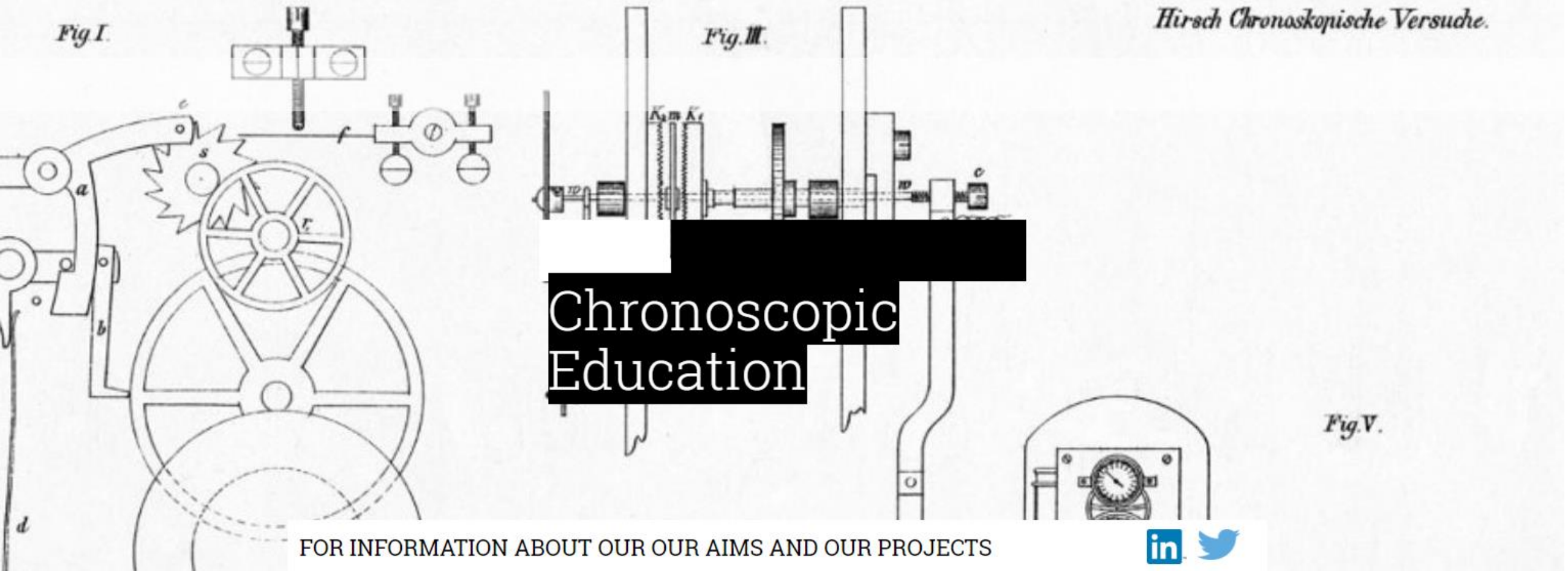


Chronoscopic Education

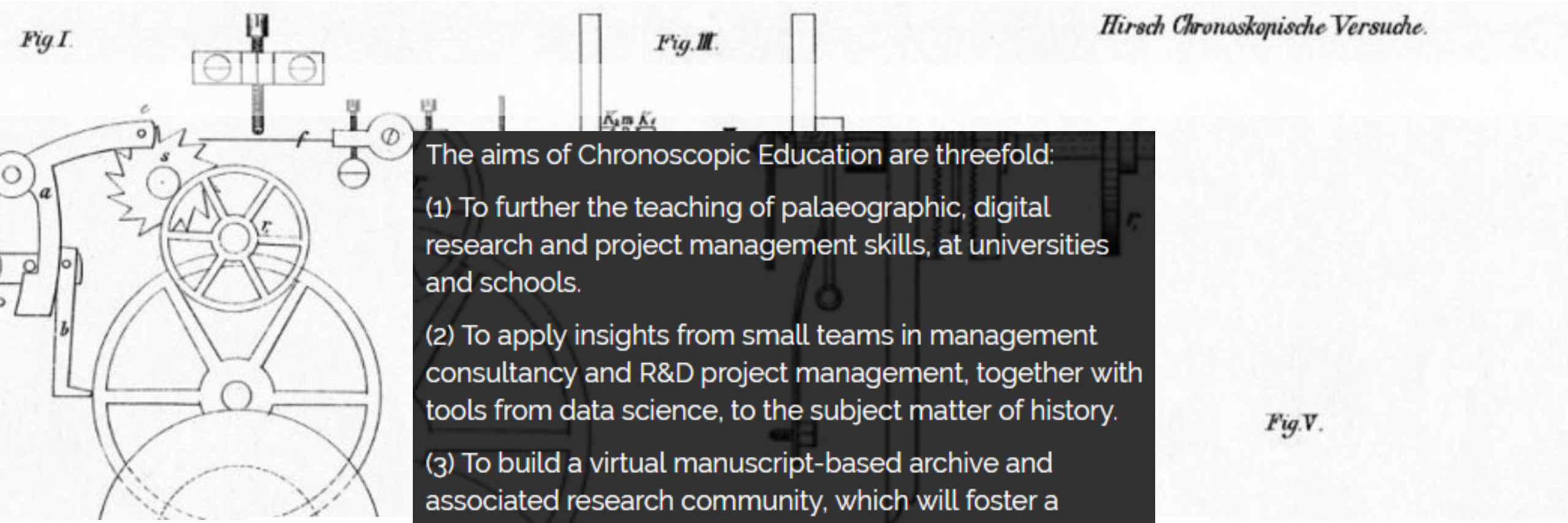
FOR INFORMATION ABOUT OUR OUR AIMS AND OUR PROJECTS

Signs of Literacy
Kaggle Research Competition

Colin Greenstreet

Tuesday, June 26th, 2018

Our social aims



The aims of Chronoscopic Education are threefold:

(1) To further the teaching of palaeographic, digital research and project management skills, at universities and schools.

(2) To apply insights from small teams in management consultancy and R&D project management, together with tools from data science, to the subject matter of history.

(3) To build a virtual manuscript-based archive and associated research community, which will foster a culture of collaborative scholarship.

http://www.chronoscopic.org

Project portfolio

MarineLives

Signs of Literacy

Maphackathon

EM Textiles, Garments & Dyestuffs Glossary

EM Maritime & Mercantile Gazetteer

http://www.chronoscopic.org

Signs of Literacy Kaggle Research Competition, Nov 2018 – Jan 2019

Google owned Kaggle has selected us as one of a small number of pro bono competitions they support each year on the merits of our proposal, and the potential impact on the research field and community of the competition.

Kaggle will cover the running costs of the competition. We will provide the prize pool, and are now seeking to raise US $30,000 from potential sponsors and partners.

The Proof of Concept will contain two parts:

(1) Algorithmic identification of markes, initials and signatures.

(2) Algorithmic discrimination between degrees of "sophistication" within the three categories of "marke"; "initial(s)", and "signature".

Having proven the concept, we will seek out an image or vision oriented computational laboratory with which to develop a grant funded collaboration to take the work further in 2019 and beyond.

https://www.linkedin.com/pulse/proposed-signs-literacy-kaggle-research-competition-2018-greenstreet/

User requirements

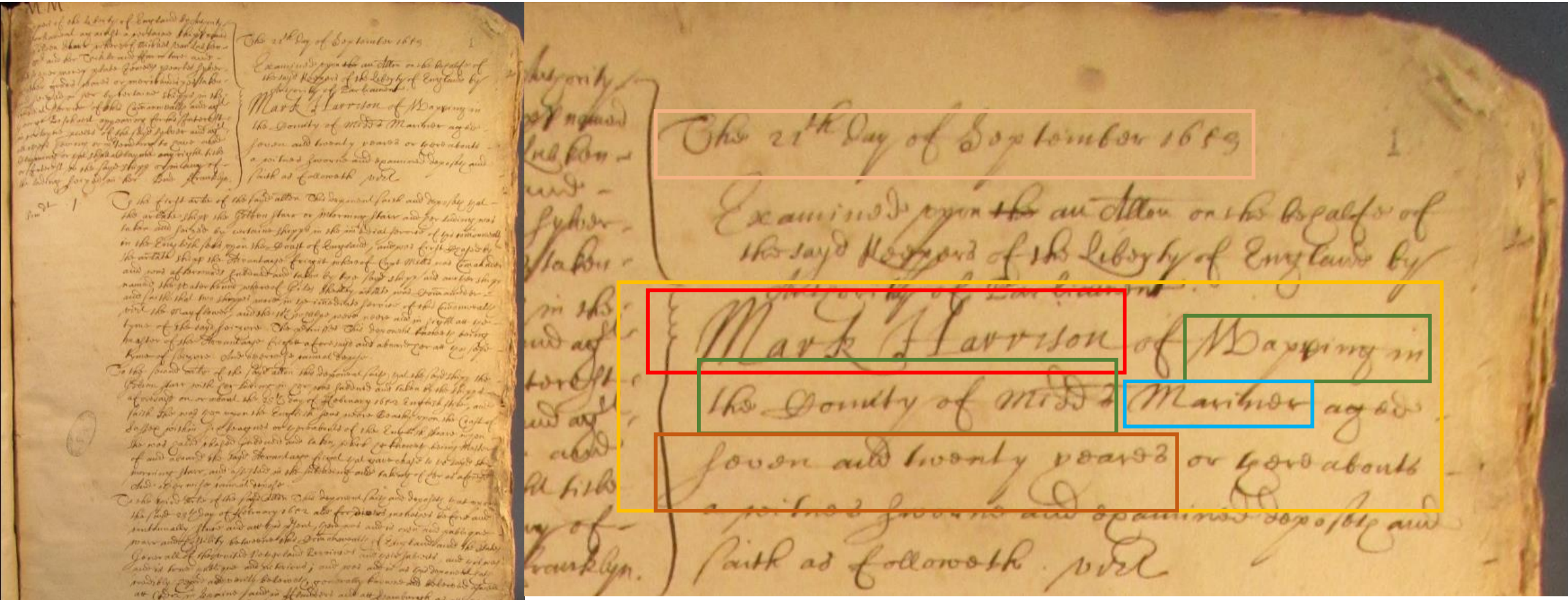
- Automatic identification of manuscript pages containing signoffs
- Markup of manuscript pages to isolate signoffs (markes, initials, signatures)
  - Hand markup of manuscript pages, but ideally automated markup
  - Signoffs can be single or multiple, for single or multiple depositions
  - Deponent signoffs; interpreter signoffs
- Automatic differentiation between classes of markes, initials and signatures
- Automatic differentiation within each class as to sophistication of execution & other parameters as a surrogate for literacy
- Automatic identification of manuscript pages containing deponent metadata (name; age; occupation; place of residence; date of deposition)
- Markup of manuscript pages to isolate deponent metadata
- Hand writing text recognition of deponent metadata and associate metadata with correct signoff
- Geo-referencing of signoffs
  - Display signoffs and/or manuscript pages containing signoffs by georeferences

