

Aggregating and Analyzing Digitized Japanese Woodblock Prints

John Resig

Visiting Researcher Ritsumeikan University

jerosig@gmail.com / <http://ukiyo-e.org/>

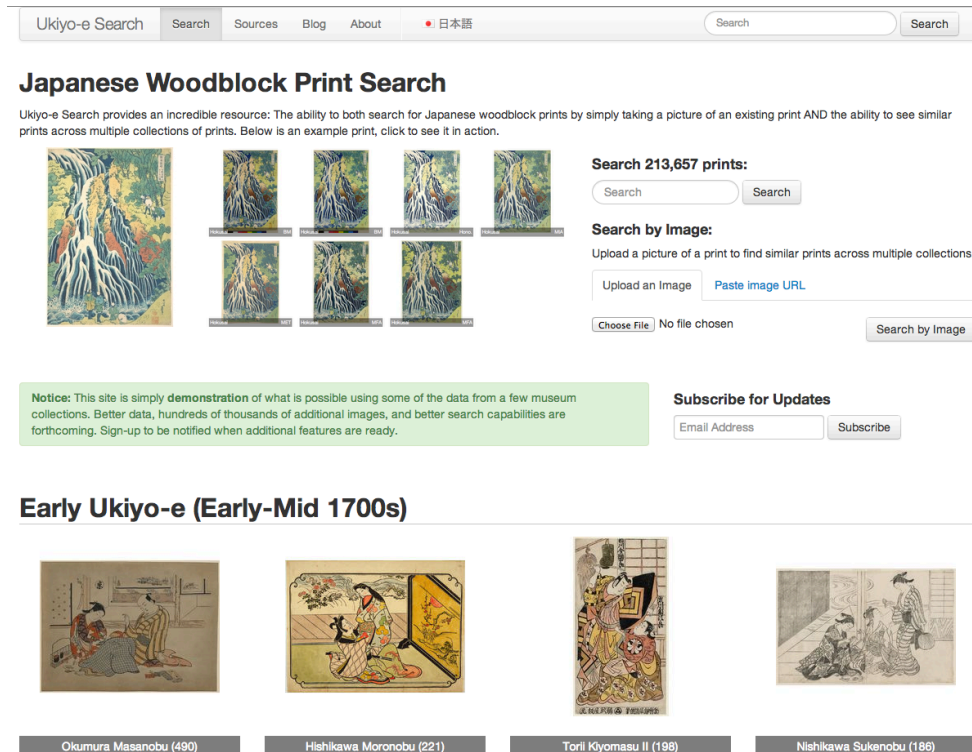
We are quickly approaching a golden age in the study of Japanese woodblock prints and printed art in general. Many museums, libraries, and universities have been successful in digitizing large portions of their collections and have made the images, and associated metadata, available for study.

This process has been a major boon to art historians, collectors, and other researchers. Instead of digging through old, expensive, incomplete, and half-forgotten books, a researcher can simply type a query into a museum database and find more information about a piece of art.

However, this approach only goes so far: while digitization efforts have greatly improved, as of late, there has not been an associated increase in the quality of cataloging. Frequently, the struggle to improve print cataloging is a result of the inability to find more information about a work of art or the lack of collaboration between institutions. The result is that while there are more works of art online than ever, it is still a major struggle for researchers to find the information they are looking for.

Ukiyo-e.org is a database of Japanese woodblock print images, and metadata, from institutions that have large numbers of digitized woodblock prints. Launched December 2012 the website is operated as a non-profit tool for Japanese woodblock print researchers to use. The prints were collected over a year from late 2011 to late 2012. The database currently contains over 213,000 prints from 24 institutions.

The prints were collected by traversing the publicly-accessible digitized collections of prints at the targeted institutions. The images were copied and saved to a separate server for fast access (a technique which avoids overburdening the institutions by loading the images indirectly from another server). The information on the website is organized broadly by artist and time period on the homepage, but is primarily designed to be used as a search engine allowing users to search both by text and by image.



The homepage of the Ukiyo-e.org website.

The textual metadata associated with the collected print images (such as print title, artist, date created, and additional descriptive text) have been aggregated and are displayed next to the image.

