Reprint

from

Proceedings of the Third International Conference on Document Analysis and Recognition

Montreal, Canada

August 14 - 16, 1995

A Framework for Validating Recognized Results in Understanding Table-Form Document Images

Toyohide Watanabe and Teruo Fukumura

IEEE
COMPUTER
SOCIETY

Washington, DC ◆ Los Alamitos ◆ Brussels ◆ Tokyo

PUBLICATIONS OFFICE, 10662 Los Vaqueros Circle, P.O. Box 3014, Los Alamitos, CA 90720-1264 USA

A Framework for Validating Recognized Results in Understanding Table-form Document Images

Toyohide WATANABE

Department of Information Engineering,
Faculty of Engineering, Nagoya University
Furo-cho, Chikusa-ku, Nagoya 464-01, JAPAN

Teruo FUKUMURA

School of Computer and Cognitive Sciences,
Chukyo University
101-Tokotate, Kaizu-cho, Toyota 470-03, JAPAN

Abstract

We address the knowledge representation for understanding table-form documents. In particular, we focus on the knowledge for establishing the correctness of recognized results in addition to the traditional syntactic/constructive-oriented understanding framework. In order to attain to this subject, we introduce the semantic information to specify domains and relationships among individual items definitely, which is complementary to the syntactic/constructive information available in the traditional framework.

1 Introduction

The knowledge-based approaches for understanding document images are successful in point of the flexibility, applicability and adaptability of recognition methods in comparison with image-processing-based approaches. Many of researches reported that the knowledge-based approaches are applicable to various kinds of documents. One of the main characteristics is to be able to explain the recognition process and recognized results logically [1, 2].

However, it is not easy to make the recognition ratio 100% even if the knowledge-based approaches were applied. Of course, the image-processing-based approaches do so. Additionally, it is difficult for both approaches to make the ratio of recognition error 0%. In many cases, the issue of low error ratio was not always one of interest. In comparison with the image-processing-based approaches, the knowledge-based approaches are more adaptable to the issue of low error ratio. This is because the knowledge-based approaches can distinguish individual recognition objects interpretatively with their own appropriate knowledge. In this paper, we address the issue of low error ratio in addition to the issue of high recognition ratio.

2 Recognition Mechanism

In the document image understanding, our framework is composed of three-layer recognition processes. These recognition processes are related to extract and classify objects stepwisely [2, 3, 4]. The layout recognition process distinguishes meaningful item blocks/fields from 2-dimensional sheet media. The item recognition process classifies individual items from 1-dimensional item blocks/fields. Also, the character recognition process unifies each character pattern and converts them to character codes from 0-dimensional item data. We show our recognition framework in Fig.1.

In Fig.1, we observe that the control flow is assigned from higher-layer to lower-layer sequently if individual recognition procedures work successfully. However, if a recognition process can not interpret objects with its own knowledge successfully, the control flow must be suspended temporarily. In this case, it is better if we can organize the system so as to first return back the control from the lower-layer to the higher-layer and then reinterpret the previously recognized objects newly with another strategy. Additionally, it is necessary to check up whether the finally recognized objects are correct or not. Even if the system was successfully developed, the recognition capability can not attain to the perfect recognition results and less recognition errors because the recognition objects(or document images) may be not constructed so as to be adaptable with the knowledge about layout structure, item sequence, item property, data format and so on.

We introduce a validation facility to check up recognized results, and a recovery facility to detect erroneously recognized results and then modify them correctly. In order to apply these functional facilities to our existing framework, we must investigate new knowledge, which can make the functionality powerful, because current knowledge is applied only to analyze document structures and recognize items.

3 Knowledge Representation

In the case of knowledge-based approaches, the knowledge specification method is very important. Hereafter, we attach to table-form documents in order to investigate the representation abilities concretely.

3.1 Physical or Logical Representation

Documents have their own geometric structures and the composition rules are defined mutually among composite items. Particularly, the layout structures of documents are dependent on application-specific usages. The difference between physical and logical representations is dominated with respect to 2-dimensional layout structures. The physical representation is defined by coordinate data of individual items such as positions, sizes, lengths, etc. While, the logical representation is specified by means of interdependent relationships among items constructively.

0-8186-7128-9/95 $4.00 © 1995 IEEE

The logical representation is abstracted more than the physical representation. So, the logical representation is applicable to various documents of the same or similar types(or classes) if the inference mechanism based on the defined knowledge is effectively provided. Consider two document fragments in Fig.2. In the logical representation "neighbor(A,B)" shows that item blocks(or fields) "A" and "B" are adjacent, and is applicable to (a) and (b). While in the physical representation the coordinate data of (a) and (b) must be checked up interpretatively whether they are the same positions for x- and y-axes.

