

HANDWRITTEN/PRINTED TEXT DISCRIMINATION USING AGENT TECHNOLOGY (HAPEDATE)

Hapedate is a software able to discriminate printed and handwritten text in document images

From invention to innovation

APPLICATIONS

- Printed text detection and extraction for Optical Character Recognition (OCR)
- Handwritten text extraction from filled up printed forms
- Automated detection of manually annotated printed documents

PUBLICATIONS

« Handwritten and printed text separation : Linearity and regularity assessment » in Image Analysis and recognition, 11th International conference ICIAR 2014, S. Hamrouni, F. Cloppet, N. Vincent.

INTELLECTUAL PROPERTY

Copyright

CONTACT

Mail : bd-3S@idfinnov.com
 Phone : +33(0)1.80.06.80.03

Agent technology ■ Optical Character Recognition (OCR) ■ Document analysis

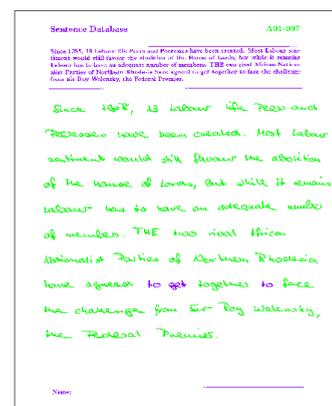
PRESENTATION

Scanned documents can contain handwritten text, such as annotations, signatures, or filled-in blanks in forms. Identifying those handwritten parts from printed parts is a preliminary step in the treatment chain of the document.

To this end, state-of-the-art approaches are generally using different classifying tools and then compare their results at the end of the classification process. Machine learning algorithms can also be used but in that case, huge databases are usually needed.

Our approach is based on the dialogue of two agents, one assessing the linearity of the writing, the other one the regularity. After an optional learning phase (for example to integrate specific characteristics of a given database), both agents interact to dynamically update their decision models and finally make a common decision, leading to better results than state-of-the-art tools. Without any learning phase, our system reaches an overall recognition rate of 92,3% on the IAM database.

Our system has been developed for Latin alphabets but may be adapted to other alphabets, such as Arabic or Asian writings.



Discrimination of printed and handwritten text in a scanned document: text recognized as printed appears in purple; text recognized as handwritten appears in green. Source : Prof. Nicole Vincent, Prof. Pavlos Moraitis

COMPETITIVE ADVANTAGES

- No need of prior knowledge of the text layout (especially in case of printed forms)
- No need of databases
- No need of deep learning algorithms
- Customizable decision function depending on context