LearnPos: a new tool for interactive learning positioning

Cérès Carton, Aurélie Lemaitre, Bertrand Coüasnon

LearnPos: a new tool for interactive learning positioning

Cérès Carton[a], Aurélie Lemaitre[b] and Bertrand Coüasnon[a]

[a]INSA - Irisa - UEB, Campus de Beaulieu, 35043 Rennes, France;
[b]Université Rennes 2 - Irisa - UEB, Campus de Beaulieu, 35043 Rennes, France

ABSTRACT

The analysis of 2D structured documents often requires localizing data inside of a document during the recognition process. In this paper we present LearnPos a new generic tool, independent of any document recognition system. LearnPos models and evaluates positioning from a learning set of documents. Thanks to LearnPos, the user is helped to define the physical structure of the document. He then can concentrate his efforts on the definition of the logical structure of the documents. LearnPos is able to furnish spatial information for both absolute and relative spatial relations, in interaction with the user. Our method can handle spatial relations compose of distinct zones and is able to furnish appropriate order and point of view to minimize errors. We prove that resulting models can be successfully used for structured document recognition, while reducing the manual exploration of the data set of documents.

Keywords: position operator, document structure recognition,user interaction, position learning

1. INTRODUCTION

Structured document analysis refers to the definition of both logical and physical layout structure. For structured document analysis, it is of major interest to model and evaluate the positioning of the components between them and within the page. Structured documents contain 2D information that requires 2D zones to locate the different components to analyze. The positioning of the components is used during the structure analysis (layout and organization) to define orientation and order of the analysis. The quality of this spatial information affects the performance of the analysis.

The 2D zones which are necessary to guide the analysis are difficult to define. Nowadays, they are often manually defined. The manual definition of 2D zones presents several drawbacks. First, it is time-consuming. A human operator has to observe document images to define an appropriate zone from these examples. Moreover, the rare cases are not observed and analysis of the errors must be done to adjust the obtained zones. The tuning of the chosen parameters is time-consuming moreover it is not an easy task.

In this paper, we present LearnPos, a new tool for the automatic determination of position operators. Using LearnPos, the user can obtain the physical structure definition without the need to manually explore the documents data set. LearnPos proceeds by an analysis of a learning data set. LearnPos allows an exhaustive and interactive study of the corpus of documents. The rare cases are easily detected, opposite to what is possible with a manual analysis. Moreover, time needed to determine a position operator is decreased.

In section 2, we present some related work on spatial relation modeling and the limitations that motivated our research. In section 3, we present how LearnPos proceeds to compute some position operators. In section 4, we present a new indicator, the confusion indicator, which allows us to make the best of the defined zones by ordering them and choosing the right point of view to minimize error. In section 5, we evaluate our new tool on the publicly available RIMES data set of handwritten business letters. Our evaluation on 1250 handwritten letters shows that our automatically defined position operators have comparable performances with manually tuned position operators, while significantly decreasing the number of manual parameters, from 102 to 66. Time spends on this task is also decreased: both general cases and rare cases are easily detected thanks to the exhaustive analysis of the corpus.

Further author information: (Send correspondence to Cérès Carton)
Cérès Carton: E-mail: ceres.carton@irisa.fr
Aurélie Lemaitre: E-mail: aurelie.lemaitre@irisa.fr
Bertrand Coüasnon: E-mail: couasnon@irisa.fr

2. RELATED WORK

In document structure analysis, we can consider two main approaches: statistical ones and syntactical ones. Statistical methods allow learning of the characteristics of each type of element but lack the ability to convey the hierarchical structure of a document. Syntactic methods segment the image in primitives and build a rule tree that describes how to compose these primitives. The definition of position operator is a part of the physical structure definition. For example, syntactic methods require defining position operators for the defined rules. In general, the parameters of the position operators are manually defined. This manual definition is not exhaustive and particularly tedious.

Spatial relations management refers to the appropriate means to express relations between the document objects and guarantee their consistency. This subject has been studied in handwritten recognition, document structure recognition and more generally in image analysis. Performance of a method depends on its expressiveness degree in the representation of the objects and on its positioning precision.

2.1 Representation of the objects

Representation of the objects depends on the method and the subject. In image analysis, objects are often reduced to representative points (centroid methods, bounding box based methods, etc.). As demonstrated in Bloch and Ralescu,[1] this representation of the objects is not adapted in the case of handwritten recognition. Handwritten recognition deals with complex objects, where rich representation and modeling of spatial information are required. For example, in Asian character recognition, you must be able to handle correctly objects with concavities.

2.2 Positioning approaches

In general, two spatial positioning approaches can be distinguished: the absolute positioning and the relative positioning.[2] For modeling relative positioning, various methods have been proposed, with different degrees of expressiveness.

A set of methods are the qualitative ones. In the case of these methods, the authors define a priori domain-dependent areas and have to manually set associated thresholds. An example of use of qualitative method for document structure recognition is introduced by Conway.[3] In this syntactical approach, a set of page relation is defined: above, leftof, over, leftside and closeto. The spatial relations that can be expressed are of two types: directional and topological. In these approaches, the objects are represented by region such as rectangles, circles and irregular shapes.

