

Encoding Writer Variability for Automatic Writer Identification

Vincent Christlein

February 25, 2014

Pattern Recognition Lab (CS 5)



[illegible]

S. ¹ ² ³ ⁴ ⁵ ⁶ ⁷ ⁸ ⁹ ¹⁰ ¹¹ ¹² ¹³ ¹⁴ ¹⁵ ¹⁶ ¹⁷ ¹⁸ ¹⁹ ²⁰ ²¹ ²² ²³ ²⁴ ²⁵ ²⁶ ²⁷ ²⁸ ²⁹ ³⁰ ³¹ ³² ³³ ³⁴ ³⁵ ³⁶ ³⁷ ³⁸ ³⁹ ⁴⁰ ⁴¹ ⁴² ⁴³ ⁴⁴ ⁴⁵ ⁴⁶ ⁴⁷ ⁴⁸ ⁴⁹ ⁵⁰ ⁵¹ ⁵² ⁵³ ⁵⁴ ⁵⁵ ⁵⁶ ⁵⁷ ⁵⁸ ⁵⁹ ⁶⁰ ⁶¹ ⁶² ⁶³ ⁶⁴ ⁶⁵ ⁶⁶ ⁶⁷ ⁶⁸ ⁶⁹ ⁷⁰ ⁷¹ ⁷² ⁷³ ⁷⁴ ⁷⁵ ⁷⁶ ⁷⁷ ⁷⁸ ⁷⁹ ⁸⁰ ⁸¹ ⁸² ⁸³ ⁸⁴ ⁸⁵ ⁸⁶ ⁸⁷ ⁸⁸ ⁸⁹ ⁹⁰ ⁹¹ ⁹² ⁹³ ⁹⁴ ⁹⁵ ⁹⁶ ⁹⁷ ⁹⁸ ⁹⁹ ¹⁰⁰ ¹⁰¹ ¹⁰² ¹⁰³ ¹⁰⁴ ¹⁰⁵ ¹⁰⁶ ¹⁰⁷ ¹⁰⁸ ¹⁰⁹ ¹¹⁰ ¹¹¹ ¹¹² ¹¹³ ¹¹⁴ ¹¹⁵ ¹¹⁶ ¹¹⁷ ¹¹⁸ ¹¹⁹ ¹²⁰ ¹²¹ ¹²² ¹²³ ¹²⁴ ¹²⁵ ¹²⁶ ¹²⁷ ¹²⁸ ¹²⁹ ¹³⁰ ¹³¹ ¹³² ¹³³ ¹³⁴ ¹³⁵ ¹³⁶ ¹³⁷ ¹³⁸ ¹³⁹ ¹⁴⁰ ¹⁴¹ ¹⁴² ¹⁴³ ¹⁴⁴ ¹⁴⁵ ¹⁴⁶ ¹⁴⁷ ¹⁴⁸ ¹⁴⁹ ¹⁵⁰ ¹⁵¹ ¹⁵² ¹⁵³ ¹⁵⁴ ¹⁵⁵ ¹⁵⁶ ¹⁵⁷ ¹⁵⁸ ¹⁵⁹ ¹⁶⁰ ¹⁶¹ ¹⁶² ¹⁶³ ¹⁶⁴ ¹⁶⁵ ¹⁶⁶ ¹⁶⁷ ¹⁶⁸ ¹⁶⁹ ¹⁷⁰ ¹⁷¹ ¹⁷² ¹⁷³ ¹⁷⁴ ¹⁷⁵ ¹⁷⁶ ¹⁷⁷ ¹⁷⁸ ¹⁷⁹ ¹⁸⁰ ¹⁸¹ ¹⁸² ¹⁸³ ¹⁸⁴ ¹⁸⁵ ¹⁸⁶ ¹⁸⁷ ¹⁸⁸ ¹⁸⁹ ¹⁹⁰ ¹⁹¹ ¹⁹² ¹⁹³ ¹⁹⁴ ¹⁹⁵ ¹⁹⁶ ¹⁹⁷ ¹⁹⁸ ¹⁹⁹ ²⁰⁰ ²⁰¹ ²⁰² ²⁰³ ²⁰⁴ ²⁰⁵ ²⁰⁶ ²⁰⁷ ²⁰⁸ ²⁰⁹ ²¹⁰ ²¹¹ ²¹² ²¹³ ²¹⁴ ²¹⁵ ²¹⁶ ²¹⁷ ²¹⁸ ²¹⁹ ²²⁰ ²²¹ ²²² ²²³ ²²⁴ ²²⁵ ²²⁶ ²²⁷ ²²⁸ ²²⁹ ²³⁰ ²³¹ ²³² ²³³ ²³⁴ ²³⁵ ²³⁶ ²³⁷ ²³⁸ ²³⁹ ²⁴⁰ ²⁴¹ ²⁴² ²⁴³ ²⁴⁴ ²⁴⁵ ²⁴⁶ ²⁴⁷ ²⁴⁸ ²⁴⁹ ²⁵⁰ ²⁵¹ ²⁵² ²⁵³ ²⁵⁴ ²⁵⁵ ²⁵⁶ ²⁵⁷ ²⁵⁸ ²⁵⁹ ²⁶⁰ ²⁶¹ ²⁶² ²⁶³ ²⁶⁴ ²⁶⁵ ²⁶⁶ ²⁶⁷ ²⁶⁸ ²⁶⁹ ²⁷⁰ ²⁷¹ ²⁷² ²⁷³ ²⁷⁴ ²⁷⁵ ²⁷⁶ ²⁷⁷ ²⁷⁸ ²⁷⁹ ²⁸⁰ ²⁸¹ ²⁸² ²⁸³ ²⁸⁴ ²⁸⁵ ²⁸⁶ ²⁸⁷ ²⁸⁸ ²⁸⁹ ²⁹⁰ ²⁹¹ ²⁹² ²⁹³ ²⁹⁴ ²⁹⁵ ²⁹⁶ ²⁹⁷ ²⁹⁸ ²⁹⁹ ³⁰⁰ ³⁰¹ ³⁰² ³⁰³ ³⁰⁴ ³⁰⁵ ³⁰⁶ ³⁰⁷ ³⁰⁸ ³⁰⁹ ³¹⁰ ³¹¹ ³¹² ³¹³ ³¹⁴ ³¹⁵ ³¹⁶ ³¹⁷ ³¹⁸ ³¹⁹ ³²⁰ ³²¹ ³²² ³²³ ³²⁴ ³²⁵ ³²⁶ ³²⁷ ³²⁸ ³²⁹ ³³⁰ ³³¹ ³³² ³³³ ³³⁴ ³³⁵ ³³⁶ ³³⁷ ³³⁸ ³³⁹ ³⁴⁰ ³⁴¹ ³⁴² ³⁴³ ³⁴⁴ ³⁴⁵ ³⁴⁶ ³⁴⁷ ³⁴⁸ ³⁴⁹ ³⁵⁰ ³⁵¹ ³⁵² ³⁵³ ³⁵⁴ ³⁵⁵ ³⁵⁶ ³⁵⁷ ³⁵⁸ ³⁵⁹ ³⁶⁰ ³⁶¹ ³⁶² ³⁶³ ³⁶⁴ ³⁶⁵ ³⁶⁶ ³⁶⁷ ³⁶⁸ ³⁶⁹ ³⁷⁰ ³⁷¹ ³⁷² ³⁷³ ³⁷⁴ ³⁷⁵ ³⁷⁶ ³⁷⁷ ³⁷⁸ ³⁷⁹ ³⁸⁰ ³⁸¹ ³⁸² ³⁸³ ³⁸⁴ ³⁸⁵ ³⁸⁶ ³⁸⁷ ³⁸⁸ ³⁸⁹ ³⁹⁰ ³⁹¹ ³⁹² ³⁹³ ³⁹⁴ ³⁹⁵ ³⁹⁶ ³⁹⁷ ³⁹⁸ ³⁹⁹ ⁴⁰⁰ ⁴⁰¹ ⁴⁰² ⁴⁰³ ⁴⁰⁴ ⁴⁰⁵ ⁴⁰⁶ ⁴⁰⁷ ⁴⁰⁸ ⁴⁰⁹ ⁴¹⁰ ⁴¹¹ ⁴¹² ⁴¹³ ⁴¹⁴ ⁴¹⁵ ⁴¹⁶ ⁴¹⁷ ⁴¹⁸ ⁴¹⁹ ⁴²⁰ ⁴²¹ ⁴²² ⁴²³ ⁴²⁴ ⁴²⁵ ⁴²⁶ ⁴²⁷ ⁴²⁸ ⁴²⁹ ⁴³⁰ ⁴³¹ ⁴³² ⁴³³ ⁴³⁴ ⁴³⁵ ⁴³⁶ ⁴³⁷ ⁴³⁸ ⁴³⁹ ⁴⁴⁰ ⁴⁴¹ ⁴⁴² ⁴⁴³ ⁴⁴⁴ ⁴⁴⁵ ⁴⁴⁶ ⁴⁴⁷ ⁴⁴⁸ ⁴⁴⁹ ⁴⁵⁰ ⁴⁵¹ ⁴⁵² ⁴⁵³ ⁴⁵⁴ ⁴⁵⁵ ⁴⁵⁶ ⁴⁵⁷ ⁴⁵⁸ ⁴⁵⁹ ⁴⁶⁰ ⁴⁶¹ ⁴⁶² ⁴⁶³ ⁴⁶⁴ ⁴⁶⁵ ⁴⁶⁶ ⁴

