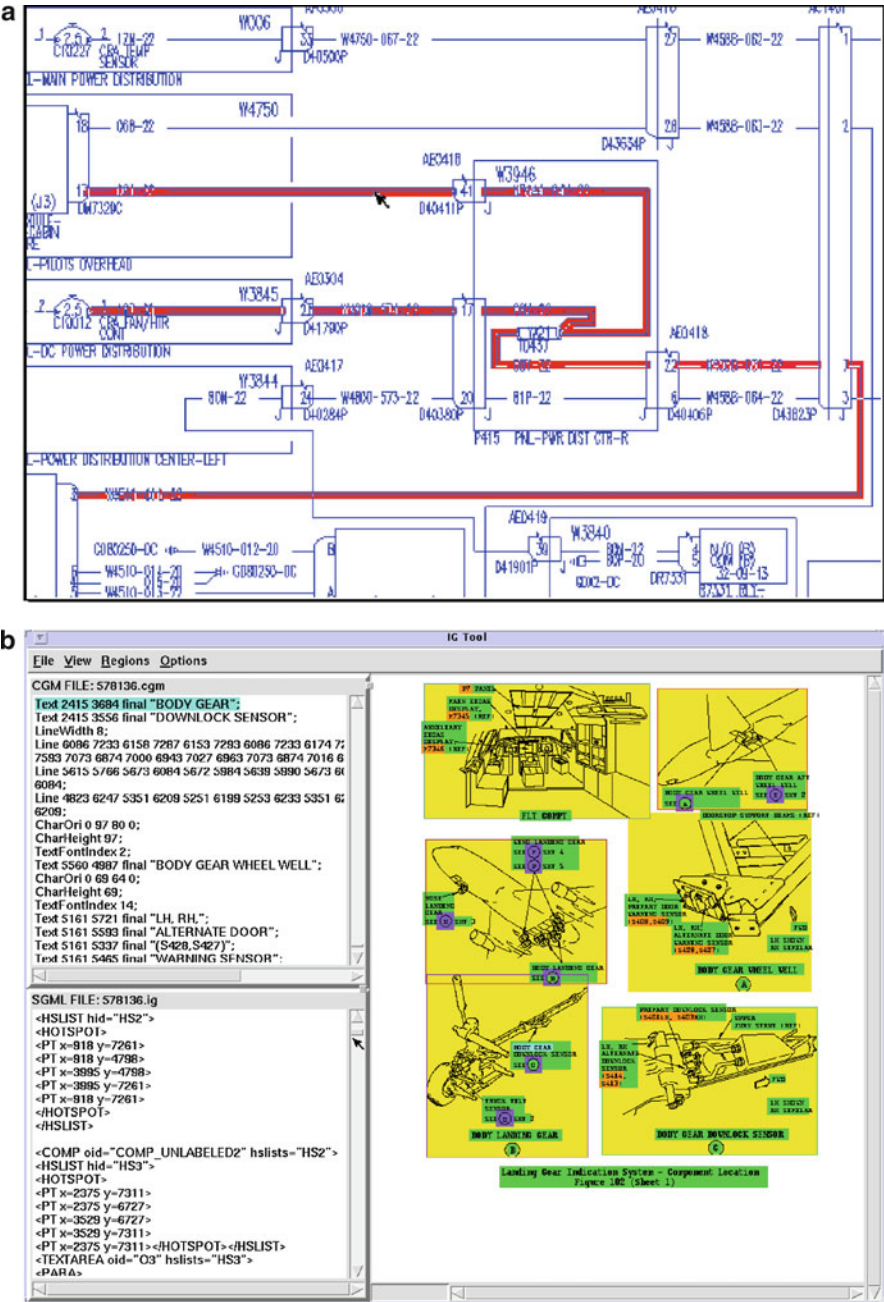
Fig. 17.2 Illustrations from [4] illustrating the results of graphical analysis tools linking textual and graphical data to aircraft maintenance semantic information
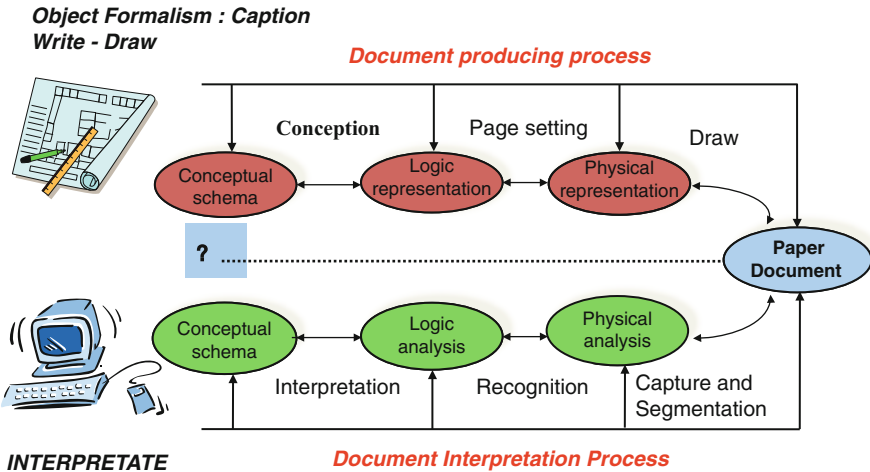
Object Formalism : Caption
Write - Draw

Document producing process

Conception — Page setting — Draw

Conceptual schema — Logic representation — Physical representation — Paper Document

?

Conceptual schema — Logic analysis — Physical analysis

Interpretation — Recognition — Capture and Segmentation

INTERPRETATE

Document Interpretation Process

Fig. 17.3 From the production of the document to its interpretation

a conceptual (semantic) level. Figure 17.3 gives an overview of this analysis process. The upper part of the figure (in red) describes the production process where an author within the context of her own conceptual schema will express her ideas using a logic (or otherwise formalized) representation language. This expression will then be transcribed and made physically available (as pixels, ink strokes, etc.).

Interpretation of the initial intentions of the author can only be achieved through the appropriate reverse analysis (in green, at the bottom of Fig. 17.3) provided that each process stage shares enough of the initial context to analyze the data. Consequently, the implementation of an analysis system requires the sequencing of many processes allowing to reach this goal and covering all the aspects required for reaching a semantic level of information, usually starting from a 2D digital image. The way these processes interact and the choice of which process to execute next, as well as the operational parameters they are fed, heavily depend on the interpretation context. Many of these processes include low-level image-processing techniques, primitive extractions, structural and statistical pattern recognition techniques, and semantic analysis. Each of these processes has already been described in great detail in the various chapters of this handbook.

The semantic analysis is principally driven by the knowledge related to the representation of symbolic objects on documents, which, in its turn, drives the different processes, tunes their parameters, stores the progressively extracted knowledge, and performs specific processing when semantic ambiguities are detected.

The main differences between the existing contributions related to analysis systems come from the way processes are organized and knowledge (and context) is (or is not) managed, from how semantic constraints are handled, and from how users are integrated in the analysis loop to achieve the required interpretation.

The state-of-the art literature contains a large variety of alternatives addressing these issues. Part of them is detailed in section "Examples of Analysis Approaches and Interpretation Knowledge Representations." The next section tries to establish a classification of the available approaches.

A special mention should be made for hand-sketched graphics recognition, some techniques of which are developed in ▶Chap. 28 (Sketching Interfaces) of this handbook. They generally apply to dynamic, on-line recognition contexts. A part from some very recent references such as [6, 23], this field of research only rarely incorporates complex interpretation goals. Most of the references of this research area deal with recognition issues and rarely consider interpretation strategies as those developed here.

Classification of Graphical Document Analysis Systems

In order to get a structured view of the existing approaches to graphical document analysis, four different axes of observation should be considered, keeping in mind that there are many possible strategies to address the general interpretation problem, and categorization is always a fairly arbitrary task:

1. The application domain: what kind of documents is concerned?
2. What specific visual representation context characterizes the documents and what features are appropriate to the problem? In some cases the set of features is fixed; in other cases, the authors propose training capabilities to the system, in order to give them the ability to learn from examples and adapt to different feature sets.
3. How is knowledge modeled? Knowledge modeling is often tightly related to how the visual representation context and features are chosen (it may, for instance, depend on the structural/statistical description of the objects). Some authors try to differentiate the knowledge representation formalism from the one used for the object recognition process as is the case for those using ontologies [33], semantic networks [2], or constraints networks [32]. Knowledge can also be distinguished in the way authors store it, according to the reasoning process used in the interpretation module.
4. How is the interpretation strategy implemented or guided? The interpretation strategy can be completely hardwired in the case of bottom-up approaches, but it can also be guided by a blackboard in the context of centralized reasoning processes [43]; it can be distributed in multi-agent systems [3, 30]. Furthermore, the reasoning process handling the interpretation strategy can either be implemented as predefined interpretation scenarios (which often rely on bottom-up strategies) or, on the other hand, rely on opportunistic approaches based on progressive assessment of the knowledge which is produced by the different modules of the system.

