Exploiting Collection Level for Improving Assisted Handwritten Word Transcription of Historical Documents

LAURENT GUICHARD, JOSEPH CHAZALON, BERTRAND COÜASNON
INSA Rennes – UMR IRISA, Université Européenne de Bretagne, France

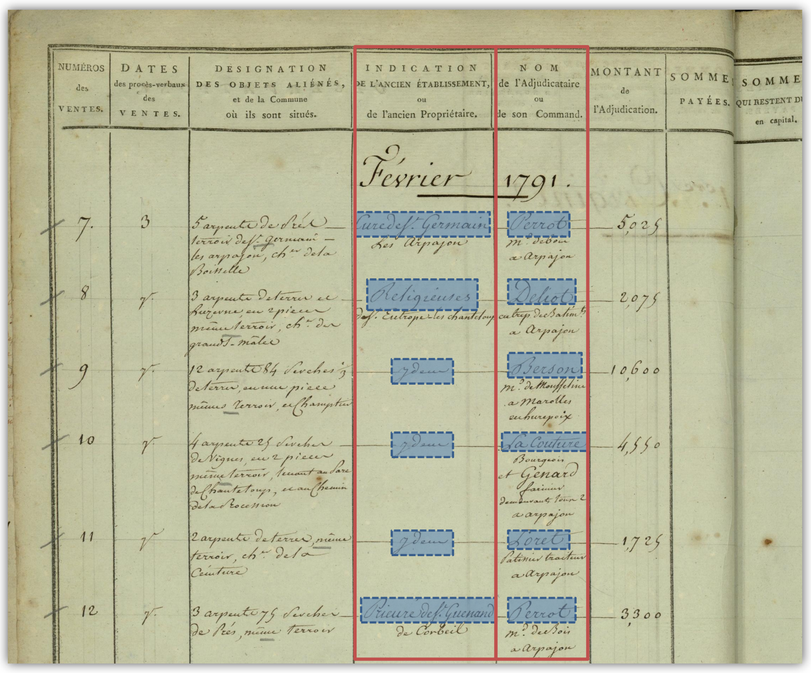
Keywords: document sets; handwritten word recognition; historical documents; human interaction

ABSTRACT

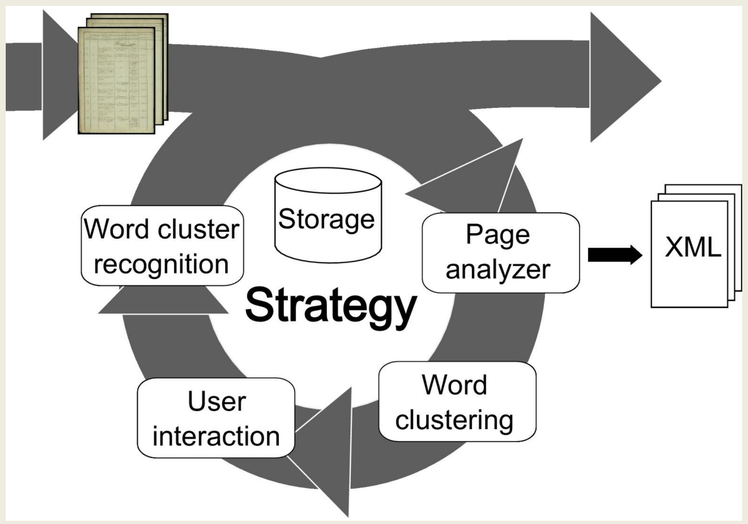
Transcription of handwritten words in historical documents is still a difficult task. When processing huge amount of pages, document centered approaches are limited by the trade-off between automatic recognition errors and the tedious aspect of human user annotation work. In this article, we investigate the use of inter page dependencies to overcome those limitations. For this, we propose a new architecture that allows the exploitation of handwritten word redundancies over pages by considering documents from a higher point of view, namely the collection level. The experiments we conducted on handwritten word transcription show promising results in terms of recognition error and human user work reductions.