Directional relations describe the order of objects in a definite space (e.g. north, south). The Papadias and al.[4] model is composed for example of 9 relations considered sufficient to express any directional relations between two regions, where regions are modeled as rectangles. Topological relations describe concepts of neighborhood, incidence and overlap and stay invariant under transformations such as scaling and rotation. In the model proposed by M.J. Egenhofer,[5] 8 topological relations between two region objects have been defined: disjoint, touch, overlap, cover, covered by, contain, inside and equal. This approach is applied to arbitrary spatial objects. In general, qualitative approaches lack of precision. For instance, directional relations do not indicate the distance between the objects. Introduction of flexible distance to enhance the expressivity and precision degrees of the model has been proposed by Mardej et al.[6] To do so, they introduce two concepts: *at nearest* and *at farther* between rectangular regions.

Methods based on CRF (*Conditional Random Fields*) are also known to be able to analyze and model the document layout. They are used to model the spatial inter-dependencies of the different regions in documents. However, the spatial inter-dependencies modeled with a CRF are limited in a small portion of the space. For example, Shetty et al.[7] model spatial inter-dependencies between neighboring patches, where a patch is approximately the size of a word. In document structure recognition, we desire to model spatial relations between blocks of text. Modeling at word level is too limited for this task.

Directional relative position aspects have been treated by several methods in different way. In the DMOS-P method,[8] point of view is manually defined by the user through the cursor parameter. This approach gives good result but needs a human operator intervention. In Bloch et al.,[1] directional fuzzy landscapes are used. In this

method, a fuzzy set is computed that assigns to any point of the plane a membership degree describing how well the point satisfies the considered spatial relation. Delay et al.[9] develop the idea presented by Bloch. They introduce a new general definition of positioning model that defines some variability in the membership degrees admitted in each directional relation. These approaches using fuzzy landscape are particularly adapted to model complex relationships, for example in handwriting recognition. However, this precision degree is not adapted for document structure recognition where the ground truth generally contains the bounding box of each element.

In this paper, we introduce an automatic learning based on learning database of position operators. In this method, objects are represented by their bounding boxes. Indeed, we want to be able to define a zone of interest where we can look for an element. Boundaries of the zone must not be used to identify the class of an element. Characteristics and rules will permit to determine which elements contained in this zone are correct and which are not. The purpose of our positioning method is to offer a good recall, precision will be after that determined by rules and conditions quality. Our tool produces a position operator similar to the ones that are produced by a human operator. One major advantage of this approach is that the user can then modify himself the position operator to introduce a priori knowledge if he desires it. Directional aspect are introduced in our analysis through the automatic computation of the best direction (referred as point of view in this article) to parse a zone.

3. INTERACTIVE LEARNING OF POSITIONING

In this paper, we consider a new tool, LearnPos, able to analyze a document ground truth to learn position operators. To do so, LearnPos needs a learning data set. For example, we have a corpus of handwritten business letters and we want to learn the position of "sender details" in our corpus. In document structure recognition, we want to be able to localize and identify the different components present in a page. In our approach, a component is represented by the smallest rectangular box that contains it (MBR: Minimum Bounding Rectangle). This representation is often used in the existing methods, as we presented it in the previous section. In order to simplify the relation specification and the spatial representation of an object, we take into account only its higher left and lower right angles coordinates, as it is shown in figure 1. As a result, for each document of the learning data set, the MBR of each element is known.

