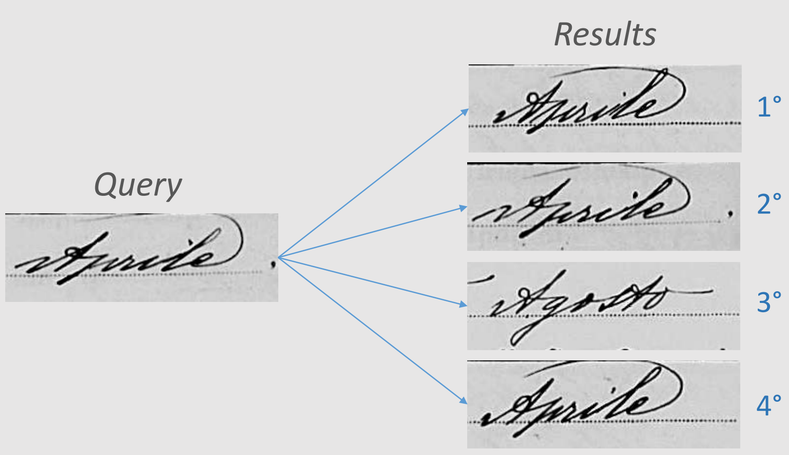
Historical Handwritten Text Images Word Spotting through Sliding Window HOG Features

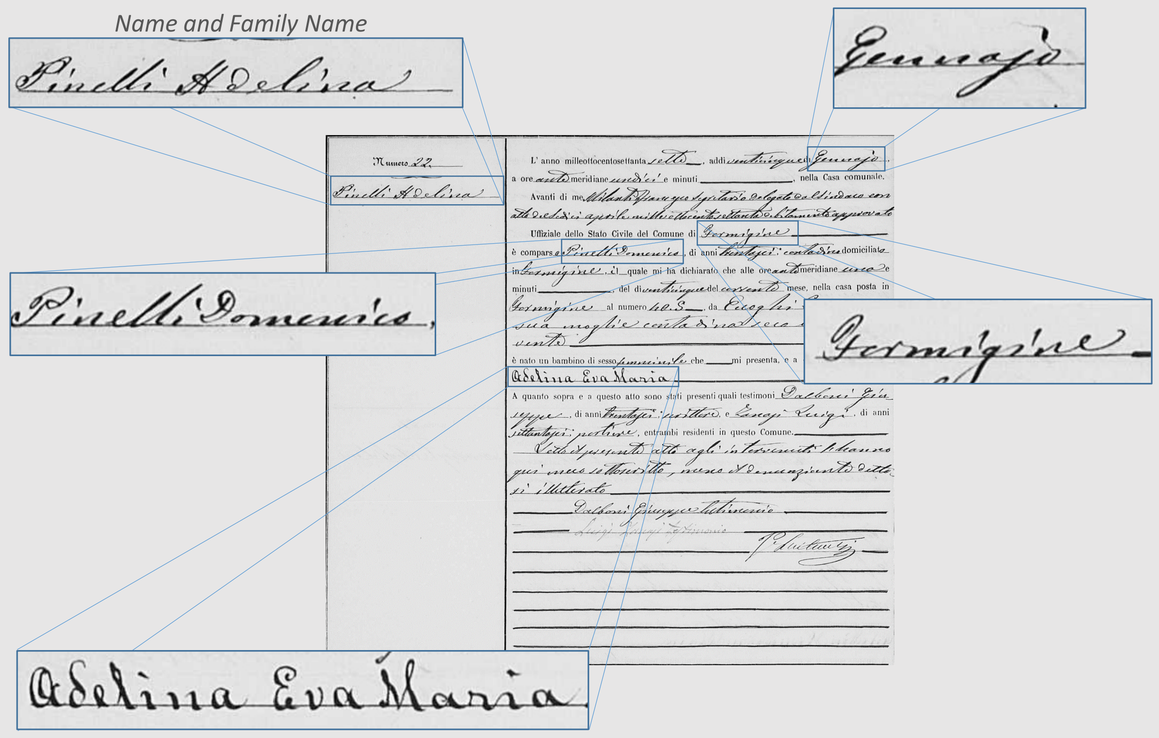
Federico Bolelli, Guido Borghi, Costantino Grana
University of Modena and Reggio Emilia

Problem Statement

The project, named XDOCS, aims at extending to a much wider audience the possibility to access a variety of historical documents published on the web. In this context, we develop a new word spotting technique able to extract document indexes in quasi-automatic mode from their handwritten contents. The devised solution is based on HOG descriptors and exploits *Dynamic Time Warping* technique to compare feature vectors elaborated from single handwritten words.

