Mapping San Francisco's Photographic History
"The San Francisco Wayback Machine"
http://SFwayback.org

Isaac Salier-Hellendag
UC Berkeley School of Information MIMS '09
dish@ischool.berkeley.edu

Project Track 3: Communication and Memory
Project Advisor: Professor John Chuang

Abstract

This project aims to combine collections of historical and current images of San Francisco in a single, simple-to-use web interface. Users may filter and search for photos, select photos within a map interface, bookmark photos for future reference, and view side-by-side comparisons of photos with Google Street View. The project is intended to be highly extensible and ultimately, easily reusable for image collections in other cities or regions. The project site is available to the public at http://SFwayback.org.

Research Problems

This project is focused on solving three major problems of information retrieval and availability.

Multiple collections, multiple interfaces

For researchers and historians searching for historical images of San Francisco, there are currently no centralized "meta-collections" available for easy retrieval of images across multiple collections. This means that one must visit each collection website to find desired images. This process would be significantly streamlined by creating a single database containing the images and data. This project aims to accomplish this task.

There are quite a few digital collections available on the web that may be included, including the San Francisco Public Library's digital collection (http://sfpl.org), the Online Archive of California (http://oac.cdlib.org/), and the Lawrence and Houseworth database by the Society of California Pioneers (http://www.californiapioneers.org/lh/).

Inadequate image retrieval interfaces

Retrieving images from these collections generally involves a simple retrieval tactic: keyword search. Keywords may be found within titles, notes, descriptions, or other metadata. However, this is a rather basic and often ineffective method of searching for images.

Some collectors, such as the San Francisco Public Library, allow users to browse by neighborhood. However, their neighborhood definitions are not standardized and the filtering interface could be improved. There are also no options to search for more exact locations (latitude/longitude, notably) because this metadata does not exist within the collection databases.

Next, these images are not easily comparable to any modern-day version of the same image. Moreover, they may not be easily viewed in relationship to other historical photos in their area or time frame.

Finally, many photographs are available through sub-collections or exhibitions specific to single events, such as 1906 earthquake exhibits. These photos may be made more available through conventional search and filtering methods.

This project aims to confront these issues of retrieval by providing the following within a publicly-accessible web interface:

- More robust filtering options: neighborhoods, specific date ranges, distinction between street-level and aerial photos, and more

- Standardized neighborhood boundaries, as defined by the San Francisco government. This standardization should allow better comparison across collections, which currently may not share neighborhood delineations.

- A simple, user-friendly interface, including a map to show geolocated images.

Insufficient user input

When collectors enter images into their digital databases, the metadata assigned is essentially permanent, and may not be committed with any input from other individuals who may have valuable expertise. While archivists and curators are experts on the history of the images they are cataloging, there may be input from other historians and researchers that could be vital to the accuracy and completeness of the historical record.

By encouraging users to contribute image locations, match photos with Google Street View images, and discuss images within a community, this project aims to combine collectors' metadata with the historical expertise of users who currently have no way to provide their own knowledge.

Target Audience

The target audience for this project includes two major groups: researchers interested in finding photographs for professional or academic purposes, and people simply curious about the history and evolution of the city. Researchers will likely already be familiar with the current methods for finding digital photos, and this interface will ideally help alleviate current issues of retrieval and collection. Moreover, for new researchers unfamiliar with the range of collections and ideal retrieval techniques, this interface should help simplify the process and eliminate the need to learn how to effectively search and browse each individual collection.

For users who simply have a curiosity about the history and growth of San Francisco, this project provides a novel method of exploration and learning about the city without carrying out intensive research. By centralizing images in one database and using familiar interface elements (maps, date dropdowns, Google Street View), users who are not familiar with retrieving data from collections may find relevant photos with minimal expertise.

Both types of users are encouraged to contribute to the project through location identification, discussion, and in the future, tagging and photo uploads. Anyone with knowledge of city or neighborhood history is welcome to take part.