Dienstag, 25. März 2014

Um Anmeldung bis 14. März 2014 wird
gebeten unter

GHW@lrz.uni-muenchen.de

Weitere Informationen unter:

<http://www.hgw.geschichte.uni-muenchen.de/aktuelles/index.html>

Benjamin Schönfeld, Benedikt Hotz



```
<p TEHeader>
  <witnessList parts="N" TEHeader
    TEHeader="db"/> * Signum
    <name key="3044" TEHeader
      <href> regis. </TEHeader
        doc>"pers">
          <name key="1792" TEHeader
            <href> consensu et subse
              type="page_db" href="2"
            <name key="2917" TEHeader
              <href> </witnessList>
        </p>
      </div>
    </body>
  </text>
```

Project Summary (1)

Data

- ≈ 1500 charters from three different archives¹
- About 30 pontificates, 11th - 12th century
- Mainly retro-digitalizations



Bundesministerium
für Bildung
und Forschung

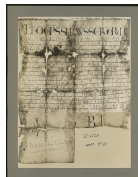


¹www.papsturkunden.de — lba.hist.uni-marburg.de — www.monasterium.de

Project Summary (2)

Team

- Chair of Medieval History (FAU)
Prof. Dr. Klaus Herbers,
Thorsten Schlauwitz, Viktoria Trenkle
- Institute of Paleography and
Diplomatics (LMU)
Prof. Dr. Fees,
Benedikt Hotz, Benjamin Schönfeld
- Pattern Recognition Lab (FAU)
Prof. Dr. Joachim Hornegger,
Dr. Elli Angelopoulou, Vincent Christlein
- www5.cs.fau.de/puhma



Project Summary (3)

Goals

- Analysis of the layout / writing style / temporal changes
- Supporting the transcription process
- Writer identification / verification



