Introduction of CODH Datasets for Machine Learning and Computer Vision such as Old Japanese Characters and Books

Center for Open Data in the Humanities (CODH)
National Institute of Informatics
Asanobu KITAMOTO
http://codh.rois.ac.jp/index.html.en
What we do @ CODH
http://codh.rois.ac.jp/

• Officially Launched on April 1, 2017.
• Located at Tokyo, promoted by National Institute of Informatics (NII) and Institute of Statistical Mathematics (ISM).
• (1) humanities research using information technology and (2) other fields of research using humanities data.
Pre-Modern Japanese Text

http://codh.rois.ac.jp/pmjt/

日本古典籍データセット（国文研所蔵）

701 Books
CC By-SA

300,000 Books
In the future
High-throughput Computer Vision Tool for Comparing Editions, Variants, and Impressions of Old Japanese Books

Match feature points

Two books for comparison

Red/blue color scale

Rigid registration of images
PMJT Character Shapes

http://codh.rois.ac.jp/char-shape/

日本古典籍データセット (国文研所蔵)

日本古典籍字形データセット (国文研所蔵・CODH加工)
IIIF Curation Viewer

http://codh.rois.ac.jp/pmjt/curation/3/
A Little Announcement

• **Poster on Wednesday**: MemoryGraph: Digital Critique of Old Photographs Using a Mobile App that Enhances the Interpretation of Landscape

• **CODH is looking for a project researcher** who is interested in applying computer vision to humanities data. Contact: [http://codh.rois.ac.jp/recruit/](http://codh.rois.ac.jp/recruit/)