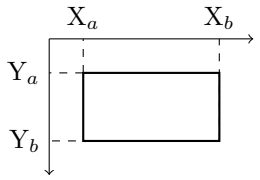


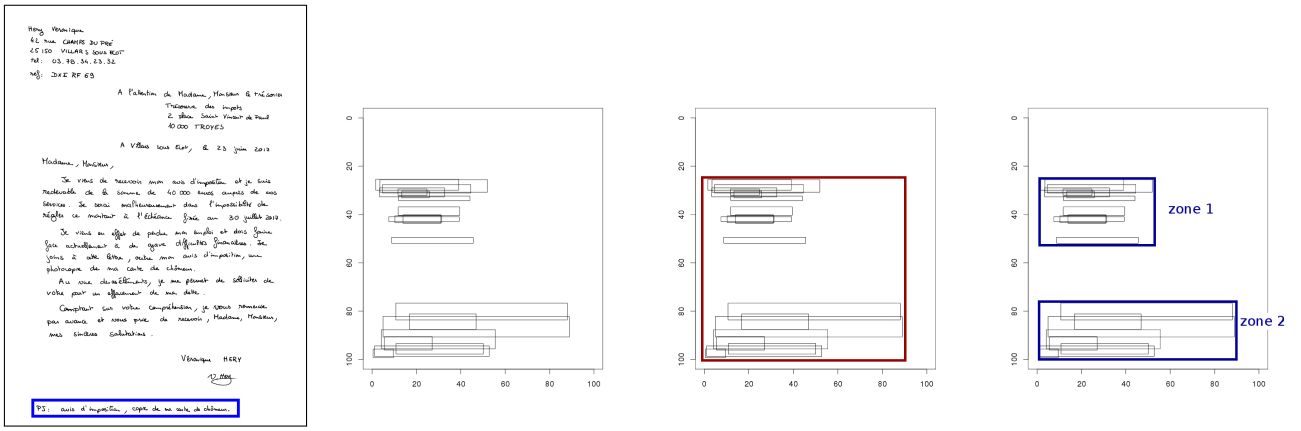
Figure 1. Object spatial representation

Using this ground truth, LearnPos then computes a position operator that user can integrate directly in his code by a simple copy and paste operation. A position operator computed using LearnPos can be composed of different separate zones. For example, in our corpus of handwritten letters, a category "attachment/postscript" (referred as att/ps) exists. Figure 2(b) is a summarized representation of all the occurrences of "att/ps" present in our learning data set. As it can be seen, two distinct groups exist. If we do not take into account these groups to compute the position operator, as it is done in figure 2(c), we nearly obtain the whole page as the zone of interest, which is not useful. On the opposite, when the two groups are taken into account and a zone is produced for each group, as it can be observed in figure 2(d), then position operator is much more precise.

The user defines the logical structure of the document and then can use LearnPos to be assisted in the physical structure definition. When ground truth is available, LearnPos is composed of different steps:

1. The user explicitly asks for the position operator P of a component
2. Computation of P (preceded by group detection and extreme values detection): one or more zone(s) boundaries are produced
3. Determination of the best point of view for each zone
4. Determination of the order of analysis of the zones
5. Modifications possible for the user on the generated position operator

6. The position operator is introduced in the logical structure defined by the user. The user can go back to the first step for another component.



(a) Example of letter containing att/ps

(b) Location of att/ps in page for a corpus of 300 pages

(c) Zones obtained when position operator does not take groups into account

(d) Zones obtained when position operator determines automatically a zone for each group

Figure 2. Example of computation of the "att/ps" position operator, showing the interest of using several zones for on position operator. Each occurrence of "att/ps" MBR is represented as one rectangle in a normalized page.

As it has already been done in the literature, we use two modes of positioning in a document: absolute and relative positioning. An absolute spatial description consists in describing the position of each component with respect to a fixed reference, here the whole page, independently of the other component positions. This approach is well adapted to elements which positions are stable in the page, but may be inaccurate for other elements. Relative positions seem to be much more associated to our perception of spatial arrangement similarity. For example, in our example on handwritten business letters, we will be able to express relations as "date/place is above sender detail". In the relative position approach, one is interested in the spatial relations between components rather than their absolute positions.

In this section, we will first detail the role of the user in our method. Then, we present the difference of treatment for absolute and relative positioning. The different steps of the processing chain are then presented: group detection, detection and treatment of extreme values and finally the density-grid method we use to compute a position operator.

3.1 User role

LearnPos is an interactive tool, it then requires user intervention. To use LearnPos, the user has to furnish two parameters: a file containing the ground truth for all the learning database and the element he wants to locate. The number of parameter is reduced to its minimum because we want our method to be fully automatic.

First, LearnPos requires an adapted ground truth that contains the different zones that we need to locate in the documents. The user must integrate in a single file all these information. A document can contain different elements belonging to the same class. For example, a document can contain two signatures. The treatment of this multiplicity of components of the same type depends on the type of positioning desired by the user. In the case of absolute positioning, no restriction exists. In the case of relative positioning, we need to know for each object which component is used as a reference. Pair association of components can be explicitly done in the ground truth or by the user, who will give a rule to do this pair association in LearnPos.

The user also has to specify which element he wants to locate. As we presented it, the user defined the logical layout of the document and he delegates to LearnPos the physical layout analysis. Our tool is here to give him the best possible position operator but it is the user role to say which element is concerned by this automatic

analysis. Indeed, he has to specify if he wants an absolute or a relative position operator and in the case of relative positioning, he must inform LearnPos which component must be locate in function of an other one.

After the analysis, LearnPos produces a position operator that the user can simply copy and paste in his code for the full document structure recognition. However, LearnPos position operators are written in the same way that manually defined position operators. It allows flexibility, the user can modify the obtained value if he has complementary information for example.

