Re-presenting Art Collections

Joon Son Chung, Relja Arandjelović, Giles Bergel, Alex Franklin and Andrew Zisserman
- Photographs of around 900 broadside ballad sheets

- Broadside ballad sheets are:
  - Cheap printed ballad sheets
  - Printed from the 16\textsuperscript{th} to the 20\textsuperscript{th} century
- The illustrations are printed with woodblocks
Dataset: The Bodleian Broadside Ballads
The life and death of Sir Hugh of the Grime.

... (text continues)
Dataset: The Bodleian Broadside Ballads
Motivation:

- Art historians now have huge digital collections for study
- Manual analysis inadequate for collections of this scale

Computer vision methods can generate new representations of the data that are better suited to their research problems
Objectives

1. Identifying woodcut illustrations
2. Clustering by semantic similarity
Objectives

3. Clustering by exact copies (from the same woodblock)
1. Identifying woodcut illustrations
   a. Finding candidate regions
   b. Matching the illustrations across the collection to refine the regions
2. Clustering by semantic similarity
3. Clustering by exact copies
4. Application: temporal ordering
Identifying woodcut illustrations

- First, identify the areas of text
Identifying woodcut illustrations
Identifying woodcut illustrations

- Areas of text highlighted
Identifying woodcut illustrations

- Remove the areas of text from the image
Identifying woodcut illustrations

- Remove all details that are outside the page boundary
Identifying woodcut illustrations

- Find the bounding boxes of the remaining objects
1. Identifying woodcut illustrations
   a. Finding candidate regions
   b. Matching the illustrations across the collection to refine the regions

2. Clustering by semantic similarity

3. Clustering by exact copies

4. Application: temporal ordering
Philbin et al. (2007), Arandjelović and Zisserman (2012)
ImageMatch

**query input**

---

**Detailed matches**

- MS. Wood E 25(13)
- 4o Rawl. 566(121)
- MS. Wood E 25(138)
- Douce Ballads 2(173a)

---

- Douce Ballads 1(128a)
- Douce Ballads 1(20b)
- 4o Rawl. 566(134)
- Douce Ballads 2(171b)

---

- Douce Ballads 1(13a)
- Douce Ballads 1(120b)
- 4o Rawl. 566(86)
- 4o Rawl. 566(58)

---

**Detailed matches**

---

**Detailed matches**

---

**Detailed matches**

---

**Detailed matches**
Boundary refinement

- Generate ImageMatch queries from all candidate illustrations
- Record all returns
Cluster the matches from other candidates that overlap the original boundary
1. Identifying woodcut illustrations
2. Clustering by semantic similarity
3. Clustering by exact copies
4. Application: temporal ordering
Clustering by semantic similarity

- Similarity of two illustrations is computed as a weighted similarities of:
  - The GIST descriptor provides a holistic description of the scene
  - The VLAD summarizes the distribution of local SIFT descriptors in an image
  - Spatially pooled VLAD encodes the spatial information
Clustering by semantic similarity

- Similarity of two illustrations is computed as a weighted similarities of the GIST, the VLAD and spatially-pooled VLAD.
- The pairwise dissimilarities are thresholded, and a graph is formed
  - Undirected edges represent sufficient similarity

Threshold is 3.0
Clustering by semantic similarity

- Similarity of two illustrations is computed as a weighted similarities of the GIST, the VLAD and spatially-pooled VLAD.
- The pairwise similarities are thresholded, and a graph is formed
- Extract connected components to find the clusters
Clustering by semantic similarity

- Similarity of two illustrations is computed as a weighted similarities of the GIST, the VLAD and spatially-pooled VLAD.
- The pairwise similarities are thresholded, and a graph is formed.
- Extract connected components to find the clusters.
- We find and break up clusters with large intra-cluster variability.
Clustering by semantic similarity

740 clusters
Outline

1. Identifying woodcut illustrations
2. Clustering by semantic similarity
3. Clustering by exact copies
4. Application: temporal ordering
Clustering by exact copies

- VLAD descriptors for these illustrations are almost the same.
- But, are these from the same woodblock?
Clustering by exact copies

- A linear SVM is trained to distinguish between a pair of *same* and *similar* illustrations. Useful features are:
  - The number of putative SIFT matches,
  - The ratio of spatially consistent SIFT matches,
  - The density of matches,
  - The spatial distribution of the matches, and …
Clustering by exact copies

- The minimum of fine-level differences in image details.

Difference images
- From the same woodblock.
- The differences are damages to the woodblock.

Difference images
- From different woodblocks
- The image differences are due to spatial inconsistencies.
Clustering by exact copies

942 woodblocks
Summary statistics

- 918 broadside ballad sheets
- 2614 illustrations
- 740 semantically similar clusters
- 942 different woodblocks used to print the illustrations
Outline

1. Identifying woodcut illustrations
2. Clustering by semantic similarity
3. Clustering by exact copies
4. Application: temporal ordering
Application: temporal ordering of the prints

- Woodblocks get damaged over time.
- We use the visual cues apparent on the illustrations to find the temporal ordering of the ballad sheets.
Application: temporal ordering of the prints

- Woodblocks get damaged over time.
- We use the visual cues apparent on the illustrations to find the temporal ordering of the ballad sheets.
Application: temporal ordering of the prints
Application: temporal ordering of the prints
Application: temporal ordering of the prints

- We combine the cues from all woodblocks
Application: temporal ordering of the prints
Conclusion

- The three contributions of this presentation:
  - Automatic cropping of illustrations
  - Semantic clustering
  - Exact clustering

- The illustrations are:
  - From the same source,
  - Or near copies,
  - Or different depictions.

- Applicable to any collection with some commonality in illustrations (e.g. The Book of Hours).
Catalogue of woodblocks

http://www.robots.ox.ac.uk/~vgg/demo/ballads/