

# IIIF 画像の詳細な利用状況の把握に向けたログ解析手法

西岡 千文 (京都大学 附属図書館)

永崎 研宣 (東京大学)

国際的な画像の相互利用の枠組みである IIIF (International Image Interoperability Framework) に対応したデジタルアーカイブが国内外で増加している。本稿では、IIIF 準拠デジタルアーカイブで公開されている画像の利用状況の解析手法ならびに解析結果の可視化手法の一例を示す。具体的には、IIIF Image API のログを解析することで、画像の各ピクセルに対するアクセス数を計算する。各ピクセルのアクセス数を参照してヒートマップを生成することで、画像の利用状況を可視化する。さらに、解析結果や可視化結果の期待される応用例を挙げる。また、利用状況の可視化についての懸念事項を検討する。

## A Methodology of Log Analysis to Understand Usage of IIIF images

Chifumi Nishioka (Main Library, Kyoto University)

Kiyonori Nagasaki (Tokyo University)

In these years, a lot of libraries and museums launched their digital collections that are compatible with IIIF (International Image Interoperability Framework), to promote interoperable use of images among different institutions. In this paper, we present a methodology to understand and visualize usage of IIIF compatible images. We analyze IIIF Image API logs and visualize which parts of IIIF images are looked using heat maps. The generated heat maps are shown over corresponding target images using Mirador's layer functionality. Furthermore, we show possible use cases where the result of the analysis and heat maps can be used. Finally, we discuss challenges to deploy the result of the analysis and heat maps as a service such as anonymization.

### 1. 序論

国際的な画像の相互利用の枠組みである IIIF (International Image Interoperability Framework) [1]は、欧米の主要な図書館・博物館・美術館を中心に導入が広まっている。IIIF は、画像とそれに付随するデータ (メタデータ、注釈) の相互利用を可能とするために、API を規定している。欧米に続いて、日本国内でも導入が広まっており、IIIF 準拠デジタルアーカイブを公開する機関は増加している。

デジタルアーカイブの利用状況の調査は、資料の電子化戦略の策定、機能の改善、ステークホルダーへの説明責任といった観点から重要である。従来のデジタルアーカイブでは、資料・画像毎のアクセス数といった指標により、利用状況が把握されてきた<sup>1</sup>。IIIF 準拠デジタルアーカイブでは、

画像は、ビューワ上で拡大・縮小する度に、IIIF Image API<sup>2</sup>で指定されている URI 構文の領域の値が更新されて呼び出されている。よって、単純にアクセス数のみで画像の利用状況を把握することができず、滞在時間といった指標が重要となってくる。一方、ビューワで拡大・縮小する度に画像がリクエストされているため、リクエストされた領域を調査することによって、画像の詳細な利用状況を把握することが可能である。

本稿では、IIIF 準拠デジタルアーカイブで公開されている画像 (以下、IIIF 画像あるいは画像) の利用状況の解析手法ならびに解析結果の可視化手法の一例を示す。具体的には、IIIF Image API の画像リクエストのログを解析し、画像の各ピクセルのアクセス数を計測する。各ピクセルのアクセス数を参照して、画像ごとにヒートマップを生成することで、画像の利用状況を可視化する。さらに、解析結果や可視化結果の期待される応用例

<sup>1</sup> たとえば、政府の知的財産戦略本部による「デジタルアーカイブの連携に関する実務者協議会 中間報告」(平成 28 年 3 月)では、「なお、コンテンツへのアクセス数は、そのコンテンツの価値を示す一つの尺度となるとの考え方もあるため、欧州の Europeana はソーシャルメディア等における発信によるアクセス数の増加に非常に力を入れている。我が国でもメタデータを発信する側にとってのインセンティブを向上するには、実際の来館数に加え、ネットワーク上のアクセスやデジタルコンテンツの利用実績等についても指標

化され、業績として評価されるような仕組みを設計する必要があるとの意見があった。」としている。

<sup>2</sup> 画像の配信方法・取得方法を規定する API である。画像は、URI 構文 `{scheme}://{server}/{prefix}/{identifier}/{region}/{size}/{rotation}/{quality}.{format}` によってリクエスト・取得することが可能である。URI 構文にて、画像の領域 (region)、回転角度・反転 (rotation)、サイズ (size)、色調 (quality)、画像フォーマット (format) をパラメータとして設定できる。

として、共同研究プラットフォームや翻刻プラットフォームでの利用を挙げる。また、一般利用者へのサービスで解析結果を活用する際に発生する、匿名性や研究の先取性の保証といった課題について検討する。

