Title	Log Analysis Methodology to Understanding Detailed IIIF Image Usage
Author(s)	Nishioka, Chifumi; Nagasaki, Kiyonori
Citation	(2019): 1-25
Issue Date	2019-06
URL	http://hdl.handle.net/2433/243235
Right	© 2019 Authors
Type	Presentation
Textversion	author

Log Analysis Methodology to Understanding Detailed IIIF Image Usage

Chifumi Nishioka, Kyoto University Library

Kiyonori Nagasaki, The University of Tokyo

2019 IIIF Conference, Göttingen, Germany

Usage Analysis of Digital Resources

> **Usage analysis of digital resources is important for various reasons…**

- Improvement of digital collections
- Plan for digitization of library holdings
- Accountability for stakeholders

Process of Quantitative Usage Analysis

1. Selection and measurement of evaluation metric
Select evaluation metrics such as the number of accesses and the number of visitors according to the purpose or guidelines in the community. Analyze server logs according to the evaluation metric.

→ So far, the following metrics have been employed: the number of accesses, the number of visitors, duration time, and the number of annotations

2. Visualization and reporting of the results
Visualize and report the results according to the characteristics of the evaluation metric, in order to make easy to understand the result.

→ For visualization and reporting, charts such as pie charts, bar charts, and histograms are used in many situations

Usage Analysis Enabled by IIIF

We present an example of usage analysis for IIIF images

In IIIF, images are requested via IIIF Image API **with specifying a region of an images**

- **Syntax of IIIF Image API:**
{scheme}://{server}{/prefix}/{id}/**{region}**/{size}/{rotation}/{quality}.{format}
- Each time the image is panned or zoomed on a viewer, the value of region is automatically adjusted and the image is requested

Thus, IIIF records detailed image usage and enables fine-grained analysis

Evaluation Metric and Viz. Method

1. Selection and measurement of evaluation metric
Select evaluation metrics such as the number of accesses and the number of visitors according to the purpose or guidelines in the community. Analyze server logs according to the evaluation metric.

→ **The number of accesses to each pixel of the image** is used as an evaluation metric. Since pixel is the smallest unit of an image, detailed usage can be grasped.

2. Visualization and reporting of the results
Visualize and report the results according to the characteristics of the evaluation metric, in order to make easy to understand the result.

→ Facilitate users to understand the result by visualizing the result in a **heatmap**

Counting the Number of Accesses

- Prepare H×W matrices for each image, where all elements are 0
    - H: height of an image, W: width of an image
    - Each element in matrices corresponds to each pixel
    - The size of images is retrieved from info.json
      {scheme}://{server}{/prefix}/{identifier}/info.json
- Retrieve the requested image and region by parsing the IIIF Image API logs
- Count the number of accesses to each pixel and record it to matrices

Generating Heatmaps

Output the results of measurement (H × W matrices for each image) as heatmaps

- Compute the RGB values for each pixel based on the number of accesses to each pixel
  - Make more accessed pixels red and less accessed pixels blue