Table 17.1 gives an overview of how the previous criteria apply to the main systems documented in the literature over the last couple of decades. It has to be noted that besides giving an overview of how interpretation strategies have been implemented,

Table 17.1 Interpretation systems and application domains

Application domain	Features of handled documents	References	Features of the systems knowledge modeling	Interpretation strategy
Engineering drawings and mechanic charts	Black and white, containing lines, symbols, and texts. Existing captions	[17]	Blackboard, modeling of knowledge based on grammars	Cycle of perception, opportunistic strategy, perceptive cycles, Neisser perceptive principles
		[43]	Blackboard-based modeling	Planified bottom-up approach
		[32]	Declarative language-based knowledge modeling, blackboard	Bottom-up approach
		[14]	Object-based approach, hierarchical class decompositions	Hypothesis emission/validation principle, dynamic planification
		[10]	Object-based approach defining the expected properties of objects	Bottom-up approach for interpretation; decomposition of the interpretation in 5 specialized subsystems specialized for 3D object reconstruction
Telephone networks, electric wiring diagrams, gas distribution network plans	Black and white, containing lines, symbols, and texts. Existing captions	[3]	Blackboard, heuristic-based knowledge introduced in the source code	Multi-operator-based approach based on cooperation between low-level operators, emission/validation strategies
		[11]	Adaptive strategies based on results assessment, perceptive cycles, semantic network for knowledge modeling	Centralized control, opportunistic strategy
Electronic diagrams	Black and white, containing lines, symbols, and texts. Existing captions	[44]	Blackboard, heuristic knowledge introduced in the source code (no externalization)	Bottom-up and planified strategy

Table 17.1 (continued)

Application domain	Features of handled documents	References	Features of the systems knowledge modeling	Interpretation strategy
Geographical maps	Colored documents, containing textures, lines, symbols, and texts Existing captions	[40]	Structure-based approach for knowledge modeling	Rule-based strategy and process assessment strategy
		[38]	Distributed knowledge on agents and blackboard for centralized information	Distributed approach, agent-based approaches
		[37]	Blackboard approach, heuristic knowledge introduced in the source code	User interaction based interpretation strategy
		[5]	Blackboard, graph-based knowledge modeling	Rule-based approach
Cadastral maps	Black and white, containing lines, symbols, and texts.	[2,41]	Blackboard, semantic networks	Bottom-up approach, centralized strategy
	Existing captions-high precision of document	[16,29]	Blackboard, object model approach	Cyclic interpretation strategy, according to the semantic consistency of the interpreted data
		[33]	Ontology-based approach, centralized modeling system	Dynamic interpretation scenario, based on adaptive strategy Graph-based strategy
Architectural floor plans	Black and white, containing lines, symbols, and texts Existing captions-high precision of document	[15]	Centralized knowledge, structure-based descriptions, knowledge based on constraints networks	Progressive solution construction, on the basis of the checking of primitives properties in a constraints network
Hand-sketched graphics	Black and white, manuscript containing text, lines, and symbols	[6]	Heuristic knowledge, blackboard	Bottom-up and planified strategy

it is quite impossible to proceed to a more thorough objective analysis of these approaches and establish quality measures or possible rankings. This is due to the fact that, unlike what has occurred for more classical pattern recognition problems (symbol recognition, signature relevance analysis, etc.) or low-level processing evaluation (binarization, segmentation, etc.), the document analysis community has never organized interpretation evaluation contests. On the one hand, this lack of objective evaluation is due to the difficulty of implementing comparison of semantics on the basis on numerical scoring, and on the other hand, it is due to the fact that many research teams abandoned this exhaustive interpretation objective in favor of spotting information problems.

However, it is possible to provide some qualitative aspects guiding the reader in the implementation of an interpretation system. These qualitative aspects concern the facility of implementation as well as the ability of the system to incrementally integrate new knowledge and to automatically analyze the semantic consistency to the interpreted data in relation with the expectations of the user.

When dealing with an interpretation problem, the first point that must be considered in the implementation of a complete system is the necessity to simply formalize and externalize domain knowledge, in order to define the expectations of the user in terms of interpreted objects. This formalization must be based on user-friendly interfaces allowing the user to define which objects she is expecting and how they are graphically represented. From this point of view, the most relevant approaches are the one based on ontologies that permit to express this kind of knowledge in straightforward ways [33].

Considering the different manners to store the information, based on a centralized (mainly based on blackboard principle) vs. more distributed approaches (often multi-agent and multi-operator based), centralized knowledge-based systems seem to be much easier to implement, compared to distributed approaches, for which it is quite difficult to maintain overall consistency.

Concerning the interpretation strategy, while bottom-up and planified approaches were the most widely developed at the beginning of these research studies, cyclic approaches [17, 29] offer some very interesting advantages related to the possibility of using opportunistic interpretation strategies, often based on automated semantic consistency analysis.

Also, the interpretation process often relies on the use of pattern recognition tools, which can be based either on a statistical description of objects or on a structural description. From this point, even if the statistics-based approaches appear to be much more interesting from the algorithmic point of view, their poorness of description gives a large advantage to structural-based approaches [32].

It is also worthwhile to mention recent approaches, which consider user interactions allowing the system to integrate corrections provided by the user in the interpretation process and sometimes providing the possibility to infer on the knowledge of the system [20, 33].

Some of these systems try to implement generic and flexible interpretation strategies, most of the time on the basis of explicit knowledge modeling. However, they generally remain quite domain specific, due to the high number of heuristics

introduced in the processing chain. In this context, the commonly accepted notion of interpretation of graphical documents is to consider it as the result of an automated analysis process converting a poor format document (paper, pdf) into a format close to human interpretation. This, of course, is only partially satisfactory, since it defines computer interpretation in terms of human interpretation, without fully assessing what the latter actually entails. The rest of this chapter will develop in detail how these various approaches and applications have been constructed and how they consider "interpretation" of graphical data.

Relations Between Machine Perception and Image Analysis Systems

The interpretation problem is a widely spread question, especially in computer vision communities. In many cases, document interpretation strategies can be partially inspired from computer vision communities, in which many image interpretation systems were also developed. Indeed, in the last 50 years, a lot of image interpretation applications have been developed in many fields (medicine, geography, robotics, industrial vision, etc.). However, image-processing specialists design applications by trial errors cycles and there is no identifiable tendency to reuse already-developed solutions. The lack of application modeling and context formalization may be a reason for this behavior:

- Accounts of full analysis systems are rare. Usually, publications focus on specific parts of the analysis pipeline, highlighting the scientific and theoretical foundations of their contributions. Very often, these reports conclude by providing experimental validation on specific data, claiming an improvement over competing approaches on the same, or similar, data. This results in a very result-focused definition of interpretation problems and obfuscates, in some way, both the actual interpretation context, on the one hand, and a formal description on how the analysis process advocates between possible choices.
- The reusability of these applications is therefore very poor because the limits of the solution applicability are not explicit. Moreover, they often suffer from a lack of modularity and the parameters are also often tuned manually without giving explanations on the way they are set. Besides, these parameters and their impact on the final interpretations hold a tight relationship with the interpretation context, as already stated before. If the context cannot be formalized on the one hand and if the parameter domain cannot be mapped to the context, reusability and generality can only occur through trial and error tuning.

There exist some approaches that try to address these issues, however. Knowledge-based systems such as OCAPI, MVP, or BORG were developed to automatically construct image-processing applications and to make explicit the knowledge used to solve such applications.

However, most a priori knowledge of the application context (sensor effects, noise type, lighting conditions, etc.) and the interpretation goal to achieve were still more or less implicitly encoded in the knowledge base. This implicit knowledge

restricts the range of application domains for these systems, and it is one of the reasons of their failure.

More recent approaches bring more explicit modeling but they are all limited to the description of business objects for detection, segmentation, image retrieval, image annotation, or recognition purposes. But they do not completely tackle the problem of the application context description and the effect of this context on the images (environment, lighting, sensor, image format). Moreover, they do not define the means to describe the image content when objects are a priori unknown or unusable (e.g., in robotic, image retrieval, or restoration applications). They also assume that the objectives are well known (to detect, to extract, or to recognize an object with a restrictive set of constraints) and therefore they do not address their specification.

Examples of Analysis Approaches and Interpretation Knowledge Representations

It should be clear to the reader, by now, that from a technical point of view, there is no formalized and standard approach nor definition of graphical document analysis and interpretation. There are merely interesting and successful approaches that have proven efficient in specific application contexts. Taking a closer look to those, there are, however, some lessons to be learned from how they integrate various levels of knowledge and what strategies are deployed to make them as flexible as possible to adapt to other contexts. This section will try and provide an overview of these strategies.

Classical Strategies: Bottom-Up Approach

Image or document analysis is a difficult task since it requires a large amount of different data-processing techniques, from low-level treatments (e.g., noise filtering, data restoration) to high-level interpretation (e.g., object identification, decision making) through intermediary operators (e.g., segmentation). In order to solve this problem, most of the different strategies available in the literature are very much based upon the hierarchical decomposition of the problem shown in Fig. 17.4.

From the acquired data, treatments are most of the time run sequentially within a bottom-up strategy. Each operator of this decomposition provides a result, constituting the entry of the next operator. Following this approach, the most sensitive points are the choice of the optimal operators, the definition of the adequate sequential ordering of these treatments, the management of the quality (or uncertainty) of their results, and the communication between the different levels. Most of the time, this kind of conventional approach relies on three main levels (Fig. 17.4), each of which manages a particular level of information:
• The first level manages the extraction of low-level primitives: it often includes prepossessing techniques and extraction of primitives (lines, circles, textures,

Fig. 17.4 Illustration from
[27] depicting the main steps
classical bottom-up
interpretation approach
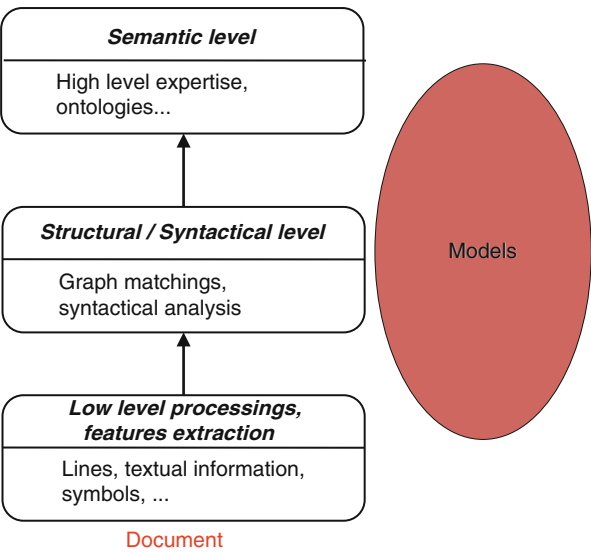
textual information, etc.). The techniques and tools related to this have been
described in ▶Chap. 15 (Graphics Recognition Techniques).