## 2. 関連研究

本章は、デジタルアーカイブならびに公開画像の利用状況の調査方法について、関連研究を示す。定量的な評価指標に基づく利用状況調査は、以下のプロセスより構成される。

### 1. 評価指標の選定・計測

調査目的、コミュニティにおけるガイドライン・慣習などに応じて、アクセス数や訪問ユーザー数といった評価指標を選定する。評価指標に応じてログデータを解析し、評価指標を計測する。

### 2. 計測結果の可視化

計測結果の受け手が直感的に把握しやすいよう、評価指標の特性に応じて、ヒストグラム、棒グラフなどで可視化する。

電子図書館では、ユーザビリティやコレクションの評価のため、利用状況の解析が行われてきた。例えば、Jones ら[2]は、電子図書館における利用者の検索行動を把握するために、セッションにおける検索クエリ数、一定期間内での電子図書館への訪問回数、クエリごとに閲覧している資料数などを集計・報告している。Hashem ら[3]は、博物館における展示物の配置場所が利用者行動へ与える影響を調査するため、各展示物における利用者の滞在時間を評価指標として採用している。Pääkkönen ら[4]はデジタルアーカイブの注釈機能やキュレーション機能の利用状況を評価するため、クリップ数（注釈数）や注釈に追加されたキーワード数を使用している。以上に挙げた関連研究では、表、円グラフ、ヒストグラム等が計測結果の可視化手法として採用されている。電子図書館やデジタルアーカイブの管理者も、表や折れ線グラフなどを使用して、サイトアクセス数といった評価指標を報告しているというのが現状であろう。

本稿では、評価指標として、「画像のピクセル単位のアクセス数」を選択し、結果の可視化手法として「ヒートマップ」を採用する。評価指標の選択の背景となる調査目的としては、画像の詳細な利用状況を把握することが挙げられる。しかし、調査結果はデジタルアーカイブ管理者等による画像の詳細な利用状況の把握という目的に限定されず、第5章において示す通り、共同研究プラットフォームといった様々なケースで利用できる可能性がある。

## 3. IIIF 画像の利用状況の計測と可視化

3.1 節では、画像の各ピクセルのアクセス数を計測し、画像の利用状況を可視化したヒートマップを生成する手法を提示する。続いて3.2 節では、計算時間や必要となるストレージ容量を短縮・減少させる手法について述べる。3.3 節では、IIIF コミュニティで開発されている画像ビューワである *Mirador* を使用して、生成されたヒートマップを解析対象である画像とともに表示する方法を紹介する。

### 3.1. 利用状況の計測手法と可視化手法

IIIF 画像の詳細な利用状況の把握に向けたログ解析手法は、以下のプロセスより構成される。それぞれ、第2章における利用状況調査のプロセスの1, 2に対応している。

#### 1. 各ピクセルのアクセス数の計測

評価指標として画像の各ピクセルのアクセス数が採用されている。IIIF Image API における画像リクエストのログを一行ずつ解析することで、各ピクセルのアクセス数を計測する。

#### 2. ヒートマップの生成

計測された画像の各ピクセルのアクセス数を、ヒートマップにより可視化する。なお、可視化手法として、二変量ヒストグラム<sup>1</sup>など、他の手法を採用することも可能である。ヒートマップが様々な分野で使用されている点と、対象画像との重ね合わせ表示に適している点（3.3 節参照）を考慮して、本稿ではヒートマップを採用した。

以下に、それぞれのプロセスの詳細について述べる。

#### 【各ピクセルのアクセス数の計測】

各画像に対して、全要素が0である $H \times W$ 行列を生成する。 $H$ ,  $W$ はそれぞれ画像の高さ、幅（ピクセル数）である。よって、行列の各要素は画像の各ピクセルに対応する。各画像の高さ、幅は、IIIF Image API により提供されている `info.json`<sup>2</sup>より取得する。続いて、IIIF Image API における画像リクエストのログをパースすることで、リクエストされている画像と領域を取得する。リクエストされている領域を参照することで、各画像の各

<sup>1</sup> 二変量ヒストグラムのプロット例として、下記のページを挙げる。

<https://jp.mathworks.com/help/stats/hist3.html>

<sup>2</sup> IIIF Image API では、URI 構文

`{scheme}://{server}/{prefix}/{identifier}/info.json` によって画像の情報の取得を可能とすることが規定されている。特に、画像の幅、高さは必須項目であり、必ず提供される。