2 MAIN CONTRIBUTIONS

1. **Iterative multi-level architecture** to make use of collection context
2. Processing documents at **collection level** leads to **better results** in terms of
   - **human user workload**
   - **recognition error**



EXPLOITING COLLECTION LEVEL: *required elements*



Consider Sets of Pages

Use collection context to
- enable handwritten word clustering
- improve user interaction…

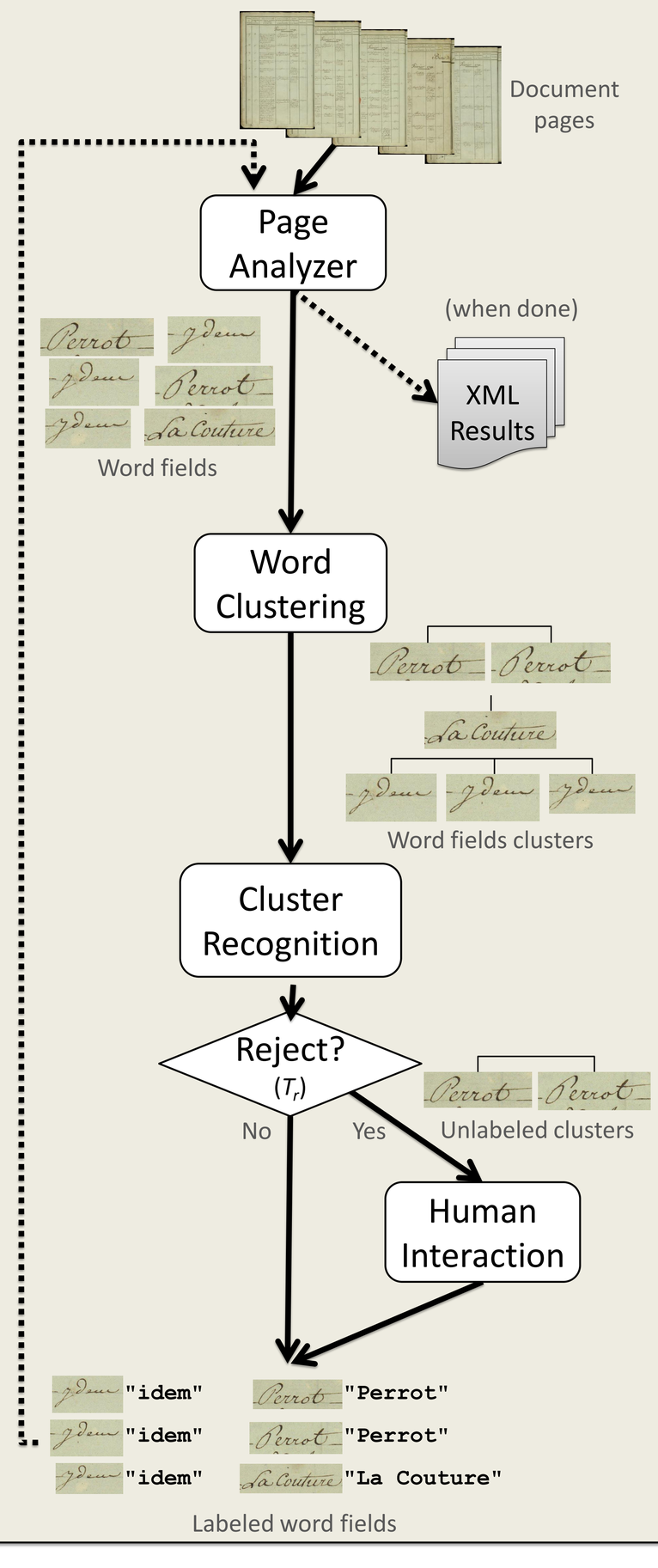
Dissociate Strategy and Tasks

Strategy module: routes data between tasks
Data: word fields, clusters, pages, user answers…
Tasks: document analysis, word clustering, interaction…

Enable an Iterative Analysis

Page analyzers are called **as many times as needed.**
They reintegrate external data to
- **check it against their document models**
- **produce final results using external information**

IMPLEMENTATION FOR A SPECIFIC DOCUMENT SET



Page Analyzer

- Word extraction using DMOS-P [1]
- Iterative analysis to use information from interaction (model validation, final result production…) [2]

[1] A. Lemaitre, J. Camillerapp, and B. Coüasnon, "Multiresolution cooperation makes easier document structure recognition," International Journal on Document Analysis and Recognition, vol. 11, pp. 97–109, 2008.
[2] J. Chazalon, B. Coüasnon, and A. Lemaitre, "Iterative Analysis of Pages in Document Collections for Efficient User Interaction," in *Proc. of ICDAR*, 2011.