• The second level generally manages statistical, structural, or syntactic informa-
tion and tries to combine low-level primitives into syntactically, structurally, or
statistically described objects, on the basis of classification techniques, graph-
matching approaches, or syntactical methods. Most of the approaches related
to this level have been described in ▶Chap. 16 (An Overview of Symbol
Recognition).

• The third level generally tries to integrate semantic constraints, in order to solve
ambiguities. This level is usually the less formalized and forms the core focus of
this chapter.

Most often, graphical document analysis follows a bottom-up strategy. Algo-
rithms are performed in a fixed sequence, usually starting with "low-level" analysis
of the gray level or black and white image (sometimes combined with noise filtering
and binarization cf. Part B (Page Analysis) in this handbook), in which primitives
are extracted by specialized operators. Generally, these primitives correspond to
segments, associated or not to polygonization algorithms, to symbols and characters,
textures, circles or circular arcs, dashed lines, arrows, etc.

In the next phase, associations between all or a part of these primitives are
detected, and higher-level graphical entities are constructed, guided by some a
priori knowledge. This knowledge is either directly written into in the source code,
or it can be declarative knowledge based on explicit rules for graphical entities.
An analysis of graphical entities and their relationships allows one to propose an
interpretation, in the case of strictly bottom-up approaches such as [5,12,19,39,41].
The main difficulty in this kind of process comes from obtaining significant and

robust graphical entities from the low-level operators and reliable association rules between each primitive in order to achieve a correct interpretation. These issues have already been partially discussed in ▶Chap. 15 (Graphics Recognition Techniques). In fact, contradictory as it may seem, these systems all extract low-level primitives the same way, using the best state-of-the-art approaches as off-the-shelf tools, without necessarily taking into account the specificity of the visual representation of each graphical object within the context they are confronted with. As a consequence, due to the variability in representation and the manual fine-tuning of many of the intervening parameters as well as the handmade and fixed combination of supporting extraction and detection algorithms, many situations in technical documents are difficult to solve by these approaches. They usually concern the connection and overlapping between different visual entities (e.g., text, lines, and texture), text identification in handwritten annotations, isolated character recognition under multi-orientation constraints (for instance, in city maps or utility maps), low image quality, and variability in the representation of graphical entities. For all such strictly bottom-up systems, the main problem is related to the poor adaptation of the parameters of the extractors and to the inadequacy of operators to the local features of certain objects. As a side note, and with respect to text identification, it has to be stressed that OCR in graphical documents is a different challenge than character recognition addressed in ▶Part C (Text Recognition) and more particularly, but not exclusively, ▶Chap. 10 (Machine-Printed Character Recognition) in this handbook. Textual references are very often very short sequences for which no "text-only" context is available, as in full text environments (where dictionaries or other linguistic heuristics can guide in solving nondeterminism). Very often the interpretation context of textual annotations is related to the graphical context on the one hand and syntactical reference conventions or encoding on the other hand.

The most emblematic bottom-up approaches in the graphics literature are [19] and an updated version, applied to architectural drawings, adapted to the evolution of low-level treatment and higher-level recognition processes [15]. It is interesting to note that [19] considers "shapes" as the ultimate level of interpretation, regardless of what these shapes may represent. This means there is a complete lack of semantics in this approach. The goal of the approach is to have a geometric description (vector image) that would be as faithful as possible to the initial raster image but that would go beyond strokes and connected lines and incorporate coherent descriptions of shapes (circles, hexagons, parallelograms, etc.). Dosch [15] extends this low-level consideration to not only integrate symbol recognition but also add an interpretation step that is targeted toward their application context: architectural drawings.

Since it lacks a more semantic verification step, the former has interpretation artifacts like those shown in Fig. 17.5. While the method (almost) correctly separates geometric shapes from text, it identifies all 2D polygon shapes independently, failing to establish that they stem from a 3D projection and thus missing the vertex co-occurrence constraints on some boxes.

In [15] the bottom-up approach is extended to incorporate more elaborate shape recognition on the one hand, but also to relate them (and their visual context) to
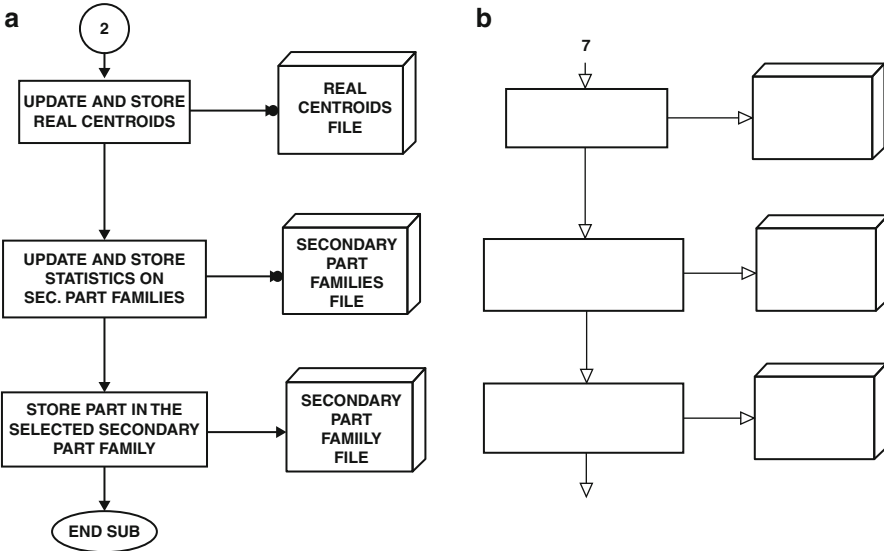
Fig. 17.5 Taken from [19] and showing the interpretation result. Note a minor segmentation error identifying the upper "2" as graphics and the vertices not overlapping for some of the 3D boxes

architectural representation knowledge, as to model 3D buildings from 2D scanned images, as shown in Fig. 17.6.

Recent approaches have tried to revisit this paradigm in the light of symbol spotting, without fundamentally changing the three levels described before. The main shift is operated at the low-level extraction where, instead of trying to extract features justified by human interpretation semantics (lines, text, textures), either non-discriminate small areas are extracted [36], at the standard image scale, scale-invariant interest points are extracted [25], or patches corresponding to regions of interest are used. These are generally based on pure signal processing techniques identifying the maximum of entropy of information-like zones [21]. They have been described in ▶ Chap. 16 (An Overview of Symbol Recognition), and although they have proven to be quite efficient in lab environments [9], they have never really been integrated in graphical document analysis contexts beyond symbol recognition. One of the current main obstacles to mainstream development of these approaches in broader analysis processes comes from the fact that there currently is no trivial approach to integrate signal-based patches with the higher levels handling syntactic and semantic constraints.

As a summary, one could say that the major drawbacks of the bottom-up approaches are due to the fact that the processes running at each of the cited levels do not sufficiently integrate contextual information, if any. At the lower level, image-processing techniques and features extraction are run globally on the whole image, without integrating local contextual knowledge. This highly conditions the quality of the extracted information. These results are transferred from one level to another

Fig. 17.6 Taken from [15] representing the results at all major bottom-up stages of the interpretation process: line segmentation, symbol recognition, context identification, 3D reconstruction

without many possibilities of coherence or quality verification. Another key problem of this kind of approach is related to the fact that many sources of knowledge are implicitly embedded in the interpretation process, and this knowledge usually is tightly linked to the data and to the targeted application. This drawback makes it difficult for this kind of approaches to be reused in other contexts. Furthermore, the lack of definition of a memorization strategy in order to apply the most adapted analysis sequence for a specific context, by using contextual information but also by using the history of the device (as would human analysis do), represents a limitation to generalization.

As a consequence the document analysis research community (as well as the broader image-processing community) has quite well identified this problem as related to the adaptability of the interpretation device. To try and address it, knowledge-based alternatives have been developed in the hope of achieving more versatile analysis processes.

An illustration of this transition is the work by Devaux et al. [13] which consists in transforming 2D ANSI representation of 3D objects (containing dimensioning lines, orthographic projections, etc.) into full 3D representations. Although their work can still be considered as very similar to the bottom-up approaches, it clearly
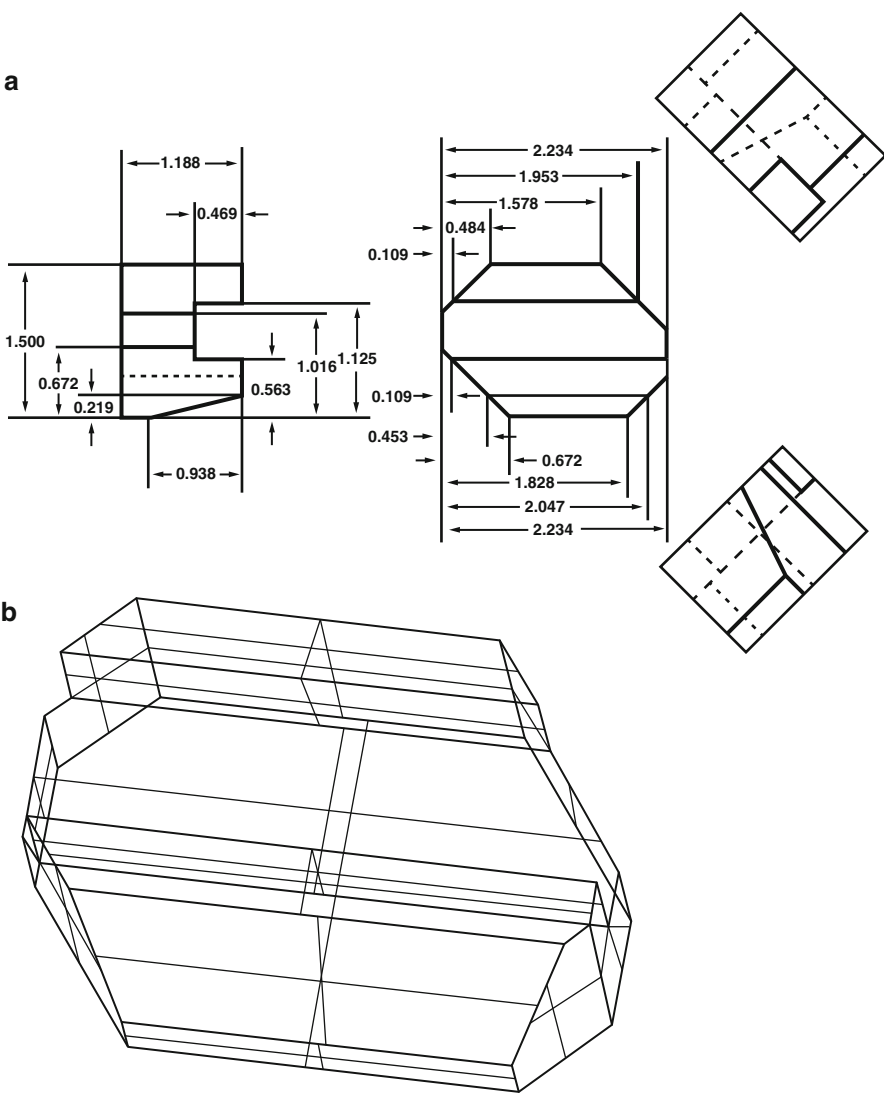
Fig. 17.7 From [13] obtaining a full 3D representation from a set of 2D ANSI orthographic projections