ピクセルのアクセス数が計算される。アクセス数は前述の行列へ記録される。

#### 【ヒートマップの生成】

各ピクセルのアクセス数の計測の終了後には、計測結果をヒートマップとして出力する。画像毎に、各ピクセルのアクセス数の最小値・最大値を参照して、各ピクセルの RGB 値を算出する。アクセスが多いピクセルを赤、少ないピクセルを青とする。RGB 値を参照して、画像として出力することで、ヒートマップが生成される。

## 2.2. 計測手法と可視化手法の高速化

以下に、各ピクセルのアクセス数の計測とヒートマップ生成を高速化させる手法を提示する。また、使用するストレージ容量を減少させる手法についても述べる。本稿では、計算機として iMac (Retina 5K, Late 2015) を使用した。プロセッサとして 4GHz Intel Core i7, メモリとして 16GB 1867 MHz DDR3 を搭載している。

#### 【 $N \times N$ ピクセル単位でアクセス数を計算】

前節では、各ピクセルに対してアクセス数を計算していた。計算に要するメモリを節約するために、 $N \times N$ ピクセル単位でアクセス数を計算する。例えば、 $1,600 \times 2,560$ ピクセルの画像に対しては、 $1,600 \times 2,560$ 行列が用意される。 $N = 10$ と設定すると、アクセス数を記録する行列は  $160 \times 256$ 行列となり、計算時のメモリの使用量を減少させることができる。この場合、1234 行目 1017 列目に位置するピクセルへのアクセスは、123 行目 101 列目の行列の要素に記録されることとなる。

実際のログを使用して、ピクセル単位  $N$  が、ピクセルのアクセス数の計測とヒートマップ生成に要する処理時間に与える影響をそれぞれ計測した。2018 年 6 月のある期間に京都大学貴重資料デジタルアーカイブ<sup>1</sup>に記録されたログから、IIIF Image API の画像リクエストのログを 10,000 行抽出して使用した。10,000 行のログでは、458 枚の画像がリクエストされている。よって、458 個の行列が生成され、458 枚のヒートマップが生成される。458 枚の画像の高さのピクセル数の平均値は 5334.80 (標準偏差: 2589.27) であり、幅のピクセル数の平均値は 3696.07 (標準偏差: 2620.34)<sup>2</sup>である。

<sup>1</sup> 京都大学貴重資料デジタルアーカイブ。  
<https://rmda.kulib.kyoto-u.ac.jp/>

<sup>2</sup> 解析対象の多くは、幅が高さより大きい画像であったが、高さが幅より大きい地図画像が含まれていた。地図画像のサイズは他の画像と比較すると圧倒的に大きく、その影響により平均値においても高さが幅より

図 1 に、ピクセル単位  $N$  が、ピクセルのアクセス数の計測に要する時間に与える影響を示す。図 1 は、縦軸が対数スケールで表された片対数グラフである。図 1 において、 $N = 10$ のときの計算時間は 27.99 秒である。対して、 $N = 100$ のときは、0.30 秒であり、 $N = 10$ のときと比較すると、93.30 倍計算時間が短縮されていることとなる。

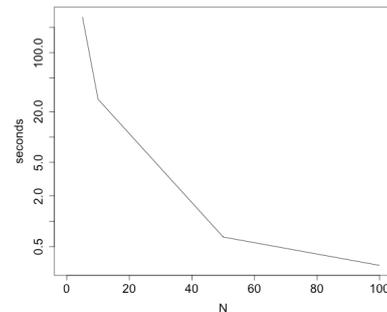


図 1 10,000 行のログにおける、各画像の各ピクセルのアクセス数計測に要する計算時間

Figure 1 Computation time of counting the number of accesses for each pixel of each image on 10,000 access logs with different values of N. The y axis in log scale.

図 2 に、ピクセル単位  $N$  が、RGB 値の算出 (ヒートマップの生成) の計算時間に与える影響を示す。図 1 同様、図 2 は、縦軸が対数スケールで表された片対数グラフである。分散は、概ね平均値の 2 倍に近い値である。RGB 値の算出に要する平均時間は、画像のサイズに依存しており、解析対象には様々なサイズの画像が含まれているため、分散が大きくなっている。図 2 において、 $N = 10$ のときの計算時間は 0.90 秒である。対して、 $N = 100$ のときは、0.01 秒であり、 $N = 10$ のときと比較すると、90.00 倍計算時間が短縮されていることとなる。

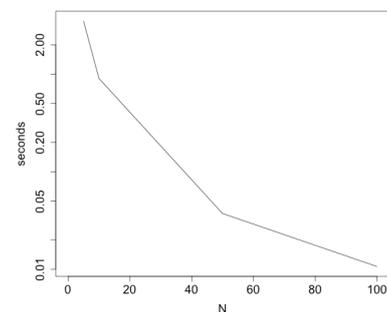


図 2 RGB 値の算出に要する平均計算時間 (画像 1 枚あたり)

Figure 2 Computation time of calculating RGB values for an image with different values of N. The y axis in log scale.