Development

The project interface was developed using the following:

- PHP 5.2.9
- MySQL 5.0.51a
- Javascript (including Ajax with JSON)
- Semantically-sound HTML, CSS

Given my previous web development experience with PHP, MySQL and Javascript, it was logical to build this interface using these technologies. The site is currently hosted by Nearlyfreespeech.net on a shared FreeBSD server using Apache. If the project scales in months and years to come -- as I hope it

will -- it will be sensible to relocate the project to a dedicated server.

At the time the writing of this paper, there are approximately 700 photographs imported into the database. The large majority are images from the San Francisco Public Library collection. Thousands more will be imported in the future, but this number is appropriate for effectively demonstrating the capabilities and features of the project.

The project was developed using Firefox 3 with Firebug, and Safari 3, in Mac OSX version 10.5.6. Internet Explorer is not currently supported, though users may find that it works without trouble. Full IE support will be added soon.

Design Overview

Before commencing development, I established a list of features I wished to include within the project interface. These features are listed and described in detail later in this paper. After forming the feature list, I sketched several design mockups for potential layouts and workflows. Key interface/coding design considerations included:

- **Ajax usage.** While I wished to include asynchronous data retrieval to streamline the user experience, I wanted to avoid excessive Ajax usage. It seemed most sensible to include Ajax in interface components where Ajax is commonly seen: filtering options, simple navigation (pagination, tabs), and the mapping interface.

- **Map implementation should seem familiar.** Regular internet users will no doubt have used Google Maps or other similar mapping interfaces. It would be a mistake to depart from the mapping paradigms such interfaces have established.

- **Logical filtering options.** Certain filtering options are especially useful: neighborhoods, date range, and keyword search. These would be included. Filtering by collection and subject could be useful for a very specific group of users, but would otherwise not be helpful in the interface.

- **Displaying photos that could not be mapped.** Some photos, notably aerial and panorama images, cannot easily be identified by latitude/longitude and can therefore not be properly noted on the map.

- **Easy bookmarking.** Photos should be very easy to bookmark and find without repeating a search.

- **Natural history/modern-day comparison.** Side-by-side comparison between photographs and Google Street View seemed to be an ideal choice. Since Street View is also a familiar web interface, it would be important to implement it in a comfortable, easily understandable manner to allow proper comparison.

- **Extensibility.** A truly vital feature of any implementation of a project like this one. The project should have very little, in terms of design or code, that would specifically tie it to San Francisco. Ideally, the project should be reusable for any city with historical photographs,

definable neighborhood boundaries, and Google Street View.

Content Management - Omeka

Recommended to me by Professor Eric Kansa as an ideal content management system for digital collections, Omeka (www.omeka.org) serves as the core CMS for the images and metadata. Omeka provides a robust administrative area allowing easy addition, removal, and editing of digital items. It allows input of Dublin Core metadata by default, as well as more specific metadata elements for item types (ie, still images, documents, oral histories). Moreover, it allows administrators to add custom metadata elements to item types. This would prove perfect for a project like this one, as I would need to add a number of new metadata elements, including latitude, longitude, Google Street View settings, and more.

Front-end APIs - OpenLayers and Google Street View

OpenLayers is a powerful Javascript library that allows developers to easily place map data from any source into a web interface. It provides several features that would be crucial to this project:

- Simple addition/removal/movement of placemarks. Used for marking and changing photo locations.

- Drawing geometrical figures. Used for outlining neighborhood boundaries.

- Zooming, panning, click detection

- Map data source flexibility. Could be used in the future to include historical San Francisco map data in conjunction with photos. For instance, it would be possible to import Sanborn Maps with plot-by-plot delineation of the city.

Google Street View is a relatively new feature from Google, allowing users to view panorama images of essentially any street location in San Francisco. Street View has expanded to include most, if not all, major US cities. Street View allows users to adjust panorama images by pitch, yaw, and zoom. This ability allows users of this project to nearly duplicate any street-level historical photograph.

The Street View API is available to developers via a Javascript interface that inserts a Flash object into a target HTML element.

Features of the Site

In this section, I describe the publicly-accessible features of the web interface itself, including design and implementation details. As noted earlier, the project is available to the public at http://SFwayback.org.

