Model Matching Based on Association Graph for Form Image Understanding

Yasuto Ishitani
Research and Development Center, Toshiba Corporation
1, Komukai Toshiba-cho, Saiwai-ku, Kawasaki 210, Japan

Abstract

A new method of image understanding for forms based on model matching is proposed in this paper as the basis of OCR which can read a variety of forms. The outline of this method is described as follows. Ruled lines are extracted from the input image of a form. These lines are used for understanding the form, taking into account their feature attributes and the relationships between them. Each line in the input image of a form is expected to correspond to a line in one of the model forms, which are described as structured features. This correspondence is represented by a node in an association graph where an arc represents compatible correspondences established on the basis of feature relationships. The best match is found as the largest maximal clique in the association graph. Experimental results show the method is robust and effective for poor quality document images and also for various styles of forms.

1 Introduction

Form documents comprising several horizontal and vertical lines play an important role in many businesses and governmental organizations. However, the cost of manual transactions with forms is usually very high. Automation is greatly desired in this area, since large amounts of documentation must be converted into computer-readable data.

In form transactions, not all information on the forms needs to be captured. However, a critical ability of a practical form reader is to obtain knowledge from certain portions with certainty. That is, to realize a practical and usable form processing system, several types of knowledge about filled blanks (or sometimes unfilled blanks), the relationships between them, and their reading order is required.

The principal issue in this kind of form image understanding is to find the best match between the processed document image and prepared models. Several attempts to develop form image understanding methods have been reported. These can be classified into two categories, as follows.

1. Methods which project physical geometric information defined in model knowledge onto an input document image and extract data directly from the results of this projection [1] [2] [3]
2. Methods which extract logical information about document layout structure independently of physical geometric information [4]

One of the problems with the first method is that documents cannot be analyzed if the input image suffers from variations such as translations, rotations, or scale changes. The second method, however, has problems if the document images are very noisy. The reason for these defects is that neither method has the qualities of both flexible matching and robust matching.

In this paper, the author proposes a new form image understanding method based on model matching which resolves these problems and realizes OCR capable of reading various types of forms. The goal is to create a form processing system which allows input of greater flexibility in terms of quality, styles, and number of blanks. The author makes several assumptions in achieving this goal.

- Form styles are known in advance.
- Document images suffer from two kinds of variation: spatial variations such as translations, rotations (skewing), and scale changes, and distortions due to noise.
- The positions of form contents may change due to revisions of document style.
- Poor document images must be readable.
- Various form types are mixed into a batch, meaning that the system must be able to identify the type of form.

The proposed method is able to identify the form type and extract data from each filled or unfilled blank from an input image according to these assumptions and using model knowledge. The model matching algorithm constructs graphs and then performs a graph search in order to find the best match between the model knowledge and a form image.

In the next section, details of the method will be discussed. Following that, I will present experimental results that demonstrate the effectiveness of the proposed method.

2 Overview of the proposed method

Since ruled lines are a major component of form documents, it is essential to compare lines in comparing two form document images. If one of the images is

0-8186-7128-9/95 $4.00

a model of a standard format and the other an input image, this comparison becomes a process called "form understanding" (Fig.1). Matching between models and an input image is the main issue discussed in this paper. In the proposed method, the model knowledge consists of various kinds of form models from which model matching procedure selects the most suitable model for an input image.

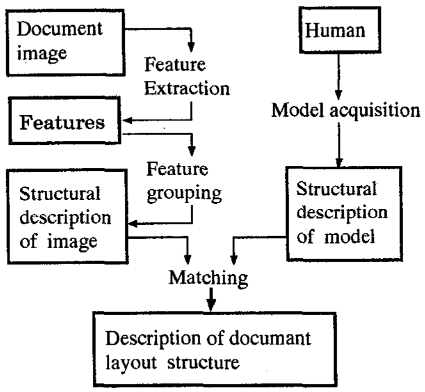


Figure 1: Framework of form image understanding

If a form document contains isolated sub-forms, as shown Fig.2, lines can be grouped in each sub-form. In this case, a document is considered to consist of several sub-forms containing several lines. As a result, straight lines, sub-forms made up of some of them, and their relationships are adopted as the basic primitives for descriptions of both images and models, and these are described structurally.

