Computer Vision-based Comparison of Woodblock-printed Books and its Application to Japanese Pre-modern Text, Bukan

Thomas Leyh[1,3] and Asanobu Kitamoto[2,3]

[1]University of Freiburg
[2]ROIS-DS Center for Open Data in the Humanities
[3]National Institute of Informatics

1 Introduction

Digital Humanities have potential to reduce the complexity of bibliographical study by developing technical tools to support comparison of books. So far, the effort has been mainly put into text-based methods on large corpora of literary text. However, the emergence of large image datasets, along with the recent progress in computer vision, opens up new possibilities for bibliographical study without text transcriptions. This paper describes the technical development of such an approach to Japanese pre-modern text, and in particular to Bukan, which is a special type of Japanese woodblock-printed book.

The authors have been approaching this problem in the past [KIT+18]. They proposed the concept of "differential reading" for visual comparison. Furthermore, [Inv19] proposed "visual named-entity recognition" for identifying family crests, using them for a page-by-page matching across different versions. This paper is a follow-up of these works and proposes a keypoint-based method for the page-by-page matching, additionally yielding an option for highlighting differences.

2 Dataset

This work is mainly concerned with extracting information from a specific type of book: 武鑑—Bukan. These are historic Japanese books from

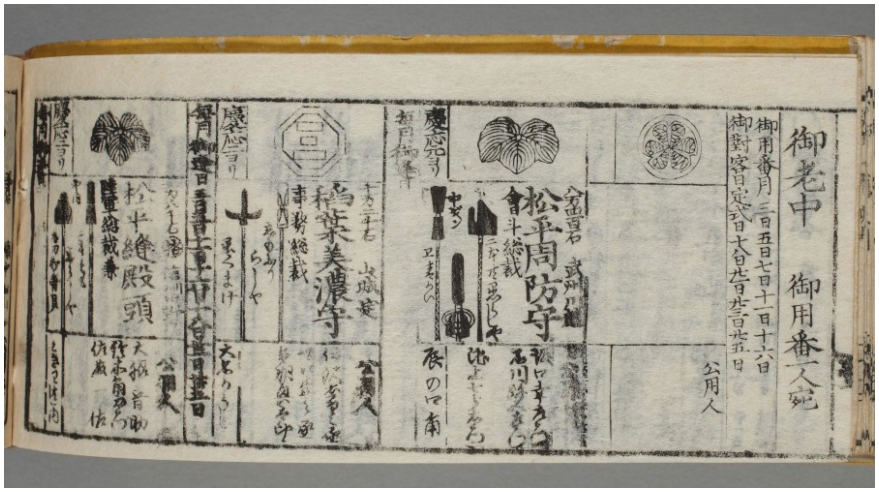


Figure 1: Shūgyoku Bukan (袖玉武鑑) from 1867, page 6; showing names, descriptions, family crests and procession items. Especially interesting are the blank areas on the right, because in other edition they contain text.

the Edo period (1603-1868). Serving as unofficial directories of people in Tokugawa Bakufu (the ruling government in Japan), they include a wealth of information about regional characteristics such as persons, families and other key factors. See figure 1 for an example. These books were created with woodblock-printing. Because the same woodblock has been reused for many versions of the book—sometimes with minor modifications—visual comparison can reveal which part of the woodblock was modified or has degraded.

ROIS-DS Center for Open Data in the Humanities and the National Institute of Japanese Literature are offering 381 of these Bukan as open data [Cen18]. The original images have a width of 5616 and height of 3744 pixels. From the open data we created a derived dataset using the following preprocessing methods. Under the assumption that this task (1) does not require this level of detail, (2) does not require information about color and (3) only compares the actual pages, not the surrounding area, all scans are resized by 25%, converted to grayscale and finally cropped, resulting in an image shape of 990×660 pixels. If there are two book pages per scan, they are split at their horizontal center, yielding a shape of 495×660 pixels per page.

3 Method

Using an approach based on Computer Vision, two techniques were applied:

1. *Keypoint Detection and Matching* for finding the same features in different images.
2. *Projective Transformations* for comparing two different images regardless of their original orientation.

We used the *OpenCV* software library [Bra00].

