Segmentation and recognition of symbols for printed and handwritten music scores

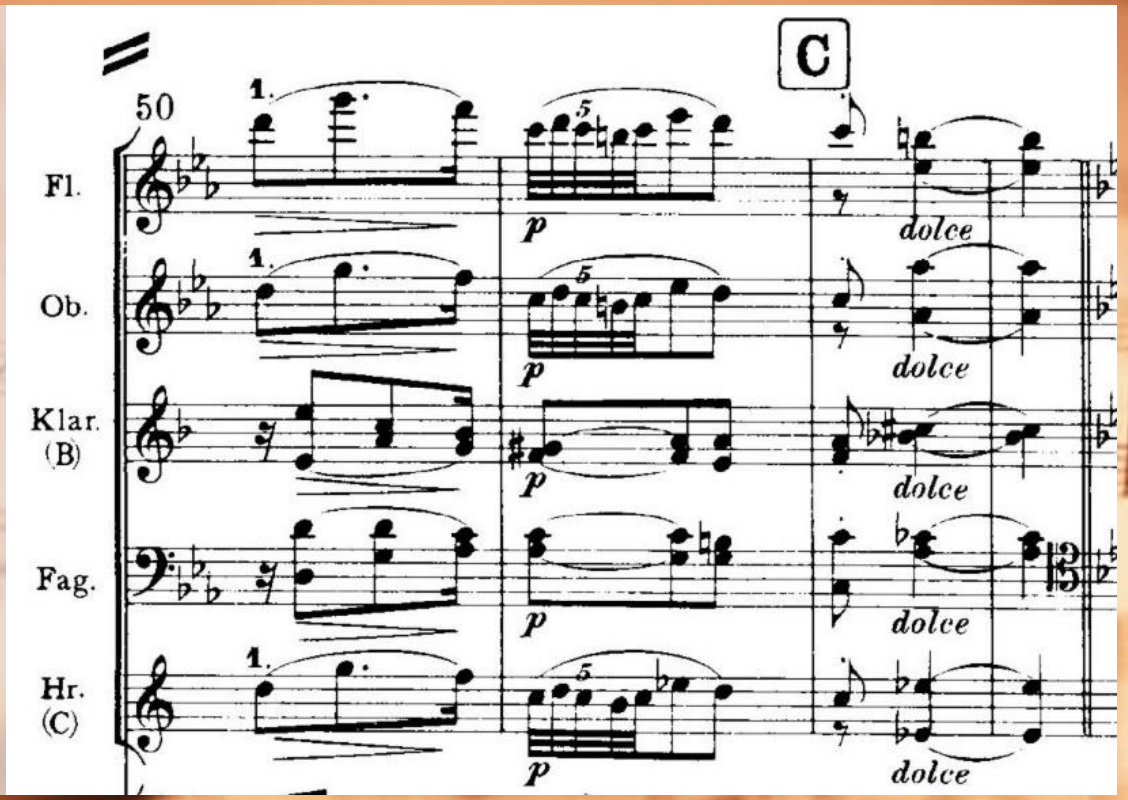
Colloque de Master Recherche Informatique