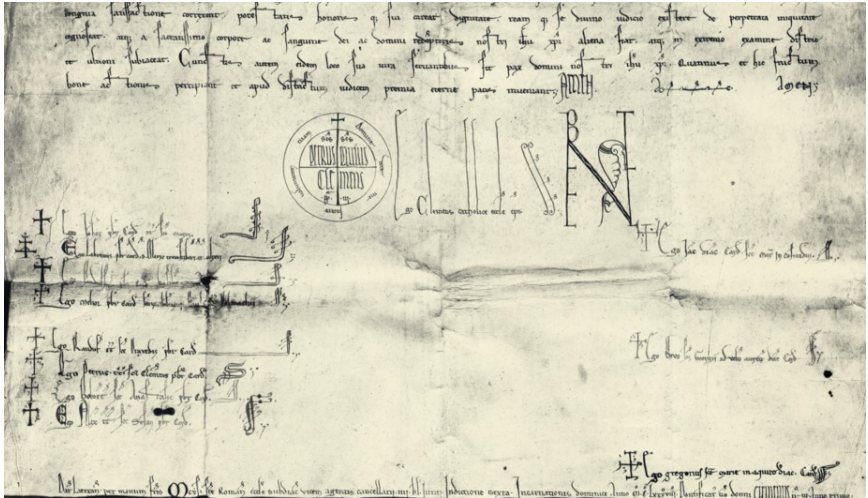
Project Summary (3)

Goals

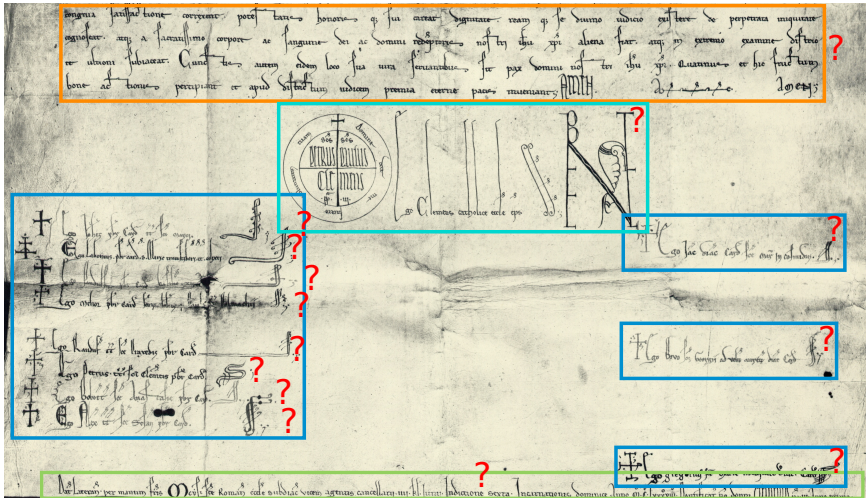
- Analysis of the layout / writing style / temporal changes
- Supporting the transcription process
- **Writer identification / verification**



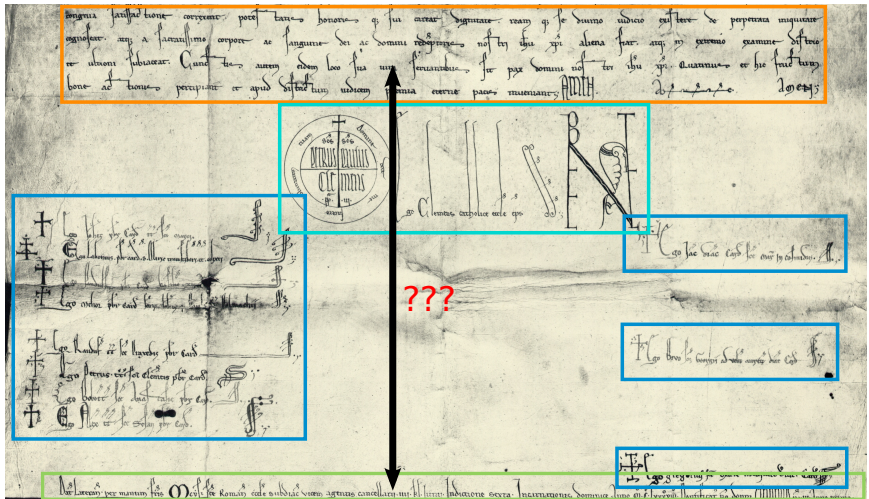
Motivation



Motivation - Writer Identification



Motivation - Writer Verification



Group of Methods

Textural based

Use global statistics to describe handwriting.