Legal deposition

Deposition of Mark Harrison; mariner and master; resident in Wapping, Middlesex; age 27;
Dated September 21st 1659 (TNA, HCA 13/68, ff. 1r-3r)

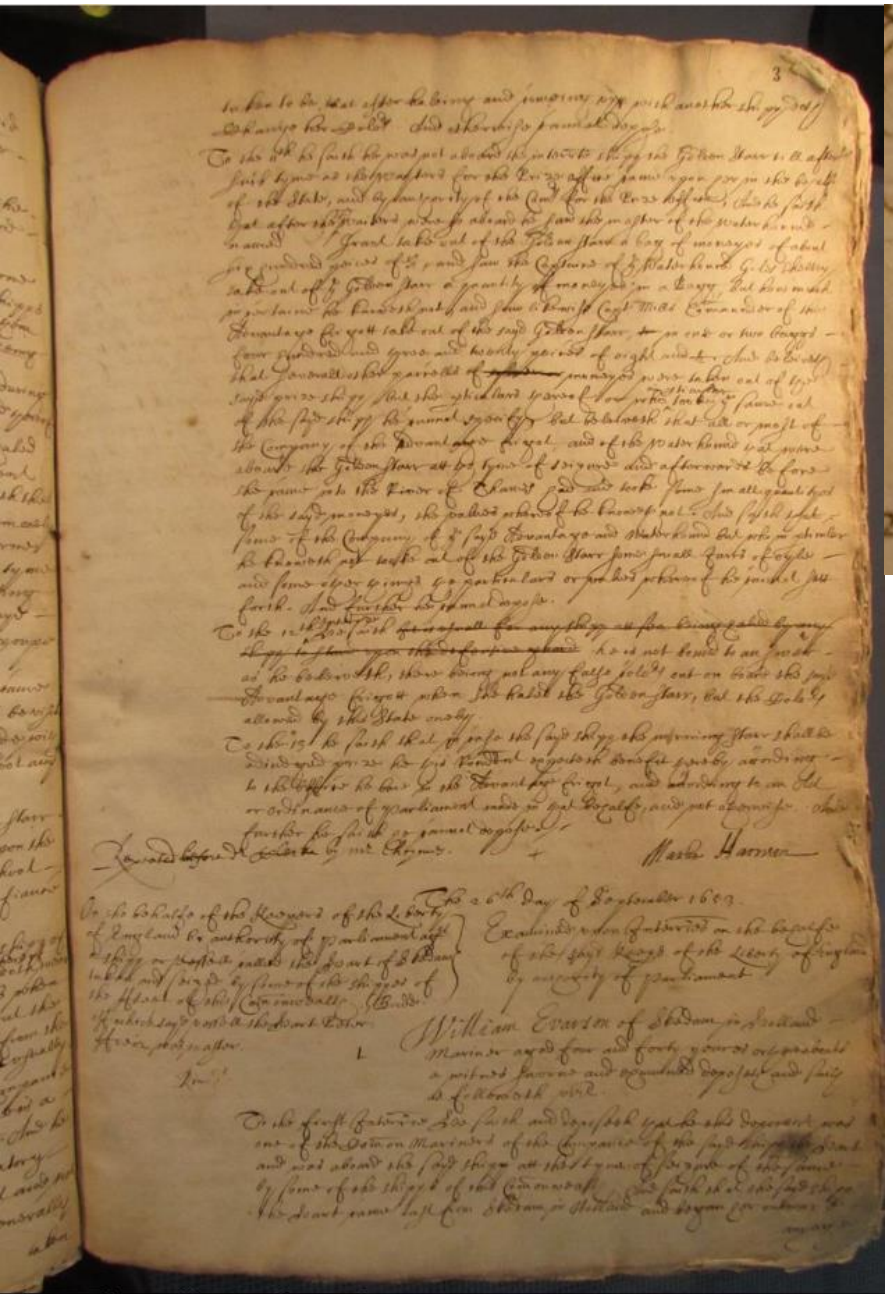
http://www.marinelives.org/wiki/HCA_13/68_f.1r_Annotate; http://www.marinelives.org/wiki/HCA_13/68_f.1v_Annotate; http://www.marinelives.org/wiki/HCA_13/68_f.2r_Annotate ; http://www.marinelives.org/wiki/HCA_13/68_f.2v_Annotate; http://www.marinelives.org/wiki/HCA_13/68_f.3r_Annotate

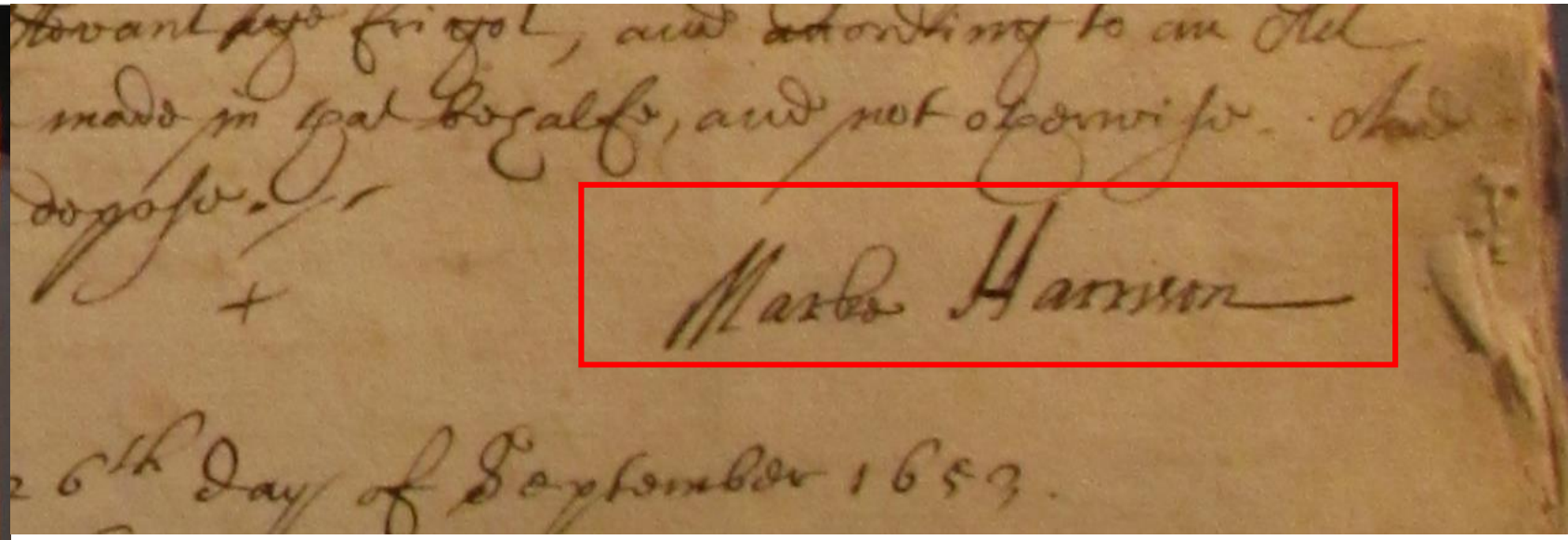
Metadata



http://www.marinelives.org/wiki/HCA_13/68_f.1r_Annotate;
http://www.marinelives.org/wiki/Mark_Harrison

Signoff

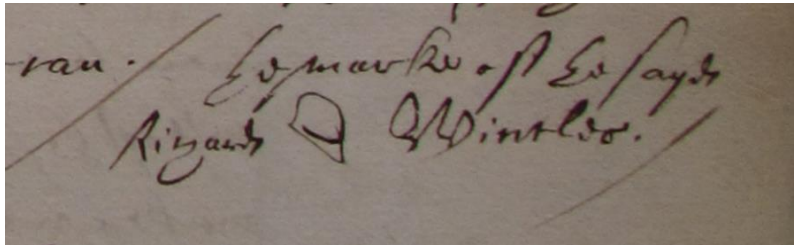




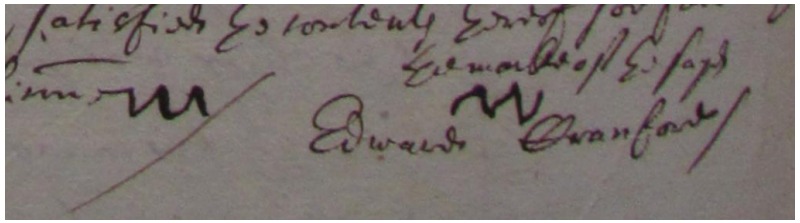
http://www.marinelives.org/wiki/HCA_13/68_f.3r_Annotate

