White Paper Report

Report ID: 106527

Application Number: HD-51581-12

Project Director: Patricia Fumerton (pfumer@english.ucsb.edu)

Institution: University of California, Santa Barbara

Reporting Period: 7/1/2012-3/31/2014

Report Due: 6/30/2014

Date Submitted: 10/1/2014

English Broadside Ballad Archive (EBBA): Ballad Illustration Archive

White Paper to the NEH Office of Digital Humanities
Digital Humanities Start-Up Grant

June 2014

Authors: Carl G Stahmer (Project Director) and Megan Palmer-Browne

Contact: cstahmer@ucdavis.edu

In 2013 the English Broadside Ballad Archive (EBBA) at the Early Modern Center, University of California, Santa Barbara received Digital Humanities Start-Up funding to begin work on the Ballad Impression Archive (BIA), a component of EBBA devoted to cataloguing and making fully searchable, both through automated image matching and descriptive metadata, the over 9,000 (and growing) individual woodcut impressions in the Archive. The project's work-plan had three primary components: 1) The development of a computer vision software application (built upon open-source computer vision libraries and algorithms) capable of performing automated searches of historical printed materials; 2) The implementation of this software in the EBBA website's public and administrative user interfaces; and 3) the computer assisted human cataloguing of a sample of the woodcut impressions currently in the EBBA archive.

The grant period was July 2012-June 2014. The project deliverables were: 1) the open-source distribution of the experimental computer vision software package (Arch-V) developed for the project; 2) The implementation of a computer vision searching interface on both the public and administrative interfaces of the EBBA website; 3) the production of metadata descriptions of a sample set of woodcut impressions; and 4) this white paper.

Table of Contents

I. Background …. 3

II. Project Participants …. 6

III. Scope ….…. 6

IV. Arch-V Overview …. 6

V. The Visual Dictionary …. 8

VI. Bag of Visual Word Creation …. 10

VII. Indexing …. 11

VIII. Querying …. 11

IX. Arch-V Components …. 12

X. Human-Machine Collaboration …. 12

XI. Supervised Cataloguing …. 13

XII. Supervised Indexing …. 15

XIII. Results …. 15

XIV. Future Development …. 18

XV. Conclusion …. 19

I. Background

Once seen as the domain of small, "boutique" projects driven by scholars interested in particular collections, the mass digitization of all forms of cultural heritage objects (text, image, sound, binary, etc.) has grown to become a major portion of both library and museum work efforts. According to *The Survey of Library and Museum Digitization Projects, 2013 Edition*[1], across the US, Canada, and the UK, major research and public libraries have a mean of 6.97 employees devoted to some form of digitization. Simply put, the digitization of printed, visual, and even 3-Dimensional artifacts is now seen as an indispensable and acceptable form of preservation, curation, and distribution of museum and library holdings.

While digitization offers the promise of increased materials access, this promise is fulfilled unevenly across collection types. As noted by Anne R. Kenney, "Additional work, which traditionally requires time-consuming descriptive cataloging or manual indexing" is required to make digital resources discoverable by potential users.[2] This problem is particularly acute when dealing with non-textual resources, such as images and sound recording. Whereas textual content can be easily searched without any special processing using a variety of readily available text searching engines, the same is not true for images and recordings.

Image collections have historically been discoverable only through the use of descriptive metadata that is both time consuming and costly to produce and maintain. Unfortunately, however, because of the labor involved in the creation of descriptive metadata, the rate of digital production has historically and continues to outpace the rate of digital description. Most institutions can simply not afford to properly catalogue visual digital collections in a manner that would be of most use to the scholarly community. The net result of this reality is that, whereas there are currently a large number of important visual collections "available" online, the usability of these collections is actually quite low as the images in these collections can, in most cases, not be navigated with anything other than the most general metadata.

There are, of course, exceptions to the above rule. Visual archives such as, for example, the Bodleian Ballad Archive and The British Printed Images to 1700 Archive (BPI1700) , have devoted significant resources to producing thick metadata catalogues of their image collections. But such efforts represent the exception rather than the rule. In both of these example cases the cost of developing such thick metadata was covered through research grant awards from major funding sources and not through the ongoing operations budgets of their host institutions. In neither of these cases could the local institutional budgets have covered the cost of this work, nor could every institution that currently manages a digital archive of visual material receive extramural funding for such efforts.

In addition to the economic realities that make thick cataloguing of visual materials cost prohibitive, there are also functional limits to this approach that argue against reliance on it as a sole solution to the problem of making digital archives of visual material properly searchable by end-users. Currently, the most widely used system for cataloguing images is Iconclass.[3] Iconclass provides a rich taxonomy for identifying both whole images and various items that appear in images. But the tree-based structure of taxonomies such as Iconclass creates a situation wherein the same exact item can be described by nodes

1 Primary Research Group, *The Survey of Library and Museum Digitization Projects, 2013 Edition* (2013).
2 Anne R. Kenney, "Technology: Mainstreaming Digitization into the Mission of Cultural Repositories," *Collections, Content, and the Web* <http://www.clir.org/pubs/reports/pub88/technology.html>, Council on Information and Library Resources (CLIR), February, 2000.
3 Http://www.iconclass.org

on multiple branches of the tree, thereby descriptively bifurcating otherwise like items. Take, for example the woodcut impression depicted in Figure X below:

Figure 1: A Man Clothed in Foliage

A cataloguer using Iconclass could easily describe the foliage that appears on the figure's head in the above image as, among many possibilities, “48A98212 foliated head ~ ornament”, “48AA9831 foliage, tendrils, branches ~ ornament - AA – stylized”, or “25H151 deciduous forest – leaf.”[4] None of these would be technically wrong. And a room full of knowledgeable scholars could argue ad infinitum about whether one of these or some other possible designation could even be preferred. As such, the likelihood that different cataloguers, or even the same cataloguer encountering similar images with a significant time gap between exposures, could use completely different designations to describe exactly the same thing is, in fact, quite high. Such realities dramatically impact the overall effectiveness of large catalogues.

This is not to say that descriptive tagging with Iconclass (or any such vocabulary) is not a valuable and worthwhile endeavor. It is. But it tends to tell us more about what a given cataloguer or scholar at a particular place or time thinks about an image, how visual forms that appear in the image are interpreted, than about the image itself. Anyone who has ever tried to catalogue an image quickly realizes the necessity of making a host of interpretive decisions even while cataloguing the simplest of images. Consider, for example, the following woodblock impression:

Figure 2: Single Male Figure Outside

4 The full scene in which the figure appears establishes a forest setting.

The above impression depicts a relatively simple scene of a solitary figure standing on a vaguely grassy plain. Even this simple scene presents descriptive challenges in which the cataloguer must make decisions that will impact the overall discoverability of the resource by scholars. For example, does the series of stacked horizontal lines that define the landscape horizon represent an artistic trope for doing so, or are they meant to communicate foliage or something about topography? It the figure standing on or in front of hilly ground? If so, is this important enough to capture in the metadata? Do the three clumps of foliage represent grass or simply a generic plan? And, in either case, do the shape differences between each matter? If so, how are they to be captured? Finally, is it enough to say that he the man is finely dressed? Or does the cataloguer need to identify specifically that he is wearing a pointed doublet with a stiffened lace collar and lace-edged cuffs?