- Generate heatmaps based on RGB values
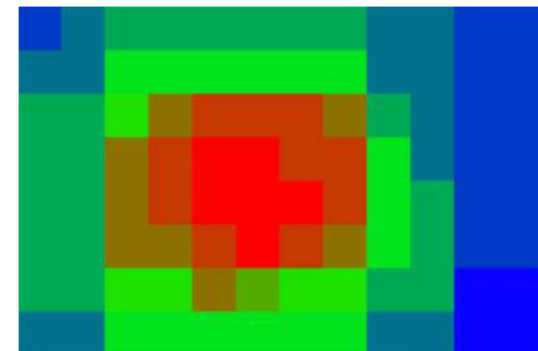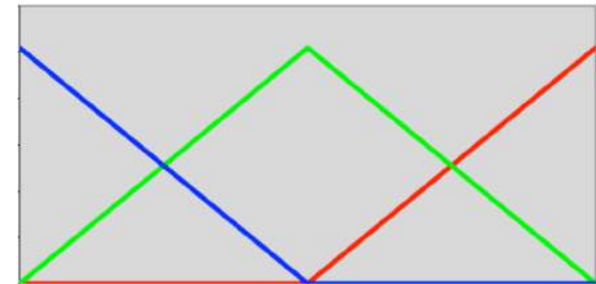
Speed Up
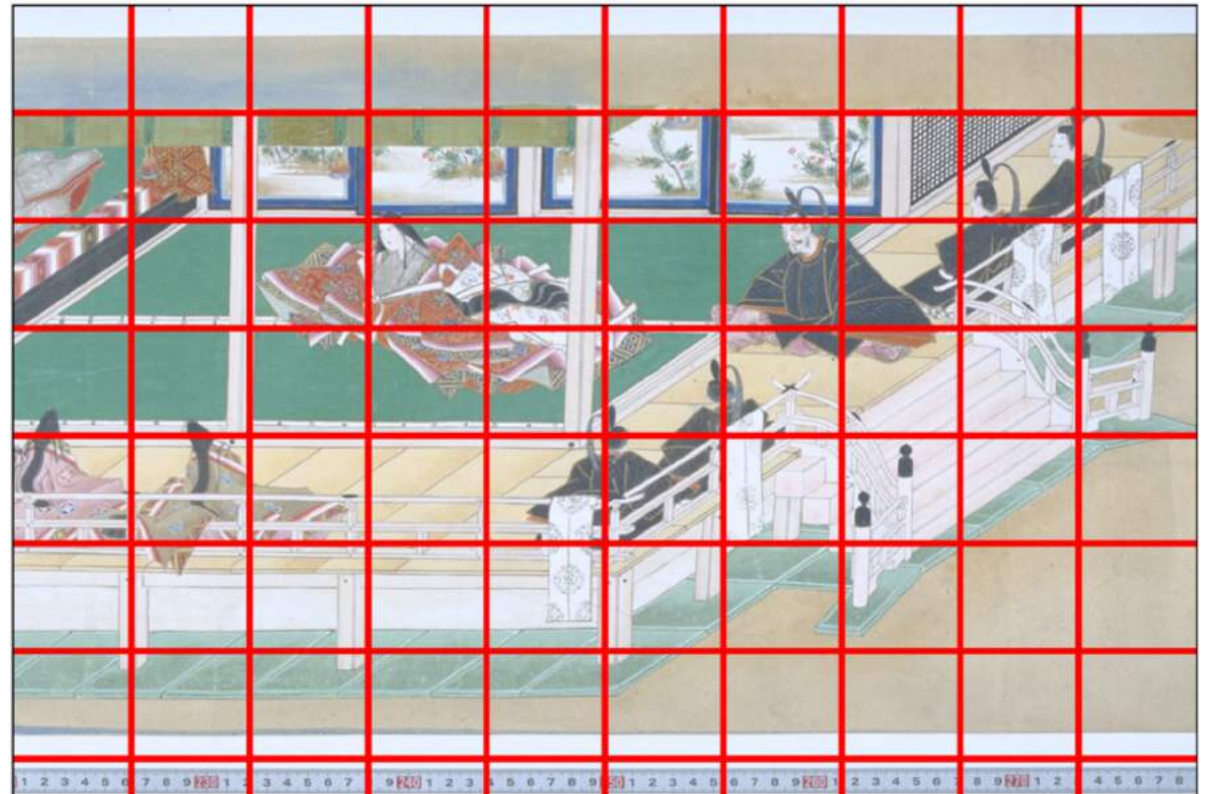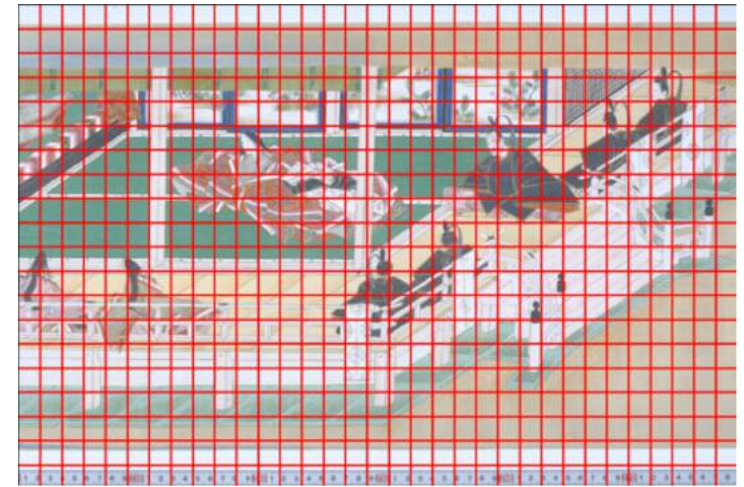
Count the number of accesses to each pixel

Count the number of accesses in N-pixel unit

Computation time for counting # of accesses (100k access logs, in which 27,736 logs are calls of IIIF image API)
10-pixel unit: 84.23 (s)
100 pixel unit: 1.09 (s)

Computer used for the experiment: iMac (macOS High Sierra version 10.13.4), Processor 4GHz Intel Core i7, Memory 16GB, 1867 MHz DDR3