も大きくなっている。

### 【領域が full と設定されているログの除去】

解析に要する時間を短縮させる手法として、領域に full という値が入力されているログを除去することが挙げられる。full は画像の全領域をリクエストするとき使用される。画像をサムネイルとして呼び出すときに頻繁に使用されるため、該当するログは多い。領域が full と設定されている場合、行列の全要素に等しく 1 が追加されるだけであるので、生成されるヒートマップに与えられる影響はない。ただ、ヒートマップに、各色が表すアクセス数を付随して表示する場合には、ログの除去は不適切であるかもしれない。目的に応じた選択が必要となる。

### 【ヒートマップの出力サイズを縮小】

生成されたヒートマップ画像を、対象画像と同サイズ ( $H \times W$  ピクセル) で出力するのではなく、 $(H/N) \times (W/N)$  ピクセルで出力することによって、計算時間が短縮される。さらに、ヒートマップ保存に必要なストレージも節約される。

## 2.3. ヒートマップの表示

本節では、生成されたヒートマップの表示方法について紹介する。

ヒートマップを対象画像と重ね合わせて表示することで、利用者<sup>1</sup>は画像の利用状況を理解することが初めて可能となる。IIIF Presentation API の仕様は、画像を重ね合わせて表示することを可能としている。具体的には、IIIF マニフェスト<sup>2</sup>の各キャンバス<sup>3</sup>で二画像 (対象画像とヒートマップ) を重ね合わせて配置するように指示する。ヒートマップを対象画像より小さいサイズで保存している場合は、ヒートマップを対象画像と同サイズで表示するように指定する必要がある。なお、各キャンバスからヒートマップのリクエストを可能とするために、対象画像と同様に、ヒートマップも IIIF 対応画像サーバに保存する必要がある。

以上のように、IIIF Presentation API の仕様では画像の重ね合わせ表示を可能としているが、重ね合わせ表示に対応した実装を提供しているビューワは多くない。IIIF コミュニティにおける代表的なビューワである Mirador は、レイヤ機能を実装しており、画像の重ね合わせ表示に対応してい

る。Mirador では、重ね合わせ表示を記述したマニフェストを読み込ませることで、各キャンバスに二画像ロードさせることができる。Mirador での重ね合わせ表示例を、図 3 に示す。図 3 左部のタブで表示されているように、それぞれの画像に対して表示状態 (visibility) ならびに透過度 (opacity) を操作することができる。ヒートマップを透過させて表示することで、対象画像の各領域におけるアクセス状況が容易に把握できる。

なお、Mirador では、キャンバスに画像が複数枚ある場合には、キャンバスの全画像の表示状態が False になった状態で表示される。画像毎に、左部のタブにて、表示状態・透過度を手動で設定しなければならない。画像とヒートマップのデフォルトの状態での表示状態ならびに透過度を設定しておくには、IIIF Presentation API<sup>4</sup>ならびにビューワの機能拡張が必要となる。

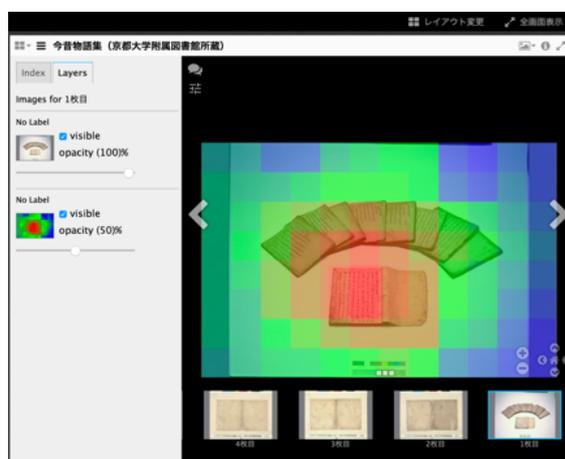


図 3 Mirador を利用した対象画像とヒートマップの重ね合わせ表示 (『今昔物語集 (鈴鹿本)』(京都大学附属図書館所蔵))

Figure 3 Overlay display of a heat map over its target image using Mirador. Photograph courtesy of the Main Library, Kyoto University - *Konjaku monogatari shuu*.