Author
Kwon-Young Choi

Supervisors
Bertrand Coüasnon
Yann Ricquebourg

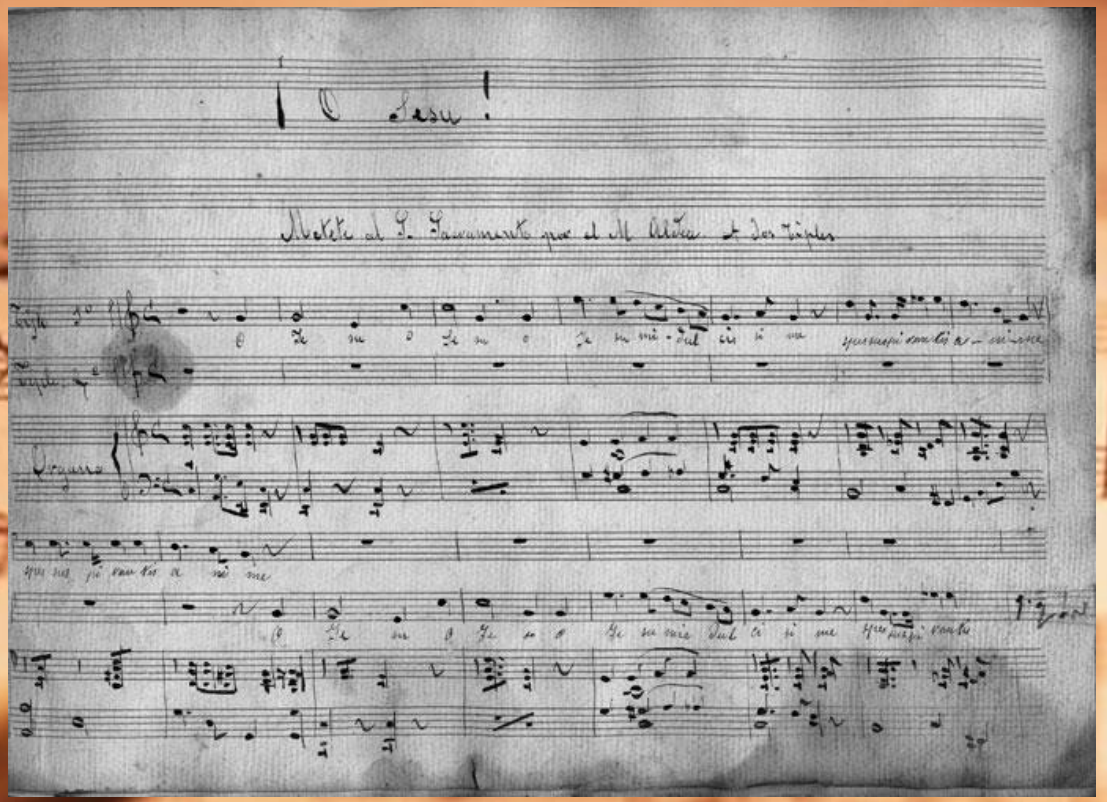
Music score definition

- Written form of a music composition
- Used to save and share music
- Characteristics
  - Complex bidimensional organization
  - High density of symbols
  - High connectivity of symbols



Music score definition

- Written form of a music composition
- Used to save and share music
- Characteristics
  - Complex bidimensional organization
  - High density of symbols
  - High connectivity of symbols



Optical Music Recognition

- Use to transform an image of a score into a machine readable format

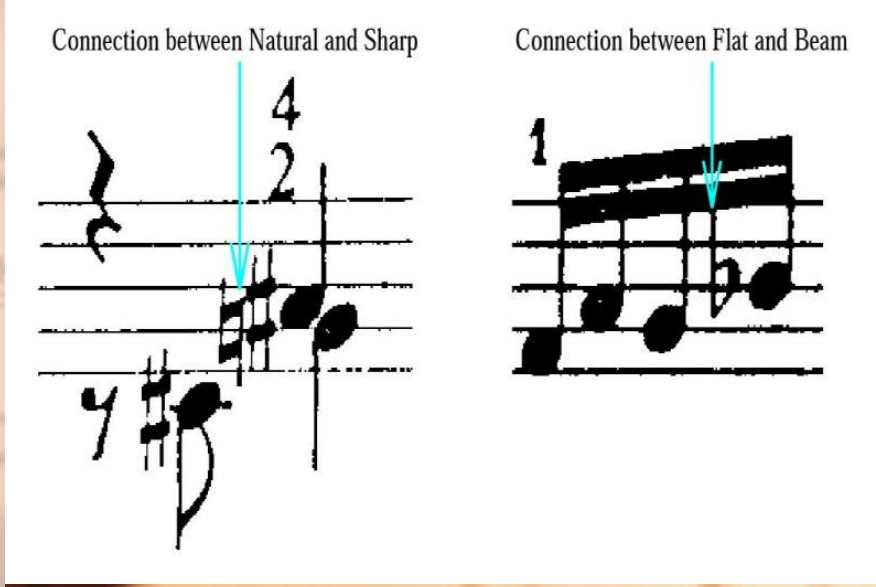
- Steps of a classic ascending OMR system [Rebelo 2012] [Fornes 2014]
  - Pre-processing
  - Staves recognition and removal
  - **Music symbols segmentation and recognition**
  - Music notation reconstruction