Speed Up Further

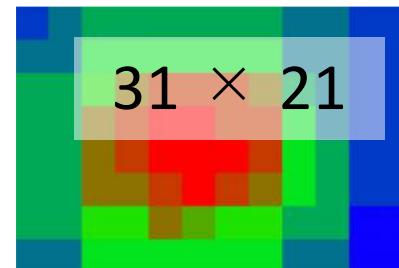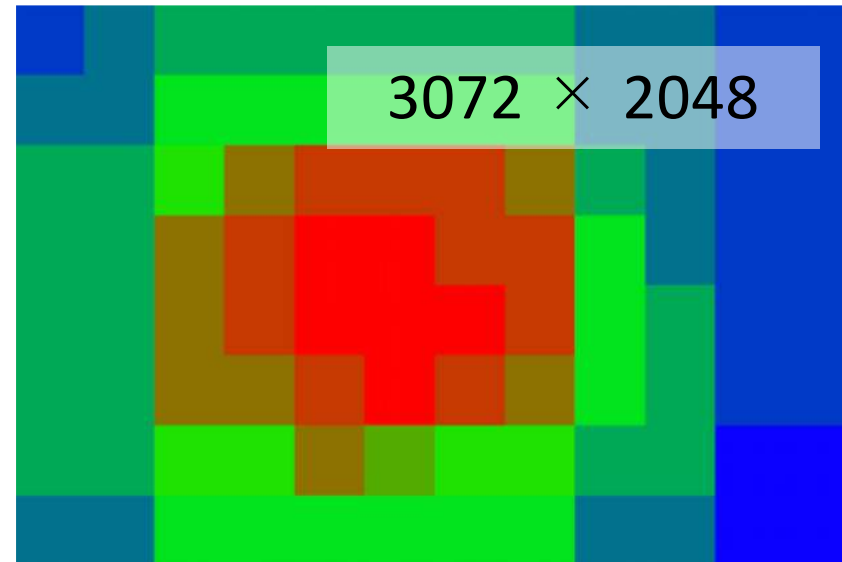Output a heatmap with the size of an original IIIF image
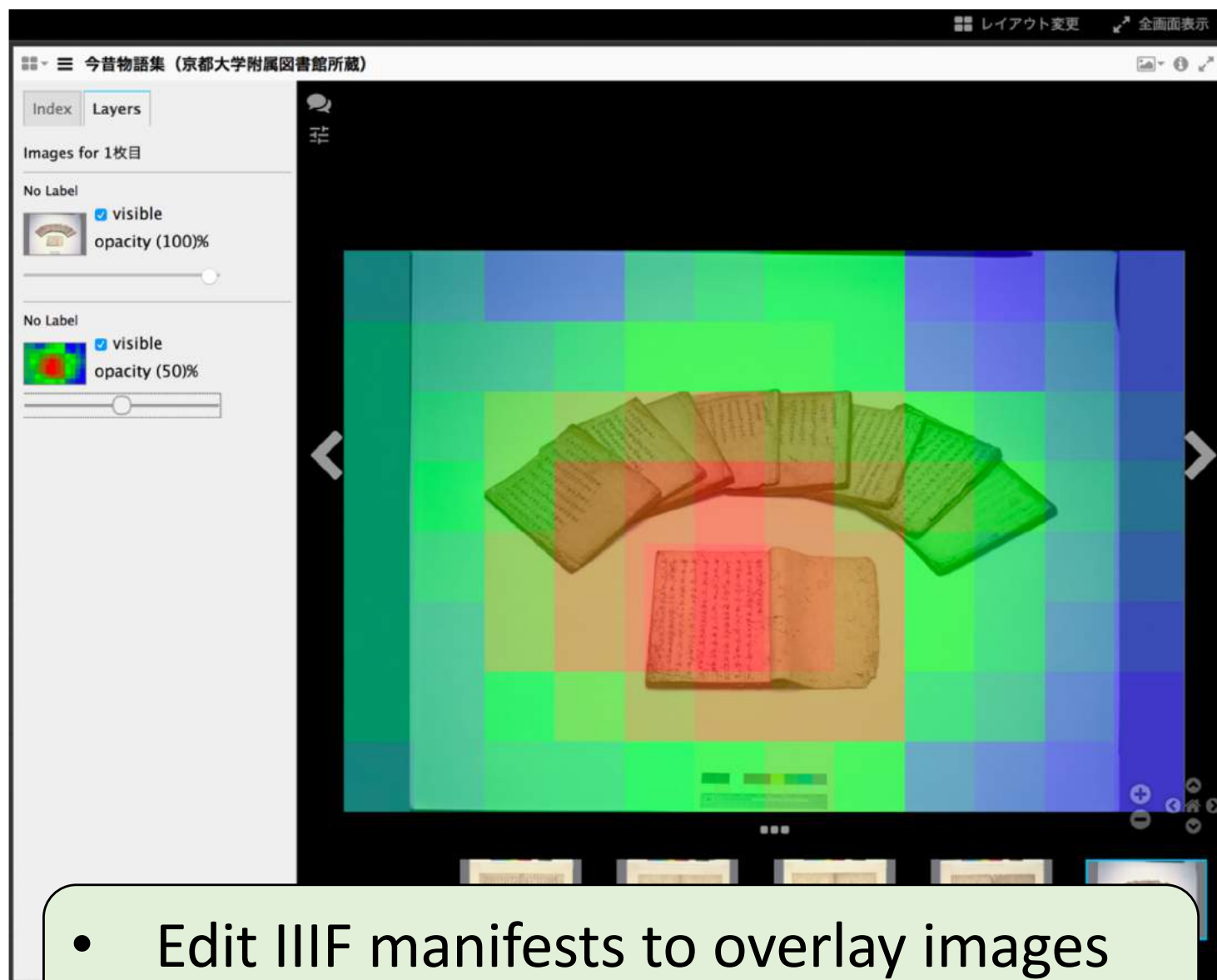
Output in a small size