It is both physically and theoretically impossible for any cataloguer to describe every aspect of an image for which a scholar could want to search. Aside from the labor involved in such an undertaking, it could only be accomplished if the cataloguer could accurately identify every bend of a line that might be of interest to scholars. This is fundamentally impossible as scholarship, as a mode of inquiry, builds on itself such that new concepts are continuously coming into being. Cataloguing efforts have typically dealt with the above problematic by tending toward descriptive generality. From a metadata perspective it is more important that figure above is wearing a hat than that the particular style of hat be identified because a scholar could find all hats and then perform his or her own filtering by type. As a general rule, this works well; however, as the number of digitized collections continues to grow, so does the number of figures with hats. What we need is a secondary system that allows the scholar to identify a previously unnoticed type of hat and then easily and reliably search all image for the hat of interest.

Content Based Information Retrieval (CBIR) speaks directly to the above need. CBIR is a search and retrieval approach wherein one image (or part thereof) serves as the seed for a search of a larger image library. In CBIR systems, the computer attempts to match visual phenomena that appear in the seed with other images in the library and returns best matches based upon the occurrence of these phenomena in the images being searched. CBIR offers a distinct advantage over descriptive metadata searches in that the user need not “know” what an image is in order to search for it. For example, consider the following ornamental image:

Figure 3: Ornamental Woodblock Impression

Certainly, there are several identifiable objects within the ornamental design presented in the impression. But there are equally as many nondescript and non-nameable design patterns. Using traditional metadata, the only means for a scholar to search for all occurrences of particular pattern or glyph of interest is for: 1) the pattern or glyph of interest to have already been given a name; 2) the scholar to know the name of the pattern or glyph; and 3) the cataloguer of the image to also have known and described the pattern or glyph accordingly. As such, a scholar's ability to search the system is completely bound by the already known—not an ideal situation for a profession whose mandate is

the production of new knowledge. Using a CBIR system, a scholar can easily identify a segment of visual interest and subsequently search the archive for other occurrences of the identified seed. In this way, CBIR facilitates the investigation and revelation of new, previously un-identified and un-named information patterns. As such CBIR offers significant promise for the study of early printed materials.

II. Project Participants

This pilot project focused exclusively on the collection of woodblock impressions that appear on the broadsides contained in the English Broadside Ballad Archive (EBBA). The project was directed by Principal Investigators Carl G Stahmer and Patricia Fumerton. Carl G Stahmer was additionally responsible for development and implementation of the project's CBIR platform, Archive-Vision (Arch-V) as well as its integration into both the EBBA website public user interface and administrative cataloguing interface. Megan Palmer-Browne was responsible for the development of EBBA's cataloguing methodologies and oversaw a team of graduate student workers in the computer-assisted cataloging of the impressions.

III. Scope

This pilot project focused its efforts on developing CBIR technologies specifically suited for application to historical prints and applying this technology in tandem with traditional cataloguing methods. The specific goals of the project were to allow users to use CBIR as a means of searching the over 9,000 printed woodblock impressions that appear on the broadside ballads in the EBBA archive and to allow human cataloguers to use the same tool to view and import the descriptions of similar impressions during the cataloguing process as a means of insuring consistent cataloguing practices across the archive.

IV. Arch-V Overview

As no fully functioning CBIR system for dealing with historical printed materials existed, a new software platform, Archive-Vision (Arch-V), was developed for this purpose. CBIR is a wide field of research and study, and there are several approaches to the CBIR problem. These include, for example, shape recognition, color histogram analysis, and feature recognition.[5] The approach implemented by Arch-V is based upon the model developed by Daniel Marcus Jang and Matthew Turk for the automated recognition of different cars in images and video.[6] In this approach points of interest known as Feature Points are extracted from each image and then indexed for later comparison as part of a query process. Figure 4 below depicts an image from the EBBA archive with its identified Feature Points highlighted:

5 For an excellent collection of articles on various approaches to image analysis see the collected lectures of Brian Morse (http://morse.cs.byu.edu/) at http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/MORSE/.

6 Jang, D. M., and Matthew Turk. "Car-Rec: A Real Time Car Recognition System." IEEE Workshop on Applications of Computer Vision (WACV) 2011 Kona, Hawaii, January 5-6, 2011. Ed. IEEE Computer Society Technical Committee on Pattern Analysis and Machine Intelligence (PAMI). Piscataway, NJ: IEEE, 2011. 599-605. Print.

Figure 4: Feature Points Identified in Woodblock Impression

In the above example, each drawn circle represents an identified Feature Point, with the size of the circle representing the scale of the feature and the drawn radial line its orientation. There are many known feature point extraction algorithms, each favoring a different type of feature definition (arcs vs. angles, continuous lines vs. angles, etc.). Additionally, each algorithm can be "tuned" to focus on particular aspects and/or sizes of features.

Figure 5: Feature Points Identified Using Differing Algorithm Tunings

Figure 5 above shows the same image with feature points identified using different extraction algorithm tunings for each image. As can be seen, variations in feature point algorithms and the tuning of these algorithms produces radically different results when applied to the same image. Feature point extraction is an art as much as a science. To produce functional results, a CBIR system must both apply the correct extraction algorithm and then tune that algorithm so that it identifies features that are likely to be meaningful when comparing images in the collection to which it is applied.

Arch-V implements Speeded-Up Robust Features (SURF) feature points. SURF offers an algorithm for both the identification of unique points of interest in images and for describing these features

mathematically in a way that is rotation- and scale-invariant, meaning that the same feature as it appears in any image of any size and/or rotation will have the same description. Additionally, as the name implies, SURF feature point extraction has been shown to be computationally faster than other known algorithms.[7]

Arch-V functions by identifying and comparing the SURF feature points found in each image of the collection of focus. Each image is understood by the system as a Bag of Features, as opposed to an actual visual representation of some object or thing. Comparison of individual images is, by extension, carried out by a bayesian comparison of the features in one bag to those of another in the same way that the Bag of Words approach has long been used in text comparison.[8]

The complete extraction and indexing process has three distinct stages of work: 1) the creation of a Visual Dictionary for the collection of interest; 2) the creation of Bags of Visual Words (BoVW) for each image in the collection of interest; and 3) the optimization and indexing of the bags of words. Each of these stages is described in detail below.