Porters handling coals, whale oil, ginger & corn

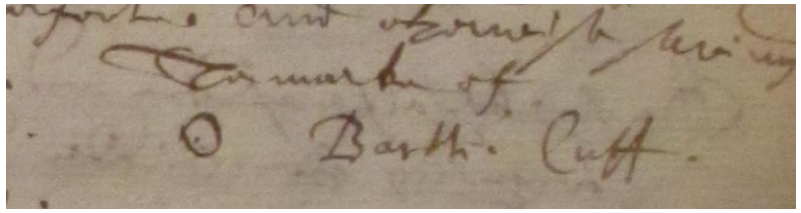
Richard Wincles, thirty-three year old porter, of the parish of Stepney, Middlesex, Dec. 15, 1656; employed as a labourer with fifteen other men to unload coals from the *Imployment* moored near Execution Dock, Wapping, into lighters for fixed rate of 12 *s* per man (HCA 13/70 f.554r)



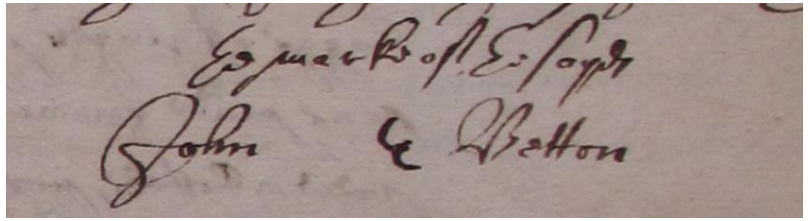
Edward Cranford, forty-four year old coale heaver or porter, of the parish of Stepney, Middlesex, Dec. 15, 1656; employed as a labourer with fifteen other men to unload coals from the *Imployment* moored near Execution Dock, Wapping, into lighters for fixed rate of 12 *s* per man (HCA 13/70 f.555v)



Bartholomew Cuff, sixty year old porter of the Stillyard, of the parish of Allhallowes the Greate, London, May 15, 1658; assisted in the landing of whale oil from lighters at the Stillyard Key and loading them away into a warehouse (HCA 13/70 f.555v)

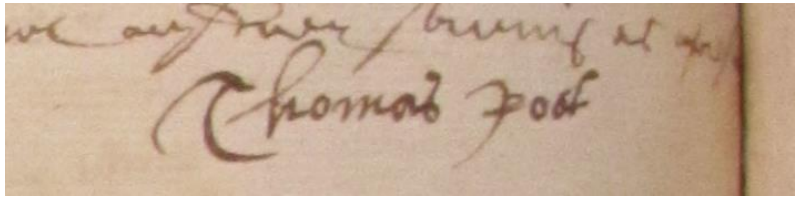


John Betton, fifty-four year old citizen and white baker of London, of the parish of Saint Buttolph Algate, London, Jul. 31, 1655; self-described as a porter employed by the Commissioners for Prize Goods to deliver ginger from a warehouse at Ralphes Key (HCA 13/70 f.449r)

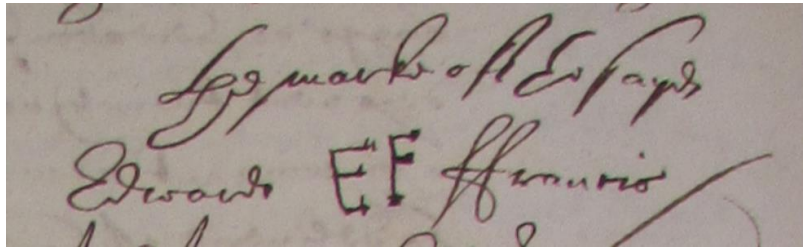


Edward Sherwin, fifty-six year old cittizen and leatherseller, of the parish of Little Allhallowes, London, Jul. 31, 1655; self-described as a porter employed by the Commissioners for Prize Goods to deliver ginger from a warehouse at Ralphes Key (HCA 13/70 f.449v)

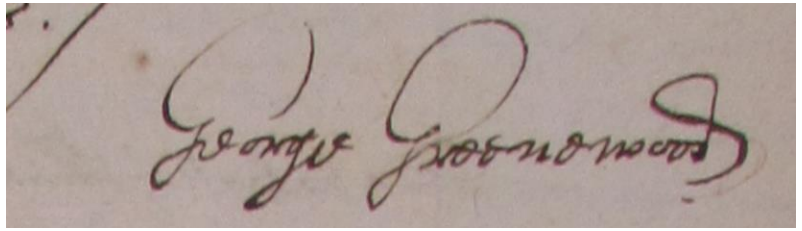
Thomas Roots, twenty-nine year old porter, of the parish of Greate Allhallowes, London, May 15, 1658; assisted in the landing of whale oil from lighters at the Stillyard Key, as one of the Stillyard porters, and loading them away into a warehouse (HCA 13/72 f.330v)



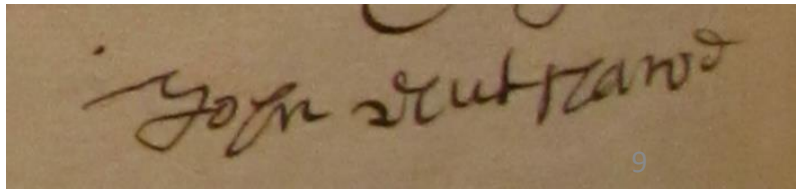
Edward ffrancis, citizen and merchantaylor of London, of the parish of Saint Olave in Southwarke, Jul. 31, 1655; self-described as a porter employed by the Commissioners for Prize Goods to deliver ginger from a warehouse at Ralphes Key (HCA 13/70 f.450v)



George Greenwood, thirty year old citizen and vintner of London, of the parish of Saint Buttolph Bishopsgate, London, Jul. 31, 1655; self-described as a porter employed by the Commissioners for Prize Goods to deliver ginger from a warehouse at Ralphes Key (HCA 13/70 f.454r)

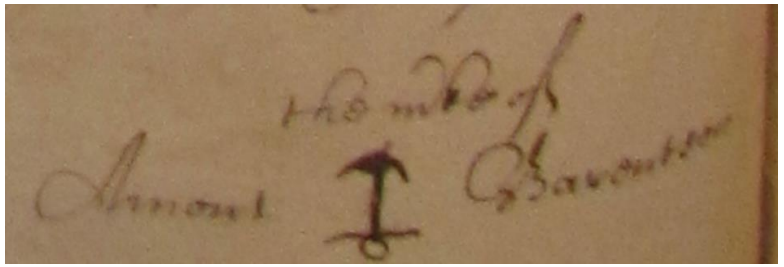


John Nutshall, fifty-five year old corne porter, of the parish of Saint Saviours Southwarke, Nov. 19, 1653; employed with a barber chyrurgeon/corne meter, an additional corne-meter, and other labourers to unlade a cargo of what in the *ffortune* of Stettin, moored against Limehouse; eight years of experience as a corne porter (HCA 13/70 f.352v)

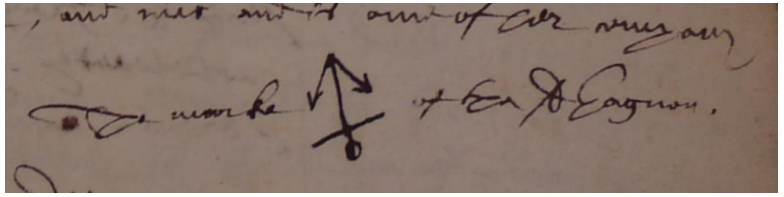


Anchors

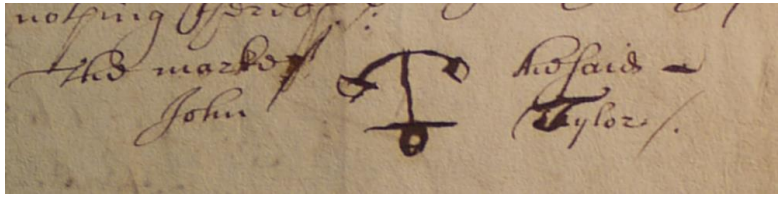
Amons Barentsen, thirty-five year old mariner, of Copenhagen, Denmark, October 13th, 1653; self-described as an "ordinary mariner", hired to sail from the Sound to Hamburg on the *Golden Hawke* of Stockholm (HCA 13/68 f.81v)



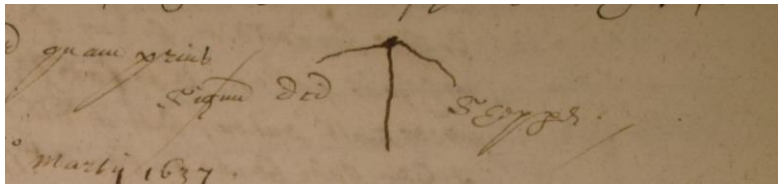
Claude de Gagnon, twenty-five year old mariner, of Melon, near Brest in Britanny, May 22nd, 1656 (HCA 13/71 f.225r)