Optical Music Recognition systems

- Problem of bottom-up system like presented in [Rebelo 2009]
  - Context information is not used during segmentation and recognition
- DMOS: « Description and MOdification of Segmentation » [Coüasnon 2001]
  - Use a grammar to describe the document organization
  - Grammar guide segmentation and recognition
  - Use the context during segmentation and recognition phase

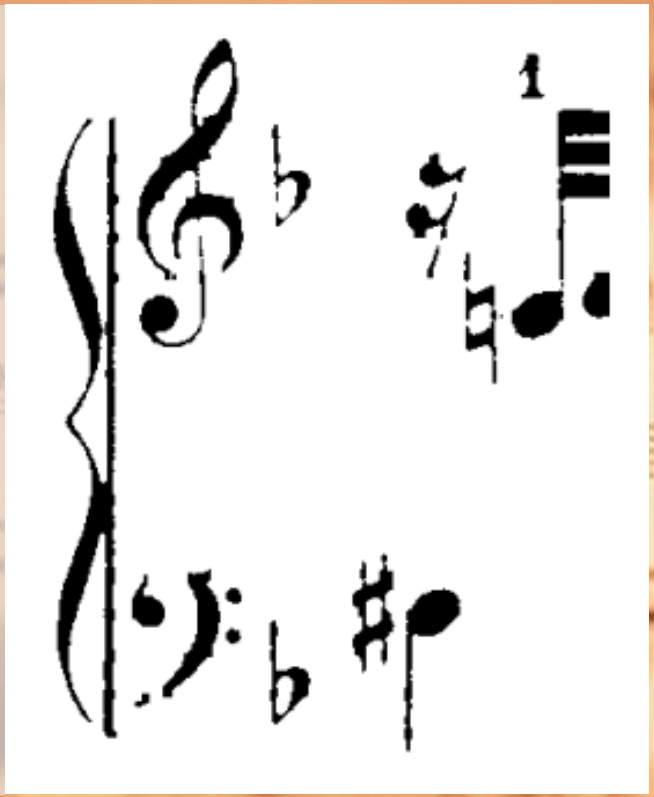
Segmentation and recognition of music symbols

- Main problematics
  - Connected and overlapping symbols
  - Broken symbols because of the staff removal
- Internship hypothesis
  - Merging segmentation and recognition will improve the recognition rate



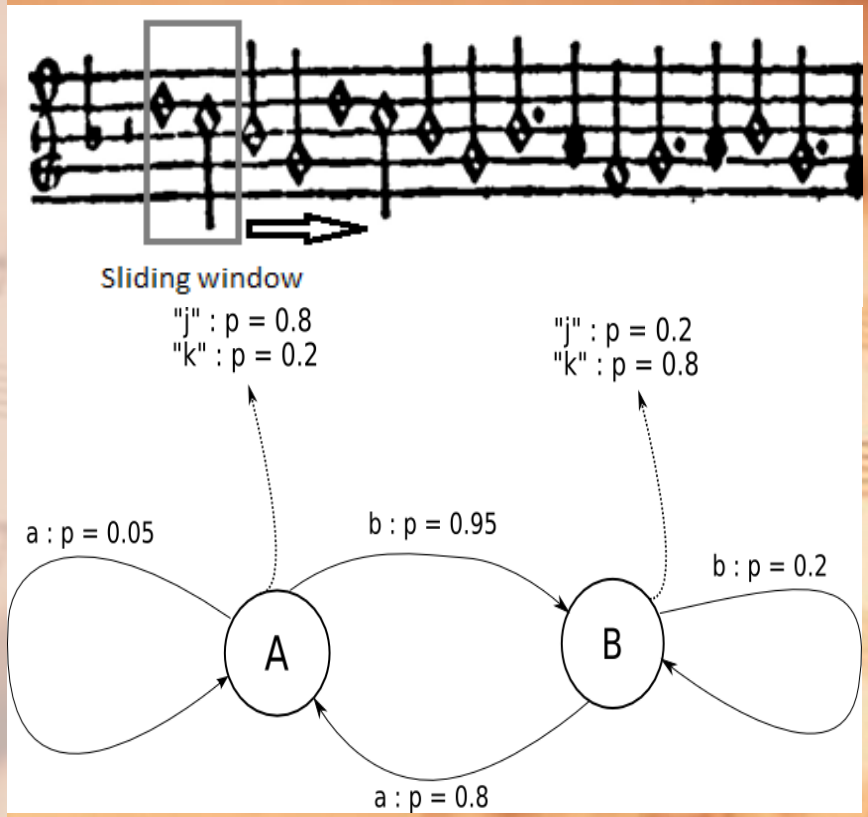
Segmentation and recognition of music symbols