V. The Visual Dictionary

The visual dictionary is a defined collection of feature points. Typically, a visual dictionary does not contain all distinct feature points found in every image in a collection. Rather, through a process known as quantization[9], a set of "ideal" features that appear in the collection is constructed. The quantization process works by grouping closely related features and then mathematically constructing an averaged feature definition from all members of each group. Figure 6 below provides a simplified depiction of the quantization process as applied to a collection of scalar numeric values:

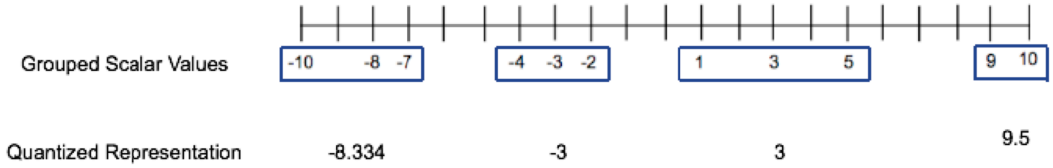

Figure 6: Simplified Quantization

In the above example, scalar values are grouped according to a calculated range of difference and then a quantized representation of each grouping is created by constructing an average or ideal value. Each item in the scalar group is then considered equivalent to its quantized representation. Actual quantization algorithms are mathematically more complicated, but the above demonstrates the basic point of quantization, which it to create an ideal representation of a group of values or, in the case of feature points, vectors describing shapes.

---

7 See Bay, Herbert, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. "Speeded-Up Robust Features (SURF)." *Computer Vision and Image Understanding* 110.3 (2008): 346-59. *ScienceDirect*. Web. 24 Sept. 2014. <http://www.cs.zju.edu.cn/~gpan/course/cv2013u/SURF.pdf>.

8 "Bag-of-words Model." *Wikipedia*. Wikimedia Foundation, 29 Aug. 2014. Web. 25 Sept. 2014. <http://en.wikipedia.org/wiki/Bag-of-words_model>.

9 Figueiredo, M´rio A. T., and Departamento De Engenharia Electrot´cnica E De Computadores,. "Scalar and Vector Quantization." (n.d.): n. pag. *Scalar and Vector Quantization*. Instituto Superior Técnico (IST), Nov. 2008. Web. 26 Sept. 2014. <http://www.lx.it.pt/~mtf/Vector_quantization.pdf>.

The dictionary creation process in Arch-V is initiated by defining the size of the dictionary. As with feature point extraction, determining the appropriate size of the dictionary for a given collection is also an art. A dictionary that is too large will result in features a human would consider identical being quantized to completely different words in the dictionary. A dictionary that is too small will produce the opposite effect, wherein features that a human reader would see as meaningfully distinct would be reduced through quantization to the same word in the dictionary. As a point of reference, the current, live implementation of Arch-V at EBBA utilizes of dictionary of 100,000 words.

Having defined the size of the dictionary, dictionary building proceeds according the process defined in Figure 7 below:

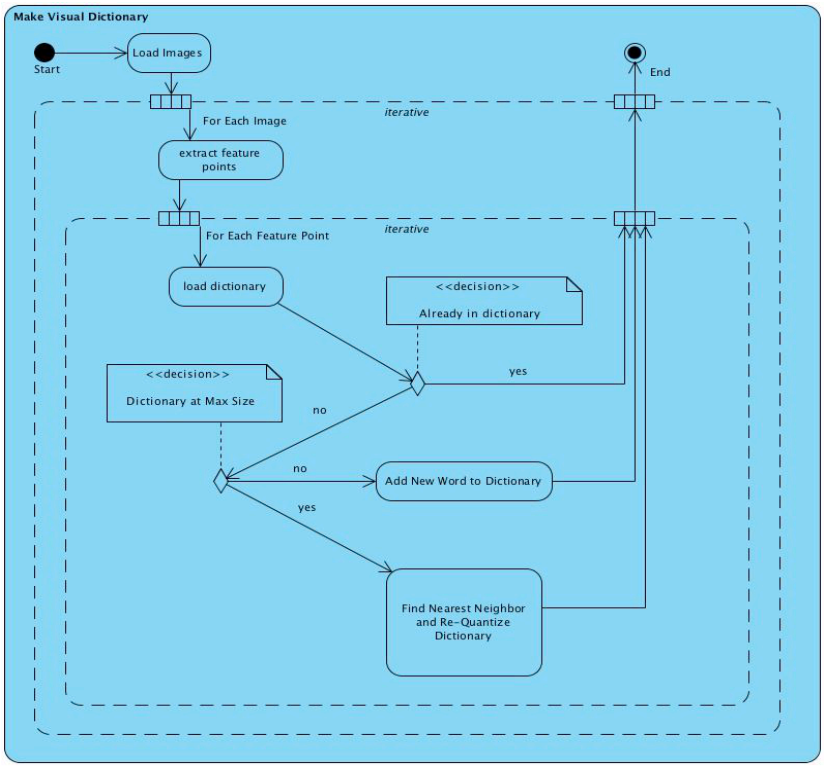

Figure 7: Visual Dictionary Creation Process

As depicted above, to create the visual dictionary the system loops through each of the images in the sample set. For each image, it extracts the feature points for that image and checks to see if, within the distance parameters of the quantization algorithm, the feature point is already represented in the dictionary. If it is not, then the system either adds the point to the dictionary or, if the dictionary has reached its maximum size, re-quantizes the entire dictionary so that the found feature is included in the list of quantized features.