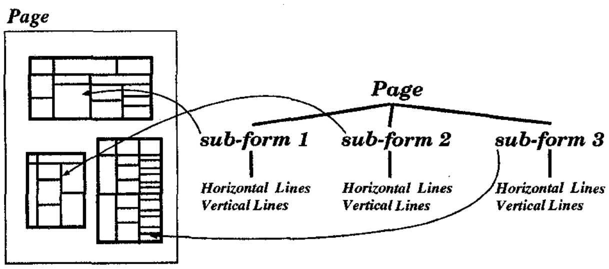


Figure 2: Example of isolated sub-forms in a form

This leads to the conclusion that structural matching is essential to understanding of form documents rather than simple matching techniques. However, if a procedure is complex and heavily heuristic, it may be difficult to develop and may suffer from problems of processing speed. From a practical viewpoint, a theoretical approach is preferable if it offers good matching performance.

The model matching problem is translated into a problem of searching for the optimal match between two structural descriptions: one is a model and the other the input image. In the proposed method, an association graph constructed from the two structural descriptions is used to search for maximal cliques and thus find matching subgraphs. Sub-isomorphisms in the two structural descriptions can be efficiently detected by maximal clique search algorithm in an association graph. This approach is able to achieve inexact matching, since the criterion for sub-isomorphisms in the two structural descriptions is less strict.

A hierarchical matching strategy based on the association graphs is adopted in the proposed method so as to realize higher and more global matching stages as well as lower and more local matching stages and the interactions between them. A uniform approach at every stage ensures that the interactions between high and low stages are consistently realized. Clearly, this leads to overall simplification of the system. In this section, the basic concept of the proposed method is described.

2.1 Outline of form image understanding

Here, an outline of form image understanding based on model matching is given. Figure 3 shows a flow chart of the whole procedure. Since the input image and each individual model are described structurally as explained above, a hierarchical matching strategy is also adopted as the model matching algorithm.

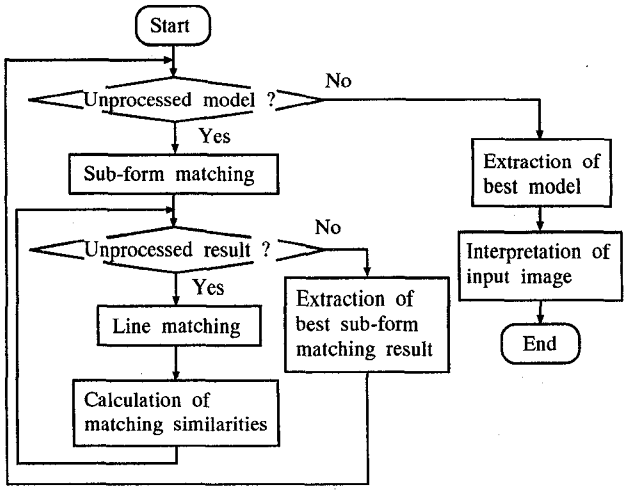


Figure 3: Flow chart of form image understanding based on model matching

One matching step in this hierarchical matching process consists of matching two sets of sub-form features between the input image and each model. Such a matching step is the highest part of the matching process and is called "sub-form matching." If sub-form matching terminates successfully, matching shifts to a lower level where two sets of line features are matched in the corresponding sub-forms obtained at the highest step. This matching step is called "line matching." The interaction between sub-form matching and line matching is realized as follows. If sub-form matching outputs several exclusive results, line matching is invoked for each result of sub-form matching. The best sub-form match is determined according to the best line match. Both sub-form matching and line matching consist of maximal clique finding. A reduction in the search space is obtained by working hierarchically.

The proposed method is flexible because global variations and local variations are handled by sub-form matching and line matching, respectively. The global variations consist of the following variations.