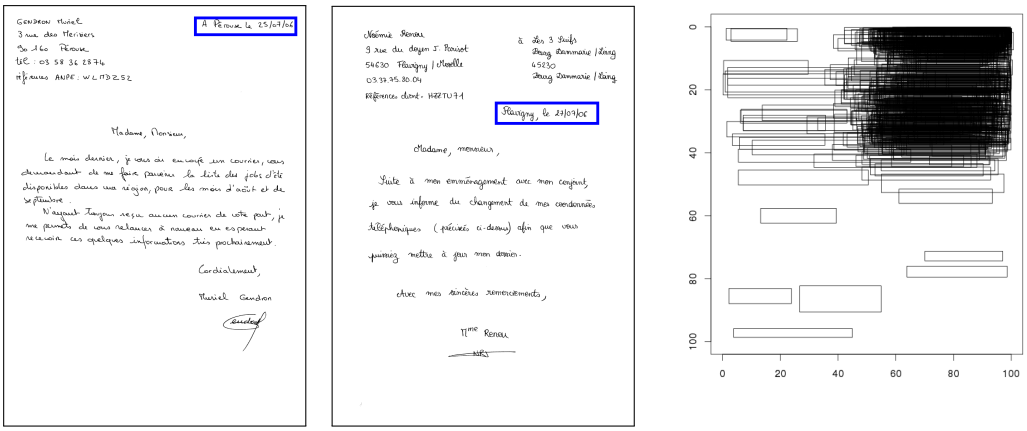
3.2 Absolute and relative positioning

As we introduced, LearnPos is able to compute a position operator for two modes of positioning: absolute and relative positioning. In this section, we detail the difference of treatments of these two modes of positioning.

Absolute positioning

Absolute positioning consists in the description of the position of a component in the whole page, independently of the other components positions. This approach is particularly useful to begin the analysis, when no other component has been found. For example, as it is presented in the figure 3, the user can use absolute positioning to locate "date/place" component in a handwritten business letter.

Absolute positioning does not require a specific pre-processing of the ground truth because the ground truth consists in the position of the MBR in the whole page. It can be used directly. In each document, the component we want to locate can be present or not. Obviously, LearnPos only uses the documents where the component is present to compute the position operator. However, the number of missing data is transmitted to the user, as it can give information about the importance of the rule or modify the layout structure defined by the user.



(a) Example of "date/place" location (b) Example of "date/place" location (c) "date/place" MBR location in a normalized page

Figure 3. Representation of the position of date/place MBR obtained from a learning data set of 300 business letters. Each rectangle represents one location

Relative positioning

Relative positioning consists in the description of the position of a component in function of another one. In this article, we call *relative* the element we want to locate in function of an other component, which is called *reference*.

In the case of relative positioning, the ground truth needs a pre-processing to be able to produce the position operator. First, in document analysis, we do not necessary have one relative for one reference in each document. A document can contain a reference but no relative or the opposite. We can also be in the case where neither relative nor reference is present in the document. As we presented it previously, in the case where several references and relatives are present in a document, matching between reference and relative must be done. It

can be done directly in the ground truth or by the user with a matching criterion (based on distance between reference and relative for example).

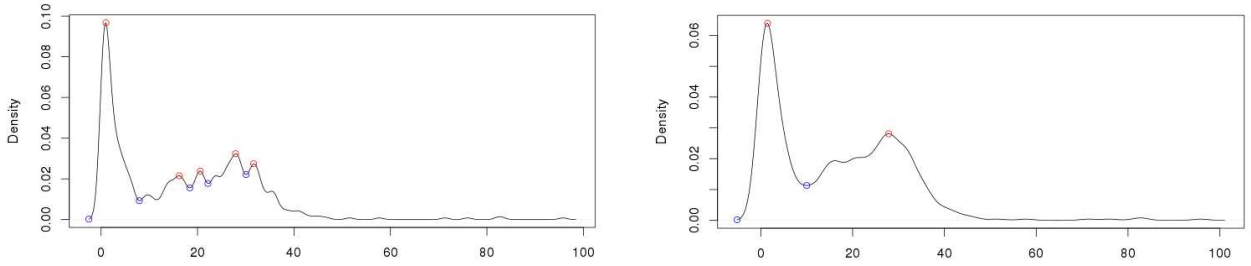
We then need to distinguish all cases that appear in our learning data set. The user is informed of the number of elements concerned by each case in the learning data set. The case where we have no reference but a relative is the simplest. We compute an *absolute positioning* for these type of documents.

The case where we have pairs of reference and relative is the one we are really concerned here. Reference is reduce to its top left point (X_a,Y_a). A translation of the axis system is made. The coordinates of the relative elements are changed so that the origin is set at this point (X_a,Y_a). Using this new coordinates for the MBR, we can proceed as for the absolute positioning.

3.3 Automatic group detection

As we presented it in the figure 2, it is of interest to have position operators that can be composed of several zones. The difficulty here is that we desire a fully-automatic method. We then need to determine automatically if several zones exist or not. Instead of manual predetermination on number of zones of interest, LearnPos automatically finds distinct groups from the histogram. The proposed method is based on local maximum detection of a histogram of positions in page. The total number of local maximum points refers to the number of different zones in the page. Such detection is achieved by a sequential search method, as presented by Lerddaradsamee.[10] In this method, we need to fix a threshold which determines if there is a true difference between the detected value and a current searching value. The value of this parameter will determine the ability of our method to find little groups in term of number of examples present in the learning data set.

Using this technique, the problem of the stability of the found local maximum points has to be treated. To do so, we use different kernel smoothing of the density and keep the local maximum points that are conserved with all the tested smoothing. As reference, we use the Sheather and Jones selector,[11] which is a data-driven bandwidth selector. This technique allows us to detect distinct groups, which require distinct zones in the position operator. The figure 4 presents an example of group detection. In the case of the SJ selected bandwidth (figure 4(a)), multiple local maximum points are detected. However, a stronger smoothing (figure 4(b)) shows that only two of these local maximum points are stable. For our analysis, only these points are kept. We then detect two distinct groups in our data set.