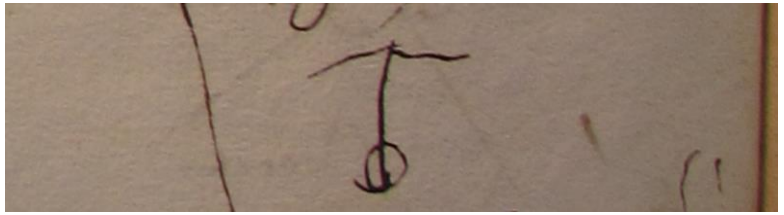
John Tylor, forty-two year old shipwright, of Lower Shadwell, in the parish of Stepney, Middlesex, February 14th, 1659 (HCA 13/73 f.36r)



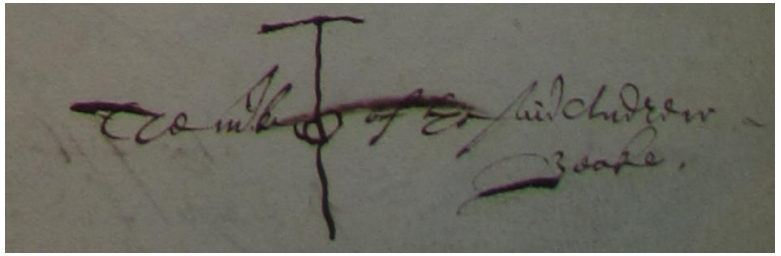
Richard Shepperd, fifty-eight year old cooke, of Brixton, Devon, March 29th, 1637; self-described cooke of the *Hope* of Ipswich (HCA 13/53 f.87r)



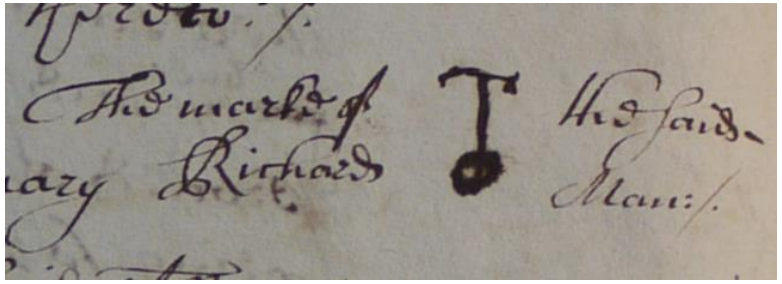
Andrew Beake, thirty-six year old lookeinglassemaker and formerly seaman, of Rose alley without Bishopsgate, London, January 21st, 1655 (HCA 13/70 f.252v)



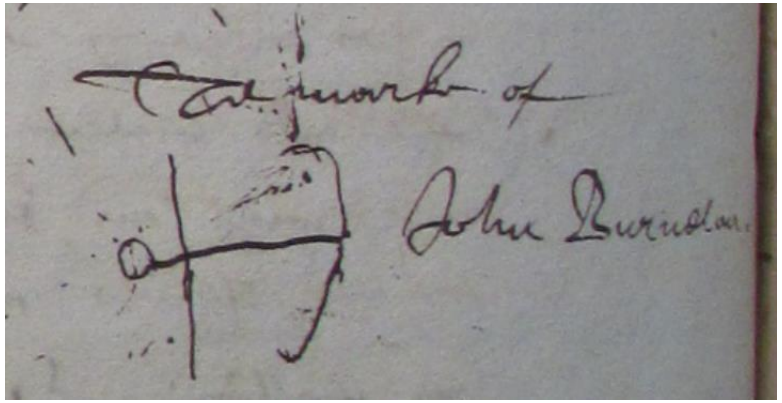
Andrew Beake, thirty-six year old looking-glasse maker, of Rose-Alley in Bishopsgate streete, London, February 13th, 1655 (HCA 13/70 f.252v)



Richard Man, thirty-one year old mariner, of Southampton, January 8th, 1659; self-described common man of the *Lisbone ffrigott* on voyage to Oratava (HCA 13/73 f.26v)



John Burnelau, twenty-eight year old sailor, of Mornar, France,March 30th, 1661 (HCA 13/73 f.486v)

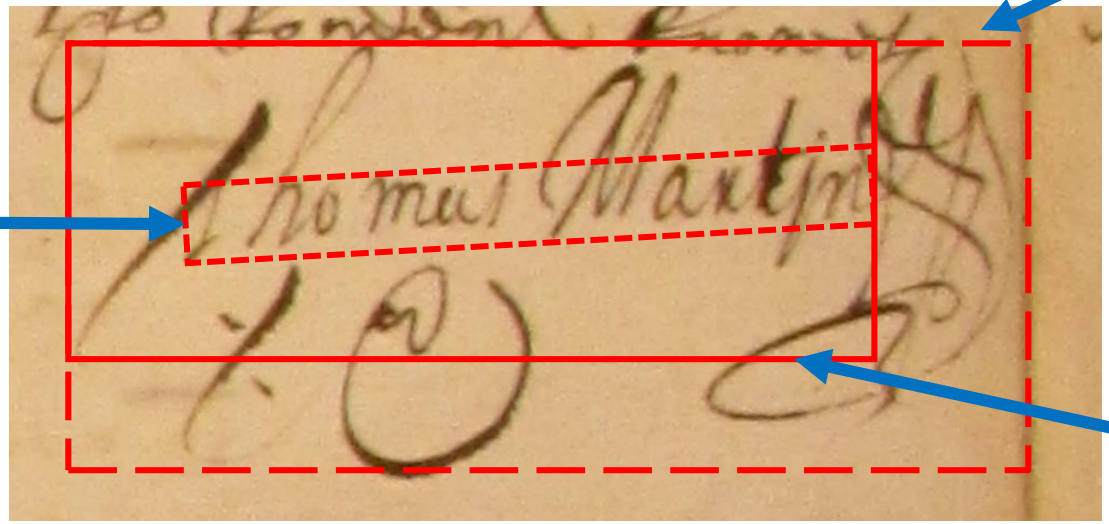


Putting boundary boxes on C17th signatures

Boundary boxes marking the visual geometry of a signature

Outside boundary box, including flourish

Inside boundary box, excluding uppers and downers

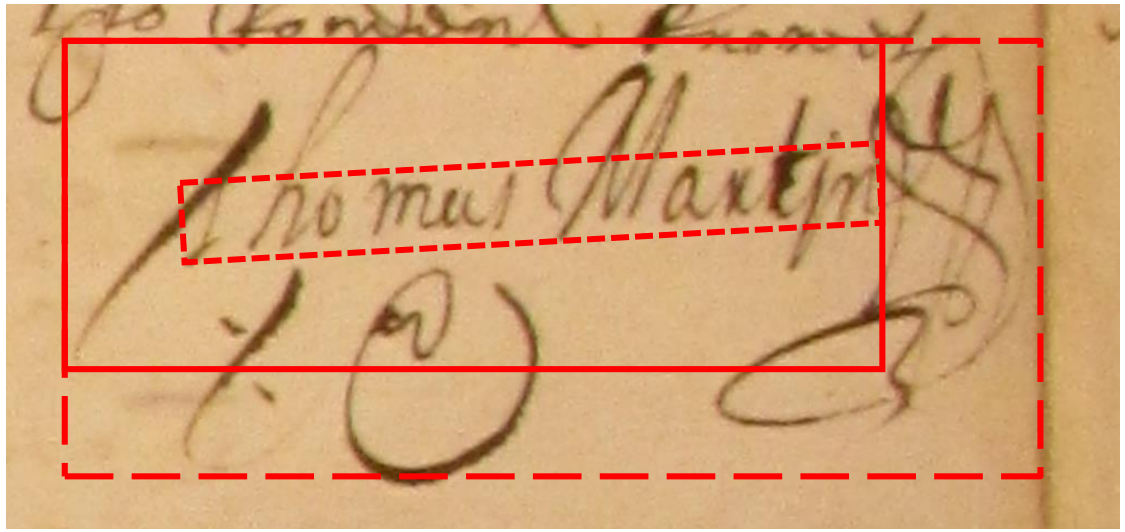


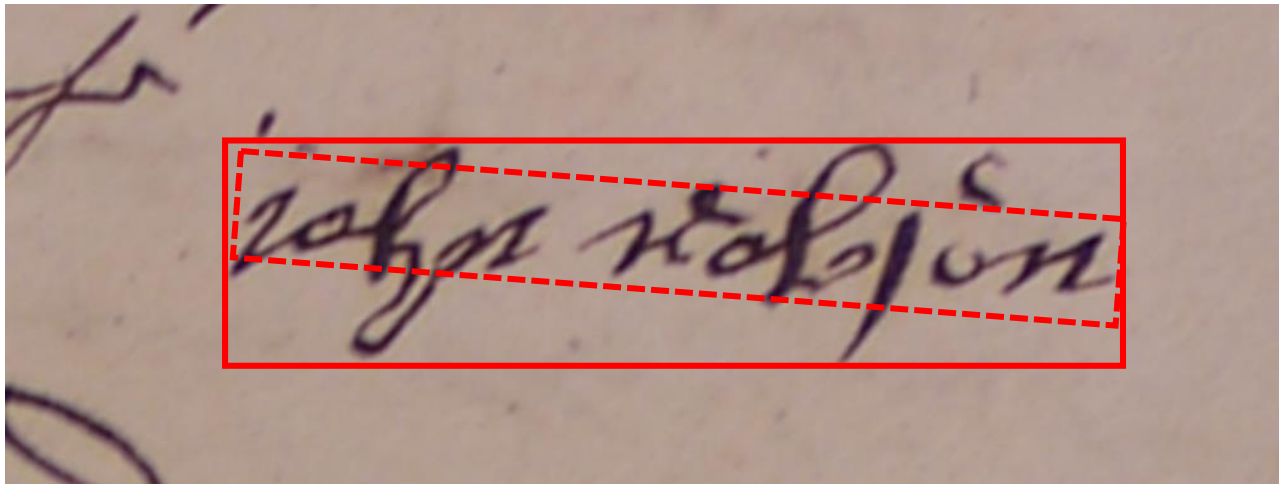
Middle boundary box, including all letters, but excluding flourish