## 4. 今後の検討事項

本章では、前章で述べたログ解析手法による解析結果を例示しつつ、今後の検討事項について述べる。4.1 節では、典型的なヒートマップを例示し、画像の各ピクセルがアクセスされる確率を考慮した解析について検討する。4.2 節では、アクセスの滞在時間を考慮した解析の必要性と課題について述べる。4.3 節では、特定箇所アクセスが集中している例を示し、アクセスされた動

<sup>1</sup> この場合、解析結果の利用者、すなわちデジタルアーカイブ管理者等を指す。

<sup>2</sup> IIIF Presentation API で規定されている資料に相当する JSON オブジェクトである。資料をビューワで表示するために必要な情報を提供する。

<sup>3</sup> IIIF Presentation API で規定されている資料の頁に相当する JSON オブジェクトである。各頁で表示する画像を指定する。キャンバスのシークエンスにより、頁の順序が表現される。キャンバスのシークエンスは、IIIF マニフェストに含まれる。

<sup>4</sup> IIIF Presentation API の各キャンバスの各画像にデフォルトの表示状態と透過度を設定できるようなフィールドを設ける。

機・背景として参照元を明らかにすることについて記す。

#### 4.1. 画像の各ピクセルがアクセスされる確率を考慮した解析

生成される典型的なヒートマップを図3に例示する。画像の中心付近のアクセス数が大きく、中心から離れるほどアクセス数が小さくなる傾向が、その他の多くの画像のヒートマップで観察された。理由として、画像の中心付近のピクセルは他のピクセルと比較すると、アクセスされる確率が高いということが挙げられる。ビューワで画像を閲覧する際に、様々な領域を拡大・縮小すると、中心付近のピクセルはリクエストされる領域に含まれやすい。

ここで、ランダムに領域を設定してアクセスを発生させたときのヒートマップを図5に示す。図5より、ピクセルが画像の中央に近ければ近いほど、アクセスされる確率が高いことがわかる。各ピクセルに対してより公平な調査結果を導出するためには、各ピクセルがアクセスされる確率に応じて、各ピクセルのアクセス数を調整することが必要となる。

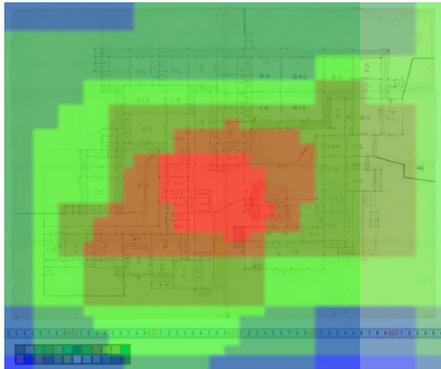


図4 典型的なヒートマップの例 (『屋敷図』中井家絵図・書類より (京都大学附属図書館所蔵))  
Figure 4 Example of typical heat map. Photograph courtesy of the Main Library, Kyoto University – Yashiki-zu (design drawing of a mansion) from Nakai Collection.

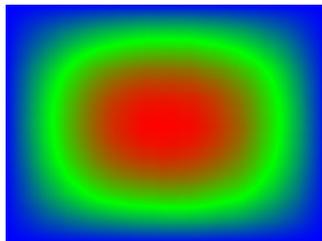


図5 ランダムに領域を選択してアクセスを発生させた際のヒートマップ  
Figure 5 Heat map for an image to which accesses are randomly generated.

ここで、画像の幅、高さのそれぞれの中心より  $a$  ピクセル、 $b$  ピクセル離れた点に位置するピクセルが、ランダムに領域が設定された場合にアクセスされる確率  $p(w, h, a, b)$  は、以下の式1より算出される。

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

式1によって、あるピクセルは画像の中心に位置するピクセルと比較してどの程度アクセスされにくいかが計算することができる。画像の各辺の中心より、それぞれ  $a$  ピクセル、 $b$  ピクセル離れた点に位置するピクセルのアクセス数を  $c(w, h, a, b)$  とする。ピクセルの位置によるアクセスされる確率の違いを考慮すると、アクセス数は以下の式2のように修正できる。