(a) Sheather and Jones selected bandwidth = 0.8762 (b) 2× SJ selected bandwidth = 1.752

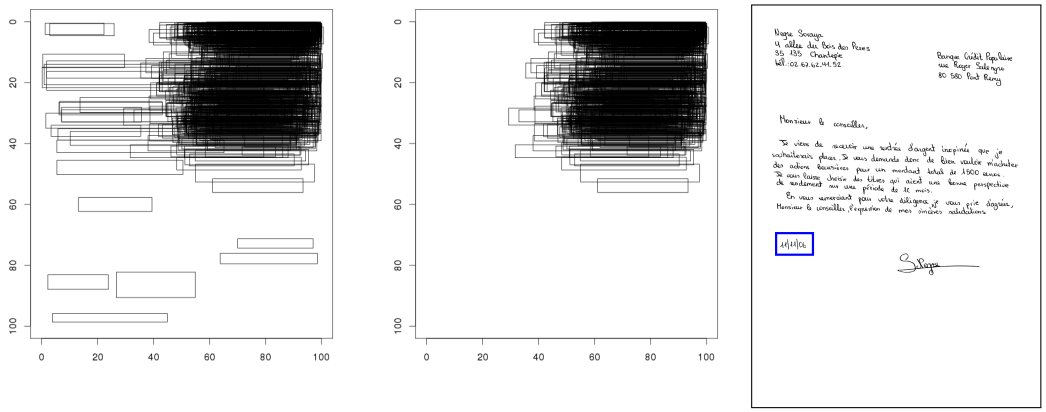
Figure 4. An example of multimodality detection: using conserved local maximum points, two distinct groups are detected

3.4 Automatic detection of extreme values

In our analysis, we do not want to take into account the extreme values. Extreme values can come either of an atypical value or of a ground truth error. We then decide not to model these elements. However, it is an important information for the user and we can communicate these suspicious zones. To detect these outliers, we use the classical ESD identifier that considers that all points that do not belong to $[mean-t\times SD; mean+t\times SD]$ are outliers, where SD is the standard deviation of the distribution observed. We use classical values for t, $t = 3$ if the number of observation is greater than 80, otherwise $t = 2.5$. The detection of extreme values step is proceeded after the group detection. The ESD identifier is applied in each detected group and not in the whole population. Mean and standard deviation of the whole population are not relevant when multimodality is detected.

Detection of outliers is done here from a univariate point of view, for position of X_a, Y_a, X_b and Y_b and for the height and width of the MBR. All zones that are detected as possible extreme values with ESD identifier are not used for computation of position operator. As ESD identifier is well known not to find all the extreme values, the risk to delete a useful element is weak. Moreover, as we are not defining strict boundaries with our position operator, but orientation of our analysis, loss of one useful element has a low impact.

The figure 5 shows the example of the position of "date/place" in handwritten business letters. As it can be seen some elements are detected as extreme values and erased for the global evaluation of the zone. However, these extreme values can be divided into two groups. One group is composed of extreme values present in the top left part of the pages, whereas another group is composed of extreme values present in the bottom of the page. Elements in the top left part are enough homogeneous to define a zone for them. Elements in the bottom part of the page are not enough frequent and homogeneous to be considered as defining a potential zone of interest. Figure 5(c) shows an example of a letter with the date component located in the bottom part of the page, which is automatically detected and erased by LearnPos. Resulting position operator will be presented in next section.



(a) "Date/place" MBR location in a normalized page for a corpus of 300 pages.

(b) Deletion of extreme values

(c) Example of detected atypical location for "date/place"

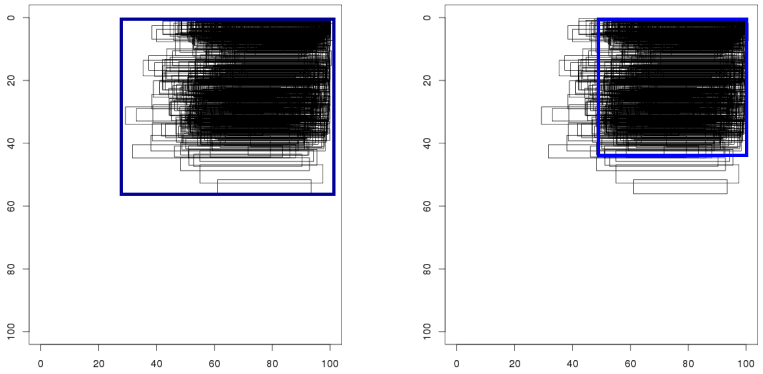
Figure 5. Representation of the position of date/place MBR obtained from a learning data set of business letters. Each rectangle represents a location.

3.5 Density-grid method to compute position operator

When the groups are detected and the extreme values are deleted of the analysis, the zones boundaries can be defined. Figure 6 presents the "date/place" MBR position in handwritten letters when extreme values are deleted by using method presented in section 3.4. If we simply define the zone that includes all the MBR as position operator, we may introduce too much error because the zone will be too large. However, we are not in the case of ground truth error but in the case of a document slightly different from the others. We cannot use the ESD identifier as we did in section 3.4.

