DMOS, It's your turn !

Bertrand B. Coüasnon, Aurélie Lemaitre

DMOS, It's your turn !

Bertrand Coüasnon
Irisa - Insa
Rennes, France
Email: bertrand.couasnon@irisa.fr

Aurélie Lemaitre
Irisa - Université Rennes 2
Rennes, France
Email: aurelie.lemaitre@irisa.fr

***Abstract*—The DMOS method (Description and MOdification of Segmentation) is a complete system for document analysis and recognition. It is a syntactical method which enables a rule based expression of knowledge. This generic method can be applied to various kinds of documents. Its adaptation to a knew kind of document simply consists in the description of dedicated knowledge. The DMOS method has been widely validated, in several domains, and at a large scale. Now, we propose to share DMOS method for research purpose. We provide a complete environment of development in Eclipse. The use of DMOS method can help you to improve your recognition systems by combining information coming from your varied sources, being guided by explicit knowledge.**

***Keywords*-DMOS, document analysis, structure recognition, rule based system, data combination**

I. Overview of DMOS method

The DMOS method is made of three blocks: *visual clues* are used as input; they are organized by the *knowledge description*; then a *compilation step* automatically produces the dedicated recognition system.

A. Visual clues

The strength of DMOS method is that it can take as input for document analysis and recognition many kinds of visual clues. Basically, it uses primitives extracted at various levels of resolution of the image: connected components, line segments, text lines. But it can also consider any output of more elaborate structure recognition systems. Concerning semantic inputs, it can use OCR outputs, or outputs from handwritten recognition systems, or specific fields localized by word spotting. The visual clues can also come from the recognition of other pages of the collection, or from an interaction with a user. DMOS method can use as primitives internal data coming from digital native pdf documents.

In order to combine that large possible variety of visual clues, the DMOS method is based on an original mechanism called perceptive layers. A perceptive layer is a structure that contains primitives. Its coordinates are expressed in a single coordinate system. Each layer can be seen as a given point of view on the document, and the analysis can combine as many layers as necessary.Then, each rule will precise in which layer it uses primitives for the analysis.

B. Knowledge description

The knowledge description if realized using a specific language called EPF (Enhanced Position Formalism). This rule based language enables to express knowledge on the physical content, but also the logical organization of the documents. Here is an example of a rule for the description of the content of a mail document.

```
letter::=
   AT(topLeftPage) &&
      senderDetails &&
   AT(middlePage) &&
      opening &&
   AT(underOpening) &&
      textBody &&
   At(underBody) &&
      signature.
```

This syntax enables to express the bi-dimensionnal position of elements inside of the document, using the position operator `AT`. Each non-terminal of the rule, such as `senderDetails` or `opening` can be detailed using terminal elements which have been stored in the perceptive layers. Thus, the EPF langage enables to set up cooperation strategies between the different sources of available visual clues: the success of each rule can be based on the presence of visual clues coming from several layers, ie several points of view on the image.

For example, the following rule expresses that a text line is detected using two points of view: the presence of a line segment at low resolution and the presence of characters at high resolution.

```
textline::=
      USE_LAYER(lowResolution)
          FOR(lineSegment) &&
   AT(lineSegmentZone) &&
      USE_LAYER(highResolution)
          FOR(setOfCharacters).
```

The role of the programmer is to describe the best strategy of recognition, that efficiently combines the visual clues coming from the different points of view of the document.

C. Compilation step

Once the EPF description of the kind of document is available, the associated recognition system is automatically produced by a compilation step.

II. Application to various document recognition problems

DMOS method has been applied to deal with various document recognition problems. We present here a few

concrete fields of application to show the interest of this method (figure 1).

