Session 4

Interpretation of drawings, music scores, tables, etc.

Chairs

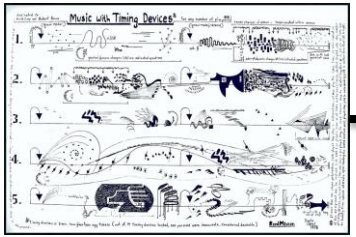
Mickaël Coustaty (L3I, Université de La Rochelle)
Véronique Eglin (LIRIS, INSA de Lyon)

Short presentation of chairs

- **Véronique Eglin**
  - Professor @ LIRIS (INSA de Lyon)
  - Handwriting analysis, word-spotting, writer identification, retrieval
  - Digital Humanities / Historical documents

- **Mickaël Coustaty**
  - Associate Professor @ L3i (University of La Rochelle)
  - Indexing of graphics in historical documents
  - Digital Humanities / Fraud Detection and Forensics

Context of graphical content interpretation



Detection
- Basic constitutive elements
- Primitives

GT#1

Composition rules
- Strings
- Cells
- Note composition

GT#2

Semantic labelling
- Hierarchical classification
- Recomposing

GT#3

Interpretating
- Transcription
- Context oriented

GT#4

Needs and open problems

- **Needs**:
  - More genericity in the detection and recognition steps
  - User knowledge in the process loop
- **Main tasks**:
  - The detection of basic constitutive elements / primitives (text, beams, note-head, stems, boxes)
  - The recognition of degraded symbols, distorted plans
  - The ground truth generation and the annotation of datasets
- **Current status**: no generic methodologies for drawing interpretation (tables, architectural plans, music scores)
- **Open problems** (not yet solved): efficient solutions with unbalanced data, high-level of heterogeneity and low quality of data, no efficient annotated datasets, no fair solution for mobility

Paper 1

Camera-based Optical Music Recognition using a Convolutional Neural Network

Adria Rico, Alicia Fornès, CVC Barcelona

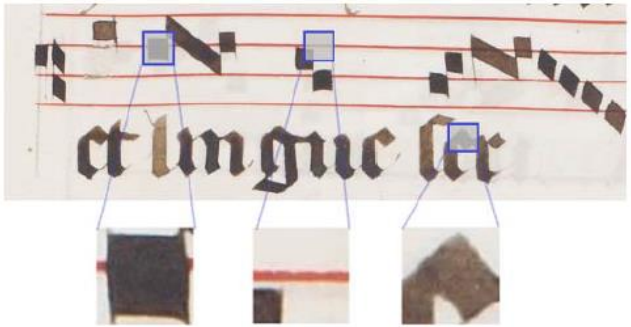
- Need for a **robust and accurate camera-based OMR** system dedicated to mobile-captured music scores.
- Proposition based on a **neural architecture using a CNN** for music symbols classification, exploration of deep features to improve low quality images.
- A proposition that relies on the **automatic generation of ground-truth** images

Paper 2

Music Document Layout Analysis through Machine Learning and Human Feedback

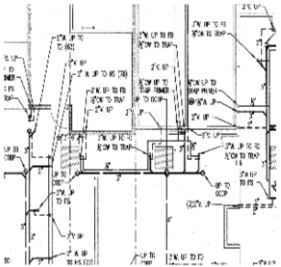
Jorge Calvo-Zaragoza, Ke Zhang, Zeyad Saleh, Gabriel Vigliensoni and Ichiro Fujinaga

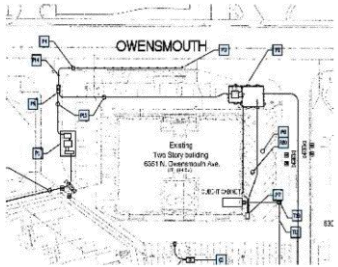
- An answer to overcome the **lack of genericity** of analysis systems that must be redesigned when working with different sources of data
  - before the step of symbols recognition
  - in a context of historical and/or real datasets (containing a high level of heterogeneity)
- Contribution to a **user-aided classification scheme** to segment efficiently categories of music scores elements and validated by a quantification of user's effort

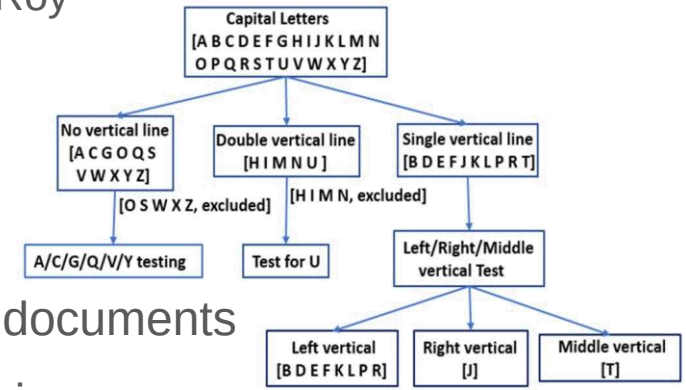


Paper 3 Automatic Orientation Correction of AEC Drawing Documents

Bhagesh Seraogi, Supriya Das, Purnendu Banerjee, Bidyut Chaudhuri, Himadri Majumdar, Srinivas Mukkamala, Rahul Roy