3.1 Keypoint Matching

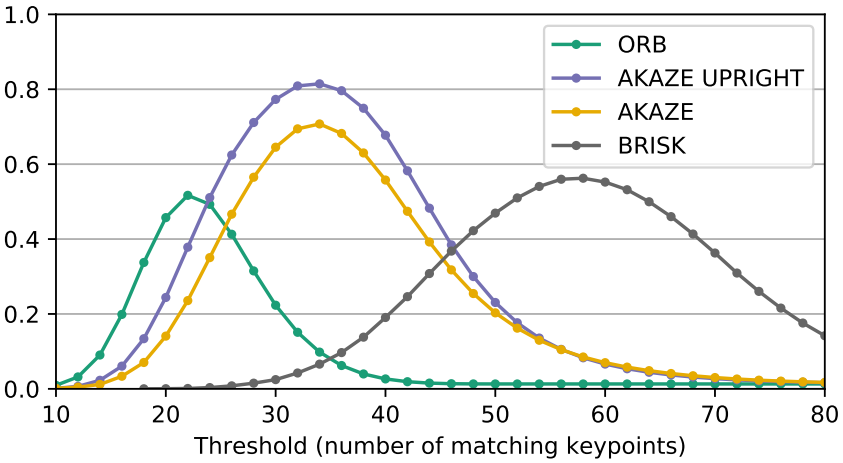
Keypoint Detection [Sze10, Ch.4] is about finding points of interest in an image that are most noticeable and give a unique description of the local area surrounding them. Computer Vision research produced various kinds of keypoints, most prominently SIFT [Low04]. For evaluating the performance of these algorithms, 12 prints of the Shūchin Bukan (袖珍武鑑) were manually annotated, in total around 1800 pages, holding information about pairs of matching pages.

Using these annotations, six keypoint algorithms were empirically evaluated[1] by trying to match over all possible page combinations. This produces a list of similar keypoint pairs for each combination. The number of these pairs is interpreted as score for the similarity of two pages. Two pages are matched if the score is above a given threshold. On the one hand, a low threshold results in more matches (higher recall), but a larger part is judged incorrectly. On the other hand, a high threshold yields a larger percentage of correct matches (higher precision), misclassifying actual matches at the same time. Using the annotations, precision and recall were calculated for a range of thresholds. At this point, AKAZE UPRIGHT has the best performance with both metrics around 0.8 for an optimal threshold value. See figure 2 for further details.

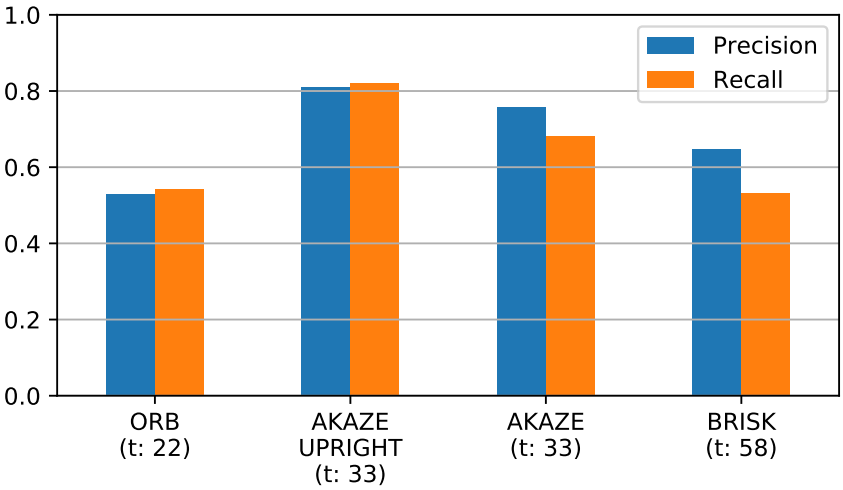
3.2 Projective Transformations

Projective Transformation (or Homography) for images is defined as a matrix $\mathbf{H} \in \mathbb{R}^{3\times3}$ for transforming homogeneous coordinates $\mathbf{H}\vec{x} = \vec{y}$ ($\vec{x}$

[1] ORB [Rub+11], AKAZE [AS11], AKAZE without rotational invariance (UPRIGHT), BRISK [LCS11], SIFT, SURF [BTV06]



(a) F1 score over a range of threshold values, which is the harmonic mean of precision and recall, thus peaking where both underlying metrics are high.