When the process is completed, feature point vectors held in the dictionary represent a collection of ideal features, each associated with its own label or "word." At the file storage level, the Dictionary exists as a YML file in which an arbitrary label is assigned to each vector containing the values used by the SURF algorithm to describe the ideal feature. Figure 8 below shows the content of the Visual Dictionary.[10]

10 For presentation purposes, in this example the contents of the Visual Dictionary have been converted from its native

1004	3.67842093e-02	-1.44417174e-02	4.17876095e-02	1.05223916e-02
1005	2.56132446e-02	2.52754427e-03	2.07309946e-02	1.33339986e-02
1006	2.65369322e-02	1.88759563e-03	4.08165623e-03	-5.44514856e-04
1007	4.85381857e-03	5.26704534e-04	3.79359978e-03	-1.37390321e-04
1008	2.96981516e-03	8.36356412e-05	1.00633409e-02	6.85110688e-04
1009	2.54524909e-02	-3.34175560e-03	2.87749469e-02	-2.28401758e-02
1010	3.74305695e-02	2.32606512e-02	1.33650050e-01	7.10960478e-02
1011	2.53416181e-01	7.18055665e-03	2.34197542e-01	-9.22785029e-02
1012	1.96215004e-01	-9.41958055e-02	3.83012772e-01	1.42335460e-01
1013	2.50786424e-01	-6.64097220e-02	2.46676311e-01	6.78411424e-02
1014	1.60762936e-01	-1.49980653e-02	4.01108935e-02	-2.03944817e-02
1015	2.70386338e-02	1.66080557e-02	2.47544423e-02	-9.18097422e-03
1016	4.45772111e-02	-7.46660680e-03	2.74042320e-02	-5.72047150e-03
1017	1.72672700e-02	-3.49596236e-03	2.32186913e-02	8.41176696e-03
1018	4.15571705e-02	2.73971371e-02	2.26367816e-01	-1.08284280e-02

Figure 8: Extract from Visual Dictionary YML File

In the above example, each line represents a single visual word. The numeric indicator to far left represents the label (word) the system has arbitrarily assigned to the feature, and the columns to the right represent variable values that, when fed to the SURF algorithm, define the shape of the feature. This list of words serves as the basis for the creation of the Bag of Visual Words for each image in the collection of interest.

VI. Bag of Visual Word Creation

The core of Arch-V's CBIR functionality is the bayesian analysis of Bags of Visual Words (BoVW), one each representing the images in the collection of interest. Figure 9 below depicts the process implemented by Arch-V to create a BoVW for an image:

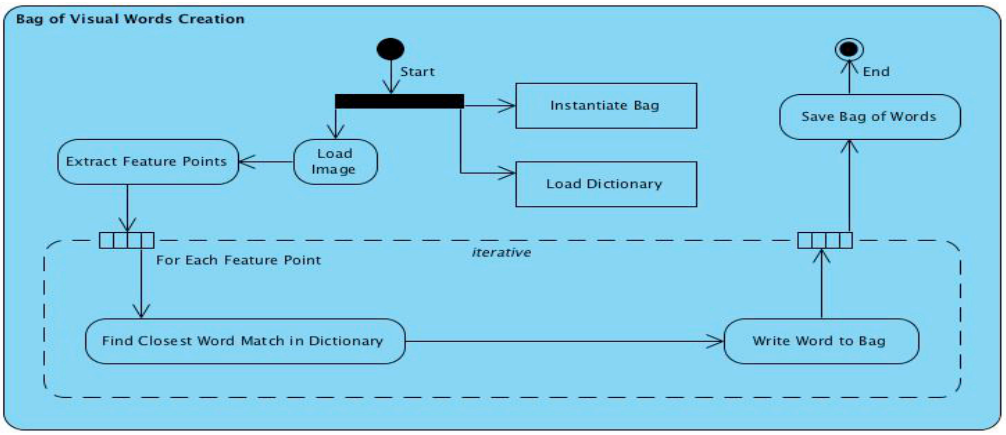

Figure 9: Bag of Visual Word Creation

As depicted above, BoVWs for an image are created by extracting the feature points from the image and then finding the closest matching word in the dictionary for each image and adding that word to the bag of words. The resulting BoVW for the image captures word frequency but not word order. As

YML to a spreadsheet in order to make it readable to those not familiar with YML. Also note that the example shows only 4 of the 128 dimensions of the actual vector for each visual word.

discussed in the *Future Development* section of this document below, considering the spatial relationship of words to their nearest neighbors (syntax) offers the potential to further improve Arch-V's results; however, as with bayesian text analysis, considering frequency presence only returns results with a high-level relevance.

Once Arch-V has constructed a BoVW for an image, the BoVW is saved as an ASCI file containing the words in the bag. In the Arch-V universe, each word, a quantized vector representing an ideal SURF feature point definition, is represented by a simple numeric label, randomly assigned as part of the dictionary creation process. As such, the resulting BoVW file for an image is a simple text file containing a string of numbers, one each for each occurrence of each word in the bag. Figure 10 below shows an extract from a visual word file. As can be seen, because features are not whole visual features (such as a hand or a leaf) but a particularly angled small shape or curve, it is common for the same feature to appear multiple times in the same image.

```
151 151 151 151 151 151 151 151 151 151 151 151 151 151 151 151 151 151 151 151 151
151 151 151 151 151 151 151 151 151 151 151 151 151 151 151 151 151 151 151 151 151
151 151 151 151 151 151 151 151 151 151 151 151 151 151 151 151 151 151 151 151 151
151 151 151 151 151 151 151 151 151 151 151 177 177 177 177 177 177 177 177 177 177
177 177 177 177 177 177 177 177 177 177 177 177 177 177 177 177 177 177 177 177 177
177 177 177 177 177 177 177 177 177 177 177 177 177 177 177 177 177 177 177 177 177
177 177 177 177 177 177 177 177 177 177 177 177 177 177 177 177 177 177 177 177 177
177 367 367 367 367 367 367 367 367 367 367 367 367 367 367 367 367 367 367 367 367
367 367 367 367 367 367 367 367 367 367 367 367 367 367 367 367 367 367 367 367 367
367 367 367 367 367 367 367 367 367 367 367 367 367 367 367 367 367 367 367 367 367
367 367 367 367 367 367 367 367 367 367 367 367 418 418 418 418 418 418 418 418 418
418 418 418 418 418 418 418 418 418 418 418 418 418 418 418 418 418 418 418 418 418
418 418 418 418 418 418 418 418 418 418 418 418 418 418 418 418 418 418 418 418 418
418 418 418 418 418 418 418 418 418 418 418 418 418 418 418 418 418 418 418 418 418
418 418 430 430 430 430 430 430 430 430 430 430 430 430 430 430 430 430 430 430 430
```

Figure 10: Extract from a Visual Word File

VII. Indexing

Once BoVWs have been created for each image in the collection of interest, the next required step is to index the BoVWs so that they are available for searching. As the BoVW library exists as a collection of text files, one each for each image in the collection, indexing these files in order to make them available for quick and low-overhead search and retrieval can be carried out using known and widely implemented processes and software developed to make text easily searchable. For its implementation, Arch-V uses Apache Lucene to index the BoVW library.[11] Lucene is a free, open source software library that is widely used in a variety of text-searching applications. Arch-V's implementation of Lucene is straightforward and completely out-of-the-box. BoVW files are frequency indexed using standard Lucene functions.