Allographic based

Writer is described by the vocabulary of letter parts.

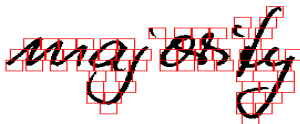
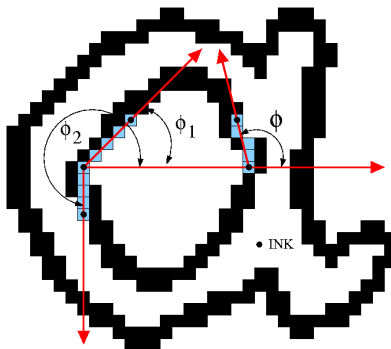


Image courtesy of Siddiqi and Vincent²

²I. Siddiqi and N. Vincent. "Text independent writer recognition using redundant writing patterns with contour-based orientation and curvature features". In: *Pattern Recognition* 43.11 (Nov. 2010), pp. 3853–3865

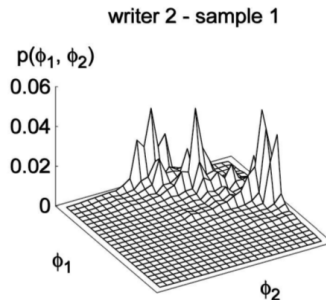
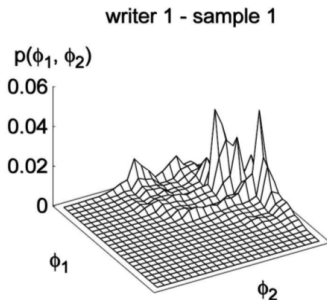
Examples for Textural based Methods



Hinge Feature³

³Marius Bulacu and Lambert Schomaker. "Text-independent writer identification and verification using textural and allographic features." In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29.4 (Apr. 2007), pp. 701–17

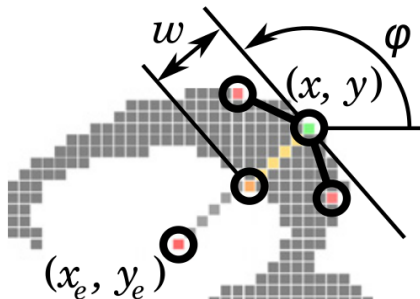
Examples for Textural based Methods



Hinge Feature - Probability Density Functions⁴

⁴A.A. Brink et al. "Writer identification using directional ink-trace width measurements". In: *Pattern Recognition* 45.1 (Jan. 2012), pp. 162–171

Examples for Textural based Methods



Quill Feature⁵

⁵A.A. Brink et al. "Writer identification using directional ink-trace width measurements". In: *Pattern Recognition* 45.1 (Jan. 2012), pp. 162–171

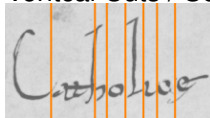
Allograph based Methods

Allograph Positions

- Grid



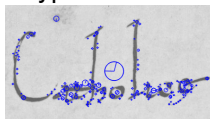
- Vertical Cuts / Seam Cuts



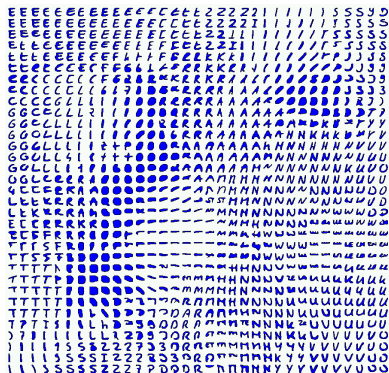
- Connected Components



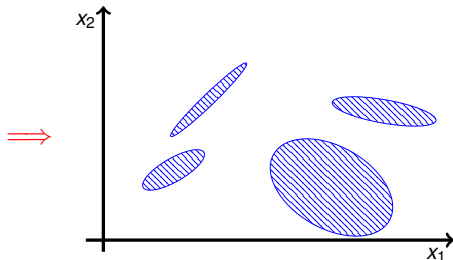
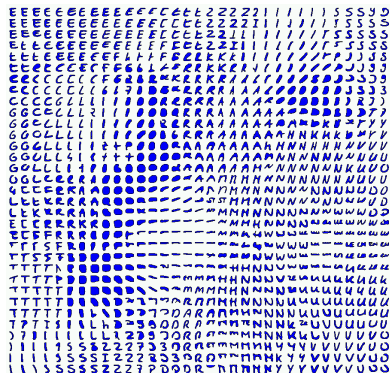
- Keypoint-based