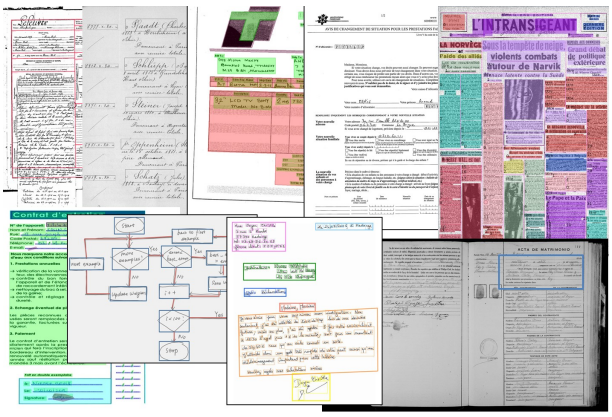

Figure 1. Various kinds of documents analyzed by DMOS method

Recognition of tabular structure in military forms: [1] The DMOS method enables to describe the tabular structure using *a priori* knowledge to overcome the difficulties presented by the strongly damaged documents.

Structure recognition of naturalization decree registers: [2] We combine the analysis of two resolutions levels of the image to obtain a more efficient recognition and make easier the document description.

Structure recognition in RIMES database: [3] We demonstrate how DMOS method can be used to create perceptive mechanisms similar to those used by human vision, by combining global vision and visual attention. We also introduced, thanks to the mechanism of perceptive layers, the scores produced by a word recognizers. This exploits the faculty of the EPF language to define local stochastic grammars.

On-line flowchart recognition: [4] DMOS can use online signal as input. The logical organization of flowchart diagrams has been described in EPF langage. An interaction was set-up with a stroke classifier to build a decision strategy, depending on both local context and classifier result.

Sale registers of the 18th century : [5] The strategy of analysis takes into account information coming from other pages of the collection. It also presents a way to interact in an asynchronous manner with a user in order to improve recognition results.

Structure recognition on Maurdor database: [6] DMOS enables to express a strategy for the analysis of complex heterogeneous contents: the grammatical rules describe the salient elements, which are progressively removed. This application is also a case of integration of the results of a commercial OCR.

Others applications has been also done on complex orchestra scores, on mathematical formulae, on the extraction of articles in old newspapers, on field localization in ancient registers... The DMOS method as been validated at a large scale on more than 700,000 pages of documents.

III. It's your turn !

In the demonstration, we will present how DMOS system can be used by the community of researchers. The main interest of DMOS method is to enable a combination of information coming from various sources produced by your systems: visual clues extracted in the image, handwriting or symbol recognition results, segmentation hypothesis, outputs of existing complete recognition systems, user-interaction mechanisms. This combination is guided by an explicit knowledge. Thus, the EPF grammatical language enables a description of physical, logical and syntactical content. The combination of data from different sources is possible thanks to the mechanism of perceptive layers. The use of DMOS method enables to improve performances and genericity. The demonstration will show the complete development environment proposed with DMOS method: a virtual machine, an Eclipse IDE with a debugger, the standard primitive extraction libraries, the EPF compiler, the tools for visualization.

IV. Conclusion

DMOS method is a complete system for document recognition. It enables the combination of heterogeneous information guided by rules. We propose an access to the complete development environment for research purposes, which provide you a way to integrate the results of your own systems guided by the expression of knowledge.

References

[1] B. Coüasnon, "DMOS, a generic document recognition method: Application to table structure analysis in a general and in a specific way," *IJDAR*, vol. 8(2), pp. 111–122, 2006.

[2] A. Lemaitre, J. Camillerapp, and B. Coüasnon, "Contribution of multiresolution description for archive document structure recognition," in *ICDAR'07*, IAPR, Ed., 2007, pp. 247–251.

[3] ——, "A generic method for structure recognition of handwritten mail documents," in *Proceedings of Document Recognition and Retrieval (DRR XV)*, 2008.

[4] C. Wang, H. Mouchère, A. Lemaitre, and C. Viard-Gaudin, "Online flowchart understanding by combining max-margin Markov random field with grammatical analysis," *IJDAR*, vol. 20, no. 2, pp. 123–136, Jun. 2017.

[5] J. Chazalon, B. Coüasnon, and A. Lemaitre, "Iterative Analysis of Pages in Document Collections for Efficient User Interaction," in *ICDAR*, China, Sep. 2011, pp. 503 –507.

[6] B. Poirriez, A. Lemaitre, and B. Coasnon, "Visual perception of unitary elements for layout analysis of unconstrained documents in heterogeneous databases," in *14th ICFHR*, 2014.