VIII. Querying

Querying through Arch-V is also accomplished through Lucene. When an image from within the library is used as the search seed, Arch-V retrieves the visual word file for the seed from the BoVW library and submits the contents of the file through Lucene as a query against the Lucene index. For images not already in the library, Arch-V first creates a BoVW representation of the image based upon the library's dictionary and then submits the new BoVW collection as the Lucene query. In each case, Lucene returns to Arch-V a relevance-ranked list of BoVW files as matches to the query. Arch-V then

11 http://lucene.apache.org/.

maps BoVW filenames to the actual images that they represent and returns the ranked list of images to the user.

IX. Arch-V Components

Arch-V consists of two distinct software libraries: 1) The Arch-V BoVW Engine (affectionately known as the Bow-Wow); and 2) the Arch-V Indexing and Search Toolset (IST).

The Bow-Wow is is a C++ application toolset built on the OpenCV computer vision library.[12] OpenCV is a widely adopted and actively supported open source computer vision library with a user community that exceeds 47 thousand people and that has been download over 7 million times.[13] While OpenCV does offer, as part of its distribution, interfaces in a variety of programming languages, the Bow-Wow was developed as a C++ application so that it could interact natively with the complete OpenCV package library. The Bow-Wow contains tools for performing all Arch-V functions up to the creation of BoVW files. It also contains a variety of additional tools for visual feature point identification as a means of assisting users in testing the tuning of feature point identification and extraction parameters. The complete source code for the Arch-V Bow-Wow is currently available online under a Creative Commons Attribution Share-Alike (CC BY-SA 4.0) license at https://bitbucket.org/cstahmer/archv.[14] Note that as the software was developed under a small Start-Up award the code is currently neither optimized nor well documented. It is our hope to attain future funding to optimize the code, improve functionality, and fully document.

The Arch-V Indexing and Search Toolset (IST) is a Java toolset built on Apache Lucene. It provides a variety of tools to assist in the indexing process and also that serve as a gateway/API for integration of query functionality into web applications. The IST's indexing tools include tools for optimizing indexes through identifying the most statistically unique features found in an image and pruning BoVW files accordingly, as well as tools for integrating descriptive cataloguing into the index as discussed in the *Human-Machine Collaboration* section of this report below. The IST toolset also contains tools for creating Lucene indexes from BoVW files. The IST search gateway provides an API for receiving search queries (in the form of an identified search seed) and returning JSON, relevance-ranked search results for processing by web applications.

X. Human-Machine Collaboration

A stated goal of the project was to investigate ways in which CBIR could enhance, rather than replace, human cataloguing of images. The *Background* section of this report documents one of the significant problems with human cataloguing systems: the fact that descriptive vocabularies are, by their very definition, bound by the limits of what is already known about both the object they describe and the community for which they are described. From the single perspective of content retrieval, this limitation is perceived as a liability. However, from the perspective of an archive's ability to participate in the scholarly ecosystem of its day, this is actually an asset. Descriptive metadata serves not only as a finding aid, but as an important means of situating objects described within a discursive universe. As such, human cataloguing should remain an important valence of digital preservation and access.

With the above in mind, BIA set out to investigate how CBIR technologies could be implemented as an

12 http://opencv.org/.
13 "OpenCV | OpenCV" *OpenCV*. Itseez, n.d. Web. 29 Aug. 2014. <http://opencv.org/>.
14 https://creativecommons.org/licenses/by-sa/4.0/.

assistive technology, working in concert with, and not as a replacement for, human cataloguing efforts. Additionally, we were interested in investigating the ways in which the idea of an assistive technology could be applied bidirectionally. CBIR research has a long history of investigation into supervised systems, in which human operators provide feedback to a machine agent in order to improve relevance of machine-generated results. For BIA, we were specifically interested in asking how existing cataloguing workflows could be leveraged to provide asynchronous supervision of the CBIR systems. The following two sections of this report, *Supervised Cataloguing* and *Supervised Indexing,* document BIA's application of Arch-V on these two fronts.

XI. Supervised Cataloguing

The idea of Supervised Cataloguing inverts the standard computer science definition of a supervised system by introducing the computer as the agent of supervision into the cataloguing system. Part of BIA's mission is to provide a rich set of descriptive metadata for each impression in the archive. In terms of the human cataloguing of the woodcut images, our team proceeded under the philosophy that users should be able to learn two things about each of these images from their keywords: first, what genre an image belonged to; and second, what particular things were being portrayed in each image. The first set of terms is abstract and generic; the second, concrete and particular, tagging all figures and objects visible in every image.

To achieve the above-described ends, keywords were divided into two categories: Genre Terms (including, for example, *narrative* and *landscape*) and Descriptive Tags (e.g., *man*, *horse*, *book*, *execution scene*). The Genre Terms were modified from the art historical hierarchy of genres[15] in order to capture all the key sorts of scenes covered by the archive. The Descriptive Keywords, by contrast, were carefully harvested from previously written descriptions of the woodcut impressions in the following way: initially, the team extracted the most significant nouns and verbs from these descriptions. We looked through the resulting list for repeated words; after gathering these, we culled the list again to represent only the most salient. So, for example, the description “court, two soldiers, man and woman; soldiers holding spears, man and woman holding scepters; crowns, cap; King William and Queen Mary” yielded the nouns *court*, *soldier*, *man*, *woman*, *spear*, *scepter*, *crown*, *cap*, *king*, and *queen*. After discussion, *court* was eliminated (the only way to tell whether an image represents a court scene is if it contains a king and/or a queen; since both *king* and *queen* can be tagged, *court* becomes redundant). Similarly, *cap* was eliminated, along with a bevy of other terms describing aspects of costume, since one quick path to madness is to tag all of the thousands of men in woodcut expressions with the term *hat*. Instead, since fashion trends tended to follow the tastes of particular monarchs, we created a set of tags for each of the major regnal periods covered by our archive, and made up a “costume book” with descriptions and images of changing fashion for our cataloguers.[16] Finally, when the term *soldier* was considered alongside similar terms in our archive, we opted to change all to the more general *military figure.* At the end of this process, our Descriptive Tags were put into subject categories (e.g., *Architecture*, *Natural Things*, *Tools / Instruments*) to make them easier to find for both cataloguers and end-users. All genre terms, keywords, and category terms are mappable onto both the well-known Getty Art & Architecture Thesaurus[17] and Iconclass,[18] increasing the BIA's interoperability with other projects.

As can be seen from the above, BIA's human cataloguing effort operates in a rich metadata universe.

15 See http://en.wikipedia.org/wiki/Hierarchy_of_genres.
16 See http://ebba.english.ucsb.edu/page/early-modern-costume.
17 http://www.getty.edu/research/tools/vocabularies/aat.
18 http://www.iconclass.nl/home.

