# Image Collation: Matching illustrations in manuscripts

Ryad Kaoua[1] Xi Shen[1] Alexandra Durr[2] Stavros Lazaris[3] David Picard[1] Mathieu Aubry[1]
[1]LIGM (UMR 8049) - Ecole des Ponts ParisTech, UPE [2]Sorbonne Université [3]CNRS

ICDAR LAUSANNE 2021

Goal: Finding correspondences between illustrations in different versions of a text

Motivation: illustrations are important for historians, in particular to study the diffusion of scientific knowledge, but aligning illustrations at scale is a difficult process

Contributions:

- Dataset and task definition
- Baselines evaluation
- Leveraging local structure and consistency in the correspondences

Dataset:

- 2 different historically important texts (*Physiologus* and *De Materia Medica*)
- More than 2000 illustrations and 1200 annotated correspondences

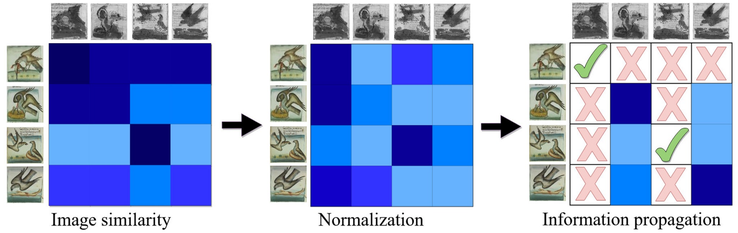
| name | code | number of folios | folios' resolution | number of illustrations | annotated correspondences |
|---|---|---|---|---|---|
| Physiologus | P1 | 109 | 1515x2045 | 51 | P2: 50 - P3: 50 |
| " | P2 | 176 | 755x1068 | 51 | P1: 50 - P3: 51 |
| " | P3 | 188 | 792x976 | 52 | P1: 50 - P2: 51 |
| De Materia Medica | D1 | 557 | 392x555 | 816 | D2: 295 - D3: 524 |
| " | D2 | 351 | 1024x1150 | 405 | D1: 295 - D3: 353 |
| " | D3 | 511 | 763x1023 | 839 | D1: 524 - D2: 353 |

Challenges: fine grained problem with large style differences

(a) Illustrations of different plants from the same manuscript

(b) Same illustration in three manuscripts with different styles

Overview:




Transformation-dependent image similarity:

$$\mathcal{T}_{k,l} = \arg\max \sum_{i=1}^{N} e^{-\frac{\|\mathcal{T}_{k,l}x_k^i - x_{k,l}(f_k^i)\|^2}{2\sigma^2}} \frac{f_k^i}{\|f_k^i\|} \cdot \frac{m_{k,l}(f_k^i)}{\|m_{k,l}(f_k^i)\|}$$

$$S(k,l) = \frac{1}{2N}\sum_{i=1}^{N} e^{-\frac{\|\mathcal{T}_{k,l}x_k^i - x_{k,l}(f_k^i)\|^2}{2\sigma^2}} \frac{f_k^i}{\|f_k^i\|} \cdot \frac{m_{k,l}(f_k^i)}{\|m_{k,l}(f_k^i)\|} + \frac{1}{2N}\sum_{i=1}^{N} e^{-\frac{\|\mathcal{T}_{l,k}x_l^i - x_{l,k}(f_l^i)\|^2}{2\sigma^2}} \frac{f_l^i}{\|f_l^i\|} \cdot \frac{m_{l,k}(f_l^i)}{\|m_{l,k}(f_l^i)\|}$$

| | |
|---|---|
| $\mathcal{T}_{k,l}$ | transformation between image $k$ and $l$ |
| $f_k^i$ | feature $i$ in image $k$ |
| $x_k^i$ | position of feature $i$ in image $k$ |
| $m_{k,l}(f_k^i)$ | feature matching $f_k^i$ in image $l$ |
| $x_{k,l}(f_k^i)$ | position of the match of $f_k^i$ in image $l$ |
| $S(k,l)$ | similarity between images $k$ and $l$ |
| $\sigma$ | scalar hyperparameter |

Normalization:

$$N_S(k,l) = \frac{S(k,l)}{\max_m S(k,m)} + \frac{S(k,l)}{\max_m S(m,l)}$$

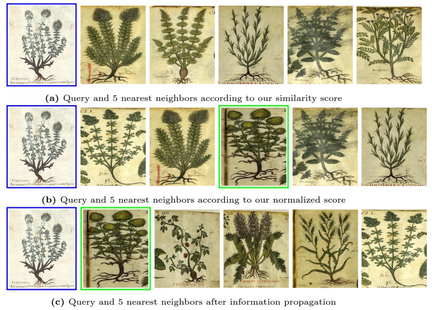
Information propagation:

$$P_S(k,l) = N_S(k,l) \prod_{(m,n)\in\mathcal{C}^*} \left(1 + \alpha \exp\left(\frac{-\|(k,l)-(m,n)\|^2}{2\sigma_p^2}\right)\right)$$

| | |
|---|---|
| $N_S(k,l)$ | similarity after normalization |
| $P_S(k,l)$ | similarity after propagation |
| $\mathcal{C}^*$ | confident correspondences (2 or 3-cycle consistent) |
| $\sigma_p$ | scalar hyperparameter |

Results:

Qualitative results



(a) Query and 5 nearest neighbors according to our similarity score

(b) Query and 5 nearest neighbors according to our normalized score

(c) Query and 5 nearest neighbors after information propagation

Quantitative analysis on the three manuscript pairs from the De materia Medica:

| Manuscripts | [1] | [2] | Our score | + normalization | +Info. Propagation |
|---|---|---|---|---|---|
| D1-D2 | 54.6 | 56.3 | 61.7 | 70.5 | 82.5 |
| D1-D3 | 69.8 | 71.3 | 77.7 | 85.3 | 88.5 |
| D1-D3 | 51.8 | 51.7 | 60.1 | 69.0 | 81.7 |

- our similarity score improves over baselines
- both normalization and information propagation provide significative boosts
- the problem remains open, but the final performance is high enough to be of practical interest
- more ablations in the paper (feature choice, alternative normalizations, selection of confident matches)

Challenging matches retrieved through information propagation

Failure cases

## References

[1] Large-scale historical watermark recognition: dataset and a new consistency-based approach. X. Shen, I. Pastrolin, O. Bounou, S. Gidaris, M. Smith, O. Poncet, M. Aubry, ICPR 2020
[2] Discovering visual patterns in art collections with spatially-consistent feature learning. X. Shen, A. Efros, M. Aubry, CVPR 2019

## Videos, code and more:

http://imagine.enpc.fr/~shenx/ImageCollation
https://github.com/Rykoua/ImageCollation