- Main problematics
  - Connected and overlapping symbols
  - Broken symbols because of the staff removal
- Internship hypothesis
  - Merging segmentation and recognition will improve the recognition rate



Merging segmentation and recognition in OMR

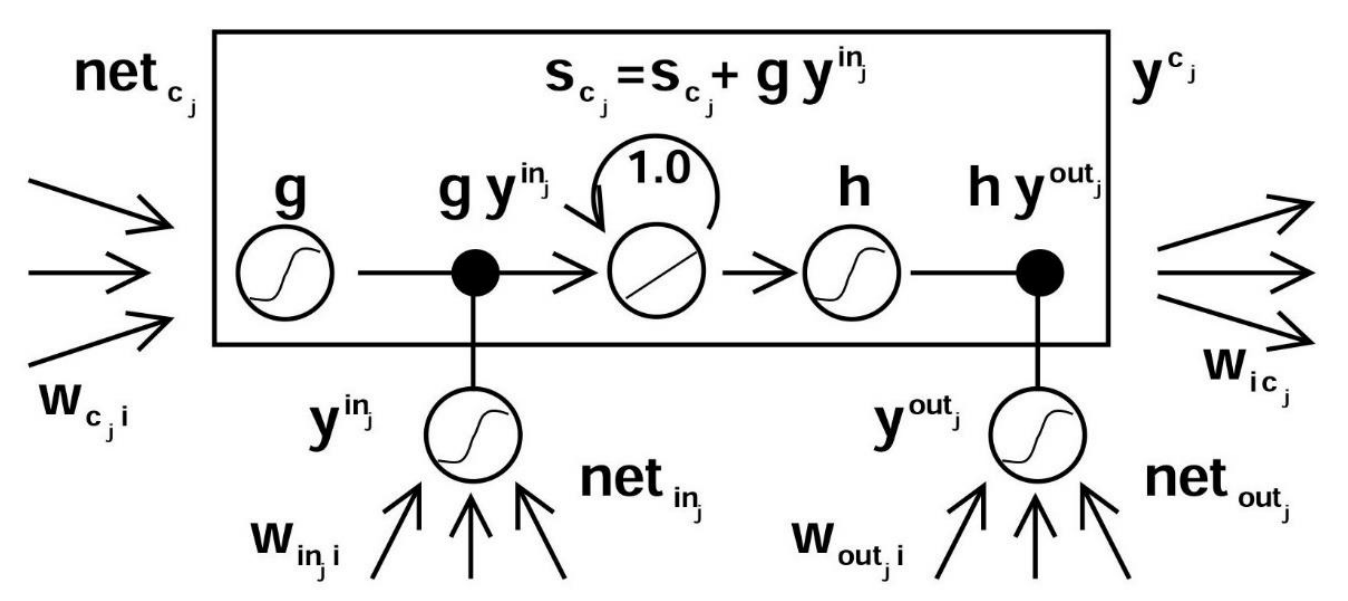
- Hidden Markov Model (HMM) on old and very simple scores [Pugin 2006]
  - Features on a sliding window
  - HMM: discrete stochastic modelling:
    - Markov property
- Problems
  - Simple and old score
    - Difficult to scale up
  - Limited use of context (HMM property)



Merging segmentation and recognition in text recognition

- By using Recurrent Neural Network (RNN) [Graves 2009]
  - Long Short-Term Memory LSTM
  - Use of context with Multi-Dimensional RNN
  - Connectionist Temporal Classification layer
    - Training and evaluating the network with no explicite segmentation
  - State-of-the-art results
- Problem: used only on classifying one dimensional sequence