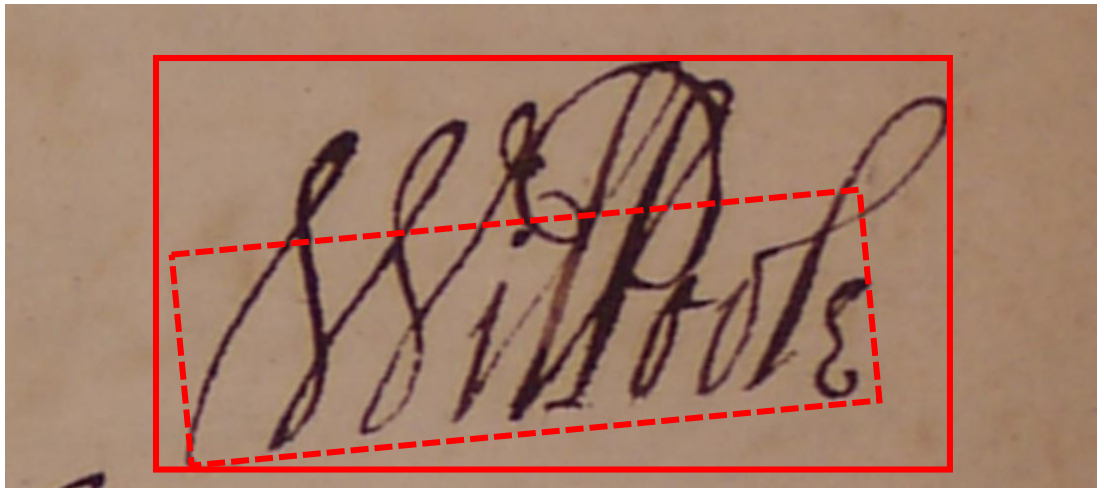
Statistics

Inside boundary box: 9.0 x 1.1
Middle boundary box: 9.75 x 4.25
Outside boundary box: 12.75 x 5.75
Rotation from horizontal: ca. 340 degrees

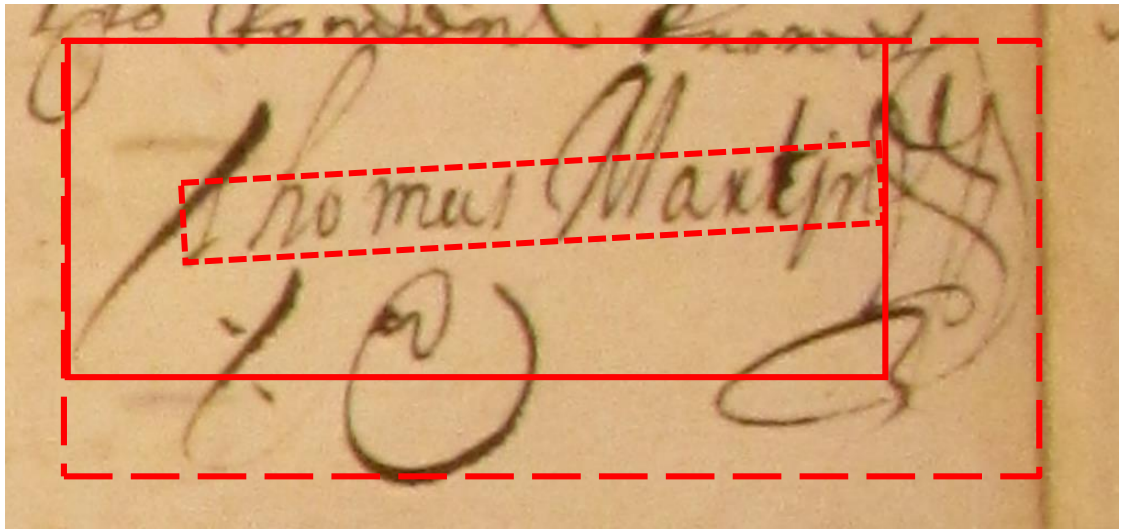
Different visual geometries of signatures

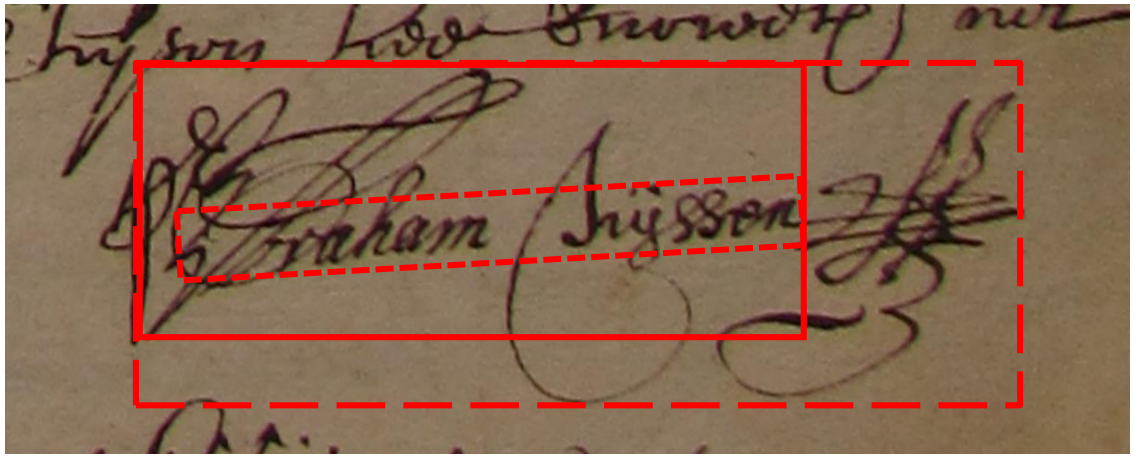


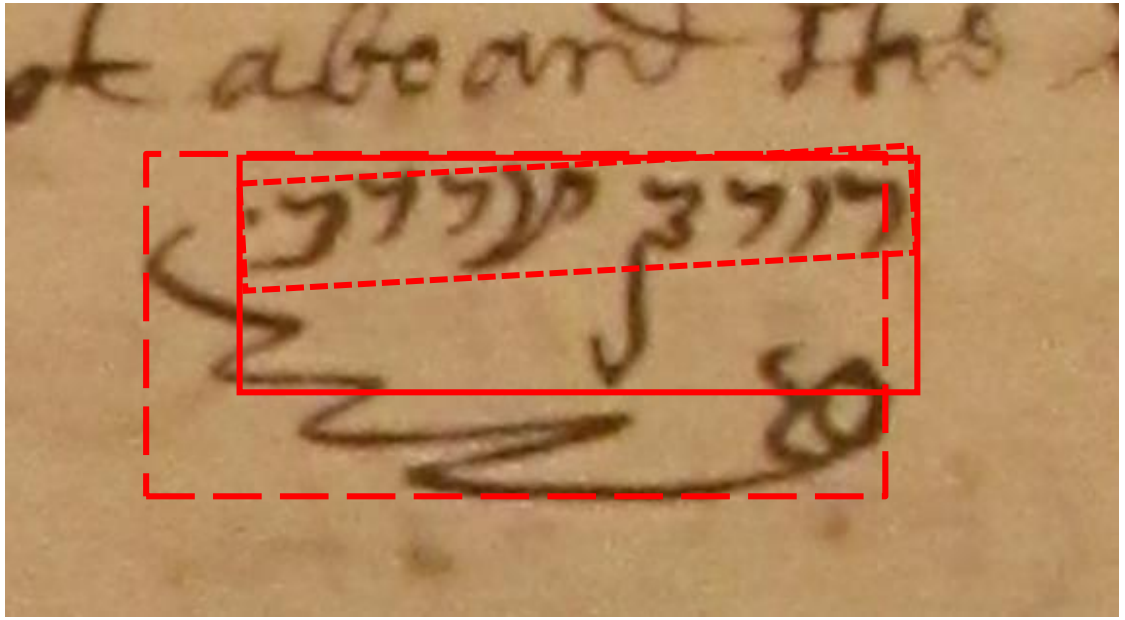


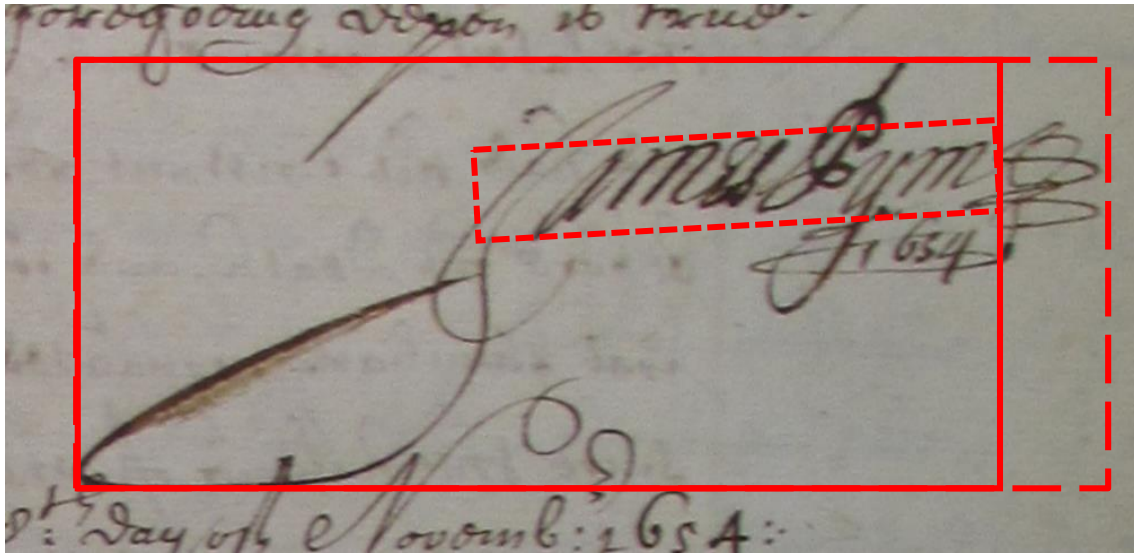


Visual geometries of flourishes – C17th Irish, Dutch, English & Moroccan merchants