Average computation time for generating one heatmaps (calculating RGB value for each pixel and output as an image)
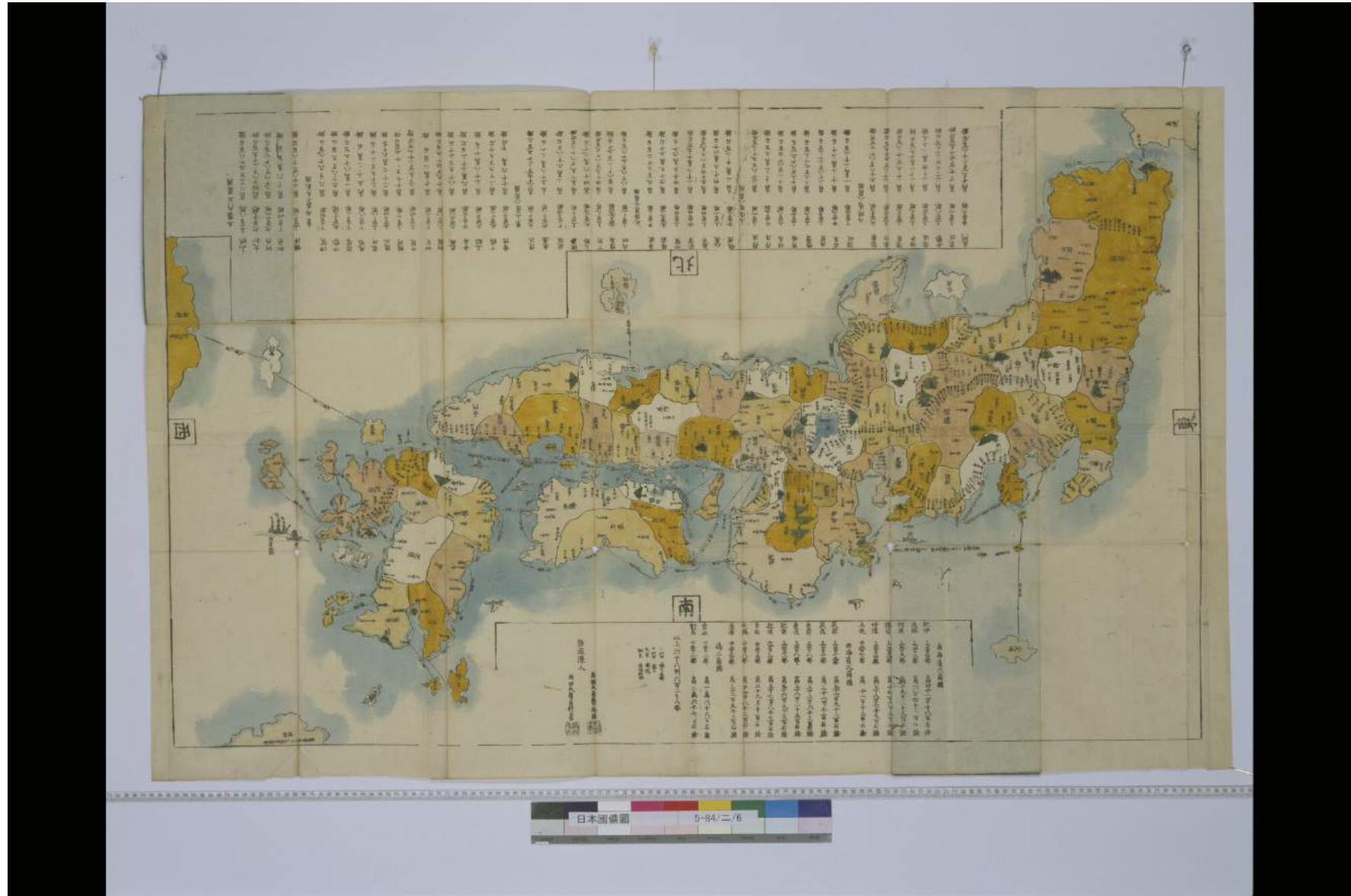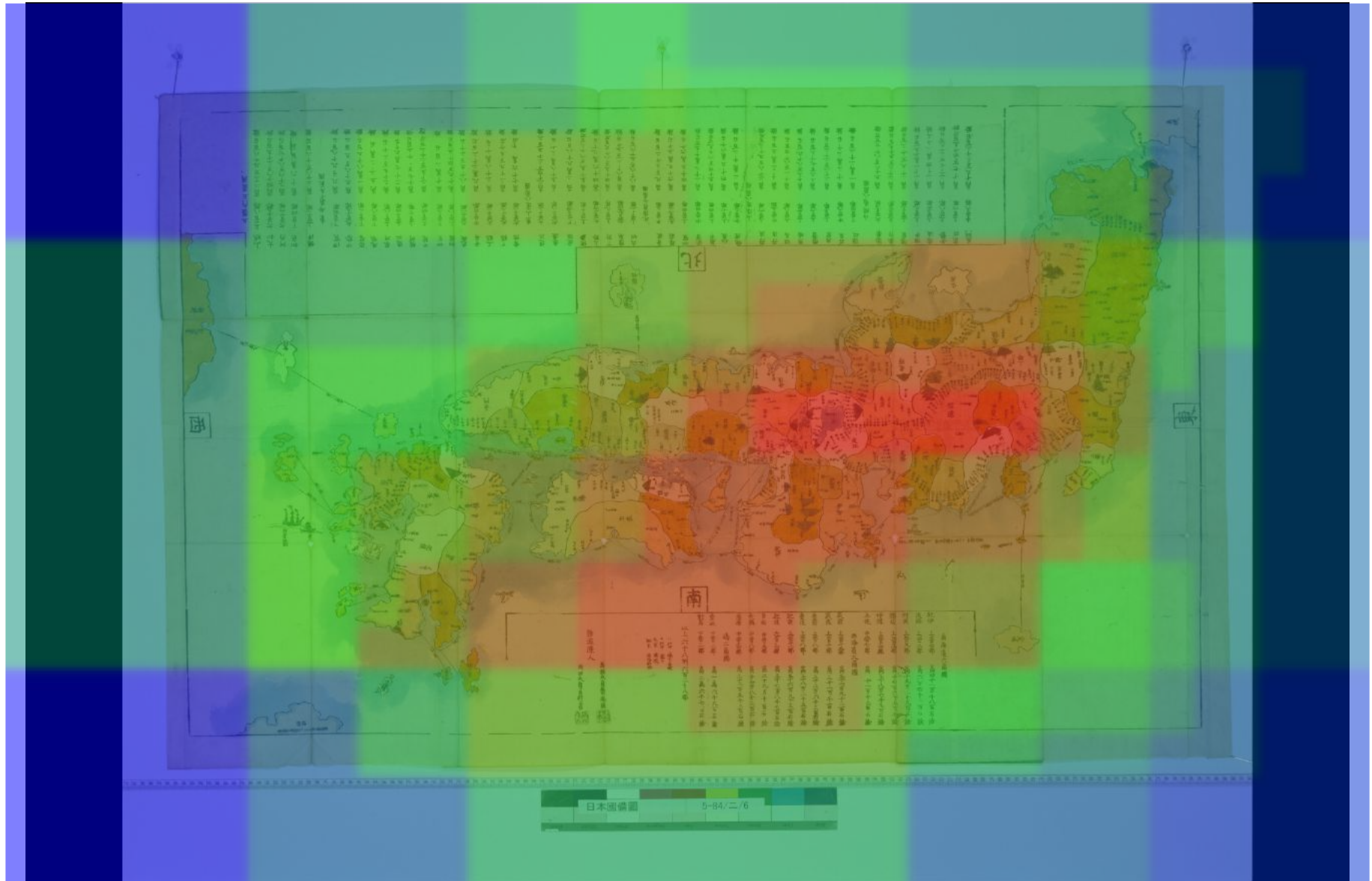**10-pixel unit: 2.05 (s) (SD: 7.01)**
100 pixel unit: 0.02 (s) (SD: 0.07)