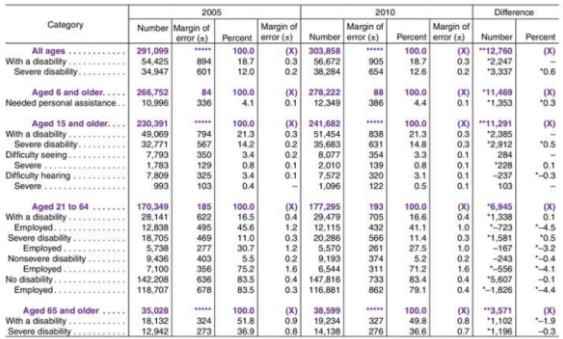


- Work on Architectural, Engineering, Construction documents
- Documents composed of texts and graphical drawings
- Need to detect skew/orientation to increase OCR system
- Propose an approach to detect and correct the orientation using new features

Paper 4 Interpreting data from scanned tables

Waleed Farrukh, Antonio Foncubierta-Rodriguez, Anca-Nicoleta Ciubotaru, Guillaume Jaume, Costas Bekas, Orcun Goksel, Maria Gabrani

- Work on table interpretation
- Extract basic objects of tables (characters, line segments)
- Conversion to table specific structural elements (strings, cells)
- Classification of table content (data vs headers)
- Extract the structure of the header cells (interpretation)
- First work on interpretation of tables



Category	2005				2010				Difference	
	Number	Margin of error (±)	Percent	Margin of error (±)	Number	Margin of error (±)	Percent	Margin of error (±)	Number	Percent
All ages	**291,099**	*****	**100.0**	**(X)**	**303,858**	*****	**100.0**	**(X)**	****12,760**	**(X)**
With a disability	54,425	894	18.7	0.3	56,672	905	18.7	0.3	*2,247	–
Severe disability	34,947	601	12.0	0.2	38,284	654	12.6	0.2	*3,337	*0.6
Aged 6 and older	**266,752**	**84**	**100.0**	**(X)**	**278,222**	**88**	**100.0**	**(X)**	***11,469**	**(X)**
Needed personal assistance	10,996	336	4.1	0.1	12,349	386	4.4	0.1	*1,353	*0.3
Aged 15 and older	**230,391**	*****	**100.0**	**(X)**	**241,682**	*****	**100.0**	**(X)**	****11,291**	**(X)**
With a disability	49,069	794	21.3	0.3	51,454	838	21.3	0.3	*2,385	–
Severe disability	32,771	567	14.2	0.2	35,683	631	14.8	0.3	*2,912	*0.5
Difficulty seeing	7,793	350	3.4	0.2	8,077	354	3.3	0.1	284	–
Severe	1,783	129	0.8	0.1	2,010	139	0.8	0.1	*228	0.1
Difficulty hearing	7,809	325	3.4	0.1	7,572	320	3.1	0.1	-237	*-0.3
Severe	993	103	0.4	–	1,096	122	0.5	0.1	103	–
Aged 21 to 64	**170,349**	**185**	**100.0**	**(X)**	**177,295**	**193**	**100.0**	**(X)**	***6,945**	**(X)**
With a disability	28,141	622	16.5	0.4	29,479	705	16.6	0.4	*1,338	0.1
Employed	12,838	495	45.6	1.2	12,115	432	41.1	1.0	*-723	*-4.5
Severe disability	18,705	469	11.0	0.3	20,286	566	11.4	0.3	*1,581	*0.5
Employed	5,738	277	30.7	1.2	5,570	261	27.5	1.0	-167	*-3.2
Nonsevere disability	9,436	403	5.5	0.2	9,193	374	5.2	0.2	-243	*-0.4
Employed	7,100	356	75.2	1.6	6,544	311	71.2	1.6	*-556	*-4.1
No disability	142,208	636	83.5	0.4	147,816	733	83.4	0.4	*5,607	-0.1
Employed	118,707	678	83.5	0.3	116,881	862	79.1	0.4	*-1,826	*-4.4
Aged 65 and older	**35,028**	*****	**100.0**	**(X)**	**38,599**	*****	**100.0**	**(X)**	****3,571**	**(X)**
With a disability	18,132	324	51.8	0.9	19,234	327	49.8	0.8	*1,102	*-1.9
Severe disability	12,942	273	36.9	0.8	14,138	276	36.6	0.7	*1,196	-0.3

Discussion and open questions

- **Contextualisation of the symbol recognition to improved interpretation**
  - Ex: Instead of training on individual segmented symbols without context, what about the direct translation of full lines of sheet music?
- **Dataset and ground-truth** : large corpora of sufficiently diverse annotated symbols are difficult and expensive to produce
  - Are deep architectures new promising solutions for graphics/symbols recognition ? For music scores are they compatible with inherent constraints : unbalanced data, regular and irregular data, unlabelled datasets…?
  - How to reduce the data requirement for training (when not available) ?
- **User-interactions: expert or non-expert ?**
  - What about the role of “expert” user to help the recognition system ? (ex of musicologists in the critical steps of validation of grouping hypothesis, of pruning of non relevant combination of compound notes…)