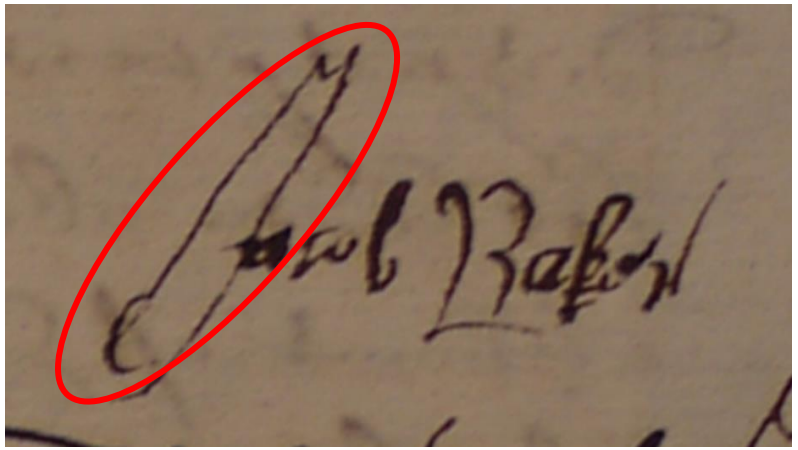


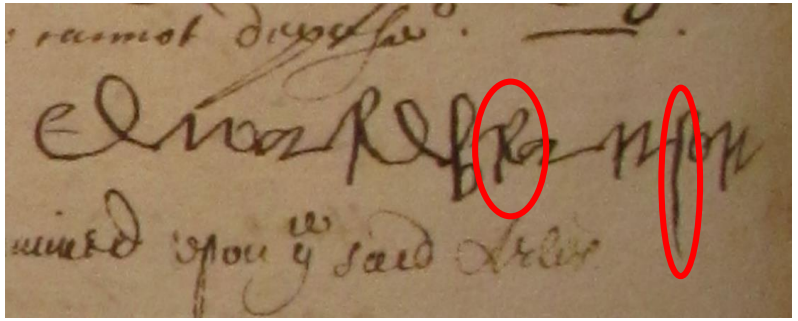


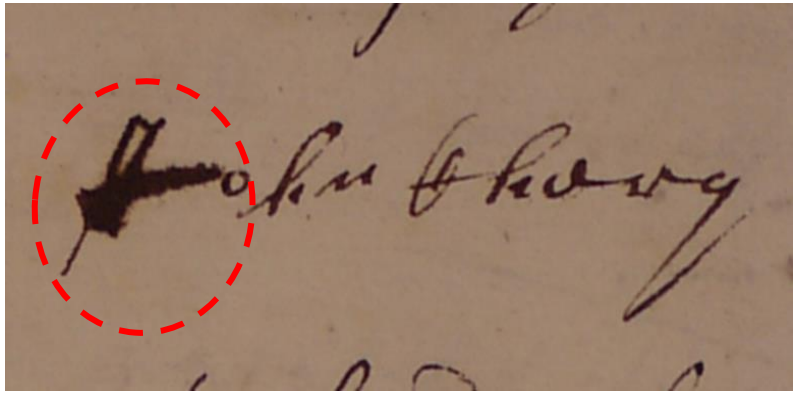


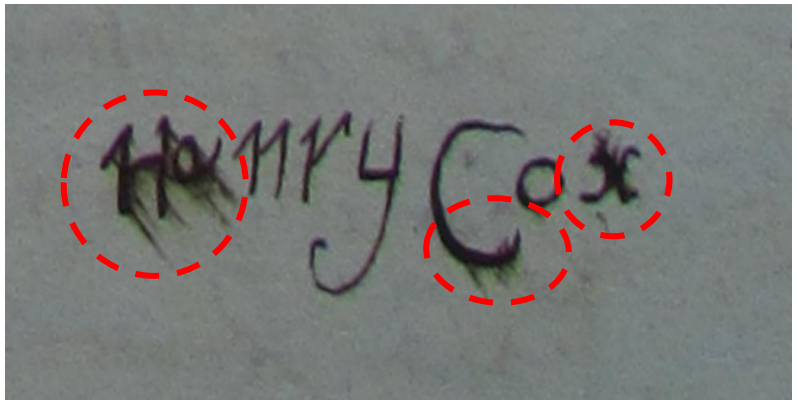
Clockwise, from top LH: KaggleTestSnippet_HCA_1368_f.34v.PNG, KaggleTestSnippet_HCA_1370_f.366r.PNG, KaggleTestSnippet_HCA_1370_f.134r.PNG, KaggleTestSnippet_HCA_1368_f.58r.PNG

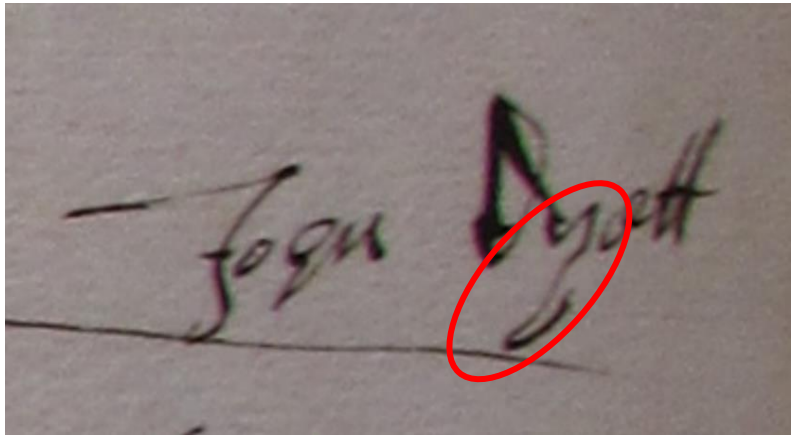
Physical characteristics of poorly executed signatures for machine detection

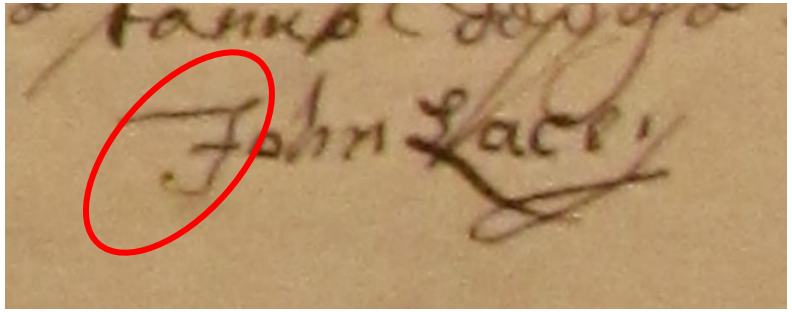










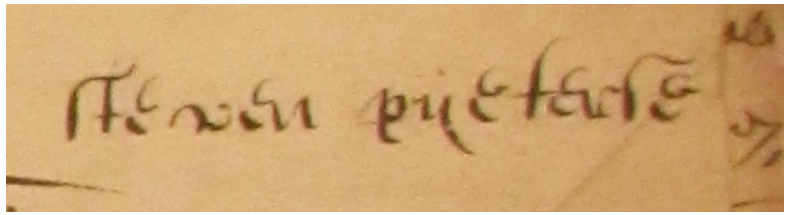


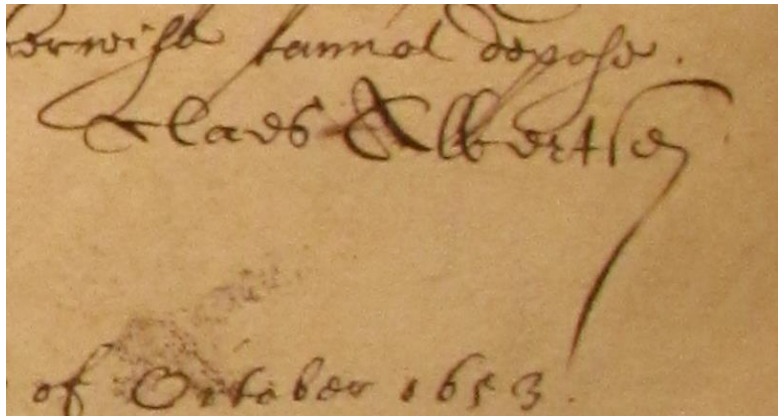
Shaky straight lines and/or loops

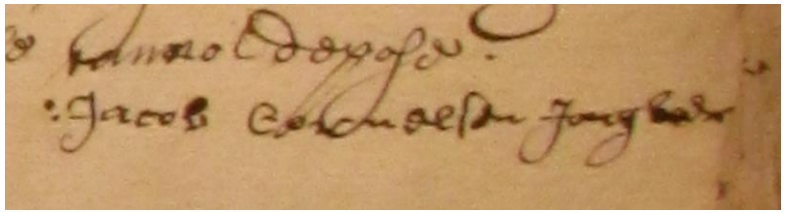
Ink blots or smudges

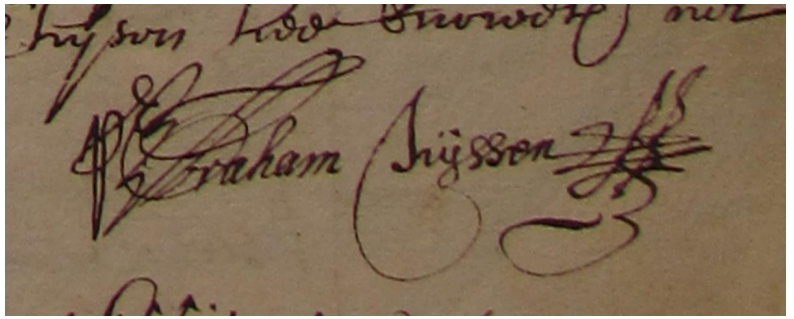
Clockwise, from top LH: KaggleTestSnippet_HCA_1371_f.263v.PNG, KaggleTestSnippet_HCA_1368_f.483v.PNG, KaggleTestSnippet_HCA_1371_f.456r.PNG, KaggleTestSnippet_HCA_1368_f.51v.PNG, KaggleTestSnippet_HCA_1370_f.168v.PNG, KaggleTestSnippet_HCA_1370_f.167r.PNG