- Extra sub-forms due to revisions of document format (for example, 1' and 5' in Fig.4 (a) are extra sub-forms)
- Differences of sub-form arrangement (e.g. 1 and 1' in Fig.4 (a))
- Scale changes between a set of sub-forms in the input image and a set of sub-forms in a model which may be independent (Fig.4 (a))

The local variations are variations such as scale changes of line length and differences of line arrangement due to revisions of document format (Fig.4 (b)).

If sub-form matching and line matching terminate successfully, matching similarities are calculated to obtain the best match in which matching similarity is maximized from various form models. As a result, the most suitable model for an input image is detected.

After the model matching process, information is extracted from filled blanks in the input image by referring model knowledge against corresponding portions of the input image. This knowledge includes properties of information, relationships between blanks, etc. The processes of data extraction and character segmentation / recognition are carried out with this knowledge. The processes described above are collectively known as "form image understanding." The model matching process is detailed in Section 4.

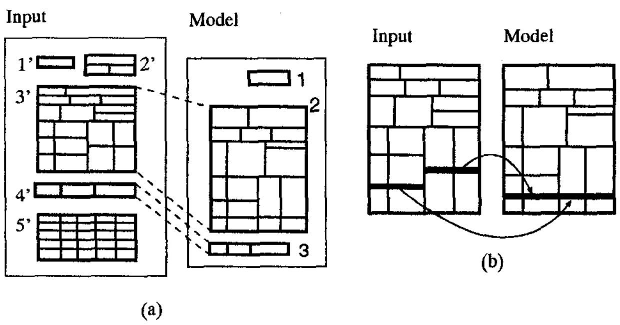


Figure 4: Examples of variations of input image

2.2 Graph theoretical matching

The most important issue in the model matching procedure described so far is how to find correspondences between two structural descriptions: one from the input image and the other a form model. The concept of an association graph **G** and a search algorithm for this graph is introduced in this section. The association graph [5] [6] is built by letting a node be a pair of features in the input image and in the model, and connecting two nodes if they are compatible; that is, the two pairs may exist simultaneously. By this definition, the best match between the input form image and a model is equivalent to the largest set of nodes that are all mutually compatible. In the association graph field, a fully connected set of nodes is called a "clique." A maximal clique is a clique that cannot be extended to include other nodes in the graph. The largest maximal clique is the maximal clique containing the largest number of nodes. Thus, the best match is determined by the largest maximal clique in the association graph **G**.

In the past, feature grouping techniques and searches for maximal cliques have been used in certain computer vision system [6] [7] [8]. These concepts are useful in the document understanding field if a maximal clique graph search is applied to document image matching.

A matching procedure based on a graph search, such as in this method, has the characteristics listed below.

- A pure graph theoretical algorithm such as clique finding is applied to finding the best match between two structural descriptions. Consequently, the matching algorithm can be simplified if it consists of a hierarchical matching strategy and has multiple matching stages.
- Ambiguities such as several exclusive results may remain unresolved at a higher stage of the matching process, where only one candidate is unsuccessfully sought as a solution. The lower part of the matching process must then resolve these ambiguities. The graph theoretical approach is suitable for the treatment of these ambiguities.

In particular, the association graph approach has certain advantages compared with other graph theoretical approaches.

- It is possible to realize inexact matching because a less strict criterion is required in sub-isomorphism between two structural descriptions.
- Information depending on domain is used in node building and node compatibility examination. This means the association graph is useful for other document image understanding applications.

3 Feature extraction and grouping

This section explains the structural description of a document image used in the model matching procedure. Before feature extraction, document skew is detected and corrected [9]. Accurate lines can be detected by this treatment of the input image. First, horizontal and vertical lines are extracted from the document image as candidates for ruled lines in the form. Secondly, these lines are grouped and recognized as data corresponding to a sub-form in the form. The relationships between sub-forms and lines in a form image are described in terms of a tree structure for the matching procedure [10]. An example of this is shown in Fig.2.

4 Matching procedure

In this section, the model matching algorithm is explained in detail by describing an actual implementation.