Registration / User Tracking

Upon first arriving at the site, the user is asked to register to use the site. I have put this requirement in place for two reasons. First, because the site encourages community input, it is important to connect contributions to individual users. An administrator or community manager -- in this case, me -- could have a tremendously difficult task at hand if anonymous contributions are allowed. Anonymous contributions are more likely to lead to vandalism issues and degraded usefulness of the service. Second, the photographs are neither my property nor the property of the project. Rights to the photos are held by the collections or other individuals. It is therefore sensible to control access.

After registering, new users will receive a confirmation email with a link to a validation page on the project site. Users who have not validated their email addresses will be unable to log into the site. After confirmation, however, users will have full access to all site features. All user-generated actions, including location identification and comments, will be attributed to the individual's user account.

Finding photos

The filter interface near the top of the site is intended to provide a simple yet powerful way for users to find specific photos and explore San Francisco history.

Photo Type

I have divided images into several different categories of "photo type," which broadly describes the subject of the image.

- **Street level.** Images of buildings, people, or other subjects from a street-level vantage point. Street level photos may be identified with a specific latitude and longitude based on camera position. Moreover, these photos can be matched with a similar Google Street View image.

- **Aerial.** Photos taken from airplanes, satellites, or other high-elevation vantage points. These photos generally cannot be applied a latitude and longitude, and cannot be compared with a Google Street View image.

- **Panorama.** Similar to aerial photos, but often with a wide field of view and a very broad range of visible areas. Like aerial photos, these often cannot be associated with a specific latitude and longitude, and cannot be compared to any Google Street View images.

- **Portrait/Indoor/Other.** These images can be catalogued within the database, but are not presently visible through the interface. This is because many of them cannot easily be applied with even basic geographic metadata, and would therefore difficult to find. They may be added to the interface in the future, if an appropriate solution becomes apparent.

Neighborhoods

Users may select up to three neighborhoods to search by toggling a dropdown checkbox menu through the "Select" link. After choosing three neighborhoods, the remaining checkboxes are disabled via

Javascript. This is in place to prevent users from attempting to request images from too many neighborhoods, which could potentially cause database, Ajax, or display issues.

Neighborhoods are specifically delineated by the boundaries defined by the San Francisco Planning Department. A link to their neighborhood map is available within the dropdown. While these neighborhood boundaries may not line up perfectly with every user's concepts of San Francisco neighborhoods, they are a good solution for standardizing boundaries. Thirty-seven neighborhoods are available for selection.

Note that some images, especially aerial or panorama, may contain multiple neighborhood identifications. This is because wide-range images often contain such a broad view that more than one neighborhood is visible. Other photos with multiple neighborhood identifications may be considered "border" locations near or between two neighborhoods. Each image has two neighborhood metadata elements available. This feature explains why filtering for a single neighborhood may result in mapped placemarks that lie outside the specified neighborhood boundary.

Date Range

Date range options are provided through a basic selection interface, with a dropdown for a "from" date and a "to" date. The dropdowns currently allow users to select as far back as the 18th century. This allows the possibility of adding in paintings, sketches, etchings, or other non-photographic image formats that may be useful. Selecting an impossible date range, such as one in which the selected "to" date is earlier than the "from" date, will disable the "Find photos" button and show a red "X" icon to warn the user.

Exactness

Users are presented with the option to show images that have exact locations (latitude/longitude) already identified or exact dates identified. "Exact date," in this case, means that the image's exact year is known, since date and month data is often unavailable.

Keyword Search

A basic keyword search feature is included, above the "Find photos" button. Users may enter any number of keywords. Keywords are sought within image titles, notes, and subjects. Note that keyword searches are currently only available as "OR" searches, and phrase searches (using quotation marks) are not available. These search features may be added in the future. Note also that keywords are sought within the range of images specified by current filtering options.