3.2 Abstraction Level

In specifying the knowledge about documents, we can concentrate on various characteristics and represent them as usable knowledge. The representation depends on the utilization way of knowledge and the interpretation way of document images. We call the representation range of knowledge as the abstraction level. The higher the abstraction level is, the stronger the applicability of knowledge becomes.

Consider the fragments of table-form documents in Fig.3. These fragments are specified as the neighboring relationships among rectangular item blocks/fields: "A", "B", "C" and "D". If we represented the adjacent relationships with commonly shared line segments among two rectangular item blocks/fields, we can specify the relationships as illustrated in Fig.4. In Fig.4, the arrows indicate the relationships, and the symbols "h" and "v", which are attended with the arrows, show that these neighboring item blocks/fields are connected horizontally or vertically. Individual fragments are represented by different neighboring relationships. While, if we make use of left-upper corners of individual rectangular item blocks/fields, the adjacent relationships are illustrated in Fig.5. Three different fragments are specified by the same neighboring relationship. In Fig.5, the arrows and the symbols "h" and "v" indicate the same as those in Fig.4. The distinction between the representation in Fig.4 and that in Fig.5 is clear.

3.3 Syntactic/Constructive or Semantic Information

Today, the knowledge usable to analyze/recognize document images has been composed on the basis of layout structures. This is because the currently proposed approaches focus mainly on the geometric characteristics of documents and distinguish individual item data constructively. However, these approaches based on geometric structures are not always applicable to various documents: layout structures are irregularly transformed from the original forms or structural characteristics are not extracted sufficiently.

It is difficult to make sure of the recognized results validly, which were distinguished only by means of syntactic/constructive knowledge. This is because the processing based on the syntactic/constructive knowledge is too heavily dependent on the representation forms of objects. The semantic knowledge is information which defines the domains of individual items and specifies the interdependent relationships among items.

Consider a fragment of table-form document in Fig.6. In this example, individual items are defined under the domains and are interrelated by the relationships. The issue of document image understanding does not only focus on the development of the effective method for extracting/classifying individual items automatically, but also must concentrate to make sure that the identified results are completely valid.

4 Knowledge for Table-form Documents

The syntactic/constructive information is useful to identify individual item blocks/fields. The semantic information is complementary to the syntactic/constructive information so as to validate objects which were interpreted syntactically or constructively.

Structure Description Tree

The structure description tree is illustrated in Fig.7. This tree is divided into a global structure tree and many local structure trees: the global structure tree represents an approximate characteristic of layout structure, using adjacent relationships among item blocks; and the local structure tree indicates a detail characteristic of composite item blocks, using connective relationships among item fields. The nodes in the global structure tree are classified into three types: vertically repeated node "D"; horizontally repeated node "R"; and non-repeated node "S". Namely, the constructive characteristics of individual item blocks are distinguished as "D", "R" or "S" when the document image is scanned from the left-upper corner to the right-lower corner. While, the nodes in the local structure tree are divided into the vertical node "v", horizontal node "h" and terminal node "t". When the parent node is vertical, the upper and lower item fields separated by the cutting line are connected through the left and right edges. When the parent node is horizontal, the left and right item fields are connected through the left and right edges. The structure description tree is a logical representation of document layouts and also is syntactic/constructive information. This is applicable to various kinds of table-form documents so that we have already addressed in our published papers [1, 2, 5, 6].

Item Sequence Rule

The item sequence rule represents composite relationships among items. Consider an address description. In Fig.8, some items may be abbreviated: the arrows indicate the connectivities among items. This item sequence rule is a logical representation and also is syntactic/constructive information [1].

Item Frame

The item frame expresses properties of item. Consider some items of an address data. The item frames "city" and "Zip-code" are shown in Fig.9. The item frame is represented as a record structure, and individual characteristics of item are composed of slot values, including slot terms. The last located slot is a condition term which judges whether the instance should be determined or not as the corresponding item. This item frame is semantic information [1].

Pattern Dictionary

The pattern dictionary may be not regarded as the knowledge. However, if we could look upon this as knowledge, this is a physical representation, and also is syntactic/constructive information.

Domain-dependency Relationship

The domain-dependency relationship defines domains of individual items and relationships among items. Consider a document fragment in Fig.10. In this case, an item is summed from other items: A=A1+A2, B=B1+B2, C=C1+C2, D=D1+D2. We define some of domain-dependency relationships:

(1) location relationship among items
  - horizon(A: B1,B2,....): The item "A" is horizontally dependent on items "B1", "B2",.... .
  - vertex(A: B1,B2,....): The item "A" is vertically dependent on items "B1", "B2",.... .