Artist: [Torii Kotondo](#)

Title: Combing

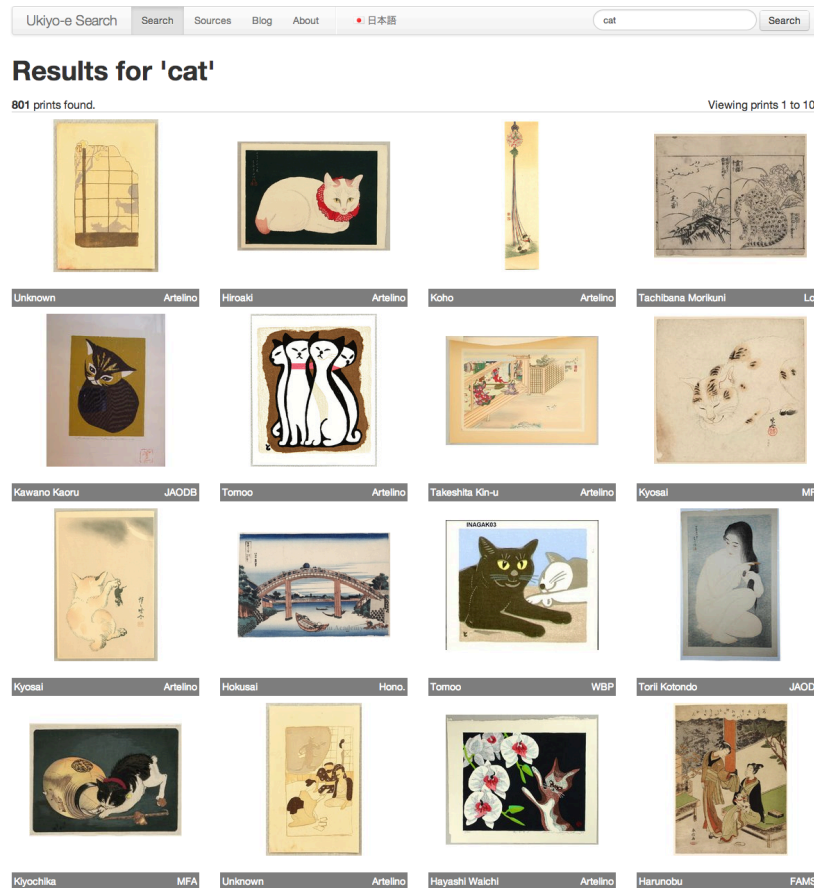
Date: October 1929

Details: [More information...](#)

Source: [Minneapolis Institute of Arts](#)
[Browse all 2,121 prints...](#)

A listing of metadata associated with a print.

The textual metadata is also indexed in a text search engine, making it possible to search for specific terms that the institutions provided with the prints.



Search results for 'cat' on Ukiyo-e.org

Any text searching that occurs is dependant upon the text titles and descriptions that the institutions provide with their prints. No visual analysis of the contents of the prints is performed.

In the current version of the Ukiyo-e.org website, there is some limited metadata rectification occurring. Specifically, most popular artists' names are analyzed and converted so that they match one another. A limited database of Japanese artist names and aliases was built (largely derived from the artist listings made available by the British Museum) and common names were detected. For example, the artist Utagawa Hiroshige was matched if any of the following names were found:

- Utagawa Hiroshige
- Ando Hiroshige
- Andō Hiroshige
- Hiroshige
- 歌川広重
- 広重

All of these names were converted to be just the default name, in this case "Utagawa Hiroshige," across all institution records. Note that for a number of the most popular artists,

there are associations between the English (romaji) version of their name and the Japanese (kanji) version. This makes it possible to search for prints by a particular artist across both English-based institutions and Japanese-based institutions: something that is prohibitively hard to do for most researchers. At the moment, there is no other correction of metadata occurring (dates and print titles, for example, are left intact).

An advantage of having most of the most popular artists' names be made available in both English and Japanese is that it becomes possible to have two versions of the website: one for those that prefer English¹ and one for those that prefer Japanese.² While the translation of artist names is done automatically, as can be seen in the following images comparing listings for the artist Suzuki Harunobu, most of the rest of the text on the site was translated manually. Note that the rest of the metadata provided by the institutions (such as print title, series, etc.) is not translated at this time.

Suzuki Harunobu

1,832 prints found.



Harunobu

BM

Harunobu

MET

Print listing for Suzuki Harunobu, in English.³

鈴木春信

1,832 件



春信

BM

春信

MET

Print listing for Suzuki Harunobu, in Japanese.⁴

浮世絵検索

検索

源

ブログ

このサイトについて

English

検索

検索

浮世絵検索

検索 浮世絵, 錦絵, 明治, 新版画, 創作版画。

検索 213,657 浮世絵:

検索

検索

画像で検索:

テキストの代わりに画像で浮世を検索します。

画像のアップロード

画像の URL を貼り付け

Choose File

No file chosen

画像で検索

The Japanese version of the Ukiyo-e.org website.

¹ English version of the Ukiyo-e.org site: <http://ukiyo-e.org/>

² Japanese version of the Ukiyo-e.org site: <http://ja.ukiyo-e.org/>

³ Print listing for Suzuki Harunobu, in English. <http://ukiyo-e.org/artist/suzuki-harunobu>

⁴ Print listing for Suzuki Harunobu, in Japanese. <http://ja.ukiyo-e.org/artist/suzuki-harunobu>

There have been a number of challenges in collecting prints from museums and other institutions. Creating the tools capable of collecting the metadata and images from the institution's websites presented the primary technical challenge. While some websites are well constructed and easy to use (making it easier to collect the images and metadata), many institutions' websites are constructed in ways that make it difficult to collect data from.