3072 × 2048

31 × 21

Displaying Heatmaps over Images

- Edit IIIF manifests to overlay images
- Use Mirador's layer function

Example 1 [1/2]

『日本国備図』（京都大学附属図書館所蔵）
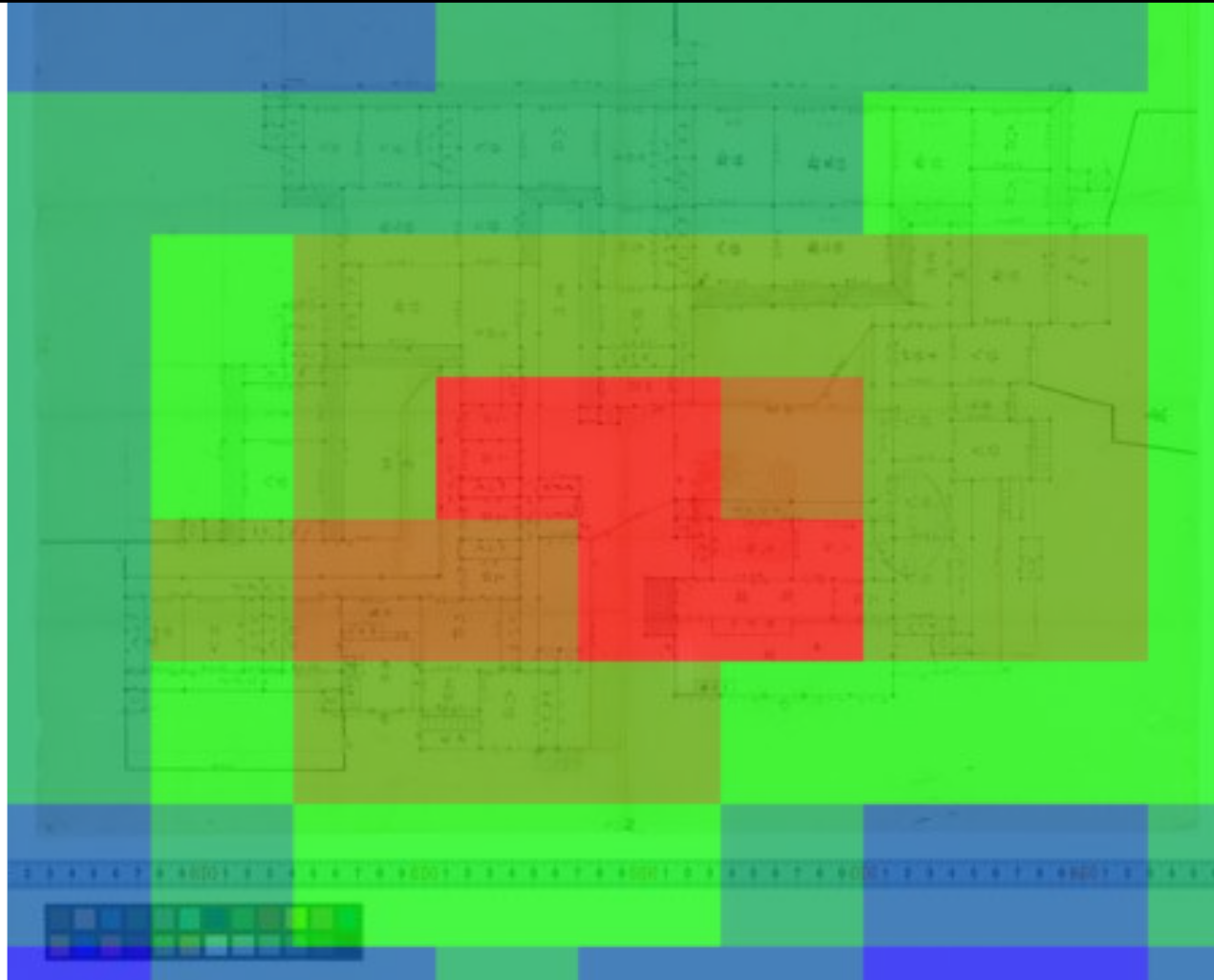
Example 1 [2/2]