Our method uses the notion of density and grid developed in density-based clustering. We do not focus on data points but on cells. Our method is based on the following steps. First, we divide space into a grid where each case width represents one percent of the original document width and each case height represents one percent of the original document height. In each case, we then count the number of MBR which have an intersection with the case which is not empty. This is used as an approximation of the density. Then, if a case contains less than a fixed parameter number of elements, we set its effective to zero. All cases that have a non null density are taken into account to define position operator. The zones are unions of connected high-density units within a subspace. We finally merge each group of connected units to a zone.

The final shape of the zone can be various. With LearnPos, for each zone found, we take into account is MBR for the final version of our position operator, as it can be seen in figure 6. A position operator is then



(a) Zone obtained including all MBR (b) Zone obtained using density-grid method

Figure 6. Example of adjustment of zone using density-grid method

composed of one or more rectangular zones.

4. HOW TO DEAL WITH CONFUSION

After the treatments presented in section 3, our tool has computed one or more zones for the asked position operator. As we previously explained, zone boundaries are computed not to be strict boundaries. It means that even if the zone actually contains the searched component, it can also contain other components that can bring confusion in our analysis. Indeed, these other components can be chosen instead of the one we are looking for.

One way to deal with this confusion is to establish rules and conditions that discriminate the component we are looking for from the others. However, characteristics may not differ enough to allow this possibility. Another way to deal with this confusion is presented here. It consists in exporting the generated position with a notion of order and point of view. In this purpose, we introduce a new indicator in our analysis, *the confusion indicator*. This notion is introduced at two points in our analysis. First, we use it to order our zones of interest. Then, we use it to know how to scan the elements in the zone.

4.1 The confusion indicator

A position operator is computed using the all learning data set. A position operator is chosen to be sufficiently large to allow a good recall. That means that other elements that the one we are looking for can be included in the defined zone. The confusion is an indicator that allows us to know how many other types of components are present in this zone in the learning data set. It is defined as follows:

$$confusion(list, zone) = \sum_{i=1}^{n} list[i] \cap zone \neq \emptyset$$

list contains all the elements except the one we are locating with position operator. In the case of relative positioning, we need to exclude both reference and relative to obtain the best possible results.

4.2 Ordering the zones

When more than one zone is discovered, we will look for the elements in the different zones, until we have found it. To minimize errors, we begin with the zones where little confusion with other elements is possible. Proceeding this way, we limit the possibility that another component that the one we are looking for is chosen and stops the analysis of the different possible zones. This is the information given by the confusion indicator. We then

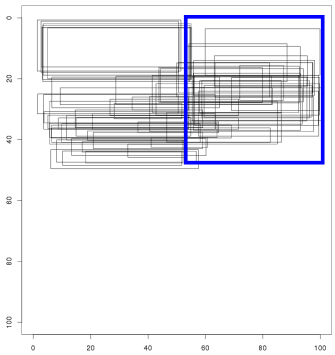
compute the confusion indicator in each zone found for the position operator. The zones are sorted by ascending confusion indicator.

For example, LearnPos detects two different zones for “expedient detail” in function of “date/place” location. One zone refers to “expedient detail” above “date/place”. The other one refers to “expedient detail” beyond “date/place”. The analysis will begin with zone which is above “date/place” because few confusions are possible whereas the zone which is beyond “date/place” can contain the opening or the text of the letter.

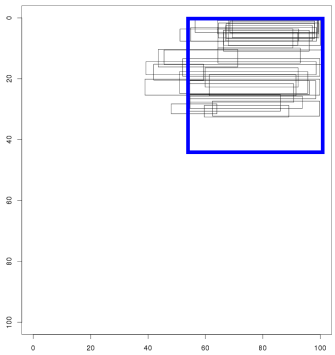
4.3 Choosing the best point of view

When we are looking for an element in a document, we not only need to know in which part of the document we can search for it, but we also need to define from which point of view we will look in this zone. The elements present in this part of the document will be tested from the closest to the furthest. We then use the confusion indicator to define from which point of view we have the best chances not to do mistakes. To be able to define the best point of view, we compute our confusion indicator using a set of typical point of view. We then choose the point of view that minimizes the confusion indicator.

In the example of the “date/place” component in business letters, the figure 7(a) shows a representation of all the components MBR that have a non empty intersection with the zone defined by our position operator. To increase readability, this example is based on a subset of 30 letters. Figure 7(b) represents the “date/place” MBR. Point of view is used to maximize the chance to find the researched element while minimizing the risk to select another element of the page. If we adopt a top-down point of vie, we will minimize errors as the “date/place” elements are frequently at the top of the zone, the other elements are more frequently located at the bottom of the zone. If we find elements in the top of the zone, the risks to do a recognition error are weaker than when we find an element in the bottom of the zone.