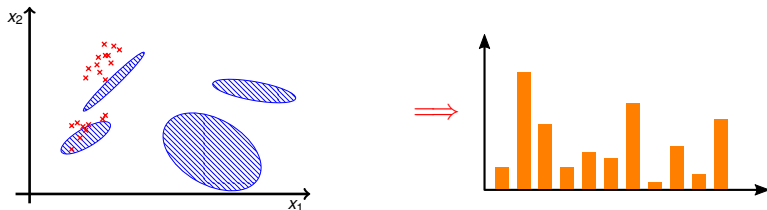
Universal Background Model (UBM)



Universal Background Model (UBM)



Encoding new Data



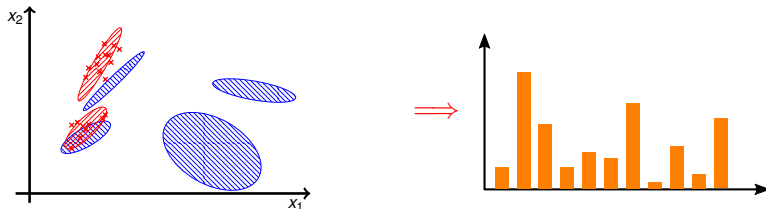
Methods

- Bag of (visual) Words⁶⁷

⁶Marius Bulacu and Lambert Schomaker. "Text-independent writer identification and verification using textural and allographic features." In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29.4 (Apr. 2007), pp. 701–17.

⁷Rajiv Jain and David Doermann. "Writer Identification Using an Alphabet of Contour Gradient Descriptors". In: *Document Analysis and Recognition (ICDAR), 2013 12th International Conference on*. Washington, NY, Aug. 2013, pp. 550–554.

Encoding new Data



Methods

- Fisher Vector⁸
- GMM Supervector⁹

⁸Stefan Fiel and Robert Sablatnig. "Writer Identification and Writer Retrieval using the Fisher Vector on Visual Vocabularies". In: *12th International Conference on Document Analysis and Recognition*. Washington DC, NY, Aug. 2013, pp. 545–549.

⁹V. Christlein et al. "Writer Identification and Verification Using GMM Supervectors". In: *(To Appear): IEEE Winter Conference On Applications of Computer Vision*. Steamboat Springs, CO, Mar. 2014.

Writer Identification (hard)

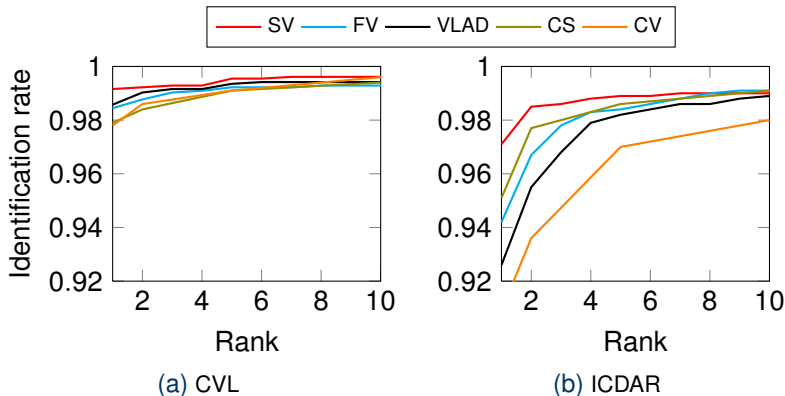
CVL

	Top-1	Top-2	Top-3	Top-4	mAP
CV ¹⁰	0.978	0.956	0.894	0.758	—
CS ¹¹	0.979	0.90	0.71	0.483	—
VLAD	0.986	0.954	0.871	0.720	0.936
FV	0.984	0.952	0.880	0.756	0.940
SV	0.992	0.981	0.958	0.887	0.971