4.1 Sub-form matching

We consider the matching of a set of sub-forms **IT** in an input structural description with the set of sub-forms **MT** in a model structural description. **IT** is defined as $\mathbf{IT} = \{it_k \mid k = 1, 2, ..., it\_num\}$ and **MT** is defined as $\mathbf{MT} = \{mt_j \mid j = 1, 2, ..., mt\_num\}$, where it_k and mt_j are elements; that is, they are sub-forms. Figure 5 shows a flow chart of sub-form matching.

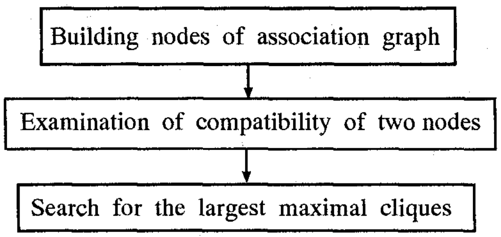


Figure 5: Flow chart of graph theoretical matching

First, a certain number of pairs which consist of two elements of **IT** and **MT** is obtained. These pairs are replaced by nodes on the association graph $\mathbf{G_T}$ (Fig.6). The nodes of $\mathbf{G_T}$ are obtained by looking for similarities between the two sub-forms it_k and mt_j. The number of intersections in each sub-form is used to calculate these similarities, because this approach is robust and insensitive to scale changes, differences of sub-form arrangement, and distortions due to noise. The similarity S_T between mt_j and it_k is defined as the equation, $S_T = \alpha_T - \beta_T \times \frac{|mj-ij|}{mj+ij}$, where α_T and β_T are constant, mj is the number of intersections of mt_j, and ij is the number of intersections of it_k. If S_T is greater than the threshold TH_T, a pair (mt_j, it_k) is defined as a node of $\mathbf{G_T}$. If an actual sub-form is broken into several pieces due to poor image quality, these pieces are merged into a virtual sub-form and correspond to one model sub-form.

The compatibility of every combination of two pairs is examined. (mt_j, it_k) and (mt'_j, it'_k) are two nodes in $\mathbf{G_T}$. Let R be the relation between mt_j and it_k and R' be the relation between mt'_j and it'_k. If the relation R is compatible with the relation R', (mt_j, it_k) and (mt'_j, it'_k) are compatible. Whenever two such pairs are compatible, an arc is set between the two corresponding nodes in the association graph. Sub-form matching is achieved by finding all the largest maximal cliques in the association graph $\mathbf{G_T}$. In Fig.6, the set whose members are $(1,1')$, $(2,3')$, $(3,4')$ and the set whose members are $(1,2')$, $(2,3')$, $(3,4')$ are the largest maximal cliques. If there is more than one largest maximal clique, the line matching process described below is invoked for all results of sub-form matching. In this case, the best match is determined by finding the best match of the lines.

After sub-form matching, parameters of affine transformation in each correspondence between **IT** and **MT** are detected [10]. The line matching invoked after this sub-form matching is carried out in an environment such that a set of lines in it_k and a set of lines in mt_j are compared in the same coordinate system.

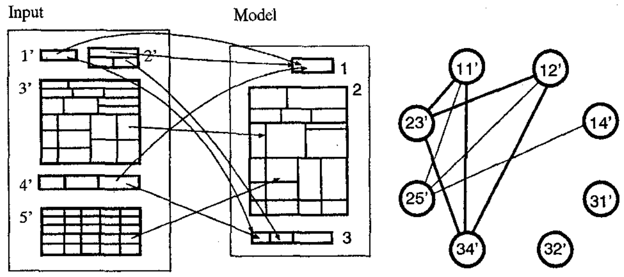


Figure 6: Sub-form matching based on association graph

4.2 Line matching

This subsection introduces the algorithm used for line matching. In the line matching procedure, horizontal lines and vertical lines are handled independently, as shown in Fig.7. Both matchings are based on the scheme shown in Fig.5.

In the following, **IVL** denotes a set of vertical lines in **IT** and **MVL** denotes a set of vertical lines in **MT**. **IVL** is described as $\mathbf{IVL} = \{ivl_g \mid g = 1, 2, ..., ivl\_num\}$ and **MVL** is described as $\mathbf{MVL} = \{mvl_f \mid f = 1, 2, ..., mvl\_num\}$.