After the launch of the website a number of institutions saw the immediate value in having a tool of this quality being made available to them. For many institutions, it helped them to find other copies of prints in their collection or correct cataloging errors.

Image Similarity

One of the most important features of the Ukiyo-e.org website is its ability to do real-time analysis on the images it holds for comparison and searching. When pulling together images and metadata from dozens of databases, it has become apparent that there is frequent disagreement between major institutions regarding the attribution, dating, titles, and other information associated with a print. Because of this incongruous metadata, it becomes virtually impossible to find similar prints among multiple institutions.

The one piece of information that is never under contention is the image of the print itself. The image that is presented by most institutions usually includes a full, straight-on photograph of the print. If we ignore the metadata provided by the institutions and only compare the actual contents of the images, we can find similar-looking prints at different institutions.

Computer Science research into image analysis and comparison techniques has been going on for decades now. Research and implementation is finally at the point where image analysis can be performed against thousands to millions of images simultaneously.⁵ Tools for performing image similarity analysis are also commercially available.

Finding the right tool that would work for the print images that were collected was especially tricky. The features needed for an effective print image search are:

- The ability to search by uploading an image and finding other images.
- The process of adding in a new image, and performing a search with an image, must be fast.
- The engine should be able to find exact matches, rather than just prints that have a similar composition.
- The engine must be able to ignore differences in color, even differences between a color photograph and a black-and-white photograph.
- It must be possible for an image of a single print page, that is part of a multi-page print, to match an image of the complete print.

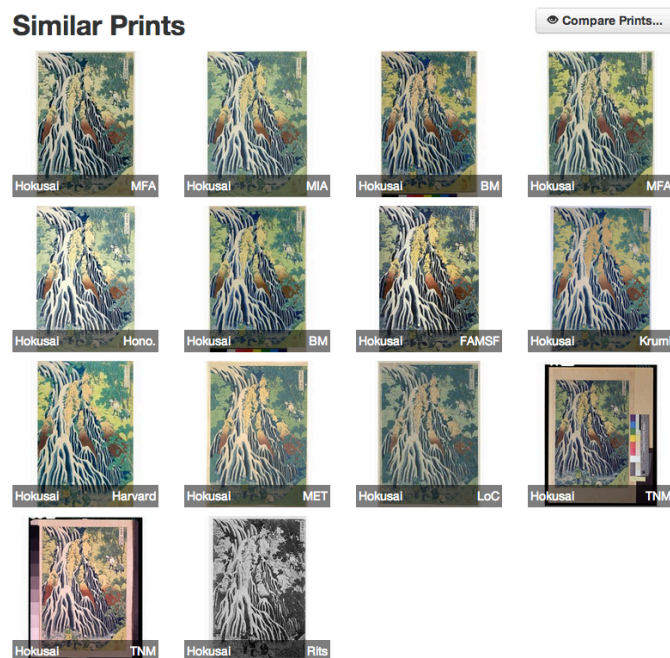
After analyzing a number of image analysis technologies TinEye's commercial MatchEngine⁶ service came to the top as the best possible tool for performing image analysis

⁵ Google, Yahoo, and Microsoft all provide publicly-accessible image search engines that are capable of searching millions, if not billions, of images.

⁶ MatchEngine <https://services.tineye.com/MatchEngine>

across the collected print images.

As an example of what is possible with the image similarity technology, let's look at a popular print created by Katsushika Hokusai around 1832⁷:



Similar print results from the image analysis engine.

The image matches found by the image analysis engine are very impressive. It was able to find matches from 11 different institutions instantly. This includes images of prints that are slightly faded, images that include photographic color bars, and even a black-and-white image of a print.

As a more-advanced example of what is possible with image similarity matching, let's look at one particular diptych created by Torii Kiyonaga in 1784⁸:

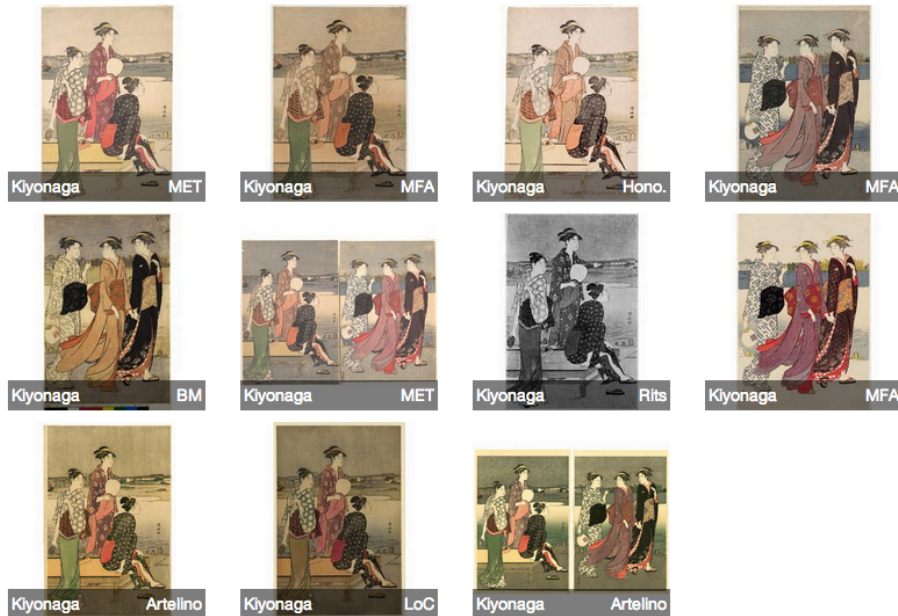
⁷ Katsushika Hokusai, Kirifuri Waterfall at Kurokami Mountain in Shimotsuke, from the series A Tour of Waterfalls in Various Provinces, at the Metropolitan Museum of Art
<http://ukiyo-e.org/image/met/DP141256>

⁸ Torii Kiyonaga, Summer Twilight on the Banks of the Sumida, at the Art Institute of Chicago
http://ukiyo-e.org/image/aic/59250_268632



Similar Prints

[Compare Prints...](#)



Similar print results from analyzing a diptych.

The primary diptych that we're looking at is one presented by the Art Institute of Chicago. The image analysis was able to find similar prints at seven other institutions. There are some interesting things to note about the matches that were made:

- The prints, as presented by the Art Institute of Chicago, are actually in the wrong order. (The panels of the diptych should be reversed.) Even with this mistake, the engine is still able to find the same diptych at the Metropolitan Museum of Art and Artelino, both of which are in the correct order.
- Even though the original search was done with a diptych, individual prints were found at other institutions (such as the Museum of Fine Arts and the Honolulu Academy of Arts).
- The coloration on the prints tends to vary based upon how badly faded they are (or if they are a later reproduction). Even with these differences in color, the similar prints are still found.

- The search is even able to find one of the prints of the diptych in black-and-white from the Ritsumeikan ARC database.

A tool like this can be of immediate help to researchers, curators, and other scholars. Being able to quickly find the same print at other institutions can be indispensable for finding out more information about the print, when it was created, who created it, and what edition the print is from.

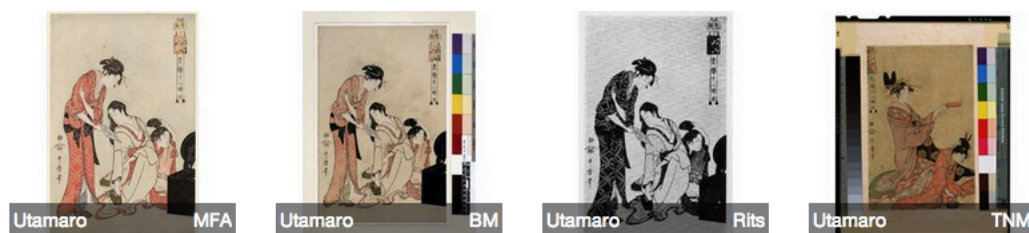
One additional point to consider is that due to the speed of the image analysis, new images (either from existing resources or from newly-added institutions) will be instantly indexed in the engine and be made available for search.

There is a relatively minor problem with the image analysis that is important to note. Since exact matches between images are attempted, any sufficiently large portion of the image will serve as a match against other images.

This particular problem can be seen when there are large, obtrusive color bars used in the images of the print. For example, in the following print⁹ by Utamaro from the Tokyo National Museum the print matches other, similar, prints first (including a black-and-white photo of the print). However, it then matches a different print from the same institution (largely due to the fact that the prominent gradient and color bars in the photograph are causing the match to occur).