While on the one hand this increases the specificity of cataloguing, on the other hand it increases the likelihood of the same or similar images being tagged differently. This possibility increases when multiple cataloguers are employed or when the time between dealing with similar images by the same cataloguer is extended. In order to protect against this difficulty, BIA implemented Arch-V as a supervisor to the cataloguing process.

Woodcut impressions in BIA are catalogued using EBBA's web-based administrative interface. The interface provides a point and click UI in which cataloguers are able to identify the impression to be catalogued and then select the terms that apply to that impression from EBBA's various controlled vocabularies as described above. Arch-V was integrated into this system such that any time a cataloguer is working on an impression the system automatically locates other similar images in the system and exposes to the cataloguer the way these images were encoded. Figure 11 below provides a screenshot of the cataloguing interface:

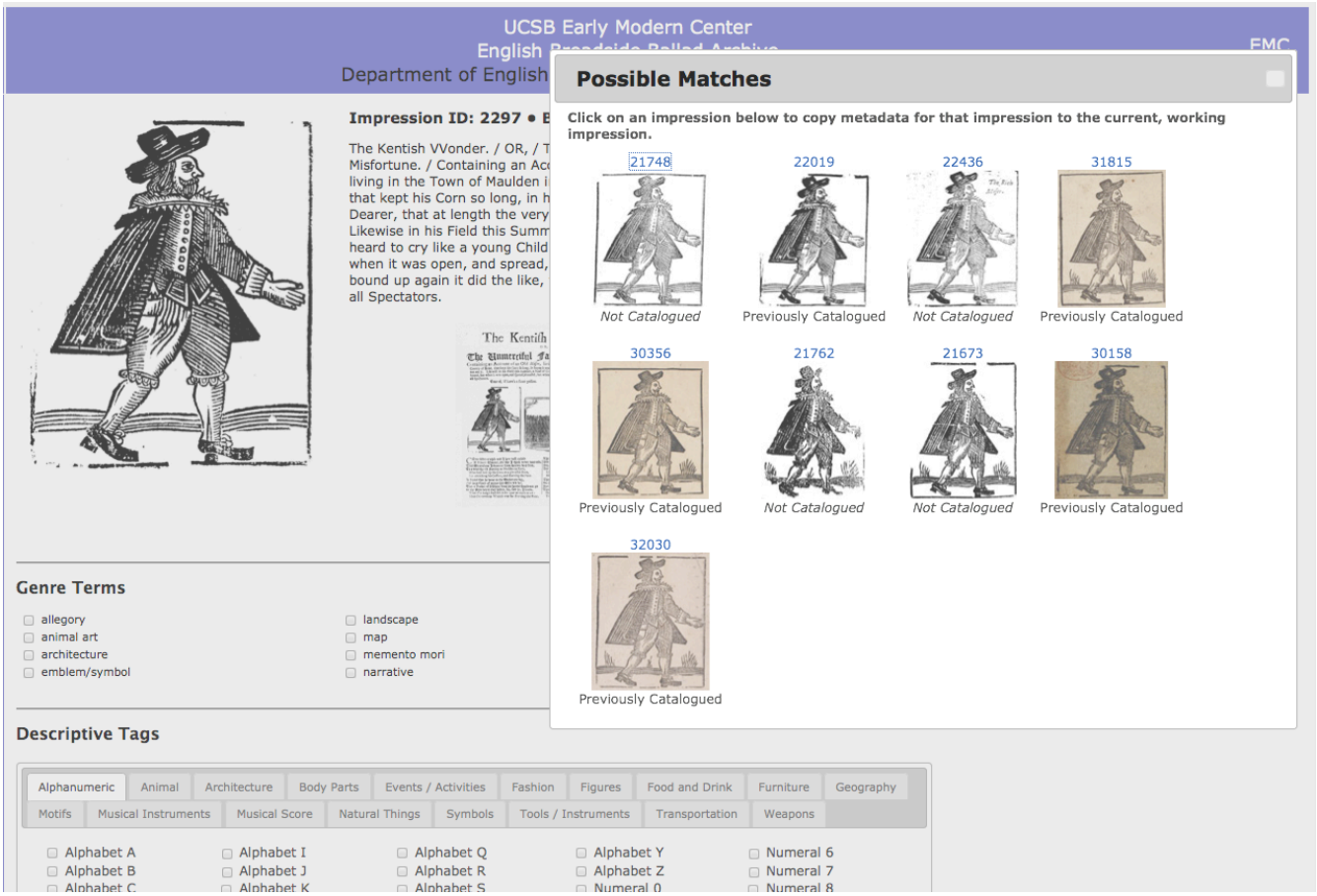

Figure 11: BIA Cataloguing Interface

Clicking on any of the found similar images will reveal the metadata with which the image was previously described. Additionally, the cataloguer can choose to import that data to the current record and/or make changes and normalize across selected known instances. Arch-V thus serves as the supervising agent in the system, helping the human cataloguer to refine her or his work, normalizing descriptive practice across the entire archive, and generally improving the quality of metadata.

XII. Supervised Indexing

As noted in the *Human-Machine Collaboration* section of this report above, another goal of the project was to test the possibility of providing asynchronous human supervision of the CBIR process. Typically, supervised CBIR implementations are designed using the following two process architectures (either independently or in combination): 1) creation of "ground truth" collections of similarity data; and/or 2) implementation of a result-rating feedback loop.

In a ground truth system, prior to the commencement of the CBIR process, a human supervisor seeds the system with representative collections of similar images. For example, if one were building a system to identify particular makes of car, the user would create a folder for each make of car and then put a representative sample of images of each make of car (ideally taken from different angles) into their respective folders. The CBIR system would then use this ground truth data as a basis for comparison during the process of feature point extraction and comparison. In a result-rating feedback loop system, expert human users are engaged to examine the results of the CBIR system and identify True and False results. The CBIR system then uses this data to re-index its results accordingly, thereby refining its decision making processes.

Both ground truth and result rating supervision systems significantly improve overall functionality of CBIR systems. However, both of these modes of supervision require an additional labor effort that can present an implementation barrier, as most digital archival efforts already struggle to support themselves financially. As such, one of BIA's goals was to architect a system of human feedback/supervision that would work as seamless and effortlessly as possible with current cataloguing workflows.

In order to achieve this goal, The Arch-V Indexing and Search Toolset (IST) includes a simple tool for merging cataloguing metadata into the BoVW file for an image. As previously discussed in detail, Arch-V relies upon the Lucene indexing of the ASCI BoVW files that represent each image in the collection to provide CBIR. The IST exploits the text-based nature of this system by adding human-catalogued metadata for an image to its BoVW. Once this addition has been completed and the BoVW file re-indexed, the index (and subsequent search returns) represent a combined visual-dictionary-words and metadata-words representation of the image. This has the effect of skewing results toward humanly identified matches. The effects of this skew are not sufficient to suggest a high relevance match where metadata matches but visual words do not, but it will move human-identified metadata matches to the front of group of visual word matched images.