Performing the Request

Images are retrieved via an Ajax request performed when clicking the "Find photos" button in the filter interface. Performing a request will show a "Searching..." dialog and an animated icon within the map area. The Ajax request returns a JSON object containing metadata for the images that match the user's specified filter settings. A successful request results in neighborhoods and/or placemarks being drawn on the map, an active photo list under the "List" tab, and the option to export the image data to KML. (These features are described below.) Additionally, the number of retrieved photos and the number of

mappable photos among them is displayed above the map area. An unsuccessful request returns an appropriate message, and suggests that the user broaden his or her filtering criteria.

Viewing photos

Map View

The map view is generated using the OpenLayers API with Yahoo! Maps tile data. It is designed to resemble familiar map interfaces like Flickr Maps (http://www.flickr.com/map/) and Google Maps' Panoramia photographs. The map displays two types of data for images returned by a filter request.

- **Mapped images.** Images that have been marked with specific latitude and longitude are depicted on the map with a gray "photo" icon. Clicking the icon brings up a small version of the selected photo to the left of the map, along with a brief view of the image's metadata and a few options for using the photo. Clicking the icon also changes it from gray to red, to allow the user to quickly note which of the mapped photos he or she has selected.

- **Selected neighborhoods.** Neighborhoods selected by the user will be drawn -- in blue, red, and purple -- on the map. This allows the user to clearly see the boundaries between the neighborhoods. Clicking within a neighborhood does not currently have any effect.

List View

Because not every photo is mappable, it is crucial to provide an interface to browse images outside the mapping interface. The "List" area performs this function. Each returned image is shown in thumbnail format, with its date (approximate if necessary), title, collection, and neighborhood. Selecting an item in the list area displays the image and brief metadata to the left of the list. If the selected item is also mappable, it will be selected in the map if the user goes back to the map pane.

There are four icons the user may notice in the list area. The "book" icon indicates that the user has bookmarked the image. The "heart" icon denotes an image marked as a "favorite" by the user. The "globe" icon is used for images that have been assigned a latitude/longitude and are both mapped and available for Google Street View comparison. The "speech bubble" icon indicates that the photo has had comments, and the number next to this icon shows how many comments there have been.

The list view is broken into groups of five images due to layout considerations, and because it may be unreasonable to perform a large number of thumbnail requests for photos that are not immediately required by the user.

Export to KML

The "Export" area provides a link to download the current set of requested images in KML (Keyhole Markup Language) format. KML is the XML-based schema used in Google Earth. Users may thus export their data to KML to view these images within Google Earth. The Google Earth timeline feature also allows users to specify date ranges and watch as images appear on the Google Earth map at the appropriate time.

Google Street View

For images that have been applied a latitude and longitude, the Google Street View comparison feature is available. As mentioned, selecting an image from the map interface or an image with a "globe" icon in the list interface will bring up the image itself, a small amount of metadata, and the option to view the image in Google Street View comparison mode. Clicking the latter option will bring open a new window with an interface containing the image, a Google Street View at the image's latitude and longitude, and an OpenLayers map with a placemark indicating the location of the photo. In this interface, users can easily make before/after comparisons between the historical image and the modern-day view at the identical location.

Google Street View is available for applicable photographs at:
http://sfwayback.org/streetview/[photo_id].

Using photos for research or interest

Bookmarks

For researchers and historians who may need to find the same images repeatedly over a period of time, a bookmarking feature is available. Within the left-hand area with a selected image and its metadata, a "Bookmark this" option is available. Clicking this link bookmarks the image for the user via an Ajax request and adds a "book" icon to the image within the List view. An image that has already been bookmarked will be noted as such within this area, in place of the link.

In the upper right section of the site, a "My Bookmarks" link is available. This area shows a full list of all bookmarks selected by the user over time. This will ideally allow researchers to quickly find the images that interest them, without requiring brand new searches.

Users may remove bookmarks by clicking "Remove bookmark" in the "My Bookmarks" area. The removal is also performed via Ajax.

Bookmarks are available at http://sfwayback.org/bookmarks.

Favorites

The Favorites feature is very similar to the Bookmarks feature in format, though it is intended more as an interest or entertainment feature. Users may "favorite" the photos that they find coolest or most interesting. Favoriting an image adds to its favorites tally within the database, and it would be a simple task in the future to create a "Top Favorites" feature on the site.