distinguishes itself from them by representing analysis knowledge as rules. An illustration from their work is shown in Fig. 17.7. One should consider [13] only from the perspective of graphical document analysis and the ways in which it extracts higher-level information from image pixel data. It does not intend to be a state-of-the-art reference to the problem of reconstructing 3D shapes from 2D projections. This problem has been addressed elsewhere and goes far beyond image analysis; it does not fall within the scope of this handbook.

Knowledge-Based Approaches

In order to solve the difficulties mentioned in the previous section, mainly that low-level segmentation and extraction methods can be used off-the-shelf, but that domain knowledge and analysis "intelligence" is embedded in the underlying algorithms, many references in the literature consider *knowledge-based* approaches. This kind of approaches generally tries to solve the adaptability and genericity problems by formally representing some of the contextual knowledge needed for the interpretation. It thus becomes "externalized" from the analysis process, where it was implicitly embedded, before. This externalization of knowledge together with dynamic links between the process and the knowledge database opens the possibility to implement flexible and adaptable interpretation devices implementing a generic analysis that is capable of adapting itself to contextual information. This section provides an overview of some of the best-known knowledge models used in graphical document analysis.

About Knowledge Modeling

Some authors propose using models of different knowledge categories that would contribute to the analysis process [1] and that these models be formalized as much as possible as to obtain a truly adaptable and context-independent interpretation system. A classification of the required knowledge in four categories can be found in [30]:

- The most obvious category concerns ***descriptive knowledge***. It covers the knowledge over the physical or conceptual domain (semantics) represented in the document on the one hand, as well as the graphical conventions (semiotics) and the rules that govern them (syntax) to represent concepts. It may also include semantics of the document's conventions like captions, legends, and references. Most often, however, it represents the rules for representing objects within the document and generally relies on structural/syntactic representations, such as graph, trees, grammars, semantic networks, or ontologies, some of which have already been described in ▶ Chaps. 15 (Graphics Recognition Techniques) and ▶ 16 (An Overview of Symbol Recognition). The representation of this knowledge allows describing the hierarchical organization of elementary primitives, as well as their topological and geometric relationships. This hierarchical description then allows to further define semantic consistency rules that can be used to check whether the information generated from a bottom-up strategy is consistent or not.

  Furthermore, it can also be used to define the sequences of tasks and subtasks to be run in order to progressively extract the information from the image and organize it according to the model. Recent developments [33] have started introducing ontologies to formalize this knowledge, since it allows not only to model hierarchies of concepts but also the relations that connect them.
- Another category covers the Image-Processing ***operator's knowledge***. It corresponds to the knowledge that is used by an expert to construct the analysis

process in the context of image-processing or pattern recognition problems. It is related to the behavior of the various image-processing operators that are used to implement the analysis strategy: they allow to describe in which context these operators are most appropriate, how they must be tuned versus a specific context, etc. For instance, they can correspond to the choice of the best image-processing operator for segmenting a texture or to the best couple (features vectors/classification process) that has to be used for recognizing a specific symbol. This knowledge is generally implicit and is rarely formally modeled. It finds itself embedded into the way algorithms are combined together (cf. next knowledge category) to form the overall analysis process, what parameter intervals are used and how they are obtained, which error or decision thresholds are used, etc. However, many papers mention the necessity to integrate this aspect when trying to implement generic systems. One of the steps in the direction of capturing the operator's knowledge, although not sufficient by itself, is to provide clear descriptions of input and output parameters and execution semantics of algorithms [20]. A more formal experiment toward integrating expert operator knowledge can be found in [7] although, strictly speaking, it does not fall within Graphics analysis; it does, in a general sense, apply to the issues described here. It presents the BORG system, aiming to generate image-processing programs and for which the proof of concept was established for cytology in medical imaging. Besides the grammar-based control mechanisms developed in the next sections (ANON, ADIC, etc.) and implementing selection strategies for finding the correct rule set to apply, BORG also allows to integrate quality measures to the image analysis steps expressing conditions like "*if the standard binarization algorithm gives rise to too many small connected components, revise the set of used thresholds in a previous step, or switch to and alternate binarization algorithm.*" The approach is not an image analysis method in a strict sense but an image analysis generator (i.e., given constraints, knowledge, and a set of training images, it will generate an image analysis program satisfying the given conditions and operating on the provided class of images).

- **Strategic knowledge** is a complementary level of implicit knowledge that is used by the image/pattern recognition expert when implementing an interpretation strategy. It deals with the sequential ordering of a set of image-processing operators in order to reach the analysis goal: how to sequence a set of image-processing operators and pattern recognition processes, in a particular context, and in order to achieve a specific level of interpretation. This kind of knowledge is far more difficult to formalize and is of a much higher level than the previous one, since it concerns the way to organize the process and not exactly how to tune each of them. The formalism that can be used for modeling this kind of knowledge can be inspired from Petri networks or serious games.

All these knowledge categories are implicitly involved in the building of an analysis device. It should be obvious to the reader that the genericity of analysis systems necessarily requires some level of formalization and external representation falling into these categories. While there is currently no existing consensus or formal theory

on how to ideally achieve this, there are however tentatives in this direction and the mentioned levels of knowledge are more or less formalized in hybrid systems, which are presented below.

## Hybrid Approaches for Graphics Analysis

Hybrid approaches use a subset of the knowledge categories mentioned in the previous section for analyzing technical documents by leading the low-level processes as a function of the context. From a historic point of view, two approaches constitute interesting contributions to these approaches and can be taken as representative tokens of a more comprehensive state of the art.

ANON and Grammar-Based Derivatives

One of them was proposed by Joseph [17] and concerns mechanical engineering drawing analysis. It was called ANON and used the "cycle of perception" proposed by Neisser, the basis of the approach being a continuous loop in which a constantly changing world model direct perceptual exploration determines how its finding are to be interpreted and is modified as a result. In ANON, this role is taken by an instance of one of a number of schema classes. The system is structured in three layers in order to separate spatial and symbolic processing. The first is composed of a large image analysis library associated to both search-tracking functions and management processes. The information extraction is adapted to the context by the second level, the "schema" (prototypical drawing construct), which receives the entities from the lower layer and interprets the result as a function of the current schema. A cycle of hypothesis verification is thus proposed by the schema to the control system (highest layer). On each cycle, the controlling instance invokes appropriate members of ANON's library of image analysis routines and informs a higher-level control of the results of its actions. This control system analyzes the proposition as a function of the current state of the proposed schema and may modify it if needed. Applied in the context of graphical recognition, the system maintains classes corresponding to solids, dashed and chained lines, solid and dashed curves, cross hatching, physical outlines, junctions, letters, words, witness and leader lines, and certain restricted forms of dimensioning. Each schema instance represents a particular example of some prototypical drawing construct.

  ANON's control module comprises a set of strategy rules written in the form of a grammar. These rules define methods by which high-level drawing entities may be obtained by hierarchical combination of low-level constructs. On each cycle, the control system determines an appropriate modification to the current schema. Modifications may correspond to an updating of an internal variable, the adding of new subparts, or the replacing of the instance with a new one representing a different type of construct. Strategy rules, like string grammars, describe acceptable sequences of events ANON's control model is represented in Fig. 17.8.

Fig. 17.8 Control structure
for ANON, as represented
in [17]
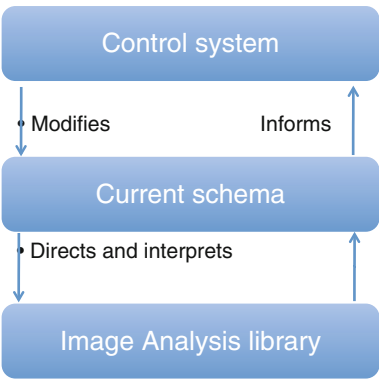
The results obtained in [17] are shown in Fig. 17.9 and clearly illustrate the limitations of the knowledge-based approaches in their beginnings. They have difficulties accounting for noise or for configurations that would be slightly deviating from the conventional configuration.

The knowledge-directed image analysis and the construction cycle according to the context are two interesting concepts that are applied to 15 different schema classes.

A similar approach is ADIK [31]. Its approach is very much related to syntactic symbol recognition and addresses the interpretation of technical diagrams by representing visual knowledge in the form of a grammar, expressing the various relationships and hierarchies between primitives and shapes as well as tolerances on allowable perturbations. The approach is conceptually similar to ANON, but is more flexible where its grammatical expression is concerned. The LR-grammar is extended with "placeholders," "triggers," and "constraints" which give it more possibilities to express local contextual conditions, making its behavior on triggering interpretations more flexible. Figure 17.10 gives some examples of detection results and the resulting exploration tree that results from the interpretation of the drawing.