Merging segmentation and recognition in text recognition



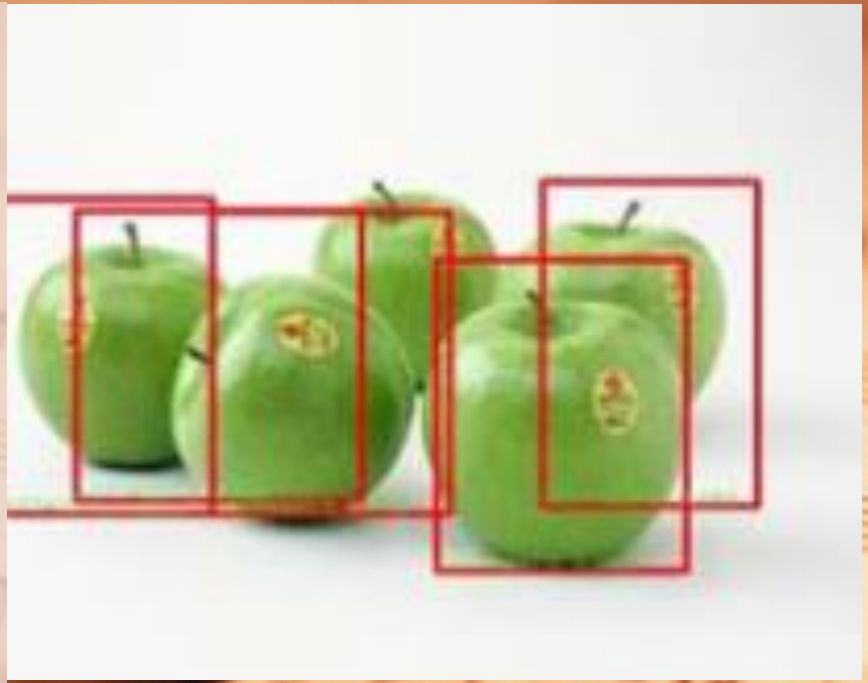
Resolving bidimensionality

- Maurdor international contest: complex document recognition
- Segmentation/recognition using previous RNN [Moysset 2014]
  - Input: block of text (bidimension)
  - Prior explicit segmentation of text lines
- Similarly, use of DMOS to roughly segment the score
  - Contextual selection of the symbol recognizer input

Object localization and identification

Another way to resolve bidimensionality:

- Deep Convolutional Neural Network [Erhan 2014]
  - Supervised training: learn features
  - DNN generate bounding box with confidence score
- Interests
  - multiple objects localization
  - With bounding boxes
- However
  - Need annotated data for training
  - None available dataset with groundtruth on music scores



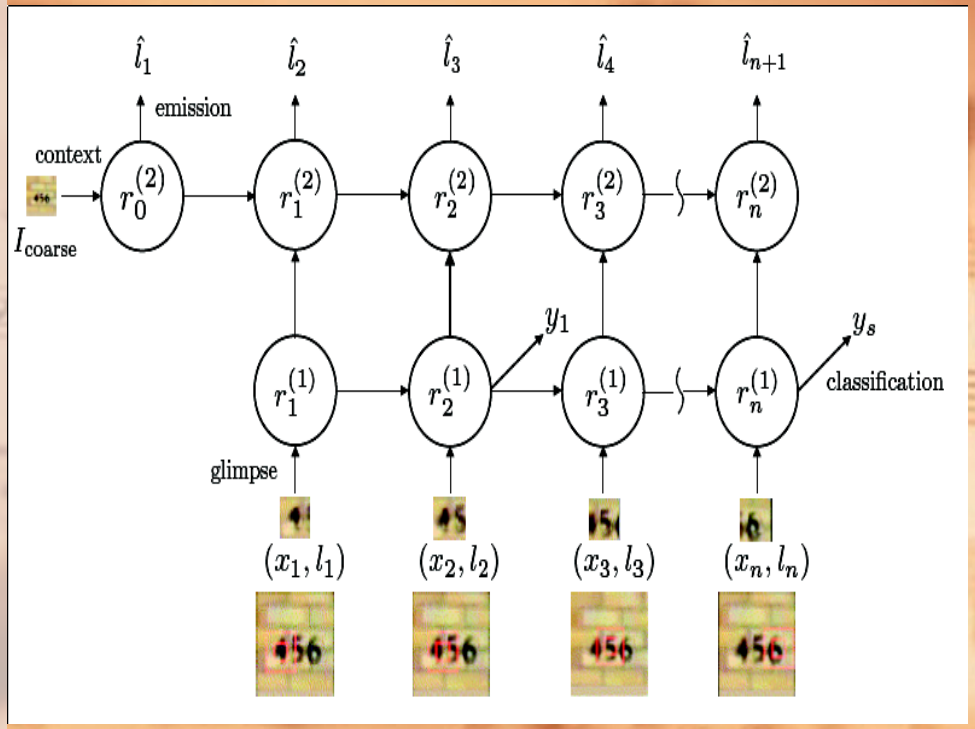
Object localization and identification

Another way to resolve bidimensionality:

- Deep Convolutional Neural Network [Erhan 2014]
  - Supervised training: learn features
  - DNN generate bounding box with confidence score
- Interests
  - multiple objects localization
  - With bounding boxes
- However
  - Need annotated data for training
  - None available dataset with groundtruth on music scores

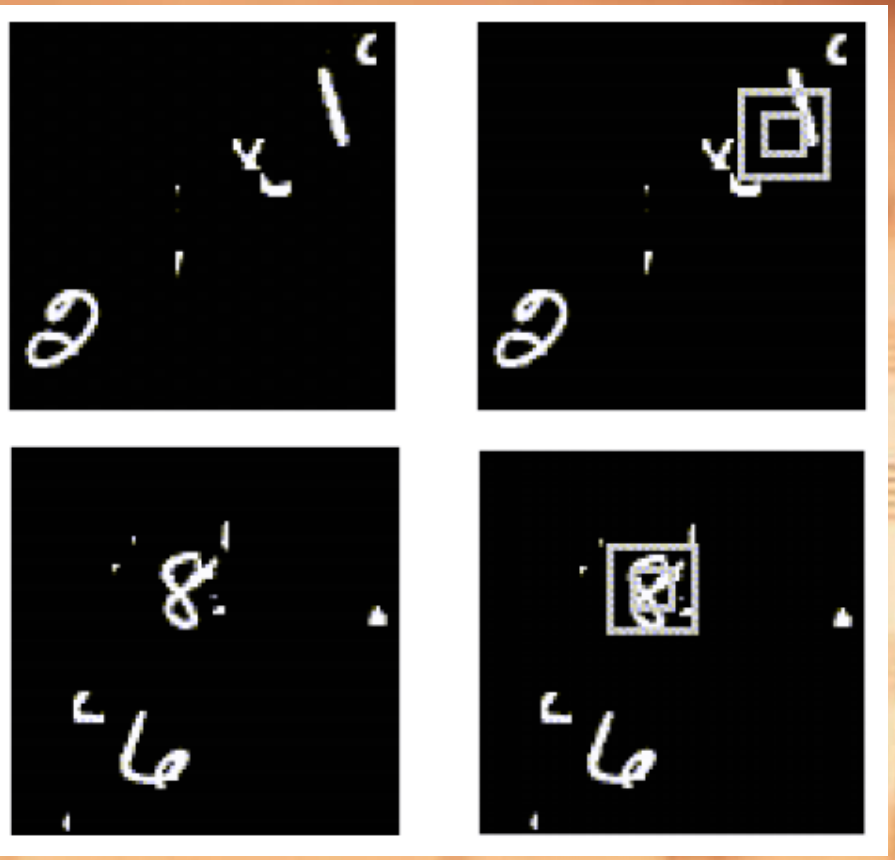
Visual attention model

- Use multiple recurrent neural networks layers to localize an object in an image (Mnih 2014)
  - Use of glimpses to explore an image
  - RNN used to predict next glimpse location, to recognize digit
- Pros : can localize an object in a two-dimensional space
- Not tested on complicated data



Visual attention model

- Use multiple recurrent neural networks layers to localize an object in an image (Mnih 2014)
  - Use of glimpses to explore an image
  - RNN used to predict next glimpse location, to recognize digit
- Pros : can localize an object in a two-dimensional space
- Not tested on complicated data



Conclusion

- Problematic: joint segmentation and recognition of music symbol
  - MD LSTM RNN
    - use of context
    - only handle one-dimensional sequence
  - DCNN
    - multiple object recognition, produce bounding boxes
    - need annotated data for training
  - Visual attention model
    - can localize a symbol in an image, robust to noise
    - only tested on simple recognition tasks
- Internship:
  study collaboration Neural Network/DMOS musical grammar
  - Rough segmentation with DMOS grammar
  - Fine segmentation and recognition with Neural Network