(a) Representation of all components except “date/place”: elements that can bring confusion



(b) Representation of “date/place”: elements we are looking for

Figure 7. “Date/place” position operator is analyzed from top to bottom to maximize the chance to select the date/place while minimizing the risk to select another element

5. EVALUATION AND RESULTS

To evaluate our proposal, we used our tool to obtain position operator for an existing grammar-based method: the DMOS-P method. Two aspects have been examined. First, on the same grammar, we compare the performance with the position operators obtained with our tool with the performance with the manually defined position operators. Then, we assess the interest of the confusion indicator by comparing the results obtained with the confusion indicator and results obtained with a fix strategy. These aspects have been evaluated using a handwritten business letters data set. First, we present the context of these experiments, then we detail the performed evaluations.

5.1 Context of the experiments

RIMES : French handwritten business letter database

The RIMES* French national evaluation campaign[12] established a publicly available database containing thousands of handwritten letters and faxes. These documents can be used for different tasks related to document recognition: layout analysis, writer identification, handwritten text recognition, etc. Images are 300 dpi grayscale scanned pages. All the images have been manually annotated with a ground truth.

The document structure recognition task in RIMES consists in the identification of up to eight different zones in each image and their assignment to one of the following labels: *sender details* (return address), *recipient details* (inside address), *date/place*, *subject*, *opening*, *message body*, *signature* and *attachment/postscript*. The recognition rate for this task is defined as the recall per class: $Recall = \frac{\text{number of assigned black pixels}}{\text{number of expected black pixels}}$. Only black pixels are counted to avoid including the background.

Existing DMOS method

The DMOS (Description and MOdification of the Segmentation) method[8] is a grammatical method for structured document recognition. This grammatical method is used to illustrate the efficiency of LearnPos. LearnPos is independent in its implementation of the DMOS method and can be used with another recognition system. The analysis is guided by position operators. A position operator is composed of the zone of analysis (higher left and lower right angles coordinates) and the location of the cursor. The cursor is used to scan the zone's component from the closest to the furthest.

For each position operator of the grammatical description, six parameters must be defined. In the standard DMOS method, these parameters are manually defined by user.

5.2 Evaluation of position operators defined with LearnPos

Adapting an existing grammar

A grammatical description for the RIMES evaluation campaign already existed, this grammar was presented by Lemaitre.[13] In this existing grammar, position operator parameters were manually defined. We generated position operators with LearnPos for this grammar and introduced our inferred position operator in the existing deterministic grammar. The objective is to check that these automatically defined position operators are correct.

Learning data set is composed of the same 300 images for the manually defined position operators and the automatically defined position operators. In the case of manually defined position operators, the 300 images where observed by a human operator to determine what are the best possible parameters. Obviously, the manual observation of 300 images in an expansive task and the synthesis of this observation to produce a position operator is not an easy task.

We used LearnPos to compute some absolute position operators for *sender details*, *place/date*, *signature* and *subject*. We asked LearnPos the relative position operator to locate *recipient details* in function of *date/place*.

Preservation in recognition rate

We then compared the recognition rate on a validation database of images. The table 1 indicates the recall rate for each class. As it can be seen, using position operators defined manually or by LearnPos gives similar results. Some components recall is improved. For example, signature detection is improved which has a strong impact on body and opening detection. Some components recall is slightly decreased but globally we have an improvement of results. Major advantage of LearnPos is that time needed to compute position operator is consequently decreased. The user just needs to specify what he wants to locate and then copy and paste the position operator computed by LearnPos tool.

With this technique, we reduce the number of manual parameters as it can be observed in table 2: 48 parameters are automatically learnt with LearnPos. However, some position operators that are used in the

*RIMES: *Reconnaissane et Indexation de donnes Manuscrites et fac-similES* (recognition and indexing of handwritten documents and faxes), `http://www.rimes-database.fr/doku.php`

Class	(1) Manual position operators	(2) LearnPos position operators
Body	**91.4**	**93.9**
Sender	91.3	91.3
Recipient	85.1	83.6
Signature	**88.1**	**91.4**
Subject	66.8	64.5
Date/place	**75.6**	**76.7**
Opening	**77.1**	**80.5**
Att/ps	9.6	9.6
Total	88.6	90.2

Table 1. Recall rate for the two handwritten structure analysis on 950 documents: (1) the existing version, with manually defined position operators and (2) our modified version, with inferred position operators.

grammar cannot be automatically learned using LearnPos. It is the case of the position operators defining relations between lines of text (above a line, beyond a line, etc.). We cannot use LearnPos for these position operators because ground truth does not contain information at line level. With an adapted ground truth, LearnPos could have been used for all the position operators of the grammar.

	Manual parameters	Automatic parameters
Existing grammar	102	0
Our proposal	66	48

Table 2. LearnPos permits a decreased in the number of manually defined position operators in a grammar. The 66 remaining manual parameters have no available ground truth.

5.3 Point of view

In the previous section, we presented the results obtained by introducing position operators determined with LearnPos in an existing grammar. For this task, we use all the information that LearnPos gives: the different zones to analyze, the order of analysis of these zones and the point of view to adopt for each zone. We globally show that these information allow us to perform as good as analysis as using manually defined position operators, while time needed for position operators is drastically reduced.