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

対象ピクセルのアクセスされる確率  $p(w, h, a, b)$  によって中心ピクセルのアクセスされる確率  $p(w, h, 0, 0)$  を割った値と、計測されたアクセス数  $c(w, h, a, b)$  の積が修正後のアクセス数となる。図4における各ピクセルのアクセス数を、式2によって修正し可視化した結果が図6である。

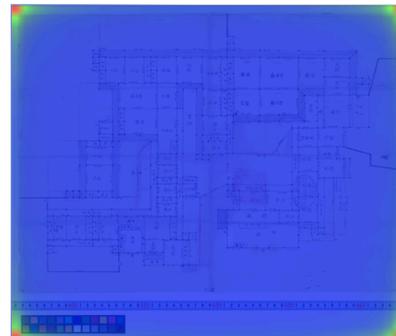


図6 各ピクセルがアクセスされる確率を考慮した解析結果例 (『屋敷図』中井家絵図・書類より (京都大学附属図書館所蔵))  
Figure 6 Example of heat map, in which the probabilities of being accessed are taken into account for each pixel. Photograph courtesy of the Main Library, Kyoto University – Yashiki-zu (design drawing of a mansion) from Nakai Collection.

図6から分かるとおり、画像の四隅のアクセス数が極端に高くなっており、利用状況が把握できるとはいえない。このような結果となった原因として、画像はピクセル単位というよりも画像サーバで設定されているタイル単位でアクセスされているということが挙げられる。さらに、画像は

実際には完全にランダムにアクセスされているわけではないということも挙げられる。アクセス数の修正方法については今後検討を重ねる必要がある。

#### 4.2. 滞在時間を考慮した解析

本稿では、各領域へのアクセスにおける滞在時間は均等であるとの前提で解析を実施している。しかし実際には、各アクセスにおける滞在時間は均等ではない。ビューワである領域を拡大する過程で必然的に発生したアクセスにおける滞在時間は至極短いであろうし、拡大結果を閲覧している場合には滞在時間は長くなる。一般的に、ある利用者のあるページにおける滞在時間は、その利用者がそのページへアクセスした時刻とその次にアクセスしたページのアクセス時刻の差分から算出される<sup>1</sup>。しかし、IIIF 画像を対象とした場合、従来の手法を適用することは難しい。IIIF 画像を大きいサイズでリクエストする際には、画像は分割され、複数の領域に同時にアクセスが発生する。よって、連続した二つのアクセスログの発生時刻の差分を滞在時間として採用した場合、そのアクセスにおいて最後に記録されたログにおける領域のみで、正確な滞在時間が算出されることとなる。同時にアクセスされたその他の領域の滞在時間は至極短くなる。サーバのログからは、それらが同時アクセスかどうか正確に判定することは難しい。判定するとなれば、連続するログの時間差に閾値を設けて、恣意的に判断することになるだろう。

正確な滞在時間を取得する方法として、ビューワにプラグインを埋め込み、ビューワで計測するといった手法を挙げることができる。もちろん、事前に利用者に許諾を得るような仕組みも必要となってくる。検索システム[5, 6]や推薦システム[7, 8]などにおいて、滞在時間による評価の正当性が主張されている。このようなことから、IIIF 画像の利用状況調査への滞在時間の考慮は、今後の検討が必要である。

#### 4.3. 画像の参照元

図7に例示されているように、画像の特定領域にアクセスが集中しているといったケースも散見された。一般的には、ビューワで拡大・縮小をすると、特定領域の周囲の部分もアクセスされる。よって、ヒートマップではグラデーションのように表現される。さらに、ビューワには拡大率に限

界があり、図7にあるように小さい特定箇所のみがビューワで表示されるということは一般的ではない。アクセスログのリファラーを参照すると、図7の場合、人文学オープンデータ共同利用センター (CODH) の顔貌コレクション<sup>2</sup>[9]から、これらの領域が参照されている<sup>3</sup>ことが判明した。顔貌コレクションは、美術作品に登場する顔の表現を IIIF Curation Platform[10]を利用して収集・公開している。IIIF は画像の相互利用を可能としていることから、統合サイトや他機関のプラットフォームから画像が参照される機会が増加することが期待される。解析結果の利用者に参照元を示すことで、アクセスされた動機・背景を知らせることができる。さらに、領域の参照元が顔貌コレクションのように完全に公開されている場合には、一般利用者に「その画像や領域がどこからよく参照されているか」ということを提示することによって、それらと関連する画像・領域を探索・発見することが可能となる。



図7 特定領域にアクセスが集中している例(『弁慶物語』(京都大学附属図書館所蔵))