『日本国備図』(京都大学附属図書館所蔵)
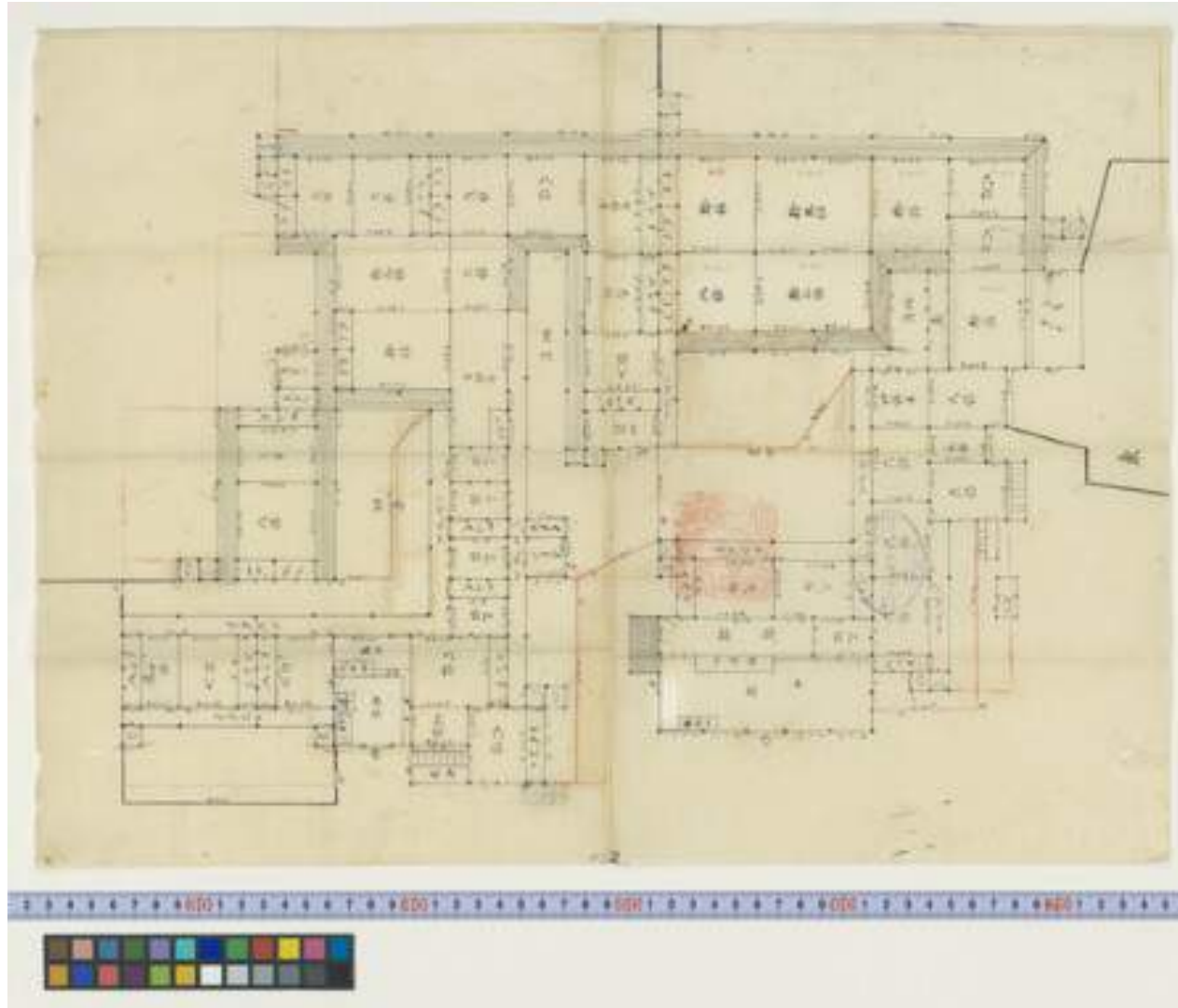
Example 2 [1/2]

Yashiki-zu from Nakai Collection (Kyoto University Main Library)

Example 2 [2/2]

Yashiki-zu from Nakai Collection (Kyoto University Main Library)

Probability of Access to Pixel [1/4]

- Pixels close to the center of an image are more likely to be accessed compared to other pixels
  - When browsing various regions on an image viewer, pixels close to the center are more likely to be included in a requested region

- Heatmap where requested regions are randomly chosen
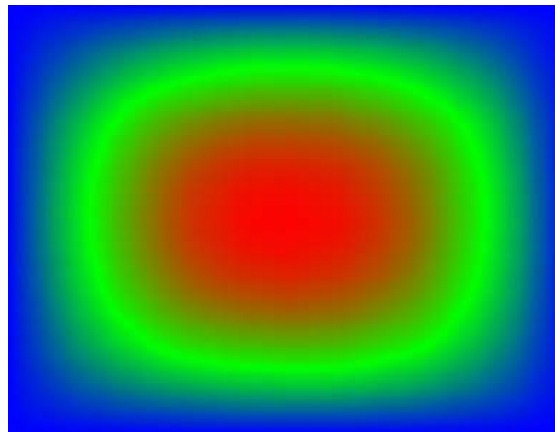
Is it necessary to adjust the number of accesses according to the probability that each pixel is accessed?