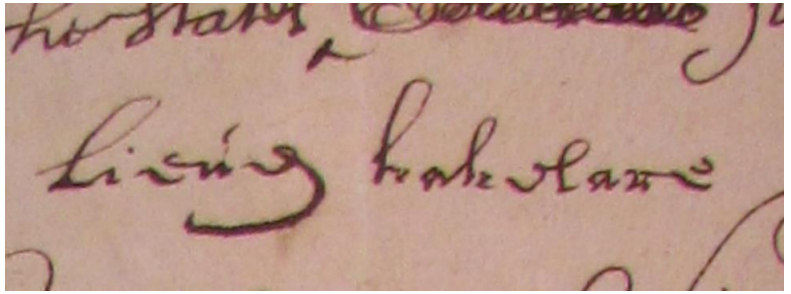
Was it less common to use capital letters in Dutch rather than English language signatures in the C17th

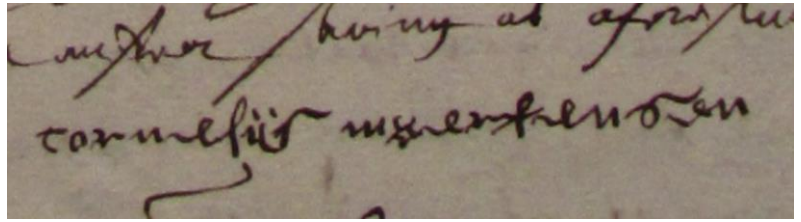


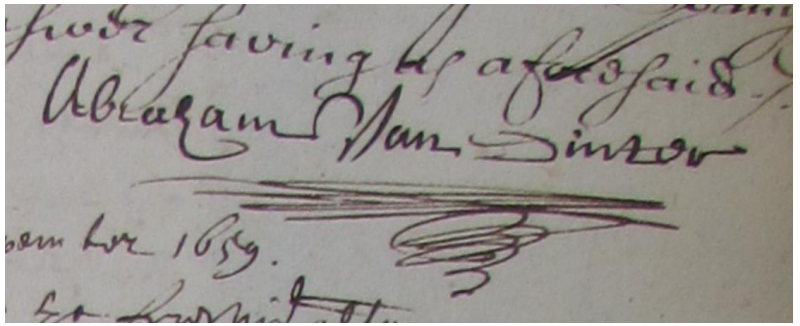


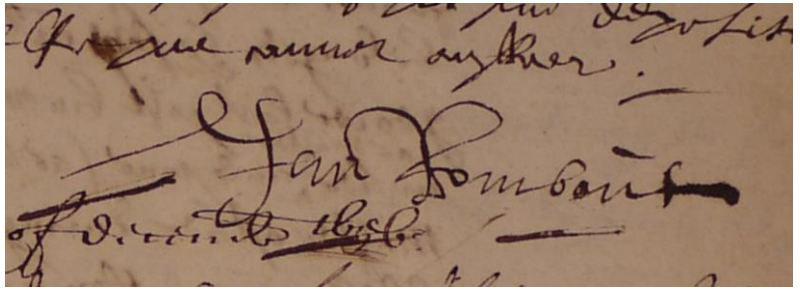




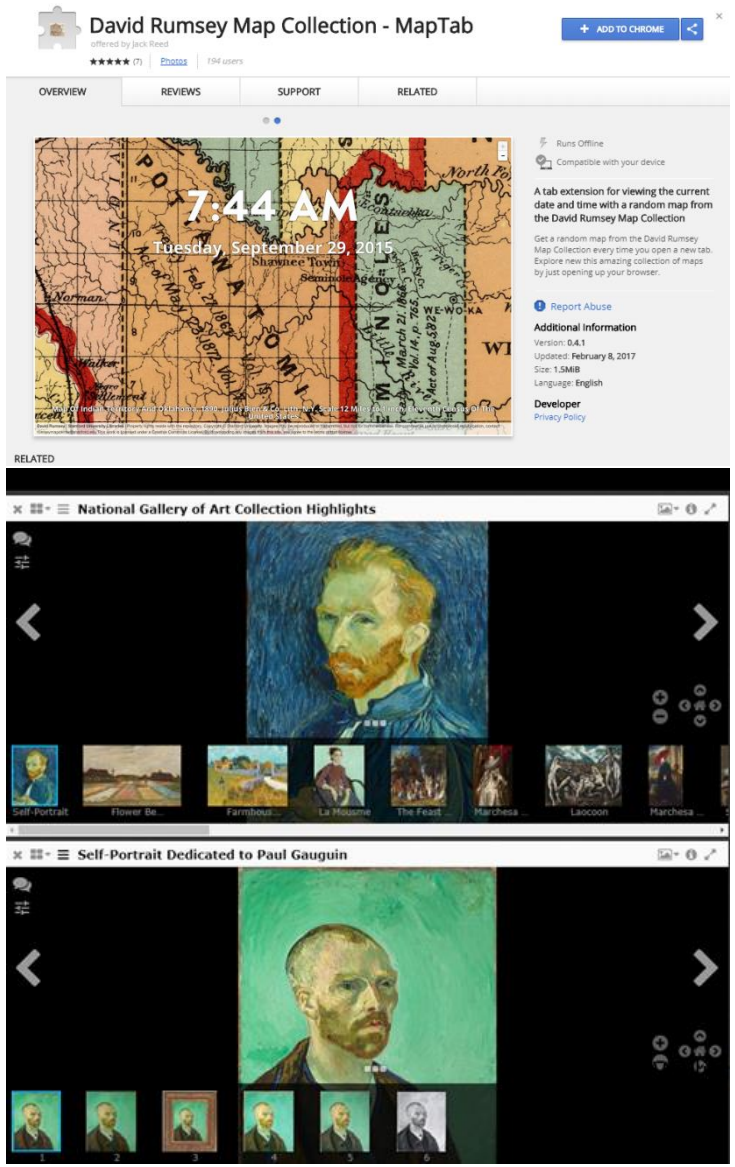


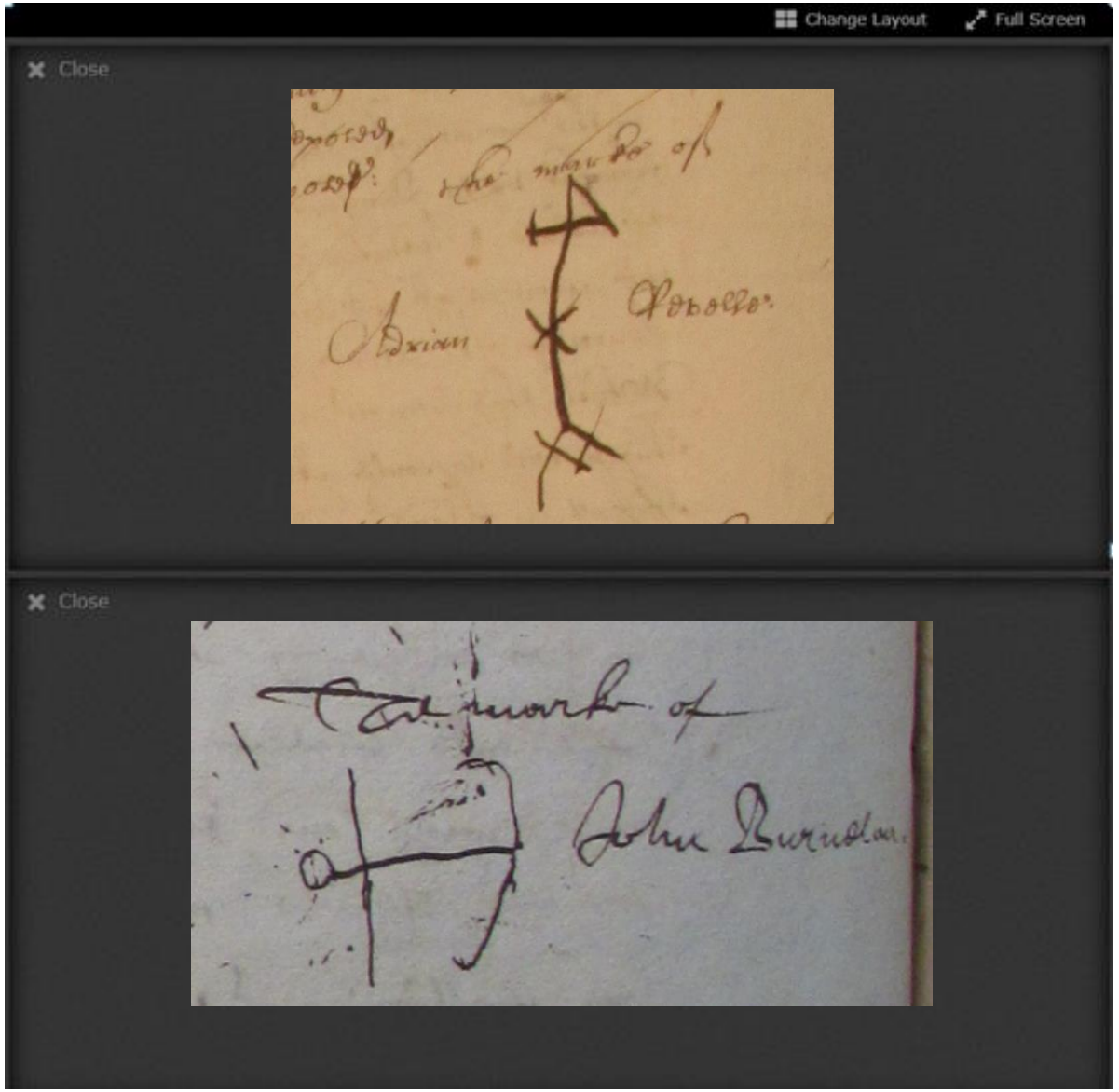






Conjoint analysis IIIF viewer plugin





Adrian Revelle, twenty-three year old mariner, of Dunquirke in fflanders, November 12th, 1653; “hee onely speaketh the flemmish speech” (HCA 13/68 f.183v)

John Burnelau, twenty-eight year old sailor, of Mornar, France,March 30th, 1661 (HCA 13/73 f.486v)

Jack Reed, David Rumsey MapTab Chrome extension; http://projectmirador.org/demo/; Conjoint analysis IIIF viewer plugin vapourware

Reading

Colin Greenstreet, Pattern recognition of signatures and markes in historical manuscriopts as the basis for sub-population recognition, March 2018 [available Signsofliteracy Github repository: Signsoliteracy/Signoff]

Colin Greenstreet, C17th alphabet of initials, 4th edn., April 4th, 2018 [available Signsofliteracy Github repository: Signsoliteracy/Signoff]

Mark Hailwood, 'The Rabble that Cannot Read', Ordinary Peoples Literacy in Seventeenth-Century England, October 13th, 2014

David Cressy, Literacy and the Social Order: Reading & Writing in Tudor and Stuart England, 1980

Reading

IAM Handwriting Database 3.0

- The IAM Handwriting Database contains forms of handwritten English text which can be used to train and test handwritten text recognizers and to perform writer identification and verification experiments.
- The database was first published in [1] at the ICDAR 1999.
- The database contains forms of unconstrained handwritten text, which were scanned at a resolution of 300dpi and saved as PNG images with 256 gray levels. The figure below provides samples of a complete form, a text line and some extracted words.
- FKI: Research Group on Computer Vision and Artificial Intelligence INF, University of Bern

U. Marti and H. Bunke. A full English sentence database for off-line handwriting recognition. In Proc. of the 5th Int. Conf. on Document Analysis and Recognition, pages 705 - 708, 1999

U. Marti and H. Bunke. The IAM-database: An English Sentence Database for Off-line Handwriting Recognition. Int. Journal on Document Analysis and Recognition, Volume 5, pages 39 - 46, 2002.

NIST Handprinted Forms and Characters Database [NIST special database 19]

- Final accumulation of NIST's handprinted sample data
- Full page HSF forms from 3600 writers
- Separate digit, upper and lower case, and free text fields
- Over 800,000 images with hand checked classifications

Reading

Siamese Convolutional Neural Networks for Authorship Verification

William Du
Stanford University
450 Serra Mall, Stanford, CA 94305
willadu@stanford.edu

Michael Fang
Stanford University
450 Serra Mall, Stanford, CA 94305
mjfang@stanford.edu

Margaret Shen
Stanford University
450 Serra Mall, Stanford, CA 94305
marshen@stanford.edu

Abstract

Determining handwriting authorship of a written text has practical significance in the realm of forensics, signature verification, and literary history. While there have been studies in signature verification and handwriting classification, a vast literature review reveals that very little work has been done in handwriting verification. Recent advances in convolutional architectures, particularly those involving facial verification, suggest that the task can be tackled effectively. In this study, we build a Siamese convolutional neural network to determine whether two pieces of handwriting are written by the same author. We examine questions such as whether long pieces of handwriting must be present to achieve good results, how many samples are needed, what features are important, and how different architectures perform on this task. We explore different convolutional architectures like VGG, GoogLeNet and ResNet, to determine which architecture produces the best encoding of each sample. We note that our best performing single model, TinyResNet, achieves a 92.08% accuracy on the held out test set.