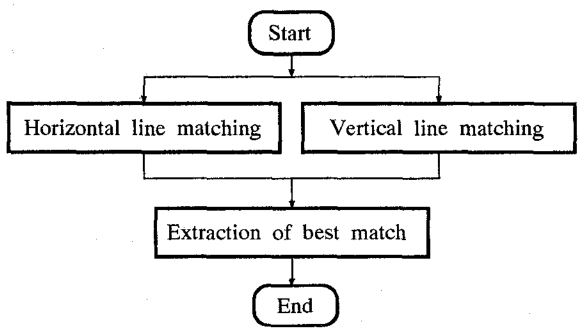


Figure 7: Flow chart of line matching

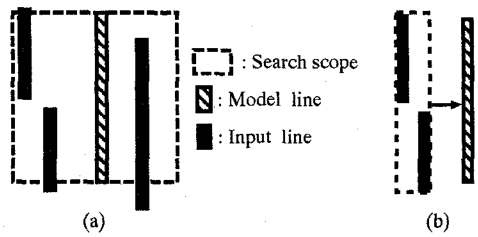


Figure 8: Search scope for line matching

A region known as a "search scope" is set up around mvl_f, as shown in Fig.8 (a), to find the lines ivl_g which may correspond to mvl_f, where $ivl_g \in \mathbf{IVL}$ and $mvl_g \in \mathbf{MVL}$. The corresponding lines ivl_g cross over or are enclosed within the search scope of mvl_f. The local variations such as the scale changes of line length and the differences of line arrangement between mvl_f and ivl_g are tolerated within the search scope of

mvl_f. This method also deals not only with one–to–one mapping but also one–to–many mapping or many–to–many mapping to resolve the differences of line arrangement as shown in Fig.9. If an actual line in **IVL** is broken into several pieces, all the broken pieces are merged into one virtual line and this corresponds to one line in **MVL**, as shown in Fig.8 (b). The nodes of the association graph $\mathbf{G_L}$ are created by referring to the similarity between the line in **IVL** and the line in **MVL** (Fig.9). This similarity S_L is defined by the equation, $S_L = \alpha_L - \beta_L \times \frac{|mlh-ilh|}{mlh+ilh}$, where α_L and β_L are constant, mlh is the length of mvl_f, and ilh is the length of ivl_g. If S_L is greater than the threshold TH_L, a pair (mvl_f, ivl_g) is defined as a node of $\mathbf{G_L}$.

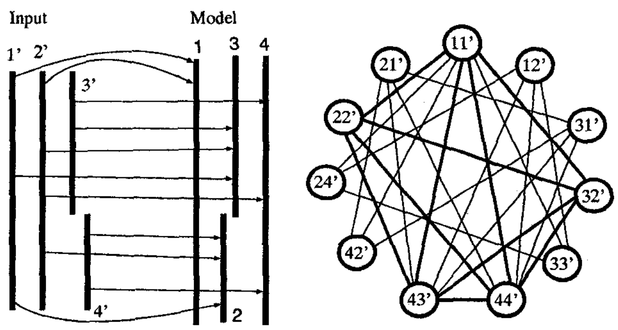


Figure 9: Line matching based on association graph

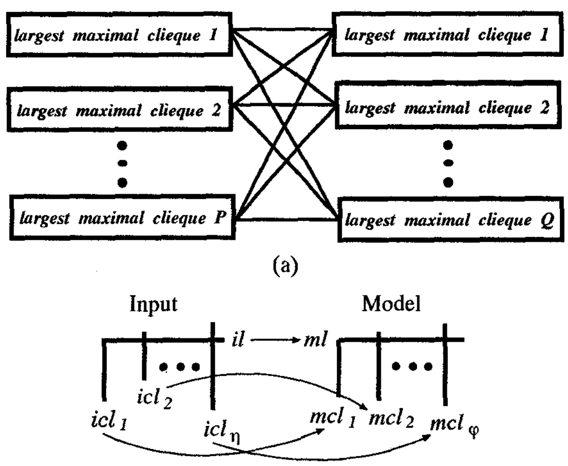


Figure 10: The best match for line matching

The compatibility of every pair of nodes is examined. As a result, arcs are created on the association graph $\mathbf{G_L}$ and the largest maximal cliques of $\mathbf{G_L}$ are searched for. In Fig.9, the set whose members are $(1, 1')$, $(2, 2')$, $(3, 2')$, $(4, 3')$, $(4, 4')$ is the largest maximal clique.