In the same category, den Hartog [11] proposed a mixed approach based on a top-down control mechanism associated with bottom-up object recognition. The system decomposes the binary image into primitives (and not vectors) having a good morphological representation of the information and uses template matching to recognize each of them. Then, contextual reasoning is performed based on a loop that includes inconsistency detection and search action generation in a region of interest (ROI). The control system defines an ordered search action list to search for a specific object type in the ROI. The user specifies priorities to define the most important search actions and to assign priority to the relationship between objects. A consistency test is applied to each recognized object in order to verify the hypothesis defined at the system's top level as a function of knowledge of the object to recognize. The knowledge framework of the methodology relies essentially on spatial relationships between primitives, without integrating and describing hierarchical relationships. In the case of particularly complex documents, this kind
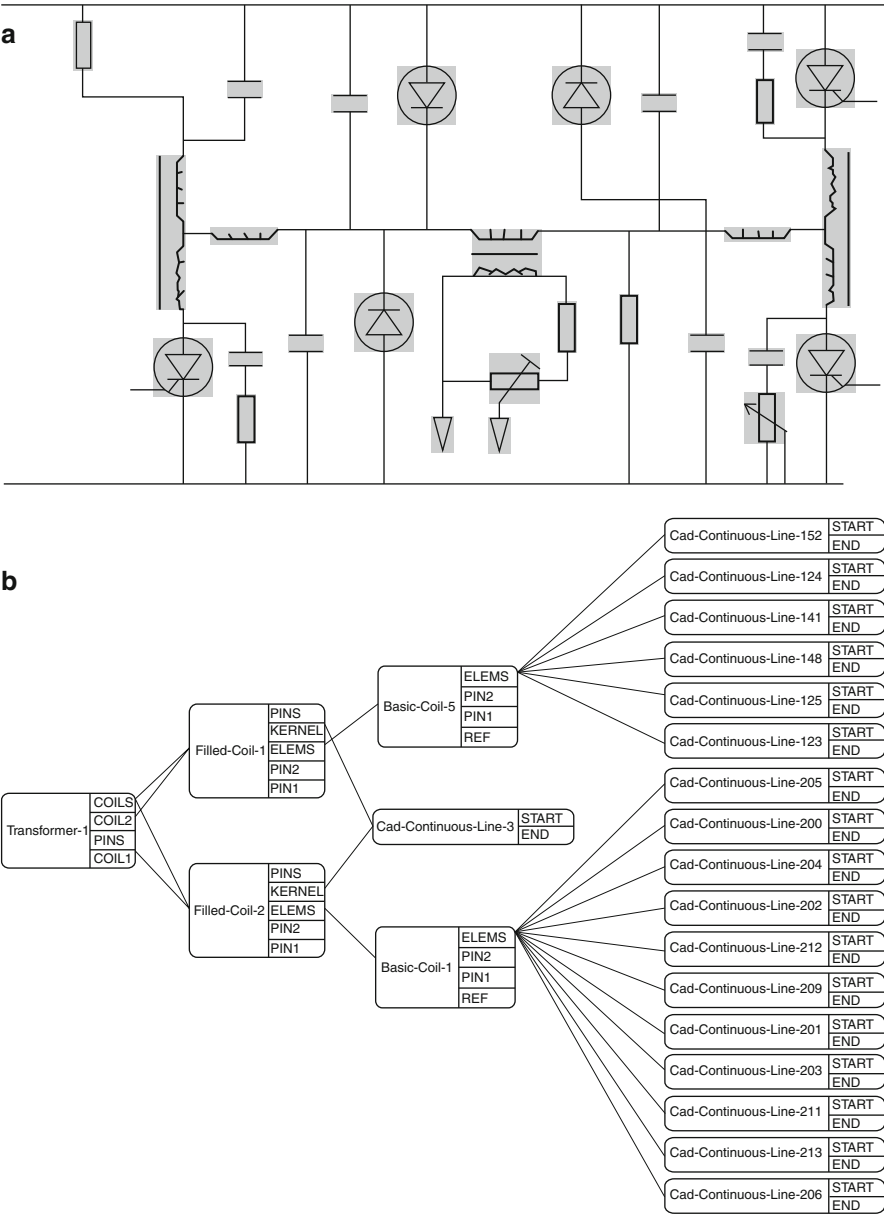
Fig. 17.9 These examples, as reported in [17], show input images and their corresponding output: (a) the original input image, (b) the raw algorithm output, (c) the manually corrected results

Fig. 17.10 Taken from [31] showing results on electrical wiring diagram analysis and the resulting interpretation tree that results from the interpretation process, connecting high-level concepts (transformers, for instance) to image features (continuous lines)

of system is penalized because of the drastically increasing number of relationships and the necessity to generate new search actions for the "designed objects."

More recent work, revisiting the previous approaches, can be found in [22]. They identify several shortcomings, the main ones being related to the fact that only graphical constraints are explicitly modeled, while, in reality, technical drawings are also governed by implicit composition rules. This has led to over-investigate approaches that are essentially "linear" in their approach to combine graphical information and achieve interpretation from a set of pixels and for which non-shape domain information is not explicitly represented or at best, according to the authors, embedded in complex rule sets.

Lu et al. [22] therefore suggest using the explicit geometric shape definitions as entries for which implicit (in the sense that they are implicit for the human interpreter, meaning they need to be made explicit for an automated process) composition rules and representation conventions are used for guiding the analysis process and to check consistency or remove ambiguity. Their architecture is based on a knowledge interpreter, a knowledge parser, and an entity searcher, very much like ANON [17]. Using the assumption that automatic interpretation is composed of a series of condition-driven processes, these conditions are represented as knowledge descriptors addressing either representation issues (recognition) or interpretation issues (control).

They identify four levels of interpretation targets: project, drawing, engineering entity, and graphical primitive for which knowledge is represented in EBNF (Extended Backus-Naur Form), all of which have external (purely graphical representations) and internal states (based on contextual and composition rules).

As shown in Fig. 17.11, the method is capable of detecting "similar" items, not only based on shape but also based on actual semantics.

The approach remains very sensitive to exhaustive visual modeling and errors introduced by noise on the one hand, as well as incoherency or missing information (local non-respect of conventions).

Among other relevant work, it is important to mention Coüasnon's DMOS system (Description and Modification of Segmentation) for analyzing structured documents and which can also be applied to graphical documents. The aim of this system is to design a generic recognition system, being able of producing either general or specific systems. The DMOS system is made of a grammatical language (Enhanced Position Formalism—EPF) and an associated parser able to deal with noise. As for the previously presented systems, the main principle of DMOS is to separate domain knowledge from source program, in order to develop the adaptability of the system. Actually, DMOS relies on a compilation phase, which, in its turn, builds an adapted recognition system, on the basis of an EPF description of the expected document structure.

As the authors state in [8], "*This method has been successfully used to produce recognition systems on musical scores, mathematical formulae and even tennis courts in videos. This. Therefore, for a same kind of document like table structures, it is possible to define with EPF, more or less specific descriptions to produce more or less specific recognition systems. For example, we have been able to produce*
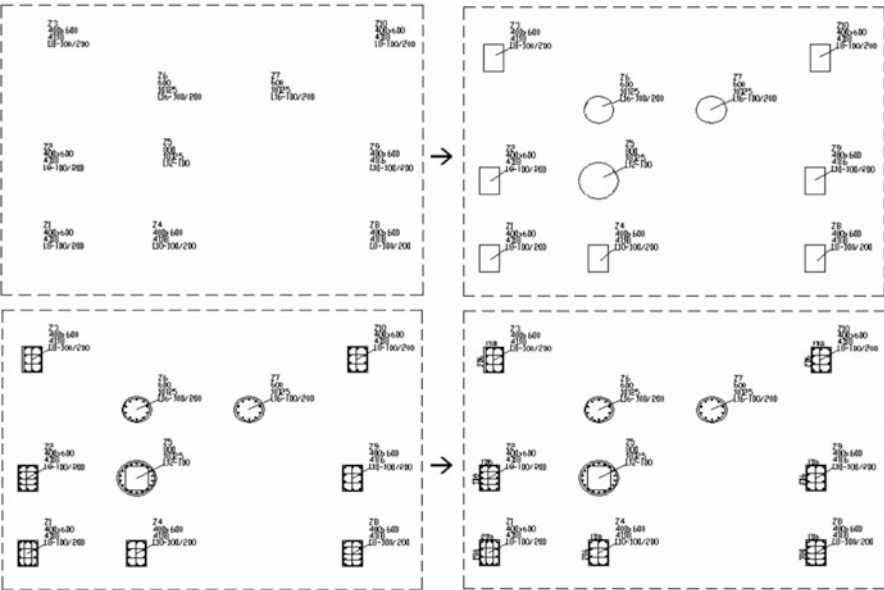
Fig. 17.11 Taken from [22] looking for reference source columns in technical construction blueprints, starting with an initial example, and progressively expanding its knowledge tree to find all occurrences in the drawing

a general recognition system of table structures. It can recognize the hierarchical organization of a table made with rulings, whatever the number/size of column/rows and the deep of the hierarchy contents in it, as soon as the document has a not too bad quality (no missing rulings for example). We will present the way the description is done using EPF to be general enough to recognize very different table organizations. With the same DMOS generic method, we have also been able to easily define a specific recognition system of the table structure of quite damaged military forms of the nineteenth century. This specific description was necessary to compensate some missing information concerning the table structure of those military forms, due to a very bad quality or hidden part of the table. This system has been successfully validated on 88,745 images, showing that this DMOS generic method can be used at an industrial level."