(b) Precision and recall, assuming an optimal threshold t was chosen. This is an upper bound on the performance of keypoint detection for this dataset.

Figure 2: Evaluation of the performance of four keypoint detection algorithms. AKAZE UPRIGHT performs best.

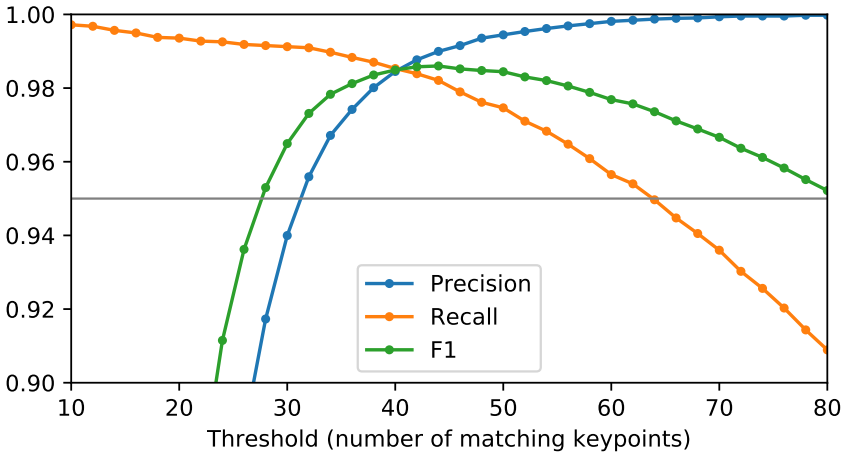


Figure 3: Applying RANSAC on the keypoint matches from AKAZE UPRIGHT results in better classification results: The F1 score—see figure 2a—is above 0.95 for a wide range of thresholds up to 80, indicated by the grey horizontal line.

and $\vec{y}$ are interpolated pixel coordinates). This operation results in linear transformations like translation and rotation, but also changes in perspective [MS15]. For finding such a transformation from matching keypoints, the heuristic *Random Sample Consensus* (RANSAC) algorithm is commonly used [FB81]. Basically, a random subset of matching keypoints is chosen, using this to compute a transformation and calculating an error metric. By iteratively using different random subsets, eventually the transformation with the smallest error is picked.

The benefit is twofold: First, the algorithm implicitly uses spatial information of the matching image candidates to filter out false positives, thus greatly boosting the matching performance. Precision and recall are close to their maximum of 1.0, see figure 3. Secondly, it directly yields the transformation matrix $\mathbf{H}$, enabling the creation of image overlays for visualizing the differences. Looking at the perspective components of $\mathbf{H}$, additional filtering is done[2], removing even more of the remaining false positives.

4 Discussion

This two-step-pipeline archives high precision and recall when looking for similar pages of different book editions. Performance seems to be robust

[2]By asserting that $|\mathbf{H}_{3,1}| \leq 0.001$ and $|\mathbf{H}_{3,2}| \leq 0.001$.

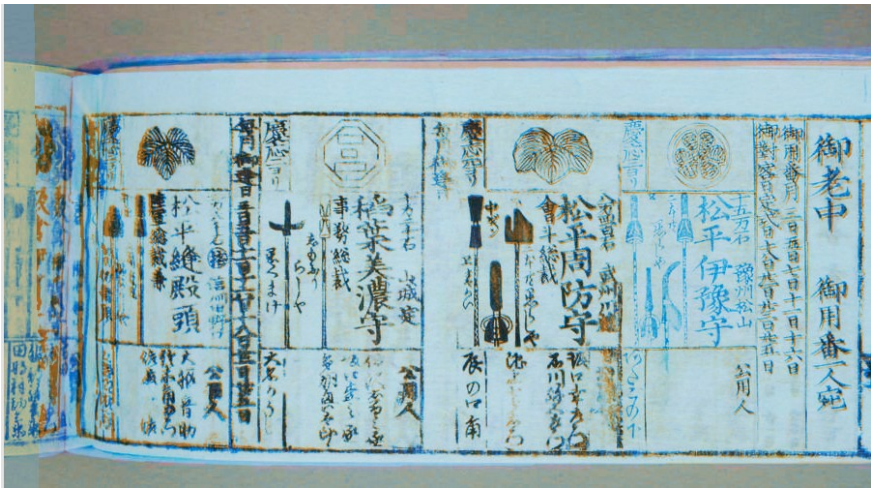


Figure 4: Visualizing page differences between two prints of the Shūchin Bukan (袖珍武鑑) from 1867. Differences are indicated by blueish and reddish coloring.