Favorites are available at http://sfwayback.org/favorites.

Full metadata view

In the left-hand pane, a link is available to view all metadata and comments for the selected image. This link opens a new window with all available metadata for the image, including Dublin Core data, any other metadata imported from a collection's website, and geographic data generated within the project

site. Comments are also available in this area, and are discussed in the section below.

The full metadata view is available at http://sfwayback.org/photo/[photo_id].

User contributions

One of the key goals of this project is to allow community input from users with historical knowledge or curiosity about the city. The following features are in place to allow and encourage this input.

Identifying Locations

Because collections do not geotag their images, essentially none of the images that are imported into the project database will initially have latitude and longitude. The task of identifying specific locations for images is left solely to the user base of the site.

The feature is implemented as follows:

1. Clicking a street-level image within the map or list view brings up the image/metadata view in the left-hand area of the site, as described above. For images without latitude or longitude, the "See in Google Street View" link instead reads "Identify this Location". Clicking this link brings up a modified view of the Google Street View interface.

2. The location-identification area has an OpenLayers map and a set of instructions for selecting the image's location. The user may click on the map to identify the correct location of the image. He/she should then click "Confirm this location."

3. The user is then taken to the now-familiar Google Street View interface. He/she may move the placemark within the map to the appropriate location, if necessary, and move the Google Street View to match the historical image as closely as possible. He/she may then confirm the location and Street View, it will be saved via Ajax request, and the page will reload. At this point, the image is geotagged and will be mappable in future filtering requests from the front page of the site. It will also be viewable within the Google Street View interface, using the user's specified Street View settings.

Google Street View revisions

Because not all users may be completely accurate in their identification of locations or Street View images for a given historical image, I have implemented a "wiki" area to allow users to make revisions to this data.

Viewing an image within the Street View comparison area, the user should notice a "View revisions/Make a change" link below the historical image. Clicking this link brings up a similar interface with a list of previous location or Street View revisions. Here, the user may revert to any of the previous revisions, or submit his/her own. He/she may move the placemark on the OpenLayers map to identify a new latitude/longitude, or move the Street View to a more appropriate view.

Note that moving the placemark within the OpenLayers map will update the Google Street View

automatically.

The user may confirm his/her selections by clicking "Save as new revision." The user's changes will be used for subsequent filtering requests from the front page of the interface.

Comments

In order to allow the community to properly discuss the history of the city and the images themselves, a Comments area is available within the full metadata view. Here, users may discuss any aspect of the image or its history. All comments are tagged with the contributing user's username and a timestamp for their comment submission.

Images with comments will be identified as such within the front page List view, as noted earlier in this paper.

Administration

Importing Images

Because historical photographs of San Francisco are almost entirely unavailable through any APIs, the photos currently in place on the project site have been imported manually.

Some collections, such as the San Francisco Public Library's extensive collection, are relatively easy to import through simple data scraping. Others, such as Calisphere, are less simple to import. This is an impediment to fully building out the database of photos for this project. However, it is an impediment that can hopefully be overcome in the future if these collections make APIs available or an administrator (other individuals or myself) are able to build efficient data-importing scripts or applications.

Managing Users

Within the time frame of this semester, it was not possible to build a robust administrative interface to handle users of the project website or their input. This will be an important step in the future to ensure the integrity of user-generated content.

Managing Images

The Omeka administrative area is a highly effective interface for allowing administrators to add, modify, or remove images from the database. Moreover, Omeka software is open-source and code revisions are readily available. I believe it will prove to be an ideal content management system for this project.

Future direction

More community input

- **Tagging.** I have not implemented user-generated tags. These could be very useful for image organization and retrieval, and allow another level of user involvement.

- **User-uploaded photos.** Users cannot currently contribute their own photographs or other images. This could change in the future, especially since users will be able to provide thorough histories and memories for the photos they submit. This could add highly valuable data to the historical record.

- **Using Flickr, Panoramia, or other photo services.** It would be useful to import photos from photo services with APIs, as these may be geotagged and simple to import into the database. I have not yet implemented this because most of these photos are not "historical" images and due to time constraints. In the years to come, however, these images will no doubt become valuable additions to the history of the city.