Probability of Access to Pixel [2/4]

- When a region is randomly selected, probability that a pixel located at a and b points apart from the middle point of the width and height of the image is accessed is:

$$p(w, h\; a, b) = \frac{w^2 + 2w - 4a^2 - 1}{2w^2} \cdot \frac{h^2 + 2h - 4b^2 - 1}{2h^2}$$

- The adjusted number of accesses adjusted is:

$$c_{rem}(w, h\; a, b) = c(w, h, a, b) \cdot \frac{w^2 + 2w - 1}{w^2 + 2w - 4a^2 - 1} \cdot \frac{h^2 + 2h - 1}{h^2 + 2h - 4b^2 - 1}$$

  - $c(w, h, a, b)$: The number of accesses of a pixel located at a and b points apart from the midpoint of each side of the image
  - $\frac{w^2+2w-1}{w^2+2w-4a^2-1} \cdot \frac{h^2+2h-1}{h^2+2h-4b^2-1}$: The probability that the center of the image is accessed divided by the probability that the point is accessed

Probability of Access to Pixel [3/4]
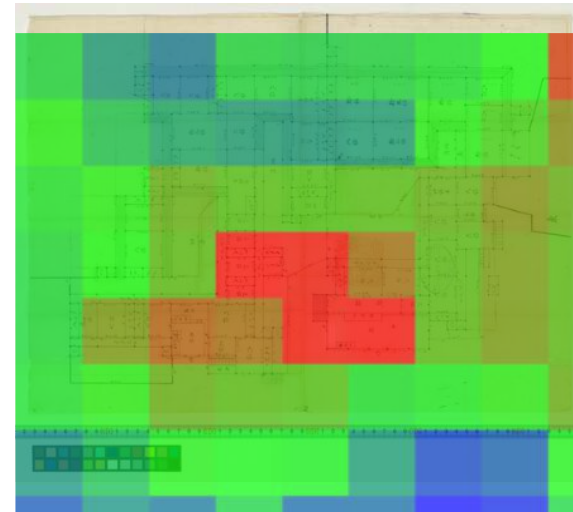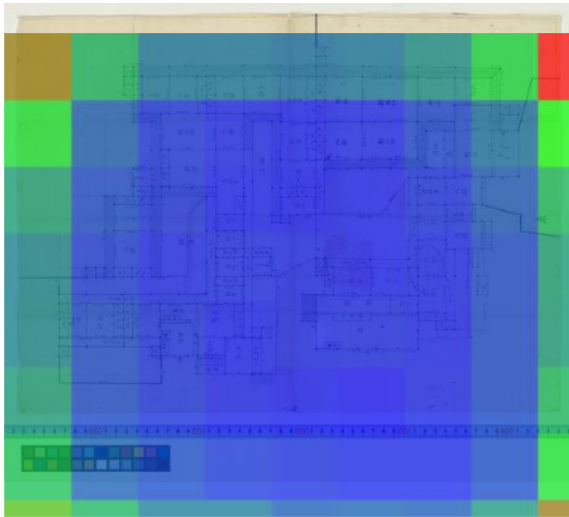
- (left) The numbers of accesses at the four corners are extremely high. It does not capture the image usage.

- The influence from adjustment by probability of being accessed is too large.

- (right) By using the logarithm, the influence is reduced.

$$c_{rem}(w, h\ a, b) = c(w, h, a, b) \cdot \log\left(\frac{w^2 + 2w - 1}{w^2 + 2w - 4a^2 - 1} \cdot \frac{h^2 + 2h - 1}{h^2 + 2h - 4b^2 - 1}\right)$$

Probability of Access to Pixel [4/4]

- Other factors needed to be considered for adjustment
  - Screen size
  - Viewer's maximum magnification ratio
  - Aspect ratio

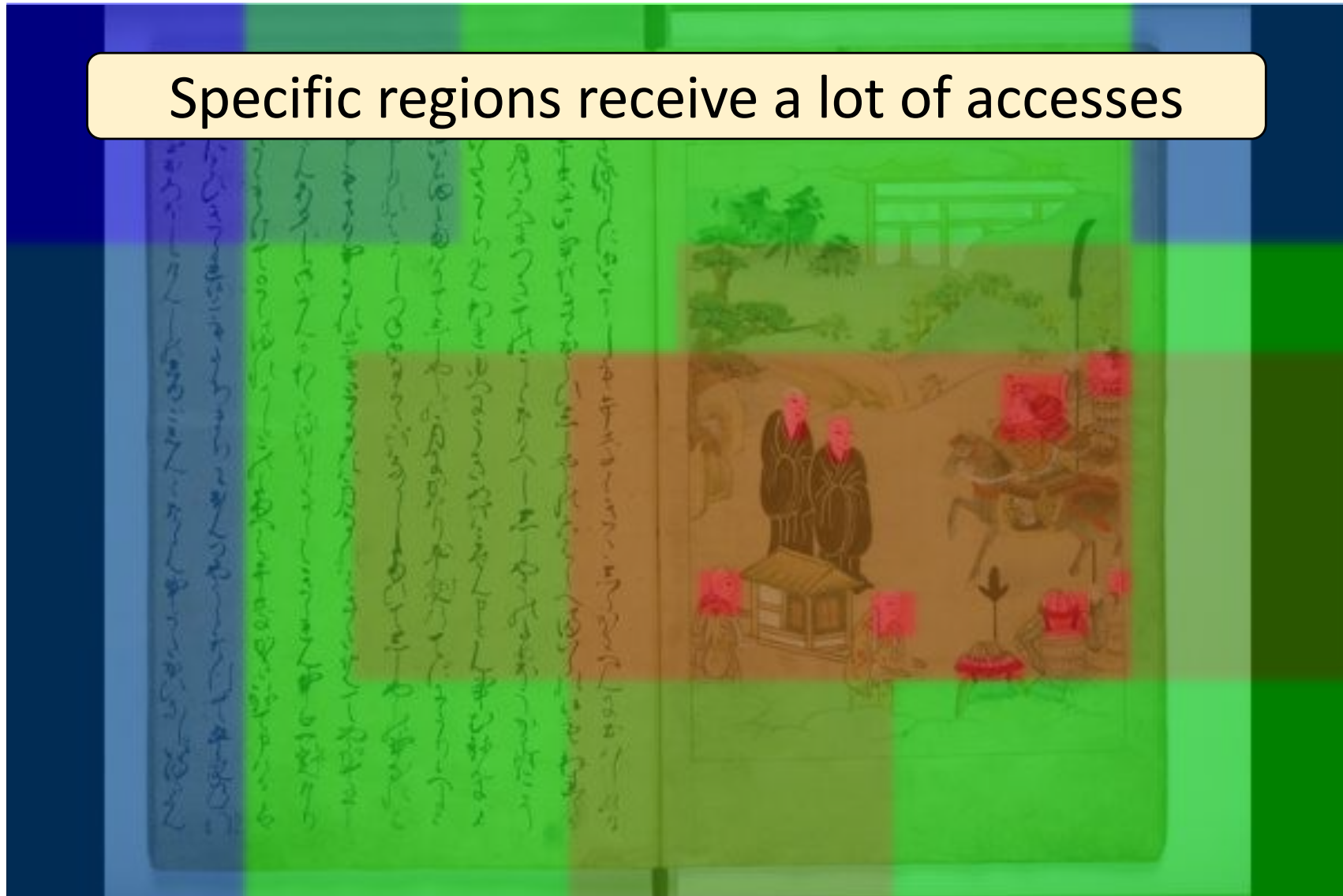Need to study more about how to adjust the number of accesses