Word Clustering

1. Build a cost matrix ***M*** using DTW (features from [3])
2. Hierarchical Agglomerative Clustering
   a. Start with each sample as a cluster
   b. Agglomerate clusters C_i and C_j $\forall\ i \neq j$ until $d_{i,j} > T_c$

$$d_{i,j} = \max_{e_k \in C_i, e_l \in C_j} M(e_k, e_l)$$

[3] T. M. Rath and R. Manmatha, "Features for word spotting in historical manuscripts," in *Proc. of ICDAR*, 2003, p. 218.

Cluster Recognition

1. **Sample recognition**
   a. Recognize each cluster sample e_i using [4]
   b. Produce a set of hypothesis (label, score) = H_i

   Ex: e_1 → H_1 = { ("Pavot", 0.6), ("Perrot", 0.4) }
   e_2 → H_2 = { ("Peirot", 0.6), ("Perrot", 0.4) }
   = H

2. **Hypothesis fusion:** using H, the union of the H_i sets, we build a well-ordered set H' defined as:

$$H' = \left\{ (\hat{l}, \hat{s}) \middle| \hat{s} = \frac{1}{|H|} \sum_{(l,s) \in H, l = \hat{l}} s \right\}$$

   where elements are sorted by descending score $\hat{s}_i$.
   Ex: H = { ("Perrot", 0.8/4), ("Pavot", 0.6/4), ("Peirot", 0.6/4) }

3. **Decision:** using the two best hypothesis of H', **if**

$$\hat{s}_1 - \hat{s}_2 < T_r$$

   then the best label $\hat{l}_1$ is assigned to the cluster
   otherwise the cluster is rejected *(has to be labeled manually by a human user)*

[4] L. Guichard, A. Toselli, and B. Coüasnon, "A novel verification system for handwritten words recognition," in *Proc. of ICPR,* 2010.

Human Interaction

- Human user reviews unlabelled (rejected) clusters
- Interface shows a subset of cluster elements
- Human user labels the cluster (or rejects it)
- Label is propagated to all elements in the cluster

EXPERIMENTS AND RESULTS

Dataset

- 70 document pages manually annotated
- **1206 handwritten word fields (= word set *S*)**
- 502 different words values

Protocol: Comparison of 2 Strategies

Baseline (no clustering, no iterative analysis)
1. Word fields extraction and recognition
2. Annotation of rejected fields by human user

Clustering (← illustrated on left frame)
1. Word fields extraction with page analyzer
2. **Word field clustering** (threshold T_c)
3. **Cluster recognition** recognized
4. Annotation of rejected **clusters** (T_r) by human user

Evaluation

MR	**M**anual annotation **R**ate
ER	**E**rror **R**ate
AR	**A**utomatic annotation **R**ate
N_M	**N**umber of **M**anual annotations
N_e	**N**umber of incorrectly annotated field

Localization and extraction were not evaluated in the experiments.
Thresholds T_c and T_r were tuned using a grid search on a validation set.

$$MR = \frac{N_M}{|S|} \quad ER = \frac{N_e}{|S|} \quad AR = 1 - \frac{N_M + N_e}{|S|}$$

Results

Strategy	AR	MR	ER	AR	MR	ER
Baseline	59	21	20	24	75	1
Clustering	65	15		38	61	

CONCLUSION

For document retrieval – good indexation coverage
(reasonable error rate is ok)
- for an overall annotation rate of 80%
- relative diminution of 28% of human workload

For the adaptation of the system through retraining
(very low error rate is necessary)
- overall annotation rate of 99%
- relative increase of 58% of automatic annotation

Our approach permits to
- **balance automatic and manual processing**
- **reduce recognition error and human workload**