¹⁰Stefan Fiel and Robert Sablatnig. "Writer Identification and Writer Retrieval using the Fisher Vector on Visual Vocabularies". In: *12th International Conference on Document Analysis and Recognition*. Washington DC, NY, Aug. 2013, pp. 545–549.

¹¹Rajiv Jain and David Doermann. "Writer Identification Using an Alphabet of Contour Gradient Descriptors". In: *Document Analysis and Recognition (ICDAR), 2013 12th International Conference on*. Washington, NY, Aug. 2013, pp. 550–554.

Writer Identification (soft)



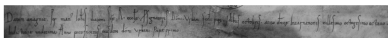
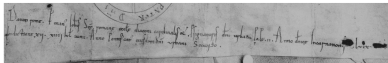
Writer Verification in Papal Charters

John of Gaeta

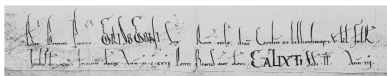
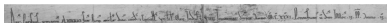
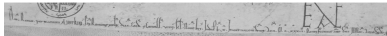
All Others

Writer Verification in Papal Charters

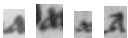
John of Gaeta



All Others

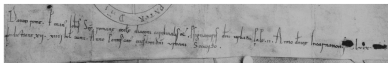


Letter based writer verification (60% accuracy)

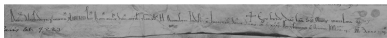
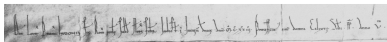
- Often bad quality 
- Letter formation not individual enough
- Features not discriminative enough

Writer Verification in Papal Charters

John of Gaeta



All Others

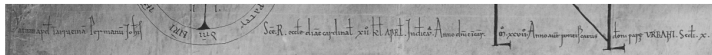
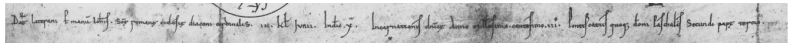


Texture based writer verification (80% accuracy)

- + Nearly no annotation needed
- May classify background instead of handwriting

Method Idea

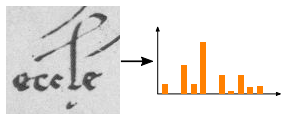
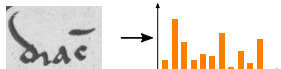
Method Idea