Example 3 [1/2]

『弁慶物語』（京都大学附属図書館所蔵）

Example 3 [2/2]
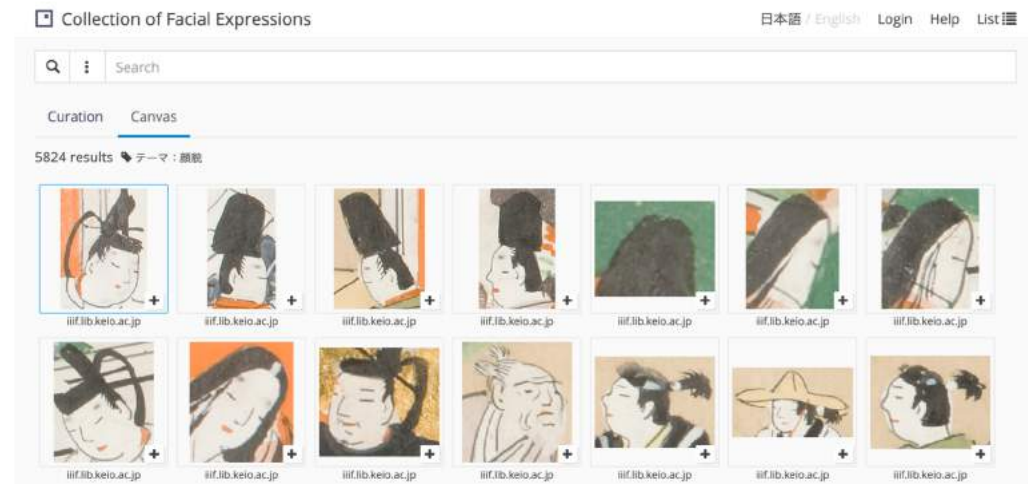

Specific regions receive a lot of accesses

『弁慶物語』（京都大学附属図書館所蔵）

Display of Referrers

- An examination of referrers of logs reveals that these regions are referred by IIIF Curation Viewer.

- It is expected that opportunities to be referred from external portals and websites will increase.

- If the websites of the referrer is public, we can show "from which websites are the images and regions often referenced? " to users. Providing a link to the referrer enables to find images and regions related to them.

Possible Applications [1/2]

Thumbnails
- Most-viewed regions of images are used as thumbnails.

Research platforms for Collaborations
- There are research platforms using IIIF.
- Heatmaps enable collaborators to understand which regions of images have been already investigated.
- A tool to stimulate motivation for crowd-sourcing

Possible Applications [2/2]

Understanding research process
- Researchers can reflect their own research process.
- Young researchers can learn research methodology by looking how experienced researchers browse images.

Transcription platform
- It is able to investigate the correlation between transcription accuracy and browsing behavior. Is there any difference in transcription accuracy between zoomed regions and not-zoomed regions?
- If there is a pattern (e.g., low accuracy of transcriptions in not-zoomed regions), it can be used for the verification of transcription.

Risks and Concerns

- Visualization of access logs is not a problem, if anonymization is conducted appropriately

- However, anonymization can be invalidated for IIIF images with few accesses
  - In the field where a small number of researchers work, peers can easily guess who accessed and investigated images

- In addition, a series of activities on IIIF images might reveal his/her viewpoint that would be a key issue of his/her academic outcome
  - Key issues can be revealed even before publication of outcome
  - Priority rights of research can be spoiled

- Therefore, we need a careful management of access logs to make services for researchers trustworthy

Wrap-up

A methodology of usage analysis for IIIF images

1. Selection and measurement of evaluation metric

→ **The number of accesses to each pixel of the image**

2. Visualization and reporting of the results

→ **Heatmaps**

Example of analysis results and future considerations

- Take into account the probability of a pixel to be accessed
- Display of referrers

Examples of applications of analysis results and issues for practical use

- Applications: Research platform, transcription platform and so on
- Several challenges for practical use including priority rights of research
- Deploy the service with users' consents

Thank you