(2) dependency relationship among items
  - dependency(A,B,C,....) : If the item "A" has a value, items "B", "C",.... have also individual domain values. If "A" has no value, "B", "C",.... have no values.
  - association(OP: A,B,C,....): The item "A" is defined as a result of the operation "OP" for items "B", "C",.... .

(3) value relationship among items
  - same(A,B): The value of item "A" is the same as that of item "B".
  - variant(A=a1: B=b1, C=c1, ... ; A=a2: B=b2, C=c2, ...): Values of items "B", "C",.... are different by depending on the value of item "A".
  - derived(A: B,C,....): The item "A" is derived from items "B", "C",.... .
  - referred([A,B,....]: F): The tuple [A,B,....] is assigned selectively from the predefined table "F".

These relationships are useful to specify individual items meaningfully. The domain-dependency relationship is complementary to the other knowledge. Fig.11 shows interrelations among various kinds of knowledge, which we addressed here.

5 Validation

Even if the recognition ratio was 100%, the method is not useful if the ratio of recognition error is larger. The validation of recognized results is important. In addition, the detection and recovery processes are applicable with a view to improving miss-recognized results correctly. The syntactic/constructive information is applied to distinguish objects, while the semantic information is effective to validate the recognized objects(and is useful to detect and recover the miss-recognized objects). Fig.12 illustrates such an interrelation between recognition and validation.

In Fig.12, the validation procedure is located as a postprocessing. This is because such a validation process checks up whether the finally recognized results are valid or not in the domains or relationships. The item frames or domain-dependency relationships are applicable to individual recognition procedures with respect to the validation and recognition cycles locally.

6 Conclusion

Currently, the approach about document image understanding have been almost established, we think. However, we can observe various kinds of documents, which are different from the originally estimated document forms: layout structures may be transformed year by year; document forms may be irregularly altered; documents are not always scanned under the same condition; and so on. In this case, it is important to validate the recognized results perfectly and isolate erroneous objects from correctly recognized results. Even if more excellent methods were developed, they can not make the recognition ratio 100% and the ratio of recognition errors 0%. The validation process is useful to recover such a recognition limit.

Acknowledgements

We are very grateful to Prof. N.Sugie of Meijyo University, and Prof. Y.Inagaki and Prof. J.Toriwaki of Nagoya University for their perspective remarks, and also wish to thank our research members for their many discussions and cooperations.

References

[1] T.Watanabe, Q.Luo and N.Sugie: "Knowledge for Understanding Table-form Documents", *IEICE trans.on Inf. and Syst.*, Vol.E77-D, No.7, pp.761-769 (1994).

[2] T.Watanabe and T.Fukumura: "An Architectural Framework in Document Image Understanding", *Proc.of ICARCV'94*, pp.1431-1435.

[3] Q.Luo, T.Watanabe, Y.Yoshida ,and Y.Inagaki: "Recognition of Document Structure on the Basis of Spatial and Geometric Relationships Between Document Items", *Proc.of MVA'90*, pp.461-464.

[4] T.Watanabe, Q.Luo and N.Sugie: "A Cooperative Document Understanding Method Among Multiple Recognition Procedures", *Proc.of 11th ICPR*, pp.689-692 (1992).

[5] T.Watanabe, H.Naruse, Q.Luo and N.Sugie: "Structure Analysis of Table-form Documents on the Basis of the Recognition of Vertical and Horizontal Line Segments", *Proc.of 1st ICDAR*, pp.638-646 (1991).

[6] T.Watanabe, Q.Luo and N.Sugie: "Structure Recognition Method for Various Types of Documents", *Int'l J.of MVA*, Vol.6, pp.163-176 (1993).

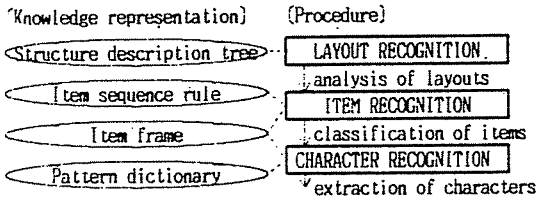
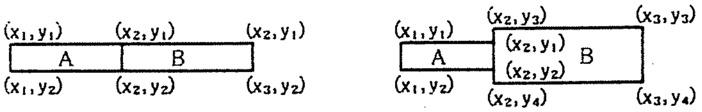


Fig.1 Framework of document image understanding



physical representation:
$(((x_1,y_1),(x_2,y_2)),((x_2,y_1),(x_3,y_2)))$
logical representation:
neighbor(A,B)

physical representation:
$(((x_1,y_1),(x_2,y_2)),((x_2,y_3),(x_3,y_4)))$
logical representation:
neighbor(A,B)