Figure 7 Example where specific regions get a lot of accesses. Photograph courtesy of the Main Library, Kyoto University - *The story of Benkei, a tragic warrior.*

### 5. 利用状況解析結果と可視化結果の応用例

本章では、IIIF 画像の利用状況の解析ならびにヒートマップによる可視化の応用例を四点挙げる。

**サムネイルの選定:** デジタルアーカイブでは、コレクション情報のページなどで、資料のサムネ

<sup>1</sup> 実際には、ページ読み込みの際にアクセスされるスクリプト、favicon、画像などへのアクセスは除外される。これらの必然的に読み込まれるページは、リファラーを参照することで判断することができる。または、スクリプトなどが除外された調査対象となるページのリストを用意され、リストが利用されることもある。

<sup>2</sup> 人文学オープンデータ共同利用センター (CODH). 顔貌コレクション (顔コレ).

<http://codh.rois.ac.jp/face/>

<sup>3</sup> 特に、顔貌コレクションの「弁慶物語」より参照されている。

<http://codh.rois.ac.jp/software/iiif-curation-viewer/demo/?curation=https://mp.ex.nii.ac.jp/api/face/curation/json/34ae534e-82e2-4973-84a8-8ea2cad951d>

イル一覧が表示されている。多くの場合、資料の一枚目に表示される画像がサムネイルとして利用されている。しかし、一枚目の画像が必ずしもその資料を代表するわけではない。また、資料に含まれるとある一枚の画像全体が資料を代表するわけではない。サムネイルの選択方法の一案として、資料全体において最も利用されている画像の領域を表示することが挙げられる。利用状況の解析結果を活用することで、最も利用されている画像の領域を特定し、表示することが可能である。

**共同研究プラットフォームでの利用:** IIF は W3C が推奨する Web Annotation に沿って、データモデルが構築されている。よって、画像のみならず、翻刻・注釈といった画像に付随する情報の共有も容易である。このような背景もあり、IIF を利用した共同研究プラットフォームについて検討・実装がなされている[11]。IIF を利用した共同研究プラットフォームで複数の研究者が作業する際に、ヒートマップを提示することで、画像のどの領域が共同研究者によって既に調査されているかということを知ることができ、作業を開始することができる。

**研究プロセスの理解:** 自己の研究プロセスの振り返りは、研究プロセスを改善させていく上で重要である。研究者による画像の利用状況をヒートマップによって提示することで、研究プロセス振り返りの機会を創出する。また、ベテラン研究者の画像の利用状況をヒートマップにより若手研究者に例示することで、彼らの研究プロセスの一部を理解し、学習することが可能となる。

**翻刻プラットフォームでの利用:** 近年、手稿等の資料について、多くの翻刻プロジェクトが実施されており[12]、翻刻プラットフォームが検討・開発されている。翻刻者は、対象画像を拡大・縮小しながら閲覧し、翻刻作業を進める。IIF に対応している翻刻プラットフォームであれば、拡大されて翻刻されている領域と拡大されないで翻刻された領域で、翻刻精度について差が存在するかといったパターンを検証できる。拡大されないで翻刻された領域では翻刻精度が比較的低いといったようなパターンを発見することができれば、翻刻の検証作業が容易となり、翻刻精度の向上に貢献することができだろう。

## 6. 実用化に向けた課題

本章では、IIF 画像の利用状況の解析結果の活

用に係る課題について述べる。

ヒートマップの提供を含む画像の利用状況の活用については、誰がアクセスしたかについての情報が匿名化され、完全に秘匿されているようであれば懸念はない。しかし、少ないアクセス数の IIF 画像では、匿名化が意味をなさない場合がある。例えば、研究者数が少ない特定分野に関連が深い資料の画像については、誰がアクセスしたかということは、同分野の研究者より容易に想像がつくだろう。

アクセスの匿名化がなされていても、ヒートマップとして解析結果を提供することは、公表前の研究成果の重要なポイントを明らかにする可能性がある。例えば、画像の従来注目されていなかった特定領域に気が付けば、研究成果が導出されるといったようなケースでは、ヒートマップによって重要なポイントが明らかになってしまう。この場合、研究者の研究の先取性が侵害される恐れがある。よって、アクセスログの適切な管理は当然のこと、サービスとして解析結果を活用するには慎重な検討が今後必要となる。

このことは、従来のテキストデータベース検索や Web 検索のサイトにおけるユーザが使用した検索語や情報探索行動のログの取り扱いと類似している点がある。そうした情報に関する利用ガイドライン等についても今後検討を行い、必要に応じて参照していきたい。