If there is more than one largest maximal clique in horizontal line matching and vertical line matching, respectively, the best combination among them is extracted as follows. First, every combination is generated from the two kinds of line matching results, as shown in Fig.10 (a), where the line crossing relationships in both the input description and the model description are extracted. Let an input line il correspond to a model line ml. Then, let **ICL** denote the set of input lines which cross il, where **MCL** denotes the set of lines which cross ml. If each element of **ICL** corresponds to an element in **MCL**, as shown in Fig.10 (b), il is considered as having a correct line crossing relationship. Each input line in each combination is examined for this correct relationship. The combination with the most correct line crossing relationships among all combinations is taken as the best line matching result.

4.3 Calculation of matching similarity

After sub-form matching and line matching, the matching similarity between the input image and each model is calculated to find the best match among all the models. This similarity S_M is defined by the evaluation function, $S_M = \alpha_M - \beta_M \times \frac{|A-A'|+|B-B'|}{A+A'+B+B'}$, where A is the number of horizontal lines in a model, A' is the number of horizontal input lines corresponding to horizontal lines in the model, B is the number of vertical lines in a model, and B' is the number of vertical input lines corresponding to vertical lines in the model. The similarity S_{max} denotes the maximum value of the evaluation function. If this S_{max} is greater than the threshold TH_M, the model which corresponds to S_{max} is chosen as the model offering the best match.

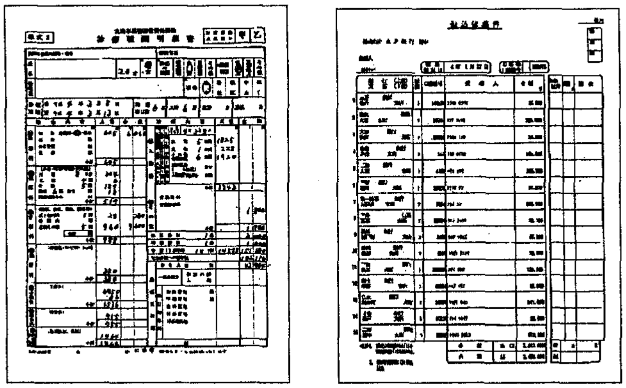
If this sub-form matching finds several exclusive results, the matching similarity is calculated for each result and the result which maximizes S is considered the best match.

5 Experimental results

The proposed method was implemented on a SPARCstation2. The algorithm was written in the C language. Experiments were carried out with 14 categories of forms. Some of the test documents are displayed in Fig.11. All documents, which were sized within the B4 format, were digitized at 400 dpi and converted into binary images. The model knowledge is defined for each category of forms. Multiple input images of different quality were generated for each category by changing the threshold for the binarization process. As a result, a total of more than 1000 input images were used.