Fig.2 Logical and physical representations

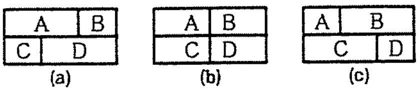
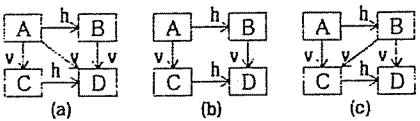
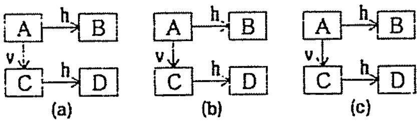


Fig.3 Fragments of table-form document



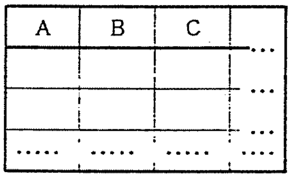
[note] A, B, C, D: item fields,
v: vertical adjacent relationship,
h: horizontal adjacent relationship

Fig.4 Representation by commonly shared line segments



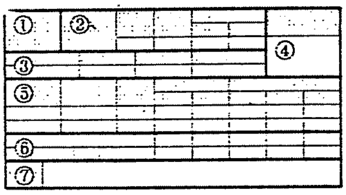
[note] A, B, C, D: item fields,
v: vertical adjacent relationship,
h: horizontal adjacent relationship

Fig.5 Representation by left-upper corners

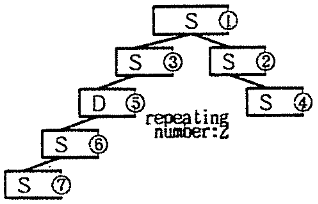


dom(A) = { ...... }
dom(B) = { ...... }
dom(C) = { ...... }
relation { C=A*B }
relation { C←A }
relation { .... }

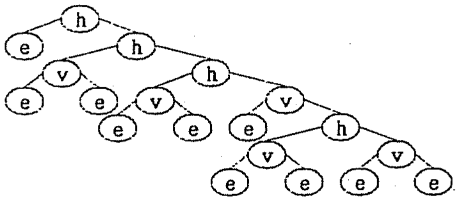
Fig.6 Specification of semantic information



(a) globally analyzed layout structure



(b) global structure tree



(c) local structure tree for block ②

Fig.7 Structure description tree

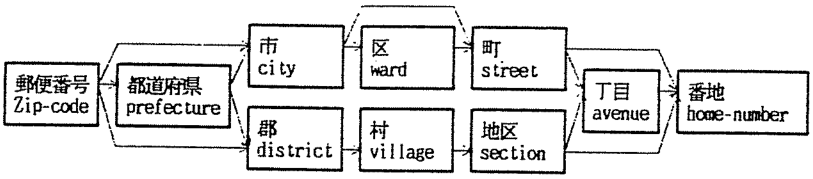


Fig.8 Item sequence rule

NO	NAME	City	Zip-code
1	SEPARATOR	"市"(after)	"〒"(before)
2	LENGTH	max(10)	(3,0) or (3,2) by "-"
3	DATA TYPE	character	numeric
4	OCCURRENCES	single	single
5	CANDIDATES	----	----
6	LOCATION	before("都道府県", "郵便番号")	top
	DECISION CONDITION	1 and 2 and 3	2 and 3

Fig.9 Item frame

	車　賃		鉄　道　賃		
	定額	実費	路程	運賃	その他
	A1	B1	C1	D1	
	A2	B2	C2	D2	
合　計	A	B	C	D	
請求額	／				

Fig.10 Fragment of table-form document

(note)
車賃: taxi/bus fare,
実費: paid money,
路程: distance,
運賃: paid money,
合計: total amount
鉄道賃: train fare,
定額 : regular account,
その他: others,
請求額: requested money,

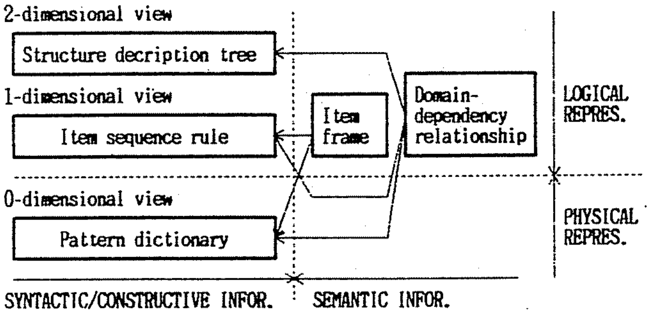


Fig.11 Interrelation between various kinds of knowledge

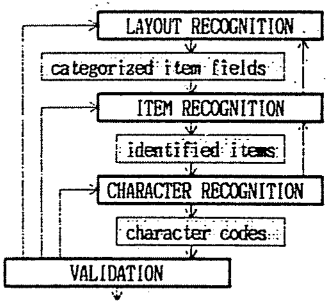


Fig.12 Recognition and validation