Similar Prints



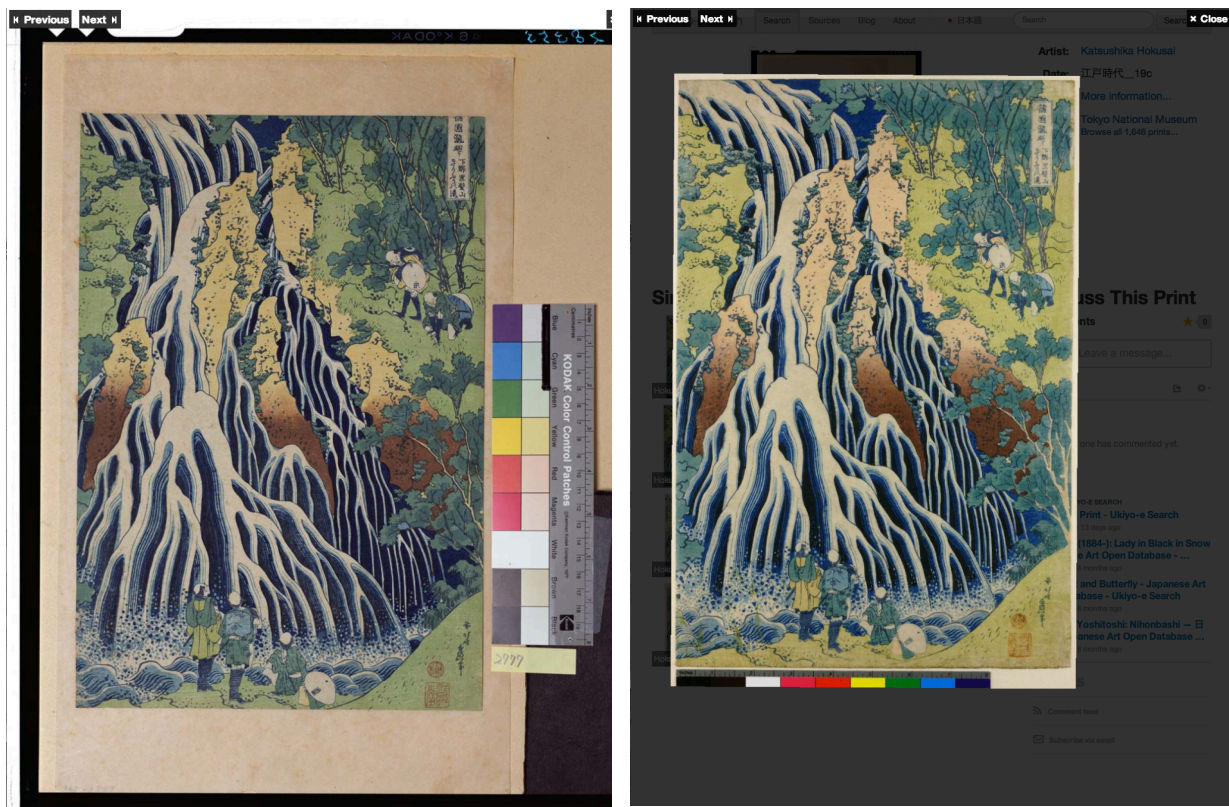
Erroneous similar print match from a color bar used by the Tokyo National Museum.

⁹ Print by Kitagawa Utamaro from the Tokyo National Museum <http://ukiyo-e.org/image/tnm/C0028053>

This problem isn't insurmountable, however. If a contributor were to go through the images from the institutions in question and crop the color bars off the image would fix most of the problematic matches.

The Ukiyo-e.org website provides one additional tool to print researchers hoping to learn more about a print and the differences between it and other copies. Frequently, researchers will visually compare two prints to see what differences in printing may exist. Even small differences in the printing may be indicative of a later edition of the print or even a complete re-carving of the original woodblock.

The image comparison tool, available on any print page, will show a full-size view of the print in question. When viewing a second, similar print, the second print will be lined up precisely on top of the original print. The contents of the similar print images are analyzed in order to ensure that even if there are other details in the image (such as color bars) or even if the photo is taken at an angle, the prints will be aligned for direct comparison.



Two images aligned for direct comparison.

Image Similarity Search

The most exciting aspect of having an image analysis engine behind the prints is that it can grossly simplify research. The Ukiyo-e.org site was originally created with the intention of being able to simplify the process of searching for additional information about a print.

Traditionally a researcher is only able to figure out the artist, and possibly the date, of a

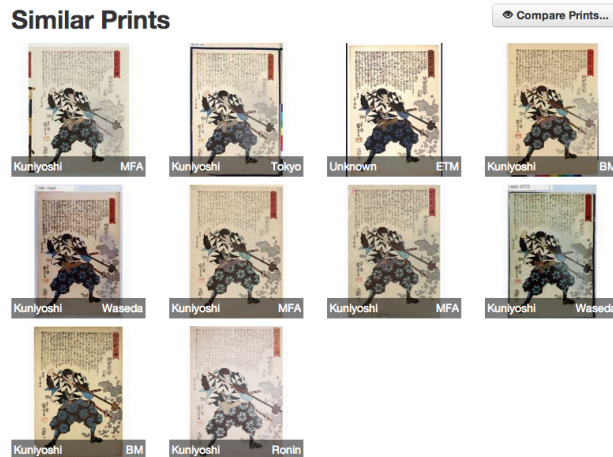
print if they are able to read the artist's signature and publisher or censor's seals. Unfortunately, this is not an easy skill to acquire and is insurmountable for many people who are casually interested in prints.