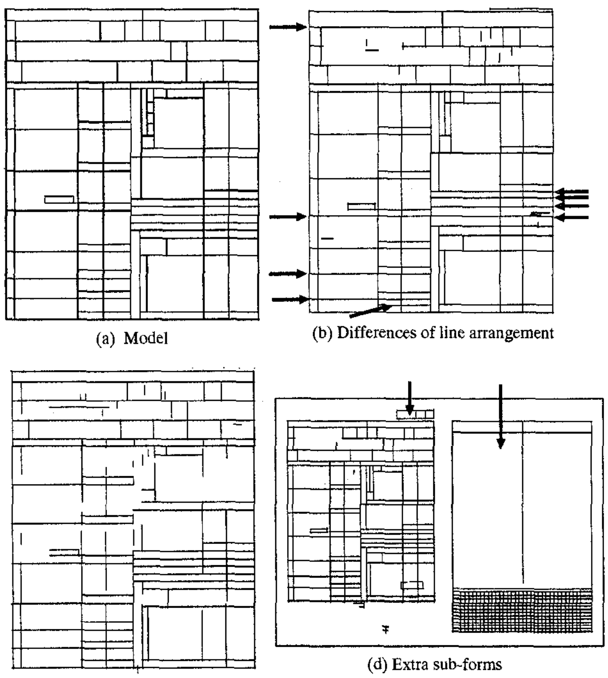
The rate of correct interpretation was on the average 98% for all test documents. The remaining two percent of the test documents correspond to failures rejected due to a low value of matching similarity attributable to poor image quality. Figure 12 (a) is the model of the form shown in Fig.11 (a). Figure 12 (b), (c), and (d) are results obtained by the feature extraction process from the input images corresponding to the model shown in Fig.12 (a) and present some variations such as the differences of line arrangement, defects in line quality, and extra sub-forms. In the matching process, these variations could be resolved.

The average processing time for all test documents was 5.76 CPU seconds. Most of this processing time was taken up by the feature extraction process. In fact, the matching time was only 0.54 CPU seconds. Experimental results show that this method is practical in terms of both accuracy and processing speed.



(a) (b)

Figure 11: Examples of test samples



(a) Model (b) Differences of line arrangement (c) Defects in line quality (d) Extra sub-forms

Figure 12: Examples of experimental results

6 Conclusion

A form image understanding method based on a model matching approach has been described as part of an OCR system which can read a variety of forms. This method can determine the form type and extract filled or unfilled blanks from an input image automatically, using models which define each category of form.

A hierarchical matching strategy based on subgraph matching is adopted in the proposed method. This consists of global matching stages (sub-form matching), local matching stages (line matching), and the interactions between them. The global variations and the local variations in the input image are treated by sub-form matching and line matching, respectively. Therefore, it is possible to obtain flexible and efficient matching between input images and form models. This approach can also achieve inexact matching in the face of distortions due to noise because it has a less strict criterion for sub-isomorphisms in two hierarchical relationships.

Experiments on a variety of documents have shown that the method is robust and effective even where document image quality is poor and where there are various styles of forms.

References

[1] S. L. Taylor, R. Fritzson, and J. A. Pastor, *"Extraction of data from preprinted forms,"* Machine Vision and Applications, vol. 5, no. 3, pp. 211-222, (1992).

[2] R. Casey, D. Ferguson, K. Mohiuddin, and E. Walach, *"Intelligent forms processing system,"* Machine Vision and Applications, vol. 5, no. 3, pp. 143-155, (1992).

[3] S. W. Lam, L. Javanbakht, and S. N. Srihari, *"Anatomy of a form reader,"* Proc. 2nd ICDAR, pp.506-509, (1993).

[4] T. Watanabe, Q. Luo, and N. Sugie, *"Structure recognition methods for various types of documents,"* Machine Vision and Applications, vol. 6, no. 3, pp. 163-176, (1993).

[5] D. H. Ballard and C. M. Brown, *"Computer Vision,"* Prentice-Hall, (1982).

[6] A. P. Ambler, H. G. Barrow, C. M. Brown, R. M. Burstall, and R. J. Popplestone, *"A versatile system for computer-controlled assembly,"* Artificial Intelligence, vol. 6, no. 2, pp. 129-156, (1975).

[7] R. Horaud and T. Skordas, *"Stereo correspondence through feature grouping and maximal cliques,"* IEEE Trans. Pattern Analysis and Machine Intelligence vol.11, no.11, pp.1168-1180, (1989).

[8] H. Ogawa, *"Labelled point pattern matching by Delaunay triangulation and maximal cliques"*, Pattern Recognition, vol.19, pp.35-40, (1986).

[9] Y. Ishitani, *"Document skew detection based on local region complexity,"* Proc. 2nd ICDAR, pp.49-52, (1993).

[10] Y. Ishitani, *"A model matching method for form image understanding,"* Technical Report of Japanese Inst. of Electron., Inf., and Comm. Eng., PRU94-34, pp.57-64, (1994).