あるいはまた、この課題を解決することを目指すのであれば、サーバに画像を毎回リクエストして利用するよりもむしろ、ローカルに画像を保存して閲覧した方がよいという観点が出てくることは否めない。近年の著作権保護期間満了コンテンツの利用条件緩和の傾向はその方向性を後押しすることにもなり得る。とは言え、利用動向の把握がデジタルアーカイブの維持・発展のための数少ない鍵であることから、可能な限りそのような事態は避けることが望ましい。現状ではローカルで大量の高精細画像を自在に操作することは困難であるため、しばらくはサーバ上の画像にアクセスするというスキームが続くと思われる。しかし、ローカルという選択肢が現実のものとなったときに、それでもサーバ上の画像にアクセスしてもらえるようにするには、利用者からのデジタルアーカイブへの信頼感の醸成が重要であるということになるだろう。

## 7. 結論

本稿では、IIF 画像の利用状況の解析手法と、解析結果の可視化手法としてヒートマップの生成手法を示した。従来の評価指標である画像のアクセス数からは何うことができない、画像のどの

領域がアクセスされているのかといった詳細な利用状況を把握することが可能となる。解析結果ならびに可視化結果の応用例として共同研究プラットフォームや翻刻プラットフォームでの活用の可能性について述べた。一方、匿名性、研究の先取性に係る課題については今後慎重に検討する必要がある。

## 謝辞

原稿を注意深くお読みいただき、建設的なコメントをいただいたことに対して、匿名査読者に感謝する。また、各ピクセルがアクセスされる確率の数式(4.1節)について確認していただいた、西岡宗氏に謝意を表す。

## 参考文献

- [1] Snyderman, S., Robert S. and Tom C.: The International Image Interoperability Framework (IIIF): A community & technology approach for web-based images, *In Proceedings of the Archiving Conference*, pp.16-21, Society for Imaging Science and Technology (2015).
- [2] Jones, S., Cunningham, S. J., McNab, R. and Boddie, S.: A transaction log analysis of a digital library, *International Journal on Digital Libraries*, Vol. 3, No. 2, pp.152-169 (2000).
- [3] Hashemi, S. H., Hupperetz, W., Kamps, J. and van der Vaart, M.: Effects of position and time bias on understanding onsite users' behavior, *In Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval*, pp. 277-280, ACM (2016).
- [4] Pääkkönen, T.: Crowdsourcing metrics of digital collections, *LIBER Quarterly*, Vol. 25, No. 2, pp.41-55 (2015).
- [5] Kelly, D. and Belkin, N. J.: Display time as implicit feedback: understanding task effects, *In Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, pp.377-384, ACM (2004).
- [6] Fox, S., Karnawat, K., Mydland, M., Dumais, S. and White, T.: Evaluating implicit measures to improve web search, *ACM Transactions on Information Systems (TOIS)*, Vol.23, No.2, pp.147-168 (2005).
- [7] Oard, D. W. and Kim, J.: Implicit feedback for recommender systems. *In Proceedings of the AAAI workshop on recommender systems*, Vol. 83, (1998).
- [8] Yi, X., Hong, L., Zhong, E., Liu, N. N. and Rajan, S.: Beyond clicks: dwell time for personalization. *In Proceedings of the 8th ACM Conference on Recommender Systems*, pp. 113-120, ACM (2014).
- [9] 鈴木親彦, 高岸輝, 北本朝展. IIIF Curation Viewer が美術史にもたらす「細部」と「再現性」 絵入本・絵巻の作品比較を事例に. *じんもんこん 2017 論文集*, 2017, pp.157-164.
- [10] 北本朝展, 山本和明. 人文学データのオープン化を開拓する超学際的データプラットフォーム

- の構築. *じんもんこん 2016 論文集*, 2016, pp.117-124.
- [11] 佐藤正尚, 太田一行. Mirador を利用したクラウドソーシングによるコラボレーション・システム—人文学の共同研究における分析と理論の構築を支援するシステムの提案. *研究報告人文学とコンピュータ (CH)*, 2017, Vol. 2017-CH-114, No. 7, pp.1-6.
  - [12] Carletti, L., Giannachi, G., Price, D., et al.: Digital humanities and crowdsourcing: An exploration, *In Proceedings of Museums and the Web*, (online), available from <https://mw2013.museumsandtheweb.com/proposals/digital-humanities-and-crowdsourcing-an-exploration/> (2013).