with respect to most parameters. For matching two actual pages, the number of matching keypoints is important since it is acting as a score. Storing this value allows a ranking of page pairs by visual similarity. Furthermore, a threshold can be set dynamically, even after the computation of keypoint pairs. Assuming semi-automatic application with a human assessing the results, recall is of higher importance, thus a low threshold of around 40 is recommended, but can be adjusted any time. See again figure 3 for the slope of the recall curve.

This was the basis for building a web-based prototype of a comparison browser: A scholar in the humanities can browse a book while getting information about similar pages. Differences between pages can be visualized at any time, similar to figure 4. As a next step, we are designing a web-based tool that is easy to use and works for IIIF images.

Prof. Kumiko Fujizane, who is a Japanese humanities scholar specialized in this particular type of books and collaborates with us, says that the major strength of our method is in identifying differences which are hard to discern by human eye, while its weakness is in errors caused by the distortion of pages by the book binding. Her request for future work is to identify region-based difference in addition to page-based difference.

5 Conclusion

The proposed method has wide applicability in woodblock-printed books, because the method uses only visual characteristics obtained from computer vision, and does not depend on the content nor the text of books. Although image alignment is a mature research area we also see recent developments, such as RANSAC-Flow [She+20], with potential to improve our results.

References

[AS11] Pablo F Alcantarilla and T Solutions. "Fast explicit diffusion for accelerated features in nonlinear scale spaces". In: *IEEE Trans. Patt. Anal. Mach. Intell* 34.7 (2011), pp. 1281–1298. DOI: 10.5244/C.27.13.

[BTV06] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. "Surf: Speeded up robust features". In: *European conference on computer vision*. Springer. 2006, pp. 404–417. DOI: 10.1007/11744023_32.

[Bra00] G. Bradski. "The OpenCV Library". In: *Dr. Dobb's Journal of Software Tools* (2000).

[Cen18] Center for Open Data in the Humanities. 武鑑全集とは?. July 28, 2018. URL: http://codh.rois.ac.jp/bukan/about/.

[FB81] Martin A Fischler and Robert C Bolles. "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography". In: *Communications of the ACM* 24.6 (1981), pp. 381–395. DOI: 10.1145/358669.358692.

[Inv19] Hakim Invernizzi. "An iconography-based approach to named entities indexing in digitized book collections". MA thesis. École polytechnique fédérale de Lausanne, Sept. 2019.

[KIT+18] Asanobu KITAMOTO et al. "Differential Reading by Image-based Change Detection and Prospect for Human-Machine Collaboration for Differential Transcription". In: *Digital Humanities 2018*. June 2018.

[LCS11] Stefan Leutenegger, Margarita Chli, and Roland Y Siegwart. "BRISK: Binary robust invariant scalable keypoints". In: *2011 International conference on computer vision*. Ieee. 2011, pp. 2548–2555. DOI: 10.1109/ICCV.2011.6126542.

[Low04] David G Lowe. "Distinctive image features from scale-invariant keypoints". In: *International journal of computer vision* 60.2 (2004), pp. 91–110. DOI: 10.1023/B:VISI.0000029664.99615.94.

[MS15] Steve Marschner and Peter Shirley. *Fundamentals of computer graphics*. CRC Press, 2015.

[Rub+11] E. Rublee et al. "ORB: An efficient alternative to SIFT or SURF". In: *2011 International Conference on Computer Vision*. Nov. 2011, pp. 2564–2571. DOI: 10.1109/ICCV.2011.6126544.

[She+20] Xi Shen et al. "RANSAC-Flow: generic two-stage image alignment". In: *arXiv*. 2020.

[Sze10] R. Szeliski. *Computer Vision: Algorithms and Applications*. Texts in Computer Science. Springer London, 2010. ISBN: 9781848829350.