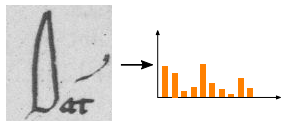
Encoding Example - Algorithm

⋮

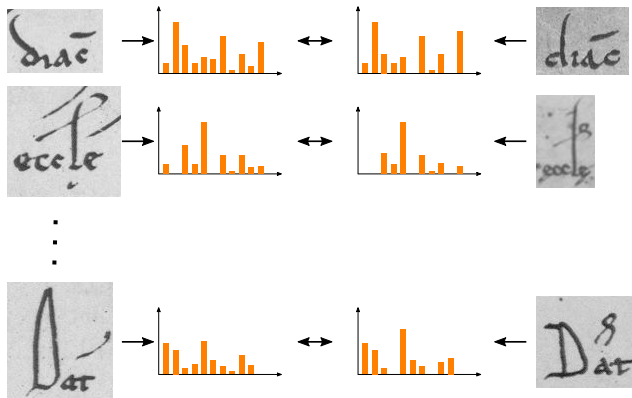
Encoding Example - Algorithm



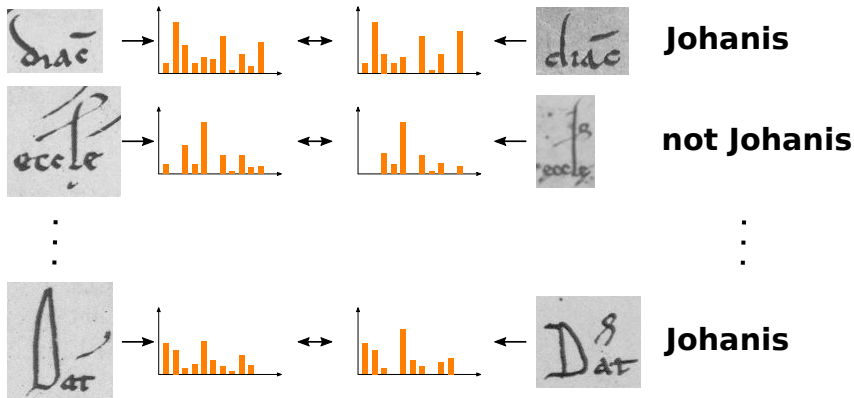
⋮



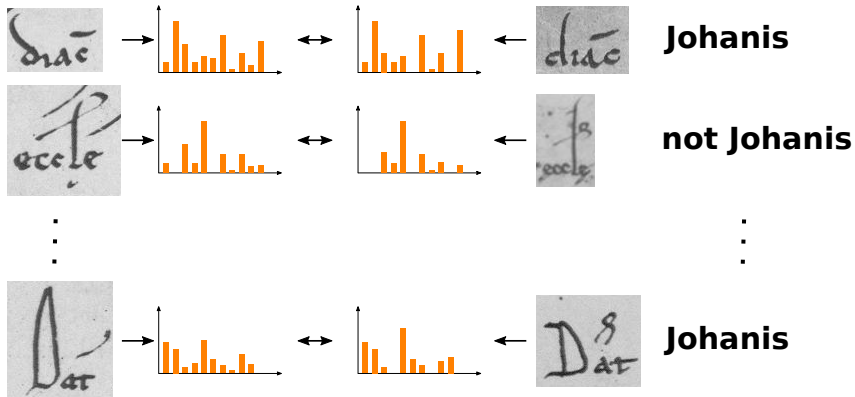
Encoding Example - Algorithm



Encoding Example - Algorithm

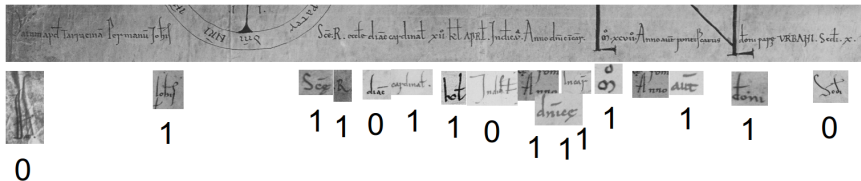


Encoding Example - Algorithm

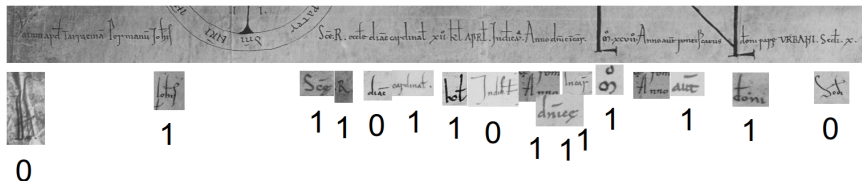


In average: 96.6% Accuracy

Failure Case



Failure Case



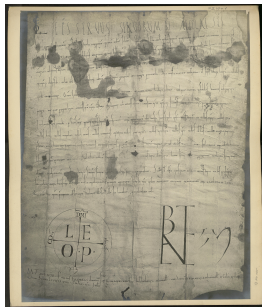
Best match might not be the correct one

- Take distances into account
- Manual inspection of in-between cases

Conclusion

Summary

- Many different writer identification methods
- Do not blindly trust the output
 - PC can only give probabilities
 - However, it can provide you with a confidence measure



Future Work

- Method combinations
- Evaluation on historical datasets → many new challenges



Thank you for your attention!