Context Modeling and Ontologies

Another hybrid approach is the system described in [28] for map interpretation. In this system, features are grouped together to constitute primitive objects, then these objects are assembled together to compose a larger object in the hierarchy and the process continues until it reaches the most global object which is the map itself (Fig. 17.12).
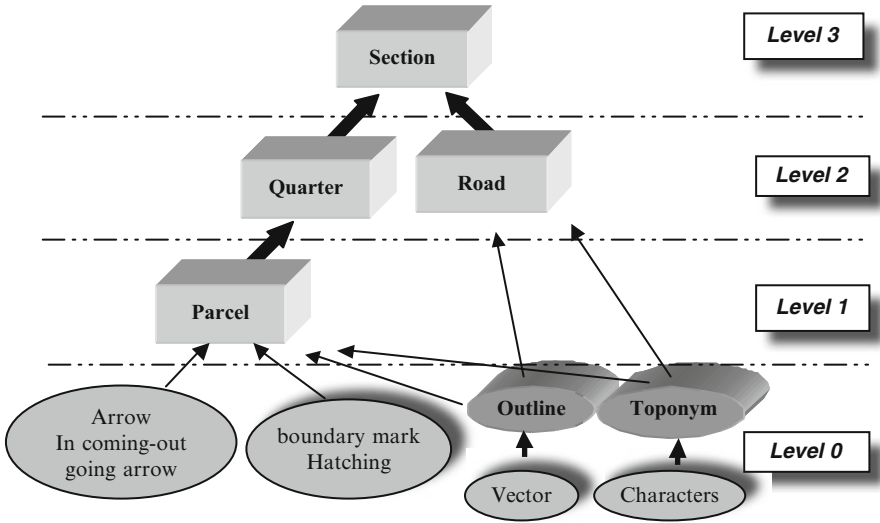
Fig. 17.12 Taken from [29] illustrating a model of knowledge representation in the context of French cadastral maps

Although this formulation is not fundamentally different than the ones expressed previously (bottom-up and/or parser-based), the focus lies more on the fact that consistency checking is performed at every level in the hierarchy. Recognized objects are analyzed to verify if they are internally and externally consistent with each other. For example, a parcel is composed of segments to set up the outline, it has a number or an arrow, and it can involve a hatched area and symbols. Internal consistency means all the components composing the object are successfully detected; if not, a forward heuristic rule is used to correct this situation by re-extracting features in this region after modifying and relaxing the parameters of the low-level image-processing tools. On the contrary, external consistency takes into account the neighborhood of the treated object. If an object has all the components and responds to the semantic of the considered level, it is defined as an internal consistent; furthermore, if all the objects adjoining it are all internally consistent, this object will equally become more reliable through the construction of the superior hierarchical level (e.g., the parcel by the block). It is then called externally consistent. This approach is summarized in Fig. 17.13.

One of the more recent approaches was proposed in [33]. In this thesis an object extraction method from ancient color maps is proposed. It consists of the localization of the frame, text, quarters, and parcels inside a given cadastral map. First, a model of what visually characterizes a cadastral map is defined by combining knowledge from various domain experts: historians and architects on the one hand, and image analysis professionals on the other hand. Next, dedicated image-processing tools aim at locating the various kinds of objects laid
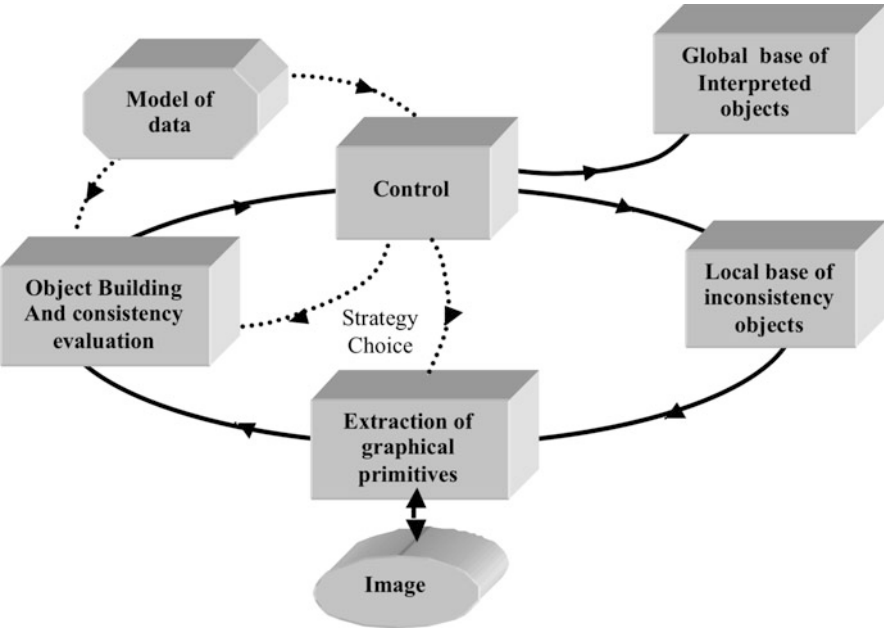
Fig. 17.13 From [29] illustrating a cyclic strategy for the interpretation of French cadastral maps

out in the raster image. These especially designed detectors can retrieve different components such as characters, streets, frame, quarters, and parcels. Thereafter, this information feeds a higher level, which elaborates a graph structure where nodes refer to the presence of objects found during the detection step and edges represent the spatial relations between them. Terms, words, and appellations to qualify node and edge labels are so called concepts. All concepts have been previously modeled by the domain experts and are represented in an ontology containing the vocabulary and the description logics of each element required to model a cadastral map. Therefore, the produced graph can be seen as a particular instance of the generic map model. On the other hand, given the relatively "bottom-up" extraction method used to obtain the graph, the latter is not constrained by the ontology and variations which are nonconforming to the knowledge base may have been introduced into the graph structure (due to defects in the detection algorithms, noise in the images, unexpected shape variations, etc.). As a consequence, a higher level of representation is required to determine up to which level the extracted graph conforms to the expert knowledge. The structure of the graph is analyzed with the joint use of a cadastral map ontology to re-engineer a meta-model corresponding to the instance data. In a last phase, the meta-model corresponding to the instance data is compared with the meta-model defined by the experts. This comparison is carried out thanks to a graph-matching algorithm. An overview of the main actions carried out at this stage is shown in Fig. 17.14.
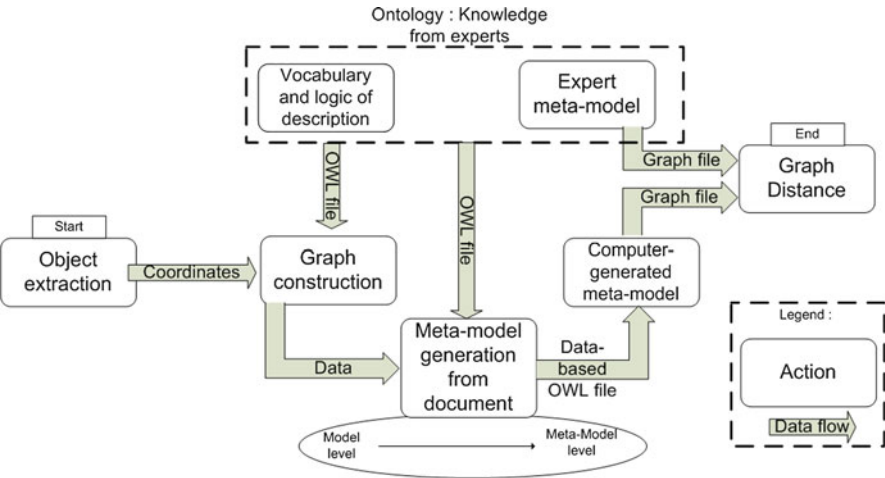
Fig. 17.14 Data flow process for meta-model inference from a model, taken from [33]

Discussion and Limitations

General Discussion

Although there is no formal evidence of the following assumption, the graphical document analysis advances seem to have reached a plateau where end-to-end generic interpretation systems are concerned. The collection of concepts and methods coming from compartmentalized research communities that are required to be integrated with one another to produce full document analysis seem to resist to all efforts trying to remove human ingenuity. Therefore, current tendencies show the evolution toward strategies that are easier to implement, based on user interaction or based on partial interpretation strategies. Among these partial interpretation strategies, many of them rely on spotting-based concepts, which consist in offering navigation services into document database, without systematically interpreting the document content. These research axes appear to be very promising since they represent a good trade-off between efficiency, genericity, and automation. Some of these approaches integrate user interactions for the management of knowledge in order to dynamically adapt to new interpretation contexts, without the need of having them formalized. A short overview of these systems is given in the following part.

Cookbook and Practical Tips for Graphics Interpretation

The essential remaining question that needs to be addressed in this chapter is "What do I do with my particular interpretation problem?" for the user who wants to implement some of the tools reviewed in this handbook.

It has already been stressed on multiple occasions throughout this chapter that there is no standard answer to the question. There are, however, some decent rules

of thumb that may guide the interested reader to a practical, operational, and efficient compromise. Table 17.2 gives an overview of configurations that are likely to occur, based on how much one knows about the graphical data at hand on the one side and the intended interpretation context on the other side.

Conclusion

Recent Evolutions: Learning, Spotting, and Indexing for Navigation into Graphical Document Repositories

Because of both the theoretical and practical considerations mentioned in the previous section related to the difficulty of developing generic analysis systems that would be able to manage any kind of document within a broad range of representations, recent approaches, principally developed during the last decade, try to approach analysis from another viewpoint. Rather than to aim for a fully automated analysis process and subsequent interpretation, with the difficulties of capturing all contextual knowledge, the research focus has shifted to leaving part of the analysis to a human interpreter and to offering efficient tools for handling large volumes of documents, mainly using "intelligent" indexing strategies.

One of the main turning points introducing this paradigm shift can be attributed to [42] spurring investigations toward the idea of "spotting" or [24] insisting on human interaction. Their overall observation is that rather than trying to fully represent contextual knowledge for solving interpretation problems, three main substitution strategies may prove just as effective:

1. Example-based or supervised learning and classification techniques can be used to replace interpretation context modeling (although there are pitfalls to be avoided if the sample population is inadequately chosen [24]).
2. Spotting and indexing can be used to (partially) replace full contextual interpretation by guiding a human interpreter rapidly to documents or document parts that have a high probability of fitting his or her interpretation.
3. Human interaction should be much more seen as a continuous part of the analysis process, rather than just participating in the knowledge modeling phase (e.g., by dynamically influencing classifier decision boundaries or indexing feature selection through relevance feedback).