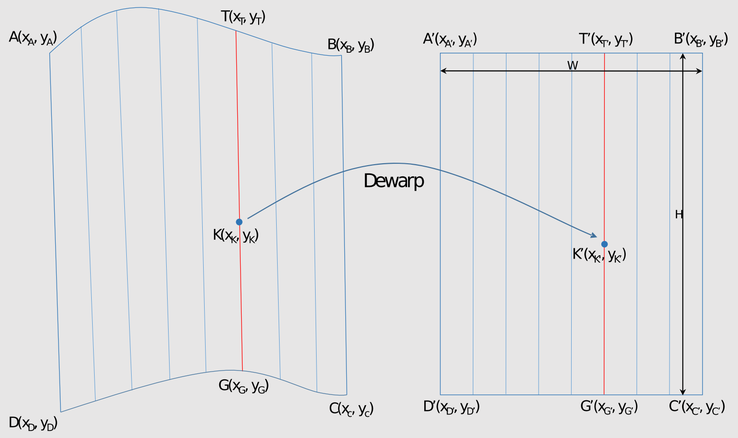


Intended Indexes and New Challenging Dataset



Indexing Pipeline

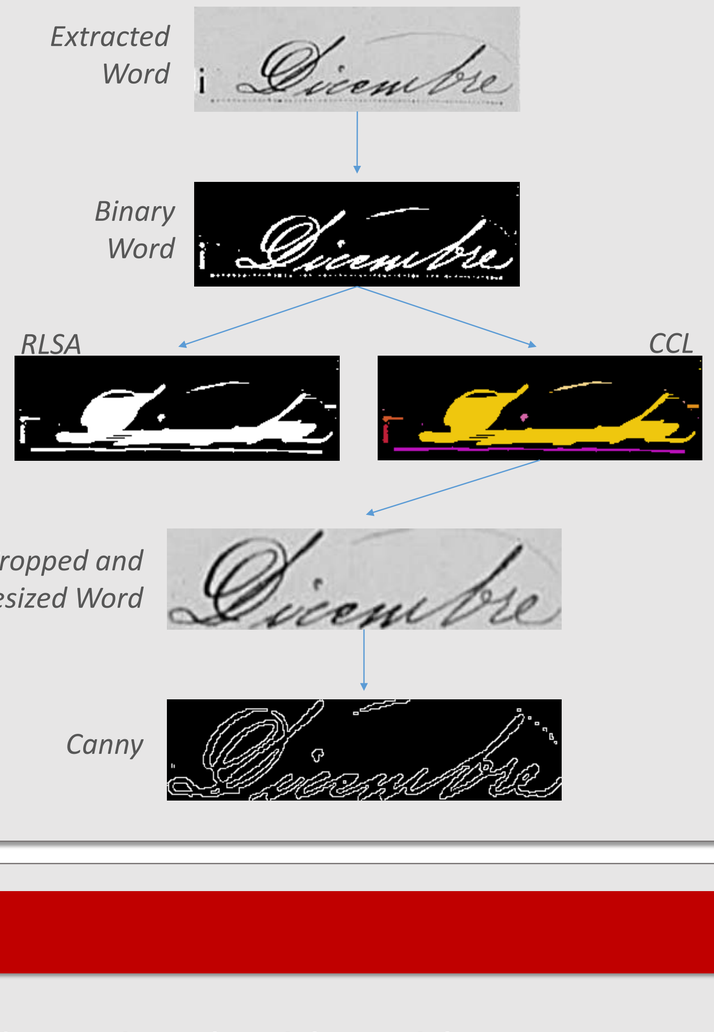
① Image Dewarping



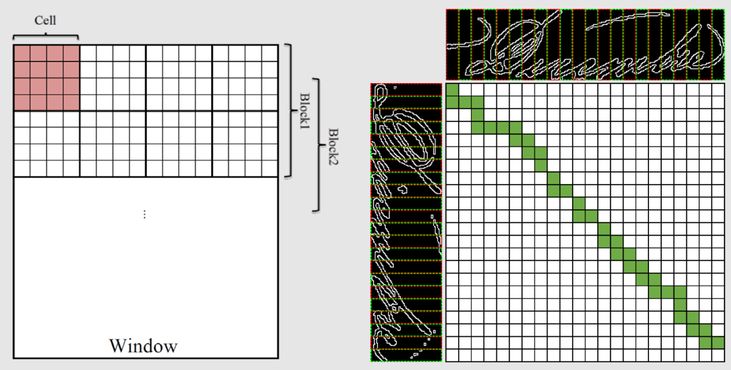
Aims at mapping the projection of the curved surface, represented by four polynomial lines, to 2D rectangular area with fixed dimensions.