1. Introduction

Determining the authorship of a written text has practical significance in the realm of forensics, signature verification, and literary history [3]. In manuscript analysis, for instance, historians frequently ask questions regarding the number of authors for a text, whether an anonymous work can be confidently attributed to a historical figure, and what time period a text might be from. These kinds of analyses are all based upon comparisons between different writing samples [1]. Techniques in the field have remained largely subjective, however, making the transition to automatic tools difficult.

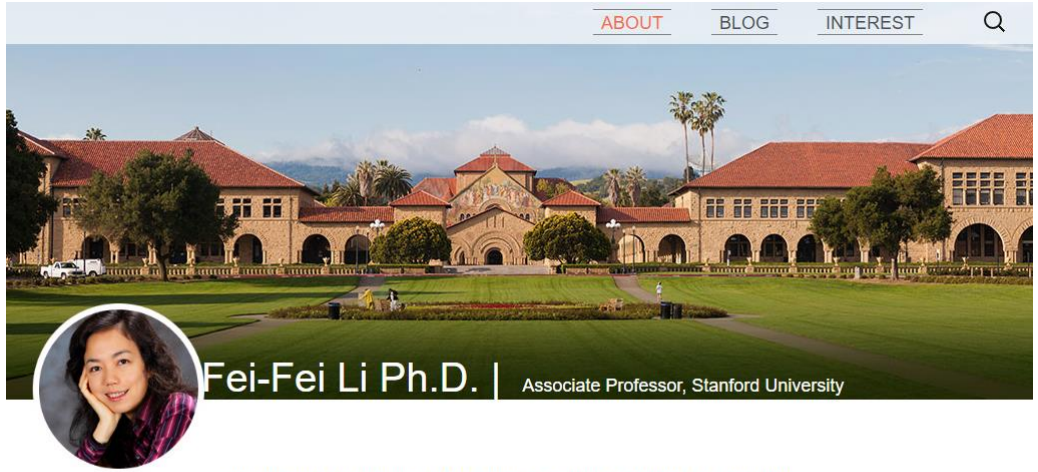
In addition, handwriting analysis is an established area of study in forensics, but there has not yet been any formal experiments measuring the accuracy of such analysis. As a result, the field is surrounded by much skepticism because of how subjective the process is (compared to, say, DNA testing) [5]. In addition, forensic handwriting analysis is time-intensive and requires two years of training for a person to obtain proper qualifications. The primary objective of this project is to develop an automatic, high-accuracy system which can determine if any two writing samples are written by the same person. In addition, our system should be able to handle authors it has never encountered before.

2. Background and Related Work

Our objective fits well with the Siamese CNN neural network architecture, which was first developed in 1993 to tackle the signature verification problem. [3] This type of architecture takes in two inputs and outputs a distance metric for the inputs. Bromley et al. was able to detect 95% of genuine signatures using this architecture. However, note that the signature verification problem expects a pair of inputs to be very similar to each other to be considered a match. This setup would not be effective for the problem we are trying to tackle, because our system should be agnostic to the actual text in a writing sample.

Other researchers have focused more closely on the authorship identification problem. A study in 2015 by Xing et al. reported an accuracy of 97% in classifying English writing samples for 657 authors. [11] They used the same dataset we will be using in this paper, the IAM Handwriting Database, and a 4-layer CNN. This study gave us confidence that we can achieve high accuracies on authorship problems using the IAM dataset. In a very recent research study from 2016, Yang et al. was able to achieve a 95% accuracy in classifying the authors for Chinese text samples,

Stanford University Department of Engineering



I am an Associate Professor at ***the Computer Science Department*** at ***Stanford University***. I received my Ph.D. ***degree from California Institute of Technology***, and a B.S. in Physics from ***Princeton University***.

I am currently the Director of the *Stanford Artificial Intelligence Lab* and the *Stanford Vision Lab*, where I work with the most brilliant students and colleagues worldwide to build smart algorithms that enable computers and robots to see and think, as well as to conduct cognitive and neuroimaging experiments to discover how brains see and think.

I am a Ph.D. student in the Stanford Vision Lab, advised by Prof. Fei-Fei Li. My research interests are in computer vision, machine learning, and deep learning. I'm particularly interested in the areas of video understanding, human action recognition, and healthcare applications.

I interned at Facebook AI Research in Summer 2016, and Google Cloud AI in Summer 2017. I also co-taught Stanford's CS231N Convolutional Neural Networks course in 2017, with Justin Johnson and Fei-Fei Li.

Before starting my Ph.D., I received a B.S. in Electrical Engineering in 2010, and an M.S. in Electrical Engineering in 2013, both from Stanford. I also worked as a software engineer at Rockmelt (acquired by Yahoo) from 2009-2011.