As an illustration, one can consider the example of an analysis system using spotting techniques and indexing. The idea of spotting essentially consists in favoring recall over precision by proposing realistic navigation and retrieving services on large document corpora, fine-tuned on the requirements of a specific use case (usually given by the organizations managing the corpora), without systematically running costly full recognition processes (i.e., both high precision and high recall). One of it main advantages is that it avoids trying to interpret the whole document (which, for specific applications, is not really useful, anyway) and allows a possible second-stage process to focus on the interpretation on relevant spotted objects. Therefore, this kind of approach allows to simplify the analysis process, since it can use

Table 17.2 Synoptic cookbook for solving practical document analysis problems in function of how well the interpretation context is known, relating to the intended public and users (horizontal) and how data variability can be controlled (vertical)

Control over interpretation context (horizontal) vs. control over data variability (vertical)	Low – interpretation context cannot be formally determined: general public and nonspecialist users	Medium – interpretation context can be formalized in some way: public or users are assumed to be knowledgeable to some level	High – interpretation context is highly constrained and can be formalized: trained public and specialized users	Invariants for given variability on data
Low Data may contain relevant patterns but they cannot be determined in advance and data of interest may be hidden in stream of potentially irrelevant documents	No reports exist of proven valid approaches in this very open context. Current consensus is that the most appropriate way of dealing with these classes of problems is to use weakly supervised classification and user relevance feedback to implement a symbol spotting approach and hope for the best [34]	This is one of the harder cases, where there is a fairly well, although insufficiently formalized, interpretation context, but without sufficient visual support in the data (i.e., there is a large *semantic gap*). In this case, the most appropriate approach is likely to investigate further whether there are classes of interpretation contexts that can be better specified and reduce the problem to a more constrained one	If the interpretation context can be formalized in explicitly, it is possible to establish a set of constraints the data must fulfill (although they need not necessarily be expressed in visually verifiable ways). This can in its turn guide top-down analysis strategies, on the one hand. On the other hand, the presence of specialized users can also lead to a two-stage coarse-to-fine approach where supervised learning techniques can be trained to reduce the number of irrelevant documents first and then separate the data "manifold" into "sub-manifolds" of lower variability [28]	For this category of graphical data, the best approach is to use statistical learning over a large set of possible relevant descriptors. There is no guarantee that the combination of classifiers and descriptors will eventually capture all possible instances of relevant data. Statistical evaluations of the power of discrimination of descriptors should be used [35]

(continued)

Table 17.2 (continued)

Medium Data can be assumed to obey general visual rules, but variations in shape or noise can be important and are difficult to model. The application context is such that outliers containing completely irrelevant data are rare	This situation corresponds to particular cases for which similarity between data can be expressed through statistical descriptors as well as visual key point approaches in the context of information spotting. This kind of approach has been recently used in the context of historic document indexing [26] in the case of weakly structured graphical images. User interaction should be used to characterize	This context is very rare and there are no relevant references available	In this situation the focus of the analysis process will be on establishing how the formal interpretation knowledge projects onto the approximate and noisy (yet globally coherent) graphical data. This is probably the most active area in the current state of the art. Coüasnon [8] and Raveaux [33] are excellent examples	When data is only approximately formalized yet most of it is relevant, the current state-of-the-art consensus is to use non- or semi-supervised classification techniques to determine optimal descriptions of the graphical data, e.g., by using visual bag of words. Depending on how much is known about the interpretation context, the supervision of the classification techniques and/or their a posteriori validation can be more or less elaborate and efficient
High Data falls into a very well understood and formally modeled scope, noise and deformation levels are within well-established bounds, and outliers are rare	Highly controlled graphical data most often comes from very well-controlled production contexts and obeying to a very well-established visual language. It seems contradictory to combine this with an undetermined interpretation context, but there are situations where this makes	Since the graphical data can be fully modeled and noise and deformation levels are known, the appropriate image description tools can be used to fully characterize the data. According to the available interpretation knowledge, the user can be presented with classes of visual	In this configuration the level of available formalization of both the interpretation and the data does not require a complex level of knowledge representation. A custom-tuned (probably bottom-up) and hardwired approach is very likely to be the most efficient and cost effective [15]	Given the high accuracy of knowledge about how the data behaves, robust and reliable visual descriptors can be easily extracted from the document images. Depending on what kind of interpretation knowledge is available, this visual (low-level) interpretation can then be more or less related to

higher-level interpretations, either fully automatically or with some user interaction, when possible

sense. For instance, electric wiring diagrams or architectural floor plans may be used for a very broad variety of interpretation contexts The most reasonable approach here is to consider "retro-conversion," without trying to capture the end user's interpretation intent, but by trying to express the visual information into a higher-level formal description that can then be used as a support for a user-driven analysis process [18]

information or partial interpretations with respect to the level of available interpretation contexts' formalization. Semi-supervised learning or user relevance feedback can be applied to converge toward a user-acceptable interpretation. Fully automatic interpretation remains out of reach [22]

The existence of a very well known and formalized interpretation context, as well as specialized users, offers the possibility to fully express acceptation/rejection criteria of interpretations. Depending on what quality of data it is applied to, these acceptation/rejection criteria can more or less be related to visual information

Invariants for given class of interpretation contexts

In this context it is impossible to make any assumption about what final interpretation is required and needs to be left to subsequent treatment or analysis; either this is done by the user itself, or enough is known about the visual description language so that a formal description of the document can be extracted for further automated analysis
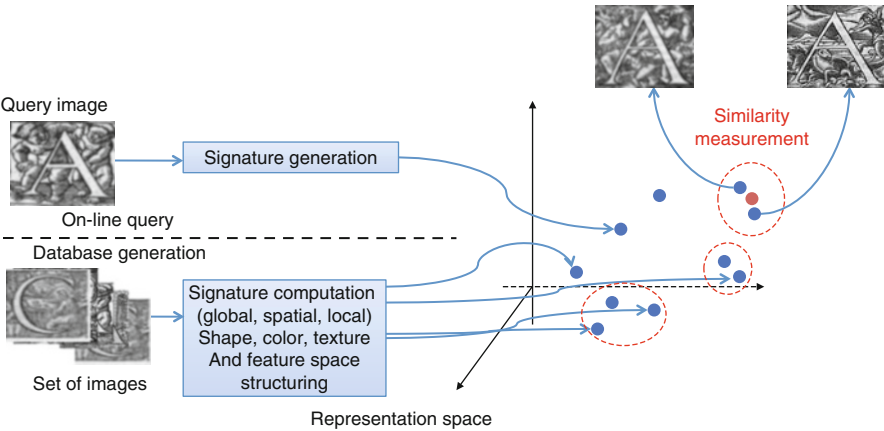
Fig. 17.15 Principle of signature computation and information spotting in the context of drop caps spotting

alternative detection and recognition processes which rely on less complex data representations and pattern recognition strategies. Furthermore, from the knowledge management point of view, this kind of system does not necessarily require exhaustive formalization since most of it can be dynamically caught through man-machine interactions or relevance feedback. The global schema of such a spotting system is presented in Fig. 17.15.

This kind of system is generally composed of two main parts, one off-line and another on-line.

The off-line part aims at analyzing the content of documents by extracting some features allowing to characterize each of them in a unique way. Until early 2000, the features that were used to describe documents generally relied on classical pattern recognition-based descriptors, i.e., based on statistical of structural approaches described in previous sections of this handbook. More recently, bag-of-words approaches were proposed, trying to characterize document contents without necessarily describing them on the basis of human knowledge.

The on-line part aims to propose interactive interfaces allowing the user to retrieve documents on the basis of queries, which can be expressed either on the basis of keywords or on the basis of images.

Challenges in Graphics Interpretation

The main conclusion drawn from the previous sections is that there is no actual commonly agreed set of best practices or globally adopted methodology for complete analysis and interpretation systems. As a matter of fact, the research community has gradually abandoned the investigation of end-to-end applications in this domain, focusing more on subparts like recognition and indexing, segmentation,

or spotting. This is the main reason why many of the references in the previous sections are relatively old. While the results of these described approaches are quite interesting in their respective application contexts, it becomes less easy to really assess their value from a more general viewpoint: how do they adapt to other contexts, how do they perform with regard to recent developments in lower level treatments, etc.? The result is that most interpretation and analysis methods remain very context specific and that, therefore, analysis problems are handled on an ad hoc basis. Therefore, the main challenge for the graphical interpretation community is to establish a classification of its methods and low-level approaches described in the previous sections and to relate their appropriateness to higher-level interpretation contexts. As made clear through the overview given in this chapter, and based on the results presented in the previous chapters, there is a significant gap between the performance of individual graphical image treatment and recognition approaches and the performance of full analysis methods. In this chapter the focus has been set on knowledge modeling as a means to bridge this gap, and several partially successful approaches have been developed in this domain, over the years. However, effective conclusions still need to be drawn from these experiments and there is no established consensus on good practices or better choices in specific application contexts.

The main challenges for graphics analysis therefore are related to performance analysis on the one hand, and context characterization on the other hand as well as fully integrating the human user in the analysis and interpretation process, helping to focus the knowledge modeling or information characterization through relevance feedback, for instance.

It remains an open question whether this is actually a realistic goal. Since there is no recent published work in what interpretation exactly means or entails from a Machine Perception point of view, the problem of measuring the state of the art remains open. However, when broadening the scope beyond Machine Perception into formal semantics and model checking on the one hand and reaching out even further into linguistics and even metaphysics, it does not seem absurd to try and relate recent advances in those domains concerning semantics and interpretation to what has been described in this chapter. While automated interpretation of perception data is a very ill-posed problem in the current state of the art, trying to formulate it in a more abstract way will probably show a number of limitations related to tractability and decidability and therefore allow the graphical document analysis domain to better grasp the reasons behind the currently observed limits of its approaches and perhaps provide means to try and overcome them.