Even if the researcher is able to read the signatures and seals on the print there is no guarantee that they'll be able to locate the print in the various institutions that have it. The print may be mis-cataloged, mis-attributed, or, more likely, just hidden in the hundreds of institutions and hundreds of thousands of prints that exist online.

This is where the advantage of having an indexed, instantly searchable, image analysis search engine becomes exceedingly apparent. Anyone can visit the Ukiyo-e.org website and upload an existing image of a print. Upon doing so, similar prints will be located amongst the current database of print images.



Positioning the phone camera to encapsulate the entire print inside the photograph.



The image similarity results from uploading a photo taken by a mobile phone.¹⁰

While a researcher can upload an image manually from their desktop computer, it can prove to be very effective to upload images while “in the field” physically examining prints. This can be done from any mobile phone with a camera and a web browser. There is an immense advantage to researching a print while in the field using a camera on a mobile phone:

- A researcher will no longer need to take a picture of a print then do additional research back at home with their research books.
- An interested user will be able to learn more about a print that they see at a museum, compare it to other prints at other museums, or even learn more about the print on the museum’s own website.
- A collector would be able to do comparisons against other prints while examining a print at an auction house.

Aiding Woodblock Print Studies Using Image Analysis

The image matching technology is able to make some very interesting original research possible, the likes of which has been challenging, if not outright impossible, to perform using traditional research techniques. There are two Japanese woodblock print research topics that the Ukiyo-e.org site has immediate impact upon: tracking the re-use of woodblocks by

¹⁰ An example search result page: <http://ukiyo-e.org/upload/788cfae9604103764e79b4521d682190>

publishers and correcting missing attribution of prints in institutions.

During the Edo period (1603-1868) woodblocks were frequently sold and reused. Locating these particular transactions and modifications to woodblocks is especially challenging. At best, it requires a substantial amount of research by a trained expert looking at thousands of prints. This is an area in which the Ukiyo-e.org image analysis engine excels. Comparing the contents of two matched images produces some interesting findings. For example, see the following prints by Utagawa Kunisada:



Utagawa Kunisada, Metropolitan Museum of Art¹¹



Utagawa Kunisada, Ritsumeikan ARC Database¹²

———One of these prints depicting a Kabuki actor in a theatrical role was created first, and presumably the woodblock was sold to another publisher (confirmed by the fact that the publisher's seal is different in the second print). Additionally, the face of the actor, the text in the upper-left-hand corner of the print, the actor's crest, the designs on the kimono, and the colors of the print have all been changed. Even with all of these changes, they are both indisputably from the same woodblock. Which came first and the context of the transaction is an area for further study.

Since only the image is ever analyzed and compared, and none of the metadata, it can't

¹¹ Utagawa Kunisada, Metropolitan Museum of Art <http://ukiyo-e.org/image/met/DP148988>

¹² Utagawa Kunisada, Ritsumeikan ARC Database <http://ukiyo-e.org/image/ritsumeikan/arcUP0086>

make the mistake of limiting its search only to prints that match a particular artist name. This is a good thing, as the similarity of the following two prints may never have been discovered.



Katsukawa Shunsho, Metropolitan Museum of Art
13



Katsukawa Shunko, Museum of Fine Arts Boston
14

These two actor prints are obviously derived from the same woodblock. The only major component that's changed is the Kabuki actor's face (presumably the head region of the woodblock was chopped out and replaced with a new carving). On this particular print there is no publisher information, so we can't be certain that this woodblock was sold or exchanged

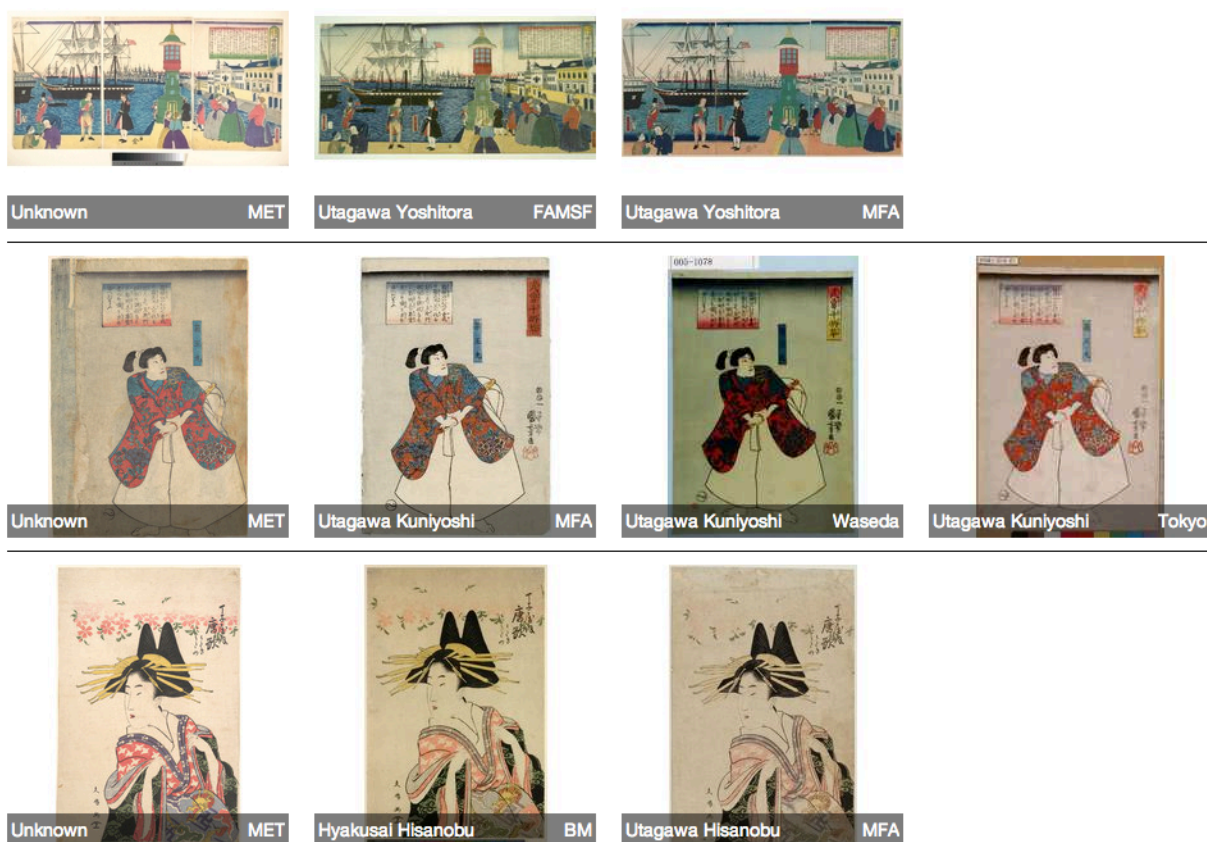
¹³ Katsukawa Shunsho, Metropolitan Museum of Art <http://ukiyo-e.org/image/met/DP134583>

¹⁴ Katsukawa Shunko, Museum of Fine Arts Boston <http://ukiyo-e.org/image/mfa/sc214530>

hands. One especially interesting point is the signature on both prints: the print on the left is signed by Katsukawa Shunsho, whereas the one on the right is signed by Katsukawa Shunko (a student of Sunsho). Why the signature on the print was changed (and from which name) is another area of potential study.

This match is especially interesting because the prints are depicting different actors, signed by different artists, and are held in different institutions. The likelihood that the prints would've been identified as similar is infinitesimal. However, this incongruity was discovered automatically by the Ukiyo-e.org attribution correction tool.

The attribution correction tool, which is still under development and not yet available for public use, was developed to utilize Ukiyo-e.org's image matching technique (based on image similarity rather than metadata) specifically to aid researchers in the discovery of attributions for unattributed prints. An initial test of the tool was made to find prints in the Metropolitan Museum of Art that are unattributed. Some of the results are seen here:



Correlation of prints with an unknown artist with prints that have a known attribution.

Here are three prints from the Metropolitan Museum of Art that have no artist attribution. These prints all have an image match at another institution where there is a correct attribution. With a quick scan, it becomes obvious that the three prints should be attributed to Utagawa Yoshitora, Utagawa Kuniyoshi, and Utagawa Hisanobu, respectively.

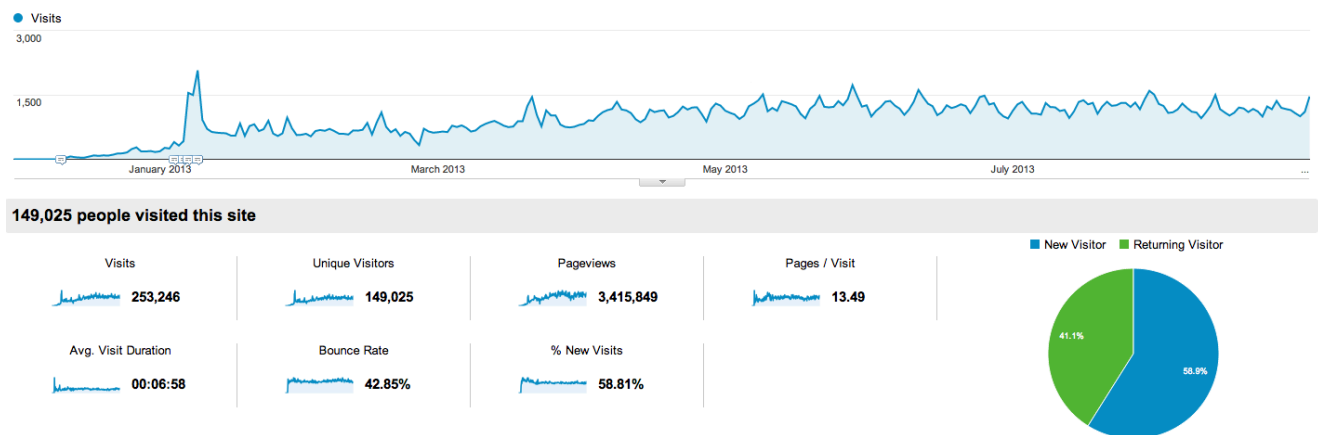
This tool can also be used to find cases where there is disagreement between

institutions, even though there is a visual match. It was in such an instance that the discovery of the Shunsho/Shunko prints was made.

The applications for this tool are very exciting. An institution's collection can be automatically scanned for missing, or conflicting, attributions, and can automatically be corrected with the aid and interpretation of a curator. It can even be used to provide attribution for prints that have never been cataloged before. Simply scanning a collection of prints and providing a basic, preliminary, attribution for as many prints as is possible becomes quite feasible with the aid of this tool.

Users

The Ukiyo-e.org website has attracted a considerable number of users since launching in December of 2012. Around 150,000 people have used the database, accounting for 3.5 million page views. There have been 12,500 images upload searches completed, 58,000 text searches, and 655,000 prints viewed.



Ukiyo-e.org usage statistics from Google Analytics.

The majority of these users are arriving at Ukiyo-e.org primarily through Google searches. Since launching the website publicly, there has been steady growth coming through Google from people searching for terms related to Japanese woodblock printing. The present ranking for the Ukiyo-e.org site on Google is relatively very high, considering how new it is:

- #3 for Ukiyo-e
- #5 for Japanese Woodblock Prints
- #15 for 浮世絵
- #2 for Utagawa Yoshitora
- #2 for Katsukawa Shunko
- #16 for Utagawa Kuniyoshi

The web site is already ranking very high for common terms such as “Ukiyo-e” and “Japanese Woodblock Prints”. It also ranks highly for less-popular artists such as Utagawa

Yoshitora or Katsukawa Shunko.

Of the people who use the web site, the majority of the users are non-English speaking. Approximately 43% of the users are English speaking, 32% speak Japanese, and 25% speak another language, primarily. The percentage of Japanese-speaking users dramatically increased when the Japanese version of the web site launched (and this number continues to grow).

Language	Visits	% Visits
1. en-us	99,747	39.39%
2. ja	52,056	20.56%
3. ja-jp	29,104	11.49%
4. fr	8,127	3.21%
5. zh-cn	6,936	2.74%
6. de-de	6,407	2.53%
7. en-gb	4,758	1.88%
8. en	4,738	1.87%
9. it	3,737	1.48%
10. es	3,618	1.43%

The languages and locales from which the Ukiyo-e.org users come.

Judging by the emails users have sent in, it has become obvious who is using the web site and for what purpose. Users can be broken down into a few categories:

- Laymen seeking to identify a print using the image search.
- Laymen interested in Japanese art and culture.
- Ukiyo-e researchers. (Using both the image search and text search.)
- Dealers. (For print identification, especially.)
- Museums. (For print identification and to improve their catalogs.)

A number of museums have been in contact asking to be included in the site's database. Unfortunately time and resources are limited, restricting the ability to add more institutions without considerable work.

The Future

There are many areas ripe for further research and improvement to the Ukiyo-e.org website. Ukiyo-e.org's greatest ongoing need is to acquire better control over, and access to, metadata such as dates, publishers, and actors. Right now, the metadata in the database is undergoing very little processing and improvements to it would result in significant benefits (such as the ability to search for prints by date or being able to plot out trends in print production over time). The collection of additional prints and metadata from institutions is ongoing.

Managing user contributions will certainly be the next largest area of expansion. At the

moment, there is no way for an individual, or institution, to add print images to the website. Nor is there any way for a scholar to fix mistakes that they see on the site. Ideally, contributors will eventually be able to add new information to the website as they see fit.

The biggest limitations for the Ukiyo-e.org project have been time and cost. The website has been purely a side project for myself, developed in my spare time. The most substantial aid to growth would be the contribution of additional developers and development time to the project. The second most important contribution would be funding to help offset the significant costs of running the server and image analysis.

Institutions would be able to greatly improve the project by digitizing more of their collections and making more print images available. Any additional metadata that could be provided would be appreciated, but not required.

Work on the web site is ongoing. Future announcements can be found on the Ukiyo-e.org site, mailing list, and blog.