$$\begin{cases} x'_k = x'_A + W * \dfrac{|\widehat{AT}|}{|\widehat{AB}|} \\ y'_k = y'_A + H * \dfrac{|TK|}{|TG|} \end{cases}$$

② Word Extraction and Preprocessing



③ Word Matching Based on HOG Descriptor and Dynamic Time Warping



$$DTW(i,j) = min\begin{cases} DTW(i-1,j) \\ DTW(i,j) \\ DTW(i,j-1) \end{cases} + \sum_{k=1}^{N} |x_{ik} - y_{jk}|$$

Experimental Results

***Intra* and *Inter* dataset evaluation of word spotting algorithm with:**

- *Mean Averages Precision* (MAP) with *cut-off* at C = {5, 10, 15}:

$$MAP@n = \frac{\sum_{i=1}^{Q} ap@n_i}{N} \qquad ap@n = \frac{\sum_{k=1}^{n} P(k)}{\min(m,n)}$$

- Correct Match First (CMF): percentage of queries with P(1) = 1.

Check out the project website

xdocs

		Vignola	*Carpi*	*Formig.*
Vignola	MAP@05	0.528	0.1	0.181
	MAP@10	0.38	0.086	0.144
	MAP@15	0.306	0.093	0.132
	CMF	75.25%	17.82%	26.73%
Carpi	MAP@05	0.135	0.466	0.095
	MAP@10	0.101	0.434	0.078
	MAP@15	0.079	0.414	0.072
	CMF	14.53%	63.25%	15.38%
Formig.	MAP@05	0.192	0.127	0.644
	MAP@10	0.156	0.114	0.541
	MAP@15	0.135	0.121	0.476
	CMF	24.69%	19.25%	77.82%

Results with 16 pixels stride

		Vignola	*Carpi*	*Formig.*
Vignola	MAP@05	0.634	0.108	0.206
	MAP@10	0.465	0.098	0.168
	MAP@15	0.376	0.101	0.156
	CMF	83.17%	18.32%	25.25%
Carpi	MAP@05	0.145	0.534	0.112
	MAP@10	0.11	0.489	0.093
	MAP@15	0.086	0.485	0.086
	CMF	17.10%	66.67%	17.95%
Formig.	MAP@05	0.268	0.15	0.775
	MAP@10	0.209	0.133	0.662
	MAP@15	0.173	0.138	0.582
	CMF	36.82%	21.34%	89.96%

Results with 8 pixels stride

		Vignola	*Carpi*	*Formig.*
Vignola	MAP@05	0.665	0.102	0.222
	MAP@10	0.493	0.093	0.189
	MAP@15	0.4	0.098	0.17
	CMF	87.13%	14.85%	27.22%
Carpi	MAP@05	0.159	0.578	0.125
	MAP@10	0.117	0.536	0.101
	MAP@15	0.091	0.527	0.096
	CMF	19.66%	73.50%	17.95%
Formig.	MAP@05	0.309	0.177	0.823
	MAP@10	0.235	0.152	0.708
	MAP@15	0.194	0.153	0.621
	CMF	40.59%	26.77%	94.14%

Results with 2 pixels stride

		Vignola	*Carpi*	*Formig.*
Vignola	MAP@05	0.468	0.042	0.077
	MAP@10	0.347	0.034	0.057
	MAP@15	0.276	0.028	0.05
	CMF	68.32%	9.90%	13.37%
Carpi	MAP@05	0.086	0.445	0.087
	MAP@10	0.06	0.411	0.067
	MAP@15	0.05	0.382	0.058
	CMF	13.78%	51.70%	15.34%
Formig.	MAP@05	0.097	0.053	0.557
	MAP@10	0.071	0.045	0.413
	MAP@15	0.06	0.042	0.342
	CMF	19.25%	9.21%	80.33%

Competitors (Rath et al.)