In this section, we want to focus the experiment on one aspect of information determined with LearnPos: the appropriate point of view. To show the interest of choosing carefully the good point of view, we compare results obtained by positioning the cursor at the middle of the zone and by choosing the cursor proposed by our tool. We limit experiments to the date/place element because as we use an existing grammar, a change of point of view requires modifying deeply the grammar rules. In the two experiments, the zones boundaries and the order of analysis of the zones are identical, extracted using LearnPos tool.

As it can be seen in table 3, the recall is improved when we select the point of view determined with LearnPos, from 75.6% to 76.7%. Pixels errors are decreased of 4.5% using the inferred point of view. Defining the best point of view allows us to minimize errors while testing candidates and improves the recall.

Class	(1) Middle point of view	(2) Inferred point of view
Body	93.9	93.9
Sender	91.3	91.3
Recipient	83.7	83.6
Signature	91.4	91.4
Subject	64.7	64.5
Date/place	**75.6**	**76.7**
Opening	80.5	80.5
att/ps	9.6	9.6
Total	90.1	90.2

Table 3. Recall rate for the two handwritten structure analysis on 950 documents: (1) the point of view is located in the center of the bounding box and (2) the point of view is determined by LearnPos tool, using the confusion indicator.

6. CONCLUSION

In this paper, we propose a new tool for interactive positioning for document analysis systems. LearnPos is a generic tool that is independent of any recognition system. Thanks to this method, the user is helped in the layout analysis. Logical structure is defined by the user and LearnPos is used by the user to define the physical structure without the need for manual exploration of the learning data set. LearnPos learns automatically position operators that can be directly integrated in a grammatical method. This analysis is driven by the user who asks explicitly the position operator he desires. The computed position operators are of the same type than the manually defined ones. These position operators are understandable and can by modify by the user who can introduce its knowledge of the domain. The concept of confusion allows us not only to be able to determine the different zones of the position operators but also to make the best of this zone by ordering them and orienting the point of view in the best possible way.

Our proposal allows a grammar writer to simplify grammar writing and to reduce time spent. LearnPos makes an exhaustive analysis of the learning set which allows the determination of both general cases and rare cases. Experiments on handwritten business letters showed that our tool can compare with grammatical description with manually tuned position operators. Position operators where obtained after a manual visualization of 300 document images, which needs several hours. Our method deals with the analysis of 300 document images in few minutes.

REFERENCES

[1] Bloch, I. and Ralescu, A. L., "Directional relative position between objects in image processing: a comparison between fuzzy approaches," *Pattern Recognition* **36**(7), 1563–1582 (2003).

[2] Boll, S., Klas, W., and Westermann, U., "A comparison of multimedia document models concerning advanced requirements," tech. rep. (1999).

[3] Conway, A., "Page grammars and page parsing. A syntactic approach to document layout recognition," in [*ICDAR*], 761–764 (1993).

[4] Papadias, D. and Theodoridis, Y., "Spatial relations, minimum bounding rectangles, and spatial data structures," *International Journal of Geographic Information Science* **11**, 111–138 (1997).

[5] Egenhofer, M. J. and Herring, J., "Categorizing binary topological relations between regions, lines, and points in geographic databases," tech. rep., University of Maine (1994).

[6] Maredj, A.-E., Nourreddine, T., Sadallah, M., and Alimazighi, Z., "A flexible distance for the spatial placement in a multimedia document," in [*Information and Communication Technologies: From Theory to Applications, 2008. ICTTA 2008. 3rd International Conference on*], 1–4 (2008).

[7] Shetty, S., Srinivasan, H., Srihari, S., Shetty, S., Srinivasan, H., Beal, M., and Srihari, S., "Segmentation and labeling of documents using conditional random fields," in [*Document Recognition and Retrieval DRR XIV*], (2007).

[8] Coüasnon, B., "Dmos, a generic document recognition method: Application to table structure analysis in a general and in a specific way," *International Journal on Document Analysis and Recognition, IJDAR* **8**, 111–122 (June 2006).

[9] Delaye, A., Macé, S., and Anquetil, E., "Modeling Relative Positioning of Handwritten Patterns," in [*14th Biennial Conference of the International Graphonomics Society*], 122–127 (Sept. 2009).

[10] Lerddararadsamee, T. and Jiraraksopakun, Y., "Local maximum detection for fully automatic classification of em algorithm," in [*Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2012 9th International Conference on*], 1–4 (2012).

[11] Sheather, S. J. and Jones, M. C., "A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation," *Journal of the Royal Statistical Society. Series B (Methodological)* **53**(3), 683–690 (1991).

[12] Grosicki, E., Carree, M., Brodin, J.-M., and Geoffrois, E., "Results of the rimes evaluation campaign for handwritten mail processing," in [*Document Analysis and Recognition, 2009. ICDAR '09. 10th International Conference on*], 941–945 (2009).

[13] Lemaitre, A., Camillerapp, J., and Coüasnon, B., "A generic method for structure recognition of handwritten mail documents," in [*Document Recognition and Retrieval DRR XV*], (2008).