Cross-References

▶An Overview of Symbol Recognition
▶Graphics Recognition Techniques
▶Imaging Techniques in Document Analysis Processes
▶Sketching Interfaces

References

1. Adam S, Ogier J-M, Cariou C, Mullot R, Labiche J, Gardes J (2000) Symbol and character recognition: application to engineering drawings. Int J Doc Anal Retr 3(2): 89–101
2. Antoine D, Collin S, Tombre K (1992) Analysis of technical documents: the REDRAW system. In: Baird HS, Bunke H, Yamamoto K (eds) Structured document image analysis. Springer, Berlin/Heidelberg, pp 385–402
3. Arias JF, Lai CP, Surya S, Kasturi R, Chhabra A (1995) Interpretation of telephone system manhole drawings. Pattern Recognit Lett 16:355–369
4. Baum LS, Boose JH, Kelley RJ (1997) Graphics recognition for a large-scale airplane information system, GREC. Lecture notes in computer science, vol 1389. Springer, Berlin, pp 291–301
5. Boatto L et al (1992) An interpretation system for land register maps. IEEE Comput Mag 25(7):25–33
6. Broelemann K, Jiang X (2012) A region-based method for sketch map segmentation. In: Graphics recognition. New trends and challenges. Lecture notes in computer science, vol 7423. Springer
7. Clouard R, Elmoataz A, Porquet C, Revenu M (1999) Borg: a knowledge-based system for automatic generation of image processing programs. IEEE Trans Pattern Anal Mach Intell 21(2):128–144
8. Coüasnon B (2006) DMOS, a generic document recognition method: application to table structure analysis in a general and in a specific way. Int J Doc Anal Retr 8(2–3): 111–122
9. Coustaty M, Uttama S, Ogier J-M (2012) Extraction of light and specific features for historical image indexing and matching. In: 21st international conference on pattern recognition, Tsukuba, 11–15 Nov 2012, pp 1326–1329
10. Das AK, Langrana NA (1997) Recognition and integration of dimension sets in vectorized engineering drawings. Comput Vis Image Underst 68(1):90–108
11. Den Hartog JE, ten Kate TK, Gerbrands JJ (1996) Knowledge-based interpretation of utility maps. Comput Vis Image Underst 63(1):105–117
12. Deseilligny MP, Mariani R, Labiche J, Mullot R (1998) Topographic maps automatic interpretation: some proposed strategies. In: Tombre K, Chhabra AK (eds) Graphics recognition algorithms and systems. Lecture notes in computer science, vol 1389, pp 175–193
13. Devaux PM, Lysak DB, Kasturi R (1999) A complete system for the intelligent interpretation of engineering drawings. Int J Doc Anal Retr 2(2–3):120–131
14. Dori D, Wenyin L (1999) Automated CAD conversion with the machine drawing understanding system: concepts, algorithms, and performance. IEEE Trans Syst Man Cybern Part A 29(4):411–416
15. Dosch P, Tombre K, Ah-Soon C, Masini G (2000) A complete system for the analysis of architectural drawings. Int J Doc Anal Retr 3(2):102–116
16. Janssen RDT (1995) The application of model-based image processing to the interpretation of maps. PhD thesis. Technische Universiteit, Delft
17. Joseph SH, Pridmore P (1992) Knowledge-directed interpretation of line drawing images. IEEE Trans PAMI 14(9):928–940
18. Kasturi R, Tombre K (eds) (1996) Graphics recognition – methods and applications. Lecture notes in computer science, vol 1072. Springer, Berlin
19. Kasturi R et al (1990) A system for interpretation of line drawing. IEEE Trans PAMI 12(10):978–992
20. Lamiroy B, Lopresti D (2011) An open architecture for end-to-end document analysis benchmarking. In: 11th international conference on document analysis and recognition – ICDAR, Beijing. IEEE Computer Society, pp 42–47
21. Li L, Tan CL (2012, to appear) Real scene graphics symbol recognition. In: Post-conference proceedings of the GREC 2011, Seoul. Springer

22. Lu T, Tai C-L, Yang H, Cai S (2009) A novel knowledge-based system for interpreting complex engineering drawings: theory, representation, and implementation. IEEE Trans Pattern Anal Mach Intell 31(8):1444–1457
23. Mas J, Lladós J, Sánchez G, Jorge JAP (2010) A syntactic approach based on distortion-tolerant Adjacency Grammars and a spatial-directed parser to interpret sketched diagrams. Pattern Recognit 43(12):4148–4164
24. Nagy G (2004) Visual pattern recognition in the years ahead. In: Invited plenary presentation, Proceedings of the international conference on pattern recognition, Cambridge, vol II, pp 311–314,
25. Nguyen T-O, Tabbone S, Boucher A (2009) A symbol spotting approach based on the vector model and a visual vocabulary. In: Proceedings of the 10th international conference on document analysis and recognition, Barcelona, pp 708–712
26. Nguyen TTH, Coustaty M, Ogier J-M (2011) Bags of strokes based approach for classification and indexing of drop caps. In: Proceedings of the 11th international conference on document analysis and recognition, Beijing, pp 349–353
27. Ogier J-M, De l'image au sens, Habilitation à diriger de recherches thesis (2000). University of Rouen, 22 Dec 2000
28. Ogier J-M, Mullot R, Labiche J, Lecoutier Y (1998) Methodology and evaluation of a document interpretation device. In: Proceedings of the SCI'98 and ISAS'98, world multiconference on systemics, cybernetics and informatics and the 4th international conference on information systems analysis and synthesis, Orlando, vol 2, 12–16 July 1998, pp 457–463
29. Ogier J-M, Mullot R, Labiche J, Lecourtier Y (2000) Semantic coherency: the basis of an image interpretation device-application to the cadastral map interpretation. IEEE Trans Syst Man Cybern B 30(2):322–338
30. Ogier J-M, Mullot R, Labiche J, Lecourtier Y (2001) Technical map interpretation: a distributed approach. Int J Pattern Anal Appl 3:88–103
31. Pasternak B (1994) Processing imprecise and structural distorted line drwings by an adaptable drawing interpretation kernel. In: Proceedings of the IAPR workshop on document analysis systems, Kaiserslautern, pp 349–365
32. Pasternak B, Neumann B (1995) The role of taxonomy in drawing interpretation. In: Proceedings of the 3rd international conference on document analysis and recognition, Montreal, pp 799–802
33. Raveaux R (2010) Graph mining and graph classification: application to cadastral map analysis. PhD thesis, University of la Rochelle, La Rochelle
34. Rendek J, Lamiroy B, Tombre K (2006) A few steps towards on-the-fly symbol recognition with relevance feedback. In: 7th international workshop, document analysis and systems, Nelson. Lecture notes in computer science, vol 3872. Springer, Berlin, pp 604–615
35. Rosenberger C, Ogier J-M, Cariou C, Chehdi K (1998) A statistical solution to evaluate image processing techniques. In: 7th international conference IPMU, Paris, vol 2, pp 1648–1645
36. Rusiñol M, Borràs A, Lladós J (2010) Relational indexing of vectorial primitives for symbol spotting in line-drawing images. Pattern Recognit Lett 31:188–201
37. Samet H, Soffer A (1994) A legend-driven geographic symbol recognition system. In: Proceedings of the 12th international conference on pattern recognition, Jerusalem, vol II, pp 350–355
38. Shimada S, Maruyama K, Matsumoto A, Hiraki K (1995) Agent-based parallel recognition method of contour lines. In: Proceedings of the ICDAR'95, Montreal, pp 154–157
39. Shimotsuji S, Hori O, Asano M, Suzuki K, Hoshino F, Ishii T (1992) A robust recognition system for a drawing superimposed on a map. IEEE Comput 25(7):56–59
40. Smeulders AWM, de Boer C (1997) Design and performance in object recognition. In: Second IAPR international workshop on graphics recognition (GREC), Nancy, pp 335–346
41. Suzuki S, Yamada T (1990) MARIS: map recognition input system. Pattern Recognit 23(8):919–933
42. Tombre K, Lamiroy B (2003) Graphics recognition – from re-engineering to retrieval. In: ICDAR, Edinburgh

43. Vaxivière P, Tombre K (1995) Knowledge organization and interpretation process in engineering drawing interpretation. In: Spitz AL, Dengel A (eds) Document analysis systems. DFKI, pp 307–317
44. Yu Y, Samal A, Seth SC (1995) A system for recognizing a large class of engineering drawings. In: Proceedings of the 3rd international conference on document analysis and recognition, Montreal, pp 791–794

Further Reading

Although the knowledge-based approaches described in the previous section seem intellectually much more satisfactory and more enhanced than classical approaches since they try to dynamically link knowledge, image analysis techniques, pattern recognition tools, and document interpretation contexts, very few production-ready systems or significant technological breakthroughs have been reported since 2000. Kasturi and Tombre [18] contains a good state-of-the-art review of the main available techniques and tools for the analysis of technical and cartographic documents and subsequent publications have not really contributed to fundamentally change the state of the art.

Exception can be made for the recent evolutions involving the development of ontologies and web semantic approaches [33]. They offer some interesting alternatives that allow to better formalize knowledge and render them quite versatile for use in analysis systems. However, considering the high level of complexity of these problems, even if scientific communities agree on the importance of the development of generic interpretation systems, and the necessity to dynamically connect knowledge management and analysis scenario, one must admit that there is still no generic system allowing to solve broad classes of interpretation problems. One of the reasons may be that ontologies themselves (or any other formal knowledge representation) are only partially capturing the underlying complexity of the required knowledge by requiring that it be expressed within the boundaries of computational description logics, thus shifting the intrinsic difficulties just a level further without actually addressing them in full.