XIII. Results

Empirically judging the results of such an effort is difficult to do without effective and proper control. As BIA's library of images has never before been catalogued, no such control exists for the current implementation. In small-scale testing using created control groups of images, we have found the current system to return 60% relevant results 82% of the time. This means that 82% of the time, 60% of the match results are images that a human user recognizes as a match. (Interestingly, in the 18% of results not considered successful, there is an immediate rather than gradual drop-off in the return of relevant results.) Figure 12 below is a screen shot of a sample return from the current, live implementation of Arch-V in BIA as part of the EBBA archive. (Note that this example is for an image that has not yet been catalogued, so the results are not skewed by human cataloguing as discussed above.) As can be seen, several of the images matched by the system are, indeed proper matches.

However, the remainder are not.[19]

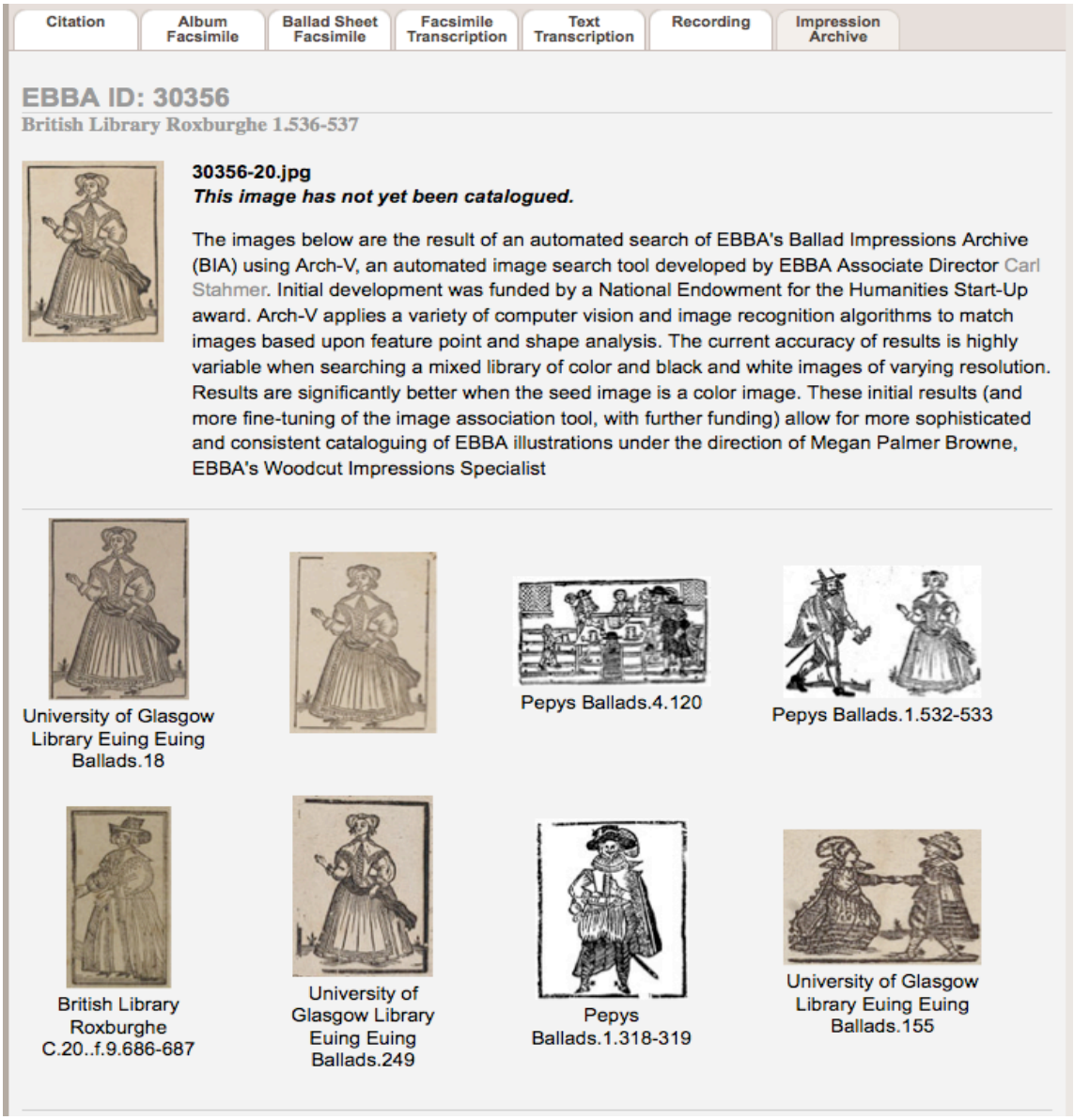

Figure 12: Sample BIA Search Return

Despite the high miss-rate, even the current system has proven useful to both scholars and cataloguers. Even a 50% success rate represents an improvement in the current state of the art, which leaves scholars to manually look through thousands of images in search of a visual match. Additionally, it is sufficient to allow cataloguers to connect one impression with several, if not all, matching images in the archive.

The failure rate of the current system can be traced to a fundamental difficulty with known algorithms

19 In fact, the inclusion of these non-matching images in the machine-produced result set is not arbitrary. To a trained user who is used to "seeing" images as a collection of feature points, the systems return of these images actually makes perfect sense.

for feature-point extraction. As noted by Relja Arandjelovic and Andrew Zisserman, the problem of automated recognition of objects has been largely solved, but only "provided they have a light coating of texture."[20]. This is because the state-of-the-art in computer vision relies upon the refraction of light across the surface texture of an object as it is captured in a digital image (or frame of video) in order to extract recognizable feature points as indexable markers of the object in the image. But in digital images of print artifacts, surface texture can serve as a distraction from rather than indicator of the objects depicted in the print. This is because the texture belongs to the delivery medium—the paper or canvas on which objects are printed—and not to the objects represented in the images. Figure 13 below depicts this problem in practice:

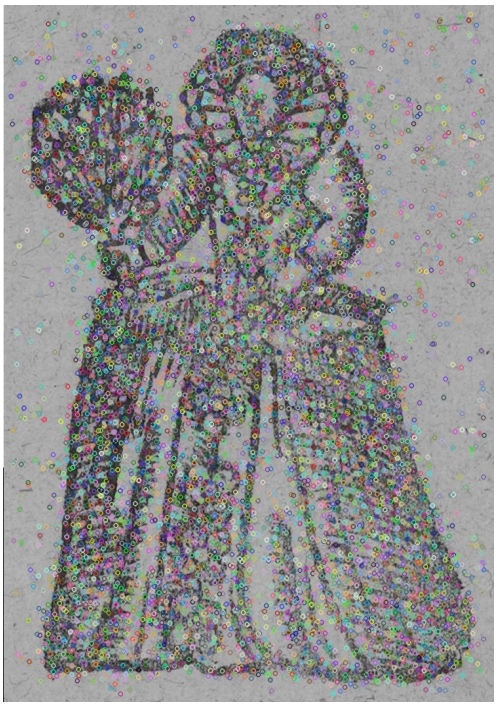

Figure 13: Feature Points Found in Paper

In the above image, the system has identified nearly as many feature points in the fibers and texture of the paper on which the woodblock impression is printed as in the image conveyed by the impression. Resolving this difficulty requires tuning the extraction algorithm to ignore smaller scale features, which necessarily results in meaningful feature points being ignored alongside the noise feature points belonging to the carrier. This problem is compounded by the fact that line inking frequently varies from print to print, thereby leading to differing textural "edges" along the boundaries of objects depicted. As a result, off-the-shelf implementations of current technologies prove less than satisfying when applied to digital archives of printed materials.

Having learned what we have learned from the present implementation, we believe that these

20 Zisserman, A. "Smooth Object Retrieval Using a Bag of Boundaries." *2011 IEEE International Conference on Computer Vision Workshops: 6-13 November 2011*, Barcelona, Spain. By R. Arandjelovi'c. Piscataway, NJ: IEEE Computer Society, 2011. N. pag. Web. 27 Sept. 2014. <http://www.robots.ox.ac.uk/~vgg/publications/2011/Arandjelovic11/arandjelovic11.pdf>.

difficulties can be successfully overcome and we have identified several avenues for future development that will greatly improve the software's overall functionality.

XIV. Future Development

There are several ways in which Arch-V's current CBIR process as described above could be improved both to overcome known difficulties with the state-of-the-art in feature point identification and extraction and also to generally improve the system to more accurately account for the ways in which a human reader determines whether or not two images are the same or similar.

First, the system needs to be enhanced to include more robust image pre-processing in order to remove the background noise of the carrier as discussed above prior to the feature point extraction process. Additionally, an improved system would normalize line width and contiguousness prior to feature-point extraction. Both of the above could be accomplished by extracting image contours, or boundaries, from the image, and then normalizing them prior to feature point extraction. During the normalization process, sharp edges could be smoothed to account for variation across print runs and small gaps in otherwise contiguous lines could be closed. The resulting collection of shapes would then be saved to a new image with a blank background, which would serve as the base image for feature-point extraction. Figure 14 below depicts a sample of an image that has been manually passed through such a process using OpenCV. As can be seen, the resulting image contains only information about the main features of the printed image itself, extracted from the noise of its carrier and with a normalized line representation.

Figure 14: Woodcut Impression Before and After Contour Extraction and Normalization

Because this extraction would result in loss not only of information about the carrier, but also of some information about internal shapes and patterns in the print, an improved indexing process should also

be implemented. An improved architecture would employ double pass feature point extraction and indexing, wherein the system would extract and index feature points from both the original image and the contour-extracted image. Results would then be calculated based upon a combined frequency analysis. The resulting final index would be significantly more focused on the images depicted in the print than on the carrier itself and also on the most unique features of each of these images.

Arch-V's CBIR capabilities could also be improved by combining feature point extraction and analysis with another proven image recognition approach: shape analysis. Whereas feature point analysis determines matching through the identification and matching of discrete features found in an image, shape matching attempts matches by comparing major shapes (such as the external, silhouette contour in Figure 14 above) found in different images. Shape analysis is a known successful form of image matching;[21] however, it is too computationally intensive to be scalable to large libraries. This limitation could be overcome by combining shape analysis with feature point analysis. A promising architecture for such a combination would use feature point analysis as a first order of comparison and then apply shape analysis only to those images in the collection that have high feature point comparison relevance. Such a system would minimize the application of computationally intensive comparison to a small subset of the library, thereby improving scalability.

A final potential area for improvement is to supplement the bayesian analysis currently performed (which considers all features as words in a bag), with an algorithm that would consider the appearance of features in an image in the context of their nearest neighbors. Another project recently funded by the NEH that has pursued this avenue of image comparison is the Paragon project at the University of South Carolina.[22] The goal of the Paragon project was to investigate using CBIR techniques as a means of performing semi-automated textual collation using digital facsimile images. As part of their efforts, the Paragon team successfully developed tools for including information about the location of feature points relative to each other in an image as part of an image matching process. As with shape analysis as discussed above, these algorithms have proved effective, but they are too computationally intensive to be scalable to large archives. This scalability problem could be solved by applying this analysis only to the few, most distinctive features of images already identified through the BoVW method as possible matches. The Arch-V Indexing and Search Toolset (IST) already contains tools for statistically identifying the most unique feature points in an image. As such, adding nearest-neighbor aware tools such as those developed by Paragon to the system would be easily achievable.

XV. Conclusion

As stated in the NEH's description of the Digital Humanities Start-Up Grant program, “Innovation is a hallmark of this grant category, which incorporates the ‘high risk/high reward’ paradigm often used by funding agencies in the sciences.”[23] By this metric, we consider the *English Broadside Ballad Archive (EBBA): Ballad Illustration Archive* to have been extremely successful. While the CBIR software developed during the grant period (Arch-V) still requires improvement, in its current form as deployed at EBBA the implementation demonstrates that CBIR technologies can be successfully leveraged to advance scholarship of historical printed materials. Additionally, the development and testing work

21 For a discussion of shape matching techniques see Zhang, Dengsheng, and Guojun Lu. "Content-Based Shape Retrieval Using Different Shape Descriptors: A Comparative Study." *Multimedia and Expo, 2001. ICME 2001. IEEE International Conference on* (2001): 1139-142. *IEEEXplore Digital Library*. IEEE. Web. 27 Sept. 2014. <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=1237928>.

22 http://cdh.sc.edu/projects/paragon

23 "Digital Humanities Start-Up Grants." *Neh.gov*. The National Endowment for the Humanities, n.d. Web. 27 Sept. 2014. <http://www.neh.gov/grants/odh/digital-humanities-start-grants>.

completed during the grant cycle allowed us to create an informed roadmap for future development to improve both functionality and ease of deployment. EBBA will continue to implement the software as part of our Ballad Impression Archive (BIA). Additionally, we hope to receive follow-up funding to continue to develop the current prototype into a fully-fledged, community-supported, open source CBIR tool.