- **Adding photos with less available metadata.** Some photos have no identifiable neighborhood or unknown subject material. These are not currently available through the project interface. Making these photos available to users could allow them to help identify images that are currently a mystery.

Mobile application for "live view" of history

While this would be a project that may require many weeks or months to implement successfully, the concept is fascinating to me as a San Francisco resident and iPhone user. By using the geolocation capabilities of a mobile device, a user could map out nearby geotagged historical photos and do in-person comparisons with the modern city.

Extensibility to other cities

One of my main goals for this project was to produce a highly-extensible interface that could be reused for essentially any city with a strong record of historical photographs. I have developed it so that only a few requirements should be necessary to reproduce the project for other cities:

- **Images.** Any city or region with historical photographs or other images would be eligible. All that would initially be required would be the task of inserting the images into the Omeka database.

- **Neighborhood boundaries.** Neighborhoods are a key metadata element within this project. An administrator would need to delineate specific neighborhood boundaries to categorize images.

- **Google Street View.** Not truly a requirement if the administrator does not wish to make Street View comparison available, but otherwise required for these comparisons.

Inspiration

As a resident of San Francisco and a longtime Bay Area resident, I have been fascinated by the city and its wild history. I decided I wanted to know more about what the city has looked like in the past and how it has evolved over time, and over time I have found that many San Franciscans who share my curiosity.

After speaking with my friend Rebecca Fogel, an architectural historian in San Francisco, it became clear that the issues of information retrieval that she faces on a daily basis would be fertile ground for a project that both solves information problems and captures the attention of the curious San Franciscan.

San Francisco is a relatively new city with an impressive photographic record that spans much of its history. Further, there are numerous collections with a very large number of digitized images of the city. As a result, it is an ideal location to create a project like this one.

The name "Wayback Machine" refers to the WABAC, the time machine created by the legendary Mr. Peabody, a canine scientist from the animated television program "The Rocky and Bullwinkle Show." The term is also well-known from its use for the web history project developed by archive.org.

Final Notes

This project has been a fascinating experience in exploring San Francisco history, designing a simple and user-friendly interface, and developing a fully-functional web application with external APIs and frameworks. I very much look forward to continuing work on the project, and it is my hope that collection owners and curators will be excited to take part in growing the project in the months and years to come. I am also hopeful that photo collections in other locations will be interested in implementing an extension of this project for their own visitors and researchers.

References

Croft, Jeff, Ian Lloyd, and Dan Rubin. Pro CSS Techniques. Berkeley: Apress, 2006.

"Developer's Guide – KML." May 2009. Google Code.
<http://code.google.com/apis/kml/documentation/topicsinkml.html>

Eason, James. Archivist, UC Berkeley Bancroft Library. Personal interview. 16 Jan 2009.

Fehily, Chris. SQL: Second Edition. Berkeley: Peachpit Press, 2005.

Fogel, Rebecca. Cultural Resouce Specialist, Page & Turnbull. Personal interview. 3 Mar 2009.

"Gecko DOM Reference." January 2009. Mozilla Developer Center. <https://developer.mozilla.org/en/
Gecko_DOM_Reference>

"Google Maps API Reference." March 2009. Google Maps API.
<http://code.google.com/apis/maps/documentation/reference.html>

Moretta, Christina. Photo curator, San Francisco Public Library. Personal interview. 14 Jan 2009.

"Omeka | Documentation." January 2009. Omeka. <http://omeka.org/codex/Documentation>

"OpenLayers Class Documentation." February 2009. OpenLayers. <http://dev.openlayers.org/releases/
OpenLayers-2.7/doc/apidocs/files/OpenLayers-js.html>

Stefanov, Stoyan. Object-Oriented Javascript. Birmingham, UK: Packt Publishing Ltd., 2008.

"The GstreetViewPanorama Object." March 2009. Services - Google Maps API.
<http://code.google.com/apis/maps